A HISTORY OF THE PRIME NUMBER THEOREM

A thesis submitted

to Kent State University in

partial fulfillment of the requirements

for the degree of Master of Arts

by

Anita Nicole Alexander

December 2014

Thesis written by

Anita Nicole Alexander

B.S., Ohio University, 2004

M.A., Kent State University, 2014

Approved by

Dr. Ulrike Vorhauer, Mathematics, Advisor

Andrew Tonge, Ph.D., Chair, Department of Mathematical Sciences

James L. Blank, Ph.D., Dean, College of Arts and Sciences

TABLE OF CONTENTS

Acknowledgements

I would like to express my appreciation and gratitude to my advisor Dr. Ulrike Vorhauer, who has been a mentor and an inspiration to me. Her expertise in the field and her passion for Number Theory have played an integral role in my research and progress. I would also like to thank my committee members, Dr. Morley Davidson, Dr. Stephen Gagola, and Dr. Scott Courtney for taking the time to thoroughly read my thesis and for providing insightful feedback at my defense. It made for an enjoyable and informative experience, for which I am grateful.

A special thanks to my friends and family for their encouragement and support throughout the research process and especially during the preparation for my defense. I would also like to thank William Hanlon for his support, and his assistance in navigating LaTex with me. I am very grateful for this experience.

CHAPTER 1

Introduction

"*Prime numbers are the atoms of our mathematical universe.*" [9]

The location of prime numbers is a central question in number theory. An integer $n > 1$ is prime if it has exactly two distinct divisors, namely itself and 1. For positive real x let $\pi(x)$ denote the number of primes not exceeding x, that is

$$\pi(x) = \sum_{\substack{p \leq x \\ p \text{ prime}}} 1 \,. \tag{1}$$

Then the Prime Number Theorem asserts that

$$\pi(x) \sim \frac{x}{\log x} \tag{2}$$

as $x \to \infty$, that is

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1$$

Here, as throughout the paper, log denotes the natural logarithm ln.

Other conjectures concerning prime numbers were formulated. For example: the twin prime conjecture states that there are infinitely many prime numbers p such that $p + 2$ is also prime. Are there also infinitely many primes of the form $2^p - 1$? These are referred to as Mersenne primes after Marin Mersenne (1588–1648). The largest currently known Mersenne prime is $2^{57,885,161} - 1$ with $17,425,170$ digits (Curtis Cooper, GIMPS) [15]. Another famous conjecture is the Goldbach conjecture, formulated by Christian Goldbach (1690–1764) in a letter to Leonhard Euler (1707–1783) in 1742. It asks if any even integer $n > 2$ can be expressed as the sum of two primes and respectively if every odd integer greater than 5 can be written as the sum of three primes. Bertrand's Postulate (1845) states that there is always a prime in the interval $[N, 2N]$ for all $N \geq 1$. Pierre de Fermat (1601–1665) in a letter to Mersenne conjectured that all numbers of the form $2^{2^n} + 1$ are prime. In 1732 this conjecture was disproved by Euler and reformulated to the conjecture that there are only finitely many Fermat primes.

Prime numbers and their significance in number theory is interesting in and of itself. However, the objective here is to explore and comprehend the distribution of prime numbers and their applications, and to understand the Prime Number Theorem; the history, relevance, and applications, and to break down and explain, as well as compare and contrast the elementary and analytic proofs.

The Prime Number Theorem carries with it a rich history, while posing mathematical questions that remain unanswered.

CHAPTER 2

History of the Prime Number Theorem

There exists quite the controversy surrounding credit for the elementary proof of the Prime Number Theorem. We define elementary in this context as the proofs that do not involve the zeta function, complex analysis, or Fourier methods. However, the methods are quite intricate [23]. The feud began in the spring of 1948, though how exactly or to whose credit depends on which party one sides with. Atle Selberg (1917–2007) [21] had said "I had the Prime Number Theorem in my thoughts; that was my goal based on the formula

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

that I had obtained." Selberg discussed his ideas with Paul Erdős (1913–1996), who had found a proof of the generalization of Chebyshev's Theorem, that there exists a prime number between x and $x(1+\varepsilon)$, for arbitrary positive ε. However, he practically discouraged Erdős from working on the proof of the Prime Number Theorem and had no intention of collaborating with him, even though he thought he could use Erdős' proof to achieve an elementary proof of the Prime Number Theorem. Following a seminar and a brief discussion on Selberg's inequality, Erdős said "I think you can also derive $\lim_{n\to\infty} p_{n+1}/p_n = 1$ from this inequality" [21]. To which Selberg responded: "You must have made a mistake because with this result I can get an elementary proof of the prime number theorem, and I have convinced myself that my inequality is not powerful enough for that." So it became clear that the insight and efforts of both Erdős and Selberg were needed in order to achieve an elementary proof of the Prime Number Theorem, yet collaboration was not on the horizon. Erdős [5] not only proved that $p_{n+1}/p_n \to 1$ as $n \to \infty$ but proved the stronger result: To every ε there exists a positive $\delta(\varepsilon) > 0$ so that for x sufficiently large we have $\pi\big(x(1+\varepsilon)\big) - \pi(x) > \delta(\varepsilon)\, x/\log(x)$, where $\pi(x)$ is the number of primes not exceeding x. Erdős is said to have communicated this proof to Selberg, who two days later had his first elementary proof of the Prime Number Theorem. This was followed by an extensive dispute over how results should be published, one that was never settled despite one of the great mathematical achievements of the twentieth century. To quote Ernst Strauss [23]: "It has always been a source of great surprise and regret to me that two such superb minds and admirable human beings, whom I both consider my friends – although Erdős is clearly a much closer friend than Selberg – have come to a permanent parting over a joint achievement that had been born with so much joy and hope." Perhaps before exploring and analyzing these and other proofs of the Prime Number Theorem that caused such a controversy among mathematicians, we should take a look further back into the history of prime numbers and number theory.

The Pythagoreans classified positive integers as even, odd, prime and composite around 600 BC. The first systematic study of prime numbers appeared around 300 BC, in Euclid's *Elements* (see [11]). In book nine of the Elements, Euclid proved that there are infinitely many primes using a rather elegant proof by contradiction, which implies that $\pi(x) \to \infty$ as $x \to \infty$. Around 200 BC, Eratosthenes (276 BC–194 BC) studied a method to produce a list of prime numbers, known as the Sieve of Eratosthenes, which is a convenient way to produce lists of primes below some rather large bound, say 10^8. Extended tables of primes were calculated in the 17-th and 18-th centuries. By 1770 Johann Heinrich Lambert (1728–1777) had compiled a table of all primes up to $102,000$. At the beginning of the 17-th century, Pierre de Fermat devised a method to factor large numbers, known as Fermat's Little Theorem which states that if p is prime, then for any integer a, $a^p \equiv a \mod p$.

In 1737 Leonard Euler [4] made a direct connection between prime numbers and the tools of analysis by proving that

$$\frac{1}{1-2^{-n}} \cdot \frac{1}{1-3^{-n}} \cdot \frac{1}{1-5^{-n}} \cdots = \sum_{k=1}^{\infty} \frac{1}{k^n}$$

which displays the connection between the Riemann zeta-function on the right to the prime numbers on the left. Mathematicians have studied this function in order to prove properties of the prime numbers ever since. Using this formula Euler showed that the sum over the reciprocals of the prime numbers diverges, therefore the number of primes must be infinite. Knowing that there are infinitely many primes, interest shifted to the distribution of the primes among the positive integers and to the prime counting function $\pi(x)$ defined in (1).

In 1798 the French mathematician Adrien Marie Legendre (1752–1833) [13] conjectured that

$$\pi(x) \approx \frac{x}{A \log x - B} \tag{3}$$

where A and B are constants. In a second paper in 1808 [14] he made this approximation more precise by proving the values $A = 1$ and $B = 1.08366$ under the assumption that (3) holds.

In 1792 or 1793, the fifteen year old Carl Friedrich Gauss (1777–1855), by counting primes in blocks of 100, 1,000, and 10,000 consecutive integers, made a remark in his diary that

$$\pi(a) \quad (= \infty) \qquad \frac{a}{\ln a}\ .$$

Many years later in 1849 Gauss conjectured in a letter to the astronomer Johann Franz

Encke (1791–1865) [24] that

$$\pi(x) \approx \mathrm{Li}(x) = \int_2^\infty \frac{dt}{\log t} .$$

In his letter Gauss confessed that he was not aware of Legendre's work but notes that it appears that the average value of B decreases with growing x. It is an easy exercise in calculus to show that

$$\mathrm{Li}(x) \approx \frac{x}{\log x}$$

as $x \to \infty$.

2.1 Chebyshev's Theorem

In 1851 Pafnuty Lvovich Chebyshev (1821–1894) [2] made a significant step in the direction of the Prime Number Theorem by giving the precise order of magnitude of $\pi(x)$. He proved the following theorem:

Theorem 1 *There exist explicitly computable positive constants c_1 and c_2 such that*

$$c_1 \frac{x}{\log x} \leq \pi(x) \leq c_2 \frac{x}{\log x} \tag{4}$$

for all sufficiently large x. He also proved that if

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x}$$

exists, it is necessarily equal to 1.

