On Erdos & Selberg's proof of Prime Number Theorem

Vatsal Limbachia

August 2020

1 Introduction

Prime numbers have been of interest to mathematicians since the dawn of mathematics. They are the building blocks of number theory and have wide range of applications ranging from cryptography to mathematical biology. Euclid, in his famous *Elements*, was the first mathematician to prove that there are infinitely many primes. Prime numbers and specifically its function π defined as

$$\pi(x) = \#\{p : p \leq (x)\}$$

where p is the no. of primes. Before 1800, Gauss and Legendre tried to prove something similar to what we call prime number theorem. It calculates how the prime are distributed among integers. Gauss had devised for large x, $\int_2^x \frac{dt}{logt}$ gives a good approximation for $\pi(x)$, where $\pi(x)$ is number of primes less than or equal to x. Today, we know the prime number theorem as

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

(1)

Euler also worked on the infinitude of primes and their sequences. His results were used in the Dirichlet Series and Landau's Theorems,and using results from complex analysis. Dirichlet also famously proved the infinitude of prime numbers in arithmetic progression. In the 1850s, Chebyshev made progress in prime number theorem by proving that the limit tends to 1 and also proving that $\frac{x}{logx}$ always lies between two positive constants.

In 1896, Hadamard and de Vallee Poussin proved the Prime Number Theorem (PNT) using the tools from complex analysis, like Riemann-ζ function. It was believed by many mathematicians that PNT was intimately connected with complex analysis and a proof by real analysis wouldn't emerge, including Hardy. However, in 1949, Atle Selberg and Paul Erdos, both working independently, gave proofs using only real variables and simple properties of logarithms. In this article, we discuss the elementary proofs of Selberg and Erdos. We begin by describing the Selberg's Symmetric formula

(2)
$$\sum_{p\leq x} log^2 p + \sum_{pq\leq x} logplogq = 2xlogx + \text{O(x)}$$

and a detailed proof due to Erdos. This paper serves as a means of revision of analytic number theory concepts and the following account of proofs and formulas is inspired by [1], [2], [3] and [4].

2 The Selberg Symmetry Formula

We first introduce some properties of asymptotic functions, which will be helpful in learning properties of $\pi(x)$. We introduce some notations which were introduced by Landau.

1) $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$ we call f asymptotic to g and denote as $f(x) \sim g(x)$

2) $\lim_{\to\infty} \frac{f(x)}{g(x)} = 0$ we write $f(x)$=$o(g(x))$ which is read as "little o"

3) If $k \in R$ s.t. $\lim_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| < k$ we write $f(x)$=$O(g(x))$. It's read as "Big O"

With the help of these notations, we can restate the PNT as

$$\pi(x) \sim \frac{x}{logx}$$

Now to begin our journey towards the symmetric formula, we introduce the *2nd Mangoldt function* defined as

$$\Lambda_2(n) = \sum_{d|n} \mu(d) log^2 \left(\frac{d}{n}\right)$$

We claim that $\Lambda_2(n)$ is an approximate indicator for primes and product of two primes. Its proof is available in [4]. We begin by understanding the left hand side of (2) using the following sum:

$$\Lambda_2(n) = \sum_{n\leq x} \sum_{d|n} \mu(d) log^2 \left(\frac{d}{n}\right)$$

(3)

We will show that the right-hand-side of (3) is connected to left-hand-side of (2). To proceed, we introduce the following lemma

Lemma 2: *For positive integer n, let the Mangoldt function $\Lambda(n)$ be defined as*

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

Then we have

$$\sum_{d|n} \mu(d) log^2 \left(\frac{d}{n}\right) = \Lambda(n) logn + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right)$$

Proof: We know from the definition of $\Lambda(n)$ that $\Lambda(n)=$ *logp* when $n = p^k$ for prime p and positive integer k otherwise $\Lambda(n) = 0$. Prime factorizing n, we can find that

$$logn = \sum_{d|n} \mu(d)$$

We can repeatedly apply the equality above, to obtain

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

We obtain the following equality by the equation above using $a, b \in n$ The equality follows by applying Mobius Inversion formula

$$log^2(n) = \sum_{d|n} \left(\Lambda(b) logb + \sum_{d|b} \Lambda(d)\Lambda\left(\frac{b}{d}\right)\right)$$

□

By Lemma