In fact Chebyshev proved (4) with $c_1 = 0.92129$ and $c_2 = 1.1056$. This enabled him to prove Bertrand's postulate, namely that for all $x > 1$ there exists a prime p such that $x < p \leq 2x$.

In order to establish (4) Chebyshev introduced the functions

$$\psi(x) = \sum_{n \le x} \Lambda(n) \tag{5}$$

where $\Lambda(n) = \log p$ if $n = p^k$ and $\Lambda(n) = 0$ otherwise, the von Mangoldt function, and

$$\theta(x) = \sum_{p \le x} \log p \ . \tag{6}$$

The motivation for considering $\Lambda(n)$ is the following fact:

Lemma 1 *For every positve integer $n \ge 1$, $\sum\limits_{d|n} \Lambda(d) = \log n$.*

Proof. Let $n = \prod\limits_p p^\alpha$, then

$$\log n = \sum_p \alpha \log p = \sum_{p,k:p^k|n} \log p = \sum_{d|n} \Lambda(n)$$

since $p^k|n$ if $k = 1, 2, \ldots, \alpha$. By applying the Möbius inversion formula we obtain

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) \ . \tag{7}$$

Therefore, by summation over n we get

$$\begin{aligned} \psi(x) &= \sum_{n \le x} \Lambda(n) = \sum_{\substack{d,t \\ d \cdot t \le x}} \mu(d) \log t \\ &= \sum_{d \le x} \mu(d) \sum_{t \le x/d} \log t = \sum_{d \le x} \mu(d) S(x/d) \end{aligned} \tag{8}$$

where $S(x) = \sum_{n \le x} \log n$. Then

$$S(x) = \sum_{n \le x} \log n = \sum_{n \le x} \Big(\sum_{d|n} \Lambda(d) \Big) = \sum_{\substack{d,t \\ d \cdot t = n \le x}} \Lambda(d)$$

$$= \sum_{d \le x} \Lambda(d) \sum_{t \le x/d} 1 = \sum_{d \le x} \Lambda(d)[x/d] \; .$$

The advantage here is that we expressed our sum over primes by $S(x)$ which involves only integers that can be estimated quite easily.

Lemma 2 *(Stirling's formula) For all $x \ge 1$ there is a real number δ, $|\delta| \le 1$, such that*

$$S(x) = x \log x - x + \delta \log ex \; .$$

Proof. Let $N = [x]$. Since $\log t$ is increasing, we have

$$\log n = \int_{n-1}^{n} \log n \, dt \ge \int_{n-1}^{n} \log t \, dt \; .$$

Hence,

$$S(x) = \sum_{n=2}^{N} \log n \ge \sum_{n=2}^{N} \int_{n-1}^{n} \log t \, dt = \int_{1}^{N} \log t \, dt$$

$$= \int_{1}^{x} \log t \, dt - \int_{N}^{x} \log t \, dt$$

$$> x \log x - x - \int_{N}^{x} \log t \, dt$$

and

$$\int_{N}^{x} \log t \, dt \le (x - N) \log x \le \log x \; .$$

On the other hand since $\log n \leq \log t$ for $n \leq t \leq n+1$, hence

$$\begin{aligned} S(x) &= \log N + \sum_{n=1}^{N-1} \log n \\ &\leq \log N + \sum_{n=1}^{N-1} \int_n^{n+1} \log t \, dt \\ &= \log N + \int_1^N \log t \, dt \leq \log x + \int_1^x \log t \, dt \\ &= \log x + x \log x - x + 1 = x \log x - x + \log ex \, . \end{aligned}$$

Theorem 2 *For $x \geq 2$*

$$\psi(x) \asymp x \, . \tag{9}$$

Moreover for $x \geq 2$

$$\psi(x) \leq 1.1056 \, x$$

and

$$\psi(x) \geq 0.9212 \, x \, .$$

Note that (9) states that there exist positive constants c_1 and c_2 such that for all $x \geq 2$

$$c_1 x \leq \psi(x) \leq c_2 x \, . \tag{10}$$

We will see that we can choose $c_1 = 0.9212$ and $c_2 = 1.1056$ for sufficiently large x. We will prove that c_1 can be chosen to be $< (7/15)\log 2 + (3/10)\log 3 + (1/6)\log 5 = 0.921292\ldots$ and c_2 can be chosen to be $> 1.2 * \big((7/15)\log 2 + (3/10)\log 3 + (1/6)\log 5\big) = 1.10555042\ldots$

Proof. When substituting Lemma 2 into (8) we obtain

$$\begin{aligned}\psi(x) &= \sum_{d\le x}\mu(d)S(x/d) = \sum_{d\le x}\mu(d)\Big(\frac{x}{d}\log\frac{x}{d}-\frac{x}{d}+O\big(\log\frac{x}{d}\big)\Big)\\ &= (x\log x - x)\sum_{d\le x}\frac{\mu(d)}{d} - x\sum_{d\le x}\frac{\mu(d)}{d}\log d + O\big(\sum_{d\le x}\log x/d\big)\ , \end{aligned}\tag{11}$$

which gets us really in trouble because we have no useful estimates for the sums $\sum_{d\le x}\frac{\mu(d)}{d}$ and $\sum_{d\le x}\frac{\mu(d)}{d}\log d$ that arise in the main term. Also the error term makes a large contribution. Especially troublesome are the large values of d. But that is where the idea to tackle this problem originates: We replace $\mu(d)$ by a function $\nu(d)$ that mimics $\mu(d)$ for small values of d, and is zero for large values of d, say for $d > D$. Let's assume that $\mathcal{D}$ is a finite set of numbers and that $\nu(d) = 0$ when $d \notin \mathcal{D}$.Then for a fixed choice of $\nu(d)$ we have

$$\begin{aligned}\sum_{d\in\mathcal{D}}\nu(d)S(x/d) =&(x\log x - x)\sum_{d\in\mathcal{D}}\frac{\nu(d)}{d} - x\sum_{d\in\mathcal{D}}\frac{\nu(d)\log d}{d}\\ &+\delta\Big(\sum_{d\in\mathcal{D}}|\nu(d)|\Big)\ \log ex\end{aligned}\tag{12}$$

for sufficiently large x. Here the implicit constant depends on the choice of $\nu(d)$. Since we want to prove $\psi(x)\asymp x$, with the notation in Theorem 2, $\nu(d)$ should satisfy

$$\sum_{d\in\mathcal{D}}\frac{\nu(d)}{d}=0 \qquad \text{and} \qquad -\sum_{d\in\mathcal{D}}\frac{\nu(d)}{d}\log d\approx 1\ .$$

On the other hand with Lemma 1

$$\begin{aligned}\sum_{d\in\mathcal{D}}\nu(d)S(x/d) &= \sum_{dm\le x}\nu(d)\log m = \sum_{dm\le x}\nu(d)\sum_{k|m}\Lambda(k)\\ &=\sum_{k\le x}\Lambda(k)\sum_{\substack{d,l\\ dl\le x/k}}\nu(d) = \sum_{k\le x}\Lambda(k)N(x/k)\end{aligned}\tag{13}$$

where

$$N(y) = \sum_{\substack{dl \le y \\ d \in \mathcal{D}}} \nu(d) = \sum_{\substack{d \le y \\ d \in \mathcal{D}}} \nu(d)[y/d]\,, \tag{14}$$

and the sum in (13) will be close to $\psi(x)$ if $N(y)$ is near 1. Note that for $y \ge 1$

$$\sum_{d} \mu(d)[y/d] = \sum_{dk \le y} \mu(d) = \sum_{m \le y} \Big(\sum_{d|m} \mu(d) \Big) = 1\,.$$

Therefore $N(y)$ will be close to 1 for small values of y and $\nu(d) \approx \mu(d)$ for small values of d. Let $\{t\}$ be the fractional part of t, then $[t] = t - \{t\}$ and with (14)

$$N(y) = y \sum_{d \in \mathcal{D}} \frac{\nu(d)}{d} - \sum_{d \in \mathcal{D}} \nu(d)\{y/d\} = - \sum_{d \in \mathcal{D}} \nu(d)\{y/d\} \tag{15}$$

since $\sum_{d \in \mathcal{D}} \nu(d)/d = 0$. The fact that $\{t\}$ has period 1, implies that $\{y/d\}$ has period d, hence $N(y)$ has period D, where D divides the least common multiple of all integers d for which $\nu(d) \ne 0$. The simplest choice for $\nu(d)$ is $\nu(1) = 1$, $\nu(2) = -2$, and $\nu(d) = 0$ otherwise. With this choice of $\nu(d)$ we obtain $\sum_{d \in \mathcal{D}} \nu(d)/d = 0$, $-\sum_{d \in \mathcal{D}} \frac{\nu(d)}{d} \log d = \log 2$, and (15) shows that $N(y) = 2\{y/2\} - \{y\}$ has period 2, $N(y) = 0$ for $0 \le y < 1$, and $N(y) = 2(y/2) - (y-1) = 1$ for $1 \le y < 2$. This implies that

$$\sum_{k \le x} \Lambda(k) N(x/k) \le \sum_{k \le x} \Lambda(k) \cdot 1 = \psi(x)$$

Since on the other hand for $x \ge 2$ and $|\delta| \le 1$,

$$\sum_{d \in \mathcal{D}} \nu(d) S(x/d) = x \log 2 + 3\delta \log ex\,,$$

we obtain the lower bound $\psi(x) \geq x\log 2 - 3\log ex$ for $x \geq 2$. For the upper bound we note that $N(y) \geq 0$ for all y and $N(y) = 1$ for $1 \leq y < 2$, therefore

$$\begin{aligned}\psi(x) - \psi(x/2) = \sum_{x/2<k\leq x} \Lambda(n) \cdot 1 \leq \sum_{k\leq x} \Lambda(k) N(x/k) \\ \leq x\log 2 + 3\delta \log ex \leq x\log 2 + 3\log ex\end{aligned}$$

for $x \geq 2$. By applying this inequality to intervals of the form $(x/2^{r+1}, x/2^r]$, for $r = 0, \ldots, R$ and $R = [\log x/\log 2]$ and summing it follows that

$$\psi(x) \leq x\log 2 \sum_{r=0}^{R} (1/2)^r + 5(\log ex)^2 < (2\log 2)x + 5(\log ex)^2\,,$$

for $x \geq 1$. Hence, (10) holds if $c_2 > 2\log 2$ for all sufficiently large x.

If we take $\nu(1) = 1$, $\nu(2) = -1$, $\nu(3) = -1$, $\nu(5) = -1$, $\nu(30) = 1$, and $\nu(d) = 0$ otherwise, then $\sum_{d\in\mathcal{D}} \nu(d)/d = 0$ and

$$\begin{aligned}c_1' = -\sum_{d\in\mathcal{D}} \frac{\nu(d)}{d} \log d = \log \frac{2^{1/2}3^{1/3}5^{1/5}}{30^{1/30}} \\ =(7/15)\log 2 + (3/10)\log 3 + (1/6)\log 5 = 0.92129\ldots\,. \qquad (16)\end{aligned}$$

Let $N(y)$ be defined as in (14), then $N(y) = -\{y\} + \{y/2\} + \{y/3\} + \{y/5\} - \{y/30\}$ which has period 30. $N(y)$ only takes the values 0 and 1. $N(y) = 0$ for $0 < y < 1$, $N(y) = 1$ for $1 \leq y < 6$, $N(y) = 0$ for $6 \leq y < 7$, $N(y) = 1$ for $7 \leq y < 10$, etc. Since $N(y) \leq 1$ for all values of y, it follows that

$$\psi(x) = \sum_{k\leq x} \Lambda(k) \cdot 1 \geq \sum_{k\leq x} \Lambda(k) N(x/k) = c_1' x - 5\log ex$$

for $x \geq 30$. Therefore $\psi(x) \geq c_1 x$ with $c_1 < c_1'$ for sufficiently large x.

For the upper bound note that $N(y) \geq 0$ for all values of y and $N(y) = 1$ for $1 \leq y < 6$. This implies

$$\sum_{r \leq x} \Lambda(r) N(x/r) \geq \sum_{x/6 < r \leq x} \Lambda(r) = \psi(x) - \psi(x/6)$$

Hence,

$$\psi(x) - \psi(x/6) \leq c_1' x + 5 \log(ex)$$

for $x \geq 30$. Calculation by hand shows that this also holds for $1 \leq x \leq 30$. Now let 6^K be the largest power of 6 less or equal to x. Replacing x by $x/6^k$ and summing over $k = 0, \ldots, K$ leads to

$$\begin{aligned} \psi(x) &= \sum_{k=0}^{K} \psi\big(\frac{x}{6^k}\big) - \psi\big(\frac{x}{6^{k+1}}\big) \leq \sum_{k=0}^{K} \big(c_1' \frac{x}{6^k} + 5 \log ex\big) \\ &\leq c_1' x \sum_{k=0}^{\infty} (1/6)^k + 5(\log ex)(K+1) \leq (6/5) c_1' x + 5(\log ex)^2 \end{aligned}$$

since $K = [\log x / \log 6] \leq \log x$. Define $c_2' = (6/5) c_1'$, then the upper bound follows for any $c_2 > c_2'$ for sufficiently large x, which proves the theorem.

Knowing the order of magnitude for $\psi(x)$, we now relate $\psi(x)$ to $\theta(x)$ and then $\theta(x)$ to $\pi(x)$, but we are not keeping close track of the constants anymore arising in the error term.

Theorem 3 *For $x \geq 1$, we have $\theta(x) = \psi(x) + O(\sqrt{x}\,)$.*

From Theorem 2 this immediately implies that $\theta(x) = O(x)$.

Proof. From the definition of both functions it is clear that $\theta(x) \le \psi(x)$ for all x. Hence $\psi(x) - \theta(x) \ge 0$. For an upper bound we have that

$$\begin{aligned}\psi(x) - \theta(x) &= \sum_{\substack{p,k\ge 1\\ p^k\le x}} \log p - \sum_{p\le x} \log p = \sum_{k\ge 2}\sum_{p^k\le x} \log p \\ &= \sum_{k\ge 2}\sum_{p\le \sqrt[k]{x}} \log p = \sum_{k\ge 2} \theta(\sqrt[k]{x}\,)\,.\end{aligned}$$

Note that the sum over k is a finite sum, since $\theta(\sqrt[k]{x}\,) = 0$ when $\sqrt[k]{x} < 2$, that is for $k > \log x/\log 2$. Hence let $K = [\log x/\log 2]$, then by Theorem 2

$$\psi(x) - \theta(x) = \sum_{k=2}^{K} \theta(\sqrt[k]{x}\,) \le \sum_{k=2}^{K} \psi(\sqrt[k]{x}\,) = \sum_{k=2}^{K} O(\sqrt[k]{x}\,)$$

where the implicit constant does not depend on k. Since $x \ge 1$, this is

$$\begin{aligned}O(x^{1/2}) + O\Big(x^{1/3}\sum_{k=3}^{K} x^{-(1/3-1/k)}\Big) &= O(x^{1/2}) + O(x^{1/3}K) \\ &= O(x^{1/2}) + O(x^{1/3}\log x) = O(x^{1/2})\,.\end{aligned}$$

Theorem 4 *For $x \ge 2$, we have*

$$\pi(x) = \frac{\theta(x)}{\log x} + O\Big(\frac{x}{\log^2 x}\Big)\,.$$

Proof. First note that for $x \ge 2$

$$\int_p^x \frac{1}{u}\,\frac{1}{\log^2 u}\,du \;=\; \frac{1}{\log p} - \frac{1}{\log x}\,.$$

Hence

$$\pi(x) - \frac{\theta(x)}{\log(x)} = \sum_{p\le x} \log p\Big(\frac{1}{\log p} - \frac{1}{\log x}\Big) = \sum_{p\le x} \log p \int_p^x \frac{1}{u}\,\frac{1}{\log^2 u}\,du$$
$$= \int_2^x \Big(\sum_{p\le u} \log p\Big)\,\frac{1}{u}\,\frac{1}{\log^2 u}\,du = \int_2^x \frac{\theta(u)}{u}\,\frac{1}{\log^2 u}\,du\,.$$

Using Chebyshef's estimate $\theta(u) = O(u)$ it remains to show that

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

We consider the range $2 \le u \le \sqrt{x}$ first. Here the integrand is $\le 1/\log^2 2$. Therefore this range contributes $O(\sqrt{x})$ to the integral. In the range $\sqrt{x} \le u \le x$ the integrand is $\le 1/\log^2(\sqrt{x}) = 4/\log^2 x$. Therefore the integral over this second range is $O\big(x/\log^2 x\big)$, which proves our theorem.

Proof of Theorem 1. Applying Theorem 3 and Theorem 2 to the above Theorem 4 gives

$$\pi(x) = \frac{\psi(x)}{\log x} + O\Big(\frac{x}{\log^2 x}\Big) \le c_2\,\frac{x}{\log x} + O\Big(\frac{x}{\log^2 x}\Big)$$

which gives the upper bound in Theorem 1 for sufficiently large x. Similarly we obtain the lower bound by using the lower estimate for $\psi(x)$ obtained in Theorem 2.

We will give a short proof of *Bertrand's Postulate*: If $x > 1$, then every interval $(x, 2x)$ contains a prime number. Note that $\log 2x = \log x + O(1)$, hence $1/\log 2x = (1/\log x)\big(1 + O(1/\log x)\big)$. If we apply the lower bound of Theorem 1 to $\pi(2x)$ and the upper bound there to $\pi(x)$ we obtain the following inequality:

$$\pi(2x) - \pi(x) \geq (2c_1 - c_2)\,\frac{x}{\log x} \;+\; O\Big(\frac{x}{\log^2 x}\Big)\,.$$

Recall that c_2 is close to $6/5c_1$, in any case $2c_1 - c_2 > 0$, which implies that $\pi(2x) - \pi(x) > 0$ for sufficiently large x, say $x \geq x_0$. The values $x \leq x_0$ have to be examined directly.

CHAPTER 3

The Riemann zeta-function and an analytic proof of the Prime Number Theorem

3.1 Properties of the Riemann zeta-function

The function that we now call the Riemann zeta-function, which plays a significant role in the proof of the Prime Number Theorem was first introduced by Leonard Euler in the 18th century. He defined the function

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$$

for real values of $s > 1$, where the sum on the right side clearly converges. Using the unique factorization of positive integers, Euler proved that

$$\zeta(s) = \prod_p \left(1 + p^{-s} + p^{-2s} + \cdots\right) = \prod_p \left(1 - p^{-s}\right)^{-1} \tag{17}$$

where the product extends over all primes p.

A few years after Chebyshev had published his paper, Georg Friedrich Bernhard Riemann (1826–1866) in 1859 presented an idea of how to prove the Prime Number Theorem in the article "On the Number of Primes Less Than a Given Magnitude", the only paper that he ever published on Number Theory. Riemann's revolutionary idea was to consider $\zeta(s)$ as a function of a complex variable. He defined

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p \left(1 - p^{-s}\right)^{-1} \tag{18}$$

for every complex number $s = \sigma + it$ with $\sigma > 1$. Here the series on the left is absolutely convergent and therefore $\zeta(s)$ is analytic in the half plane $\sigma > 1$. Also, the product on the right side is absolutely convergent for these values of s. From the convergence of this infinite product it follows immediately that $\zeta(s) \neq 0$ for $\sigma > 1$ since none of the factors of this product vanishes in this region.

Riemann expressed $\pi(x)$ in terms of a complex integral involving $\zeta(s)$. By changing the integration contour, Riemann developed a formula for $\pi(x)$ as an infinite series that involved the zeros of $\zeta(s)$ and whose leading term was $\mathrm{Li}(x)$. But it was not until the end of the 19th century that the missing essential piece was discovered: the theory of entire functions of finite order.

Riemann proved that the ζ-function has an analytic continuation to the complex plane with only one singularity, a simple pole at $s = 1$ with residue 1. By simple calculation for $\sigma > 1$ we see that

$$\begin{aligned}\zeta(s) &= \sum_{n=1}^{\infty} n^{-s} = \sum_{n=1}^{\infty} s \int_n^{\infty} \frac{dt}{t^{s+1}} = s \int_1^{\infty} \Big(\sum_{n \le t} 1\Big) \frac{dt}{t^{s+1}} \\ &= s \int_1^{\infty} \frac{[t]}{t^{s+1}}\, dt = \frac{s}{s-1} - s \int_1^{\infty} \frac{\{t\}}{t^{s+1}}\, dt \end{aligned} \tag{19}$$

where $[t]$ denotes the integer part of the positive number t and $\{t\}$ the fractional part of t. Since $0 \le \{t\} < 1$ the integral in (19) converges for $\sigma > 0$. Hence (19) defines an analytic continuation of $\zeta(s)$ to the half plane $\sigma > 0$ with the pole $s = 1$ removed.

Riemann also proved that $\zeta(s) = 0$ for all even negative integers, the so-called trivial zeros of ζ. And one of the most important properties that Riemann discovered is the functional equation for the zeta function by showing that $\pi^{-s/2}\Gamma(s/2)\zeta(s)$ will remain unchanged when we replace s by $1 - s$. With the analytic continuation (19) the functional equation provides, by induction, a continuation of $\zeta(s)$ to the whole complex plane.

It was Riemann who noticed which important role the zeros of the Riemann zeta-function play in prime number theory. He conjectured what is considered to be the most famous unsolved problem in mathematics, the Riemann Hypothesis:

All non-real zeros of the Riemann zeta-function have real part 1/2.

The term "Prime Number Theorem" likely originated in the dissertation "Über die Theorie der Hadamardschen Funktionen und ihre Anwendung auf das Problem der Primzahlen" of Hans von Schaper in Göttingen, Germany, in 1898 [20].

3.2 Analytic Proof of the Prime Number Theorem

The Prime Number Theorem was first established independently in 1896 by Jacques Hadamard (1865–1963) and Charles-Jean Étienne Gustave Nicolas de la Vallée Poussin (1866–1962). For both of them it was a big accomplishment at the beginning of their careers. Hadamard was born in 1865 in France and obtained his doctorate in 1892. Hadamard was mainly working in the field of complex function theory, partial differential equations, and differential geometry. He died in 1962, 2 months short of being 98 years old. De la Vallée Poussin was born in 1866 in Louvain, Belgium. He went to college there and joined the faculty at the age of 26, as a Professor of Mathematics. He died in 1962 at the age of 96. [1]

Hadamard's paper begins with defining the Riemann zeta-function for complex values of s with $\mathrm{Re}\,(s) > 1$ by

$$\log \zeta(s) = -\sum_{p} \log(1 - p^{-s}) \tag{20}$$

where p runs over the prime numbers. He remarks that this function is holomorphic in the entire plane except at the point $s = 1$, which is a simple pole, and that this function does not have zeros with $\mathrm{Re}\,(s) > 1$ since the right hand side of (20) is finite. But it allows infinitely many complex zeros with $0 < \mathrm{Re}\,(s) < 1$. Both Hadamard and de la Vallée Poussin proved that $\zeta(s)$ has no zeros on the line $\sigma = 1$. This fact is essential in proving the Prime Number Theorem. Hadamard argued that for $\sigma > 1$ it follows from (20) that

$$\begin{aligned}\log\left|\zeta(s)\right| &= -\mathrm{Re}\sum_p \log(1-p^{-s}) = \sum_p \mathrm{Re}\sum_{m=1}^{\infty}\frac{1}{m}\,p^{-ms}\\ &= \sum_p\sum_{m=1}^{\infty}\frac{1}{m}\,p^{-m\sigma}\cos(mt\log p)\ .\end{aligned}$$

Here higher prime powers can be ignored since that part of the series is uniformly bounded for $\sigma > 1$. He noted, because of the simple pole of $\zeta(s)$ at $s = 1$,

$$\sum_p p^{-\sigma} \sim \log\zeta(\sigma) \sim \log\frac{1}{\sigma-1} \tag{21}$$

for $\sigma \to 1+$. He also noted that if $1 + it_0$ is a necessarily simple zero, then it would follow that

$$\sum_p p^{-\sigma}\cos(t_0\log p) \sim -\log\frac{1}{\sigma-1} \tag{22}$$

for $\sigma \to 1+$. Comparing (21) and (22), Hadamard concluded that $\cos(t_0 \log p) = -1$ for most primes p, hence $\cos(2t_0 \log p) = +1$ for most primes p. Hence $1 + 2it_0$ would be a pole of $\zeta(s)$ which contradicts the fact that the only pole is $s = 1$.

Today the trigonometric inequality $3 + 4\cos\alpha + \cos 2\alpha \geq 0$ is used to establish

$$\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| > 1$$

for $\sigma > 1$, from which it follows that $\zeta(1 + it_0) \neq 0$.

Hadamard further used the fact that for $\sigma > 1$

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

and established the smoothed Mellin inversion formula for the Dirichlet series $F(s) = \sum_n a_n n^{-s}$ that is absolutely convergent for $\sigma > 1$

$$\sum_{n<x} a_n \log(x/n) \; = \; \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \frac{x^s}{s^2} F(s)\, ds$$

for $x > 0$. With the Weierstrass-Hadamard product representation for $(s-1)\zeta(s)$, the convergence of $\sum_\rho |\rho|^{-2}$, where ρ runs over the non-trivial zeros of $\zeta(s)$, and deformation and estimation of the Mellin integral, Hadamard proved

$$\sum_{n<x} \Lambda(n) \log(x/n) \; \sim \; x$$

from which the Prime Number Theorem follows immediately. In 1899 de la Vallée Poussin published the following error term in the Prime Number Theorem

$$\pi(x) = \mathrm{Li}(x) + O\big(x \exp(-c\sqrt{\log x}\,)\big) \tag{23}$$

for a positive constant c by proving that there exists a positive constant a such that $\zeta(s) \neq 0$ for $\sigma \geq 1 - \frac{1}{a \log t}$, $t \geq t_0$.

Later, in 1909 Edmund Georg Hermann Landau (1877–1938) in his book *Handbuch der Lehre von der Verteilung der Primzahlen* [12] gave a simplified proof of the Prime Number Theorem without using Hadamard's theory of entire functions and without using the functional equation of the ζ-function. He proved

Theorem 5 (Landau) *For $x \geq 2$*

$$\psi(x) = x + O\Big(xe^{-(\log x)^{1/13}}\Big) .$$

Proof. He first proved that $|\zeta(s)| > \frac{1}{c\log^7|t|}$ and $|(\zeta'/\zeta)(s)| < c\log^9|t|$ for $|t| \geq 3$ and $\sigma \geq 1 - \frac{1}{c\log^9|t|}$ with a suitable positive constant c. In particular $\zeta(s) \neq 0$ in this region.

Then he established that

$$\sum_{n\leq x} \Lambda(n)\log(x/n) \;=\; \frac{-1}{2\pi i}\int_{2-ix^2}^{2+ix^2} \frac{x^s}{s^2}\,\frac{\zeta'}{\zeta}(s)\,ds \tag{24}$$

and for $x \geq 2$ we extend this to a contour integral along the path $ABCDEFA$ where $A = 2 - ix^2$, $B = 2 + ix^2$, $C = 1 - \frac{1}{c\log^9(x^2)} + x^2 i$, $D = 1 - \frac{1}{c\log^9 3} + 3i$, $E = 1 - \frac{1}{c\log^9 3} - 3i$, and $F = 1 - \frac{1}{c\log^9(x^2)} - x^2 i$ and apply Cauchy's theorem to the integral along this contour. Since the integrand is holomorphic except for the simple pole at $s = 1$ with residue x, the main term is x, and the other integrals Landau estimated to

$$\int_{ED} = O\Big(x^{1-\frac{1}{c\log^9 3}}\Big) \tag{25}$$

$$\int_{CB} + \int_{AF} = O\Big(\int_{1-\frac{1}{c\log^9(x^2)}}^{2} \frac{x^\sigma}{x^4}\,\log^9(x^2)\,d\sigma\Big) = O\Big(\frac{\log^9 x}{x^2}\Big) \tag{26}$$

$$\begin{aligned}\int_{DC} + \int_{FE} &= O\Big(\int_3^{x^2} \frac{x^{1-\frac{1}{c\log^9 t}}}{t^2}\log^9 t\,dt\Big) = O\Big(x\log^9 x\int_3^{x^2}\frac{x^{-\frac{1}{c\log^9 t}}}{t^2}\,dt\Big)\\ &= O\Big(x\log^9 x\,\Big(\int_3^{e^{(\log x)^{1/10}}} + \int_{e^{(\log x)^{1/10}}}^{x^2}\Big)\frac{x^{-\frac{1}{c\log^9 t}}}{t^2}\,dt\Big)\end{aligned} \tag{27}$$

for sufficiently large x. In (27) the first integral can be estimated to

$$\int_3^{e^{(\log x)^{1/10}}} \frac{x^{-\frac{1}{c\log^9 t}}}{t^2}\,dt \le x^{-\frac{1}{c(\log x)^{9/10}}} \int_3^\infty \frac{1}{t^2}\,dt = O\Big(e^{-\frac{1}{c}(\log x)^{1/10}}\Big)$$

and the second integral to

$$\int_{e^{(\log x)^{1/10}}}^{x^2} \frac{x^{-\frac{1}{c\log^9 t}}}{t^2}\,dt = O\Big(\int_{e^{(\log x)^{1/10}}}^\infty \frac{1}{t^2}\,dt\Big) = O\Big(e^{-(\log x)^{1/10}}\Big)\,.$$

This leads to an upper estimate

$$\int_{CD} + \int_{FE} = O\Big(x(\log x)^9 e^{-\frac{1}{c}(\log x)^{1/10}}\Big) = O\Big(xe^{-(\log x)^{1/11}}\Big)\,. \tag{28}$$

With (25), (26), and (28) we obtain that

$$\frac{-1}{2\pi i}\int_{2-ix^2}^{2+ix^2} \frac{x^s}{s^2}\,\frac{\zeta'}{\zeta}(s)\,ds \;= x + O\Big(xe^{-(\log x)^{1/11}}\Big)\,.$$

Hence

$$\sum_{n\le x}\Lambda(n)\log(x/n) \;=\; x + O\Big(xe^{-(\log x)^{1/11}}\Big)\,. \tag{29}$$

Let $\delta = \delta(x) = e^{-(\log x)^{1/12}}$, then δ is a decreasing function, $0 < \delta < 1$. We apply (29) with $x + \delta x$ and subtract (29). This gives

$$\sum_{n\le x+\delta x}\Lambda(n)\log(\tfrac{x+\delta x}{n}) - \sum_{n\le x}\Lambda(n)\log(\tfrac{x}{n}) \;=\; \delta x + O\Big(xe^{-(\log x)^{1/11}}\Big)\,.$$

Here the left hand side simplifies to

$$
\begin{aligned}
&\sum_{n\le x}\Lambda(n)\log(\tfrac{(1+\delta)x}{n}) - \sum_{n\le x}\Lambda(n)\log(\tfrac{x}{n}) + \sum_{x<n\le x+\delta x}\Lambda(n)\log(\tfrac{(1+\delta)x}{n}) \\
&= \log(1+\delta)\sum_{n\le x}\Lambda(n) + O\Big(\sum_{x<n\le x+\delta x}\Lambda(n)\log(\tfrac{(1+\delta)x}{x})\Big) \\
&= \log(1+\delta)\psi(x) + O\Big(\log(1+\delta)\sum_{x<n\le x+\delta x}\log n\Big) \\
&= \log(1+\delta)\psi(x) + O(\delta(\delta x)\log x) \;=\; \log(1+\delta)\psi(x) + O(x\delta^2\log x)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\psi(x) &= \frac{\delta}{\log(1+\delta)}\,x + O(\delta x\log x) \\
&= \big(1+O(\delta)\big)x + O(\delta x\log x) \\
&= x + O\Big(xe^{-(\log x)^{1/13}}\Big)\,.
\end{aligned}
$$

Later, Landau proved that $\zeta(s)\neq 0$ for $t\geq 1-\frac{b}{\log t}$, $t\geq t_0$. With a similar contour as above he proved that

$$
\sum_{n\le x}\Lambda(n)\log(x/n) \;=\; x + O\Big(xe^{-c(\log x)^{1/2}}\Big)\,.
$$

Using $\delta=\delta(x)=e^{-\frac{c}{2}(\log x)^{1/2}}$, similar arguments as above imply

Theorem 6 (Landau) *For $x \geq 2$ and suitable positive constants α*

$$\psi(x) = x + O\Big(xe^{-\alpha\sqrt{\log x}}\Big) \ ,$$
$$\pi(x) = \mathrm{Li}(x) + O\Big(xe^{-\alpha\sqrt{\log x}}\Big) \ .$$

We see that the absence of zeros on the line $\sigma = 1$ already provides a very good asymptotic estimate for $\psi(x)$ and $\pi(x)$. It follows from the functional equation that all the zeros of $\zeta(s)$ lie in the critical strip, a region of the complex plane consisting of the complex numbers s with $0 < \mathrm{Re}\,(s) < 1$. The complex numbers with $\mathrm{Re}\,(s) = 1/2$ form a line in the center of the critical strip, known as the critical line. Riemann in his 1859 memoir suggested that all non-trivial zeros of $\zeta(s)$ are on the critical line, known as the Riemann Hypothesis. Assuming the Riemann Hypothesis one can show that

$$\psi(x) = x + O\big(x^{1/2}\log^2 x\big) \ ,$$
$$\pi(x) = \mathrm{Li}(x) + O\big(x^{1/2}\log x\big) \ .$$

In a 1997 email message that was forwarded around the world, Enrico Bombieri announced that a young physicist, inspired by a lecture on the topic at the Institute for Advanced Study, had proved the Riemann Hypothesis. This sensational message fooled many people who didn't notice that it had been sent on April 1.

There has been a lot of research done to prove the Riemann Hypothesis. Nevertheless, it is still an open conjecture. Numerical data supports the Riemann Hypothesis. Godfrey Harold Hardy (1877–1947) proved in 1914 that an infinite number of zeros lie on the critical line [9]. J. Brian Conrey proved that at least 40% of all non-trivial zeros are simple and lie on $\sigma = 1/2$ [3]. Xavier Gourdon [8] proved in 2004 that the first ten trillion zeros all lie on the critical line.

CHAPTER 4

Proof of the Prime Number Theorem using Fourier Analysis

The Prime Number Theorem was first established using the fact that $\zeta(s) \neq 0$ for $\sigma = 1$ and certain very weak growth conditions for $\zeta(s)$ for $\sigma > 1$, $|t| \to \infty$. Also the Prime Number Theorem implies that $\zeta(s) \neq 0$ for $\sigma = 1$. In 1930 Norbert Wiener (1894–1964) proved the Prime Number Theorem using Fourier Analysis. Wiener created an approximate integral formula for $\pi(x)$ involving a compactly supported smoothing function. This proves the equivalence of the Prime Number Theorem and $\zeta(s) \neq 0$ for $\sigma = 1$. The following tauberian theorem provides one of the most direct proofs known for the Prime Number Theorem:

Theorem 7 (Wiener-Ikehara) *Suppose that f is a non-decreasing, real valued function on $[1, \infty)$ such that*

$$\int_1^\infty |f(x)|\, x^{-\sigma-1}\, dx$$

converges for every $\sigma > 1$ and that for $\mathrm{Re}\,(s) = \sigma > 1$

$$\int_1^\infty f(x)\, x^{-s-1}\, dx = \frac{\alpha}{s-1} + g(s)$$

where α is a real number and g is the restriction to $\sigma > 1$ of a continuous function on the closed half plane $\sigma \geq 1$. Then

$$\lim_{x\to\infty} \frac{f(x)}{x} \;=\; \alpha\,.$$

CHAPTER 5

Elementary Proofs of the Prime Number Theorem

The fact that the Chebyshev approach did not produce the Prime Number Theorem and the analytic proofs using properties of the Riemann zeta-function led to the opinion (Hardy and others) that the Prime Number Theorem could only be proved using the Riemann zeta-function. It came as a great surprise when in 1949 Paul Erdős and Atle Selberg independently announced that they had found an elementary proof of the Prime Number Theorem. Elementary in this context does not mean "trivial"; it means that no methods from the theory of complex-valued functions nor from Fourier Analysis are used. These proofs were considered to be so ingenious and subtle that Selberg in 1950 was awarded the Fields medal and later Erdős in 1952 received the Cole prize. The essential tool here is the asymptotic formula

Theorem 8 (Selberg's Asymptotic Formulae)

$$\psi(x)\log x + \sum_{n\le x}\Lambda(n)\psi(x/n) = 2x\log x + O(x) \tag{30}$$

$$\sum_{n\le x}\Lambda(n)\log n + \sum_{m\ell\le x}\Lambda(m)\Lambda(\ell) = 2x\log x + O(x) \tag{31}$$

$$\theta(x)\log x + \sum_{p\le x}\log p\,\theta(x/p) = 2x\log x + O(x) \tag{32}$$

$$\sum_{p\le x}(\log p)^2 + \sum_{pq\le x}\log p\log q = 2x\log x + O(x)\,. \tag{33}$$

Proof. The statements above are all equivalent. We will only prove this for the first two asymptotic formulae. To see that 30 and 31 are equivalent note that

$$\sum_{n\le x}\Lambda(n)\psi(x/n)=\sum_{n\le x}\Lambda(n)\sum_{m\le x/n}\Lambda(m)=\sum_{mn\le x}\Lambda(m)\Lambda(n) \tag{34}$$

and with Chebyshev's estimate $\psi(x)=O(x)$

$$\sum_{n\le x}\Lambda(n)\log n=\int_{2^-}^{x}\log t\,d\psi(t)=\psi(x)\log x-\int_2^x\frac{\psi(t)}{t}dt=\psi(x)\log x+O(x). \tag{35}$$

This identity is an extension of Chebyshev's method that was based on the relation

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

Therefore with the Möbius inversion formula

$$\Lambda(n)=\sum_{d|n}\mu(d)\log{(n/d)}=-\sum_{d|n}\mu(d)\log d\;.$$

We will mimic this for a function $\Lambda_2(n)$ instead of $\Lambda(n)$.

Selberg's idea can be described as finding a new function, $\Lambda_2(n)$ such that

$$\sum_{d|n}\Lambda_2(d)=(\log n)^2,$$

and therefore again with the Möbius inversion formula

$$\Lambda_2(n)=\sum_{d|n}\mu(d)\log^2\tfrac{n}{d}\;.$$

Using these identities, we see that

$$
\begin{aligned}
\sum_{m\ell=n} \Lambda(m)\Lambda(\ell) &= -\sum_{m\ell=n} \Lambda(m) \sum_{d|\ell} \mu(d) \log d \\
&= -\sum_{d|n} \mu(d) \log d \sum_{m|(n/d)} \Lambda(m) \\
&= -\sum_{d|n} \mu(d) \log d \, \log (n/d) \ .
\end{aligned}
$$

Adding to both sides

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

gives

$$
\Lambda_2(d) = \sum_{m\ell=n} \Lambda(m)\Lambda(\ell) + \Lambda(n) \log n \ = \ \sum_{d|n} \mu(d) \log^2 (n/d) \ , \tag{36}
$$

which proves that $\Lambda_2(n) = \sum_{m\ell=n} \Lambda(m)\Lambda(\ell) + \Lambda(n) \log n$ is the Möbius transform of $\log^2 n$. Summation over n gives exactly the left hand side of (31). Therefore, in order to prove (31), we need to show that

$$
F(x) = \sum_{n\leq x} \Lambda_2(n) = \sum_{n\leq x} \sum_{d|n} \mu(d) \log^2 (n/d) = 2x \log x + O(x) \ . \tag{37}
$$

We first determine

$$
G(x) = \sum_{n\leq x} \sum_{d|n} \mu(d) \log^2 (x/d) \tag{38}
$$

by evaluating $G(x) - \gamma^2$, where γ denotes Euler's constant. In order to do that we need a few asymptotic formulas given in the following Lemmas:

Lemma 3 (Mertens, 1874) *For $x \geq 1$ we have*

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

Proof. With Stirling's formula in Lemma 2 and the Chebyshev's estimate in Theorem 2, we see that

$$\begin{aligned} x\log x + O(x) \;&=\; \sum_{n\leq x} \log n \;=\; \sum_{n\leq x}\sum_{d|n} \Lambda(d) \;=\; \sum_{d\leq x} \Lambda(d) \sum_{m\leq x/d} 1 \\ &=\; \sum_{d\leq x} \Lambda(d)[x/d] \;=\; \sum_{d\leq x} \Lambda(d)\big(x/d + O(1)\big) \\ &=\; x\sum_{d\leq x} \frac{\Lambda(d)}{d} \;+\; O\big(\psi(x)\big) \;=\; x\sum_{d\leq x} \frac{\Lambda(d)}{d} \;+\; O(x)\;, \end{aligned}$$

which proves our lemma.

Lemma 4 *For $x \geq 1$ we have*

$$\sum_{n\leq x} \frac{1}{n} \;=\; \log x + \gamma + O(1/x)$$

where $\gamma = 0.57721\ldots$ denotes Euler's constant.

Lemma 5 *Let $x \geq 1$. Then for any fixed positive α*

$$\sum_{n\leq x} \log^{\alpha}(x/n) \;=\; O(x)$$

where the implicit constant depends on α.

Proofs of these lemmas can be found in Hardy, Wright [10], p. 347.

With these lemmas in place we can evaluate $G(x)-\gamma^2$ in the following way:

$$
\begin{aligned}
G(x)-\gamma^2 &= \sum_{dm\le x}\mu(d)\log^2(x/d)-\gamma^2\\
&= \sum_{dm\le x}\mu(d)\Big(\log^2(x/d)-\gamma^2\Big)\\
&= \sum_{d\le x}\mu(d)\Big(\log^2(x/d)-\gamma^2\Big)\Big(\sum_{m\le x/d}1\Big)\\
&= \sum_{d\le x}\mu(d)[x/d]\Big(\log^2(x/d)-\gamma^2\Big)\\
&= \sum_{d\le x}\mu(d)\big(x/d+O(1)\big)\Big(\log^2(x/d)-\gamma^2\Big)\\
&= x\sum_{d\le x}\frac{\mu(d)}{d}\big(\log(x/d)-\gamma\big)\big(\log(x/d)+\gamma\big)+O\Big(\sum_{d\le x}\log^2(x/d)\Big)
\end{aligned}
$$

which with the above lemmas gives

$$
\begin{aligned}
G(x)-\gamma^2 &= x\sum_{d\le x}\frac{\mu(d)}{d}\big(\log(x/d)-\gamma\big)\Big(\sum_{m\le(x/d)}\frac{1}{m}+O\big(\frac{d}{x}\big)\Big)+O(x)\\
&= x\sum_{md\le x}\frac{\mu(d)}{dm}\big(\log x-\gamma-\log d\big)\ +\ O\Big(\sum_{d\le x}\log(x/d)\Big)\ +\ O(x)\\
&= x\big(\log x-\gamma\big)\sum_{n\le x}\frac{1}{n}\sum_{d|n}\mu(d)\ -\ x\sum_{n\le x}\frac{1}{n}\sum_{d|n}\mu(d)\log d\ +\ O(x)\\
&= x\big(\log x-\gamma\big)\ +x\sum_{n\le x}\frac{\Lambda(n)}{n}\ +\ O(x)\\
&= x\log x\ +x\log x\ +\ O(x)=2x\log x\ +\ O(x)\ .
\end{aligned}
$$

Therefore

$$G(x)\ =\ 2x\log x\ +\ O(x)\ .$$

Since

$$
\begin{aligned}
G(x) - F(x) &= \sum_{md \le x} \mu(d)\Big(\log^2 (x/d) - \log^2 m \Big) \\
&= \sum_{n \le x} \log (x/n) \sum_{d|n} \mu(d) \big(\log x + \log n - 2 \log d \big) \\
&= \sum_{n \le x} \log (x/n) \Big(\big(\log x + \log n \big) \Big(\sum_{d|n} \mu(d) \Big) - 2 \sum_{d|n} \mu(d) \log d \Big) \\
&= \log^2 x + 2 \sum_{n \le x} \log (x/n) \Lambda(n) \\
&= \log^2 x + 2\psi(x) \log x - \sum_{n \le x} \Lambda(n) \log n = O(x)
\end{aligned}
$$

where we used (35), we finally have that $F(x) = 2x \log x + O(x)$, which proves Selberg's asymptotic formula.

Let $R(x) = \psi(x) - x$, and $R(x) = 0$ for $1 \le x < 2$. Then (30) and Merten's lemma imply

$$
R(x) \log x + \sum_{n \le x} \Lambda(n) R(x/n) = O(x) \; . \tag{39}
$$

In order to prove the Prime Number Theorem, we need to show that

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

In other words, we need to prove that $R(x) = o(x)$. There is no simple argument to prove this because $R(x)$ is oscillating and the weights $\Lambda(n)$ depend on the location of the prime numbers, which we are interested in investigating. If we replace x in (39) by x/n we have

$$R(x/n)\log(x/n) + \sum_{m\le x/n} \Lambda(m)R(x/mn) \;=\; O(x/n)\;.$$

Therefore

$$\begin{aligned}\log x\Big(R(x)\log x + \sum_{n\le x}\Lambda(n)R(\tfrac{x}{n})\Big) - \sum_{n\le x}\Lambda(n)\Big(R(\tfrac{x}{n})\log(\tfrac{x}{n}) + \sum_{m\le x/n}\Lambda(m)R(\tfrac{x}{mn})\Big)\\ \;=\; O(x\log x) + O\Big(x\sum_{n\le x}\frac{\Lambda(n)}{n}\Big) \;=\; O(x\log x)\;.\end{aligned}$$

Hence

$$R(x)\log^2 x = -\sum_{n\le x}\Lambda(n)R(\tfrac{x}{n})\log n + \sum_{m\ell\le x}\Lambda(m)\Lambda(\ell)R(\tfrac{x}{m\ell}) + O(x\log x)\,,$$

which implies that

$$\begin{aligned}|R(x)|\log^2 x &\le \sum_{n\le x}\Lambda(n)\log n\,\big|R(\tfrac{x}{n})\big| + \sum_{m\ell\le x}\Lambda(m)\Lambda(\ell)\,\big|R(\tfrac{x}{m\ell})\big| + O(x\log x)\\ &\le \sum_{n\le x}\Big(\Lambda(n)\log n + \sum_{m\ell=n}\Lambda(m)\Lambda(\ell)\Big)\,\big|R(\tfrac{x}{n})\big| + O(x\log x)\\ &= \sum_{n\le x}\Lambda_2(n)\,\big|R(\tfrac{x}{n})\big| + O(x\log x), \end{aligned}\tag{40}$$

and

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

Replacing the sum in (40) by an integral gives

$$\sum_{n\le x}\Lambda_2(n)\,\big|R(\tfrac{x}{n})\big| = 2\int_1^x \big|R(\tfrac{x}{t})\big|\log t\,dt \;+\; O(x\log x)\,.$$

Hence

$$|R(x)|\log^2 x \le 2\int_1^x \left|R(\tfrac{x}{t})\right|\log t\,dt \;+\; O(x\log x)\,. \tag{41}$$

Now let $x = e^\xi$, $t = xe^{-\eta}$, and put $V(\sigma) = R(e^\sigma)e^{-\sigma}$. Then the integral in (41) becomes

$$\begin{aligned}\int_1^x \left|R(\tfrac{x}{t})\right|\log t\,dt &= e^\xi\int_0^\xi |R(e^\eta)|e^{-\eta}(\xi-\eta)\,d\eta\\ &= e^\xi\int_0^\xi |V(\eta)|(\xi-\eta)\,d\eta = e^\xi\int_0^\xi |V(\eta)|\int_\eta^\xi d\sigma\,d\eta\\ &= e^\xi\int_0^\xi d\sigma\int_0^\sigma |V(\eta)|\,d\eta\end{aligned}$$

and therefore (41) implies

$$|V(\xi)|\,\xi^2 \le 2\int_0^\xi\int_0^\sigma |V(\eta)|\,d\eta\,d\sigma \;+\; O(\xi)\;. \tag{42}$$

The function $V(\xi)$ is bounded since $\psi(x) = O(x)$. Hence

$$\alpha = \limsup_{\xi\to\infty}|V(\xi)|\;, \qquad\qquad \beta = \limsup_{\xi\to\infty}\frac{1}{\xi}\int_0^\xi |V(\eta)|\,d\eta\;.$$

Then

$$|V(\xi)| \le \alpha + o(1) \qquad\text{and}\qquad \int_0^\xi |V(\eta)|\,d\eta \le \beta\xi + o(\xi)\,.$$

If we use this estimate in (42) it follows that

$$|V(\xi)|\,\xi^2 \le 2\int_0^\xi \big(\beta\sigma + o(\sigma)\big)\,d\sigma + O(\xi) = \beta\xi^2 + o(\xi^2)$$

and therefore

$$\left|V(\xi)\right| \leq \beta + o(1)$$

which implies that

$$\alpha \leq \beta\,. \tag{43}$$

The Prime Number Theorem is equivalent with $V(\xi) \to 0$ as $\xi \to \infty$, which implies $\alpha = 0$. Assume that $\alpha > 0$. Then we will see that $\alpha > \beta$, which is a contradiction. Therefore $\alpha = 0$ and the Prime Number Theorem is proven. In order to prove $\beta < \alpha$ we need the following two lemmas:

Lemma 6 *There is a positive constant A such that for every positive ξ_1, ξ_2, we have*

$$\left|\int_{\xi_1}^{\xi_2} V(\eta)\,d\eta\right| < A\,.$$

Lemma 7 *If $\eta_0 > 0$ and $V(\eta_0) = 0$ then*

$$\int_0^{\alpha} \left|V(\eta_0+\tau)\right| d\tau \;\leq\; \tfrac{1}{2}\,\alpha^2 + O(1/\eta_0)\,.$$

Let $\delta = \frac{3\alpha^2+4A}{2\alpha} > \alpha$, and let $\sigma > 0$. Consider $V(\eta)$ in the interval $\sigma \leq \eta \leq \sigma + \delta - \alpha$. Since $V(\eta) = e^{-\eta}\psi(e^{\eta}) - 1$ decreases as η increases, except at its discontinuities, where $V(\eta)$ increases. Hence there is an η_0 in $\sigma \leq \eta \leq \sigma + \delta - \alpha$ such that $V(\eta_0) = 0$ or $V(\eta)$ changes sign at most once. In the first case it follows with Lemma 7 that

$$
\begin{aligned}
\int_{\sigma}^{\sigma+\delta} |V(\eta)|\, d\eta \;&=\; \Big(\int_{\sigma}^{\eta_0} + \int_{\eta_0}^{\eta_0+\alpha} \int_{\eta_0+\alpha}^{\sigma+\delta} \Big) |V(\eta)|\, d\eta \\
&\le\; \alpha(\eta_0 - \sigma) + \frac{1}{2}\alpha^2 + \alpha(\sigma + \delta - \eta_0 - \alpha) + o(1) \\
&=\; \alpha\delta(1 - \frac{\alpha}{2\delta}) + o(1) = \alpha_1 \delta + o(1)
\end{aligned}
$$

for large values of σ, where $\alpha_1 = \alpha(1 - \frac{\alpha}{2\delta}) < \alpha$.

In the second case, if $V(\eta)$ changes sign just once at $\eta = \eta_1$ in the interval $\sigma \le \eta \le \sigma + \delta - \alpha$ we have

$$
\int_{\sigma}^{\sigma+\delta-\alpha} |V(\eta)|\, d\eta \;=\; \Big| \int_{\sigma}^{\eta_1} V(\eta)\, d\eta \Big| + \Big| \int_{\eta_1}^{\sigma+\delta-\alpha} V(\eta)\, d\eta \Big| < 2A\,,
$$

by Lemma 6. If $V(\eta)$ does not change sign in the above interval we have

$$
\int_{\sigma}^{\sigma+\delta-\alpha} |V(\eta)|\, d\eta \;=\; \Big| \int_{\sigma}^{\sigma+\delta-\alpha} V(\eta)\, d\eta \Big| < A\,,
$$

by Lemma 6. Hence

$$
\begin{aligned}
\int_{\sigma}^{\sigma+\delta} |V(\eta)|\, d\eta \;&=\; \int_{\sigma}^{\sigma+\delta-\alpha} |V(\eta)|\, d\eta + \int_{\sigma+\delta-\alpha}^{\sigma+\delta} |V(\eta)|\, d\eta \\
&<\; 2A + \alpha^2 + o(1) = \alpha_2 \delta + o(1)\,,
\end{aligned}
$$

where

$$
\alpha_2 = \frac{2A + \alpha^2}{\delta} = \alpha\, \frac{4A + 2\alpha^2}{4A + 3\alpha^2} = \alpha\big(1 - \frac{\alpha}{2\delta}\big) = \alpha_1 .
$$

This gives

$$\int_{\sigma}^{\sigma+\delta} |V(\eta)|\, d\eta \le \alpha_1 \delta + o(1)$$

where $o(1) \to 0$ for $\sigma \to \infty$. Let $K = [\xi/\delta]$, then

$$\begin{aligned}\int_0^{\xi} |V(\eta)|\, d\eta &= \sum_{k=0}^{K-1} \int_{k\delta}^{(k+1)\delta} |V(\eta)|\, d\eta + \int_{K\delta}^{\xi} |V(\eta)|\, d\eta \\ &\le K\big(\alpha_1\delta + o(1)\big) + O(1) = \alpha_1 \xi + o(\xi)\,.\end{aligned}$$

Hence

$$\beta = \limsup_{\xi\to\infty} \frac{1}{\xi} \int_0^{\xi} |V(\eta)|\, d\eta \; = \; \alpha_1 < \alpha$$

which contradicts (43). It follows that $\alpha = 0$, which is equivalent to the Prime Number Theorem.

Erdős' proof of the Prime Number Theorem starts out from Selberg's formula (32). We will give a brief sketch of his proof. He defined

$$A = \limsup_{x\to\infty} \frac{\theta(x)}{x}\,, \qquad\qquad a = \liminf_{x\to\infty} \frac{\theta(x)}{x}\,.$$

He proved that

$$A + a = 2\,.$$

If $A = a$ it follows that $A = a = 1$ and the Prime Number Theorem is proven. Now assume that $A > a$. Then there is a sequence x, $x \to \infty$, such that $\theta(x) \sim Ax$, call this sequence

$\mathcal{X}$. Then with

$$\theta(x)\log x + \sum_{p\le x}\theta(x/p)\ \log p \ = \ 2x\log x + O(x)$$

and the fact that $A+a=2$ it follows that

$$\sum_{p\le x}\theta(x/p)\ \log p \ \sim \ (2-A)x\log x \ = \ ax\log x \tag{44}$$

Erdős used this to prove that there are many primes $p\le x$ such that $\theta(x/p)$ is approximately $a(x/p)$. Let $\mathcal{S}_x$ denote the set of these primes. Under the same assumptions let $p_1=p_1(x)$ denote the smallest prime in the set $\mathcal{S}_x$, then there exists a positive function $\eta(x)\to 0$ as $x\to\infty$, $x\in\mathcal{X}$, such that $p_1(x)<x^{\eta(x)}$, hence $x/p_1(x)\to\infty$ and

$$\theta(x/p_1)\sim a\frac{x}{p_1(x)}\ .$$

Repeating this argument with $\mathcal{X}_1=\{x/p_1(x):\ x\in\mathcal{X}\}$ shows that there are many primes $p\le x/p_1$ such that

$$\theta\big(\frac{x}{pp_1}\big)\sim A\frac{x}{pp_1}\ .$$

Let $\mathcal{T}_x$ denote the set of these primes. Now for $p\in\mathcal{S}_x$, $q\in\mathcal{T}_x$, we have

$$\theta\big(\frac{x}{p}\big)\sim a\,\frac{x}{p} \qquad \text{and} \qquad \theta\big(\frac{x}{qp_1}\big)\sim A\frac{x}{qp_1}\ .$$

Therefore x/p and x/qp_1 cannot be very close. Because otherwise $\theta(x/p)$ and $\theta(x/qp_1)$ would be close, which would imply that ax/p and Ax/qp_1 are close, which forces a and A to be close, contradicting the assumption that $A>a$. Hence for each fixed $p\in\mathcal{S}_x$ there are no $q\in\mathcal{T}_x$ such that x/p and x/qp_1 are close. This forces a thinning out of the primes $\le x$

which conflicts with

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

This completes Erdős' proof of the Prime Number Theorem.

It is puzzling that such a significant contribution to Number Theory was born from a heated dispute between two accomplished mathematicians. Selberg and Erdős not only had no intentions of working together, but were never able to come to an agreement on their respective works in accomplishing an elementary proof of the Prime Number Theorem. Selberg made a comment regarding Erdős' work, [23] "I told him that I did not mind that he try to do what he wanted to do, but I made some remarks that would discourage him." It could have been pride or ego that got in the way, or even something more personal. We look at Ernst Strauss' account of the events, as he was a close friend of Erdős and he worked with both mathematicians. Selberg took a position at Syracuse and began hearing, from several sources, that only Erdős' name was being mentioned in discussions about an elementary proof of the Prime Number Theorem. [23] Understandably this bothered Selberg, so he proceeded to write a letter to Erdős suggesting that they publish separately. Erdős most likely took this personally as, unlike Selberg who preferred to work alone, he enjoyed collaborating with other mathematicians. In fact, Erdős had more coauthors, roughly five hundred by current counts, than any other mathematician in history. [23]

Having different perspectives on how to approach problems in mathematics is another, and possibly the most significant factor in the unsettled dispute between these mathematical giants.

It was explained througout the paper that it took the combined efforts of Atle Selberg and Paul Erdős to arrive at an elementary proof of the Prime Number Theorem, after Erdős was inspired by the proof of Selberg's inequality. So it seems that for a brief time there was a friendly, and by some accounts thrilling collaboration. Then it may have been a lack of communication that spiraled somewhat out of control in terms of not neccesarily the publications, but simply getting the word out about the existence of the elementary proof, which was previously not thought to be possible. With Selberg's permission, Turan presented Selberg's proof of the Dirichlet Theorem on primes in arithmetic progressions, which states that there are infinitely many primes $p \equiv a(\mod q)$ where a and q are coprime, including Selberg's formulae at a seminar in Princeton to Erdős, Strauss, Chowla and a couple of others, while Selberg was in Canada. Erdős was inspired by Selberg's work. Selberg wrote, "Turan had mentioned to Erdős after my return from Montreal he told me he was trying to prove $\frac{p_{n+1}}{p_n} \to 1$ from my formula. Actually, I didn't like that somebody else started working on my unpublished results before I considered myself through with them." [7] Selberg even admitted to an attempt to throw Erdős off track by withholding information regarding a counterexample, but that he did not succeed in doing so. [7] The dispute became increasingly heated when the two mathematicians could not reach an agreement with regards to publication. In another letter, Selberg states "I was rather disgusted with the whole thing. I never lectured on the elementary proof of the Prime Number Theorem after the lecture in Syracuse." [7]

There have since been several other proofs of the Prime Number Theorem, elementary and analytic, but none with such historical significance on Number Theory as the elementary proof by Selberg and Erdős, despite the two mathematicians never being able to reach an agreement. One wonders what else could have been accomplished had they been able to collaborate for a longer period of time, or been able to agree on their respective contributions.

BIBLIOGRAPHY

[1] Bateman, Paul T. and Diamond H. G.; A Hundred Years of Prime Numbers. *Amer. Math. Monthly* **103** (1996), 729–741.

[2] Chebyshev, P. L.; Sur la fonction qui détermine la totalité des nombres premiers inférieurs à une limite donneé. *Mémoires présentés a l'Académie Impériale des Sciences de St. Pétersbourg par divers Savants* **6** (1851), 141–157.

[3] Conrey, J. B.; More than two fifths of the zeros of the Riemann zeta function are on the critical line. *J. für Reine und Angewandte Mathematik* **399** (1989), 1–26.

[4] Dunham, W.; Euler: The Master of Us All. The Mathematical Association of America, 1999.

[5] Erdős, P.; On a New Method in Elementary Number Theory Which Leads to An Elementary Proof of the Prime Number Theorem. *Proc. Natl. Acad. Sci. U.S.A.* **35** (1949), 374–384.

[6] Garrett, P.; Simple Proof of the Prime Number Theorem, etc. rspective. In: Number Theory, Springer Verlag, 2004, pp. 179–192.

[7] Goldfeld, D; The Elementary Proof of the Prime Number Theorem: An Historical Perspective. In: Number Theory, New York Seminar 2003, Springer, New York, 2004.

[8] Gourdon, X and Sebah P.; The $10^(13)$ First Zeros of the Riemann Zeta Function, and Zeros Computation at Very Large Height. http://numbers.computation.free.fr/Constants/Miscellaneous/zeta zeros1e13-1e24.pdf; 2004.

[9] Hardy, G. H.; Sur les zéros de la fonction $\zeta(s)$ de Riemann. *Comptes rendus de l'Académie des sciences (Paris)* **158** (1914), 1012–1014.

[10] Hardy, G. H. and Wright, E. M.; An Introduction to the Theory of Numbers. Oxford Clarendon Press, fourth edition, 1965.

[11] Heath, T.; The Thirteen Books of the Elements. Vol. 2, Books 3–9, Dover Publ., New York, 2nd ed., 1956.

[12] Landau, E.; Handbuch der Lehre von der Verteilung der Primzahlen. Teubner Verlag, Leibzig und Berlin, 1909.

[13] Legendre, A. M.; Essai sur la Théorie des Nombres. 1st ed., Paris, 1798, p. 19.

[14] Legendre, A. M.; Essai sur la Théorie des Nombres. Paris: Duprat, 1808.

[15] http://www.mersenne.org/primes/

[16] MIT OpenCourseWare; Lecture 21 and 22: The Prime Number Theorem. http://ocw.mit.edu 18.112 Functions of a Complex Variable, Fall 2008.

[17] Montgomery, H. and Wagon, S.; A Heuristic for the Prime Number Theorem. *the Mathematical Intelligencer* **28** (2006), 6–9.

[18] Newman, D.; Simple Analytic Proof of the Prime Number Theorem. *Amer. Math. Monthly* **87** (1980), 693–696.

[19] Pollack, P.; Revisiting Gausss analogue of the prime number theorem for polynomials over a finite field. *Finite Fields and Their Applications* **16** (2010), 290–299.

[20] Schaper, H. von; Über die Theorie der Hadamard'schen Funktionen und ihre Anwendung auf das Problem der Primzahlen. *Inaugural Dissertation*, Göttingen, 1898.

[21] Selberg, A.; An Elementary Proof of the Prime Number Theorem. *Ann. of Math.* **50** (1949), 305–313.

[22] Shapiro, H. N.; Introduction to the Theory of Numbers. Dover Publ., 2008.

[23] Spencer, J. and Graham, R.; The Elementary Proof of the Prime Number Theorem. *the Mathematical Intelligencer* **31** (2009), 18–23.

[24] Tschinkel Y.; About the cover: On the distribution of primes–Gauss' tables. *Bull. Amer. Math. Soc.* **43** (2006), 89–91.

[25] Zagier, D.; Newmans Short Proof of the Prime Number Theorem. *Amer. Math. Monthly* **104** (1997), 705–708.

[26] Zagier, D.; The First 50 Million Prime Numbers. *Math. Intelligencer* **1** (1977), 7–19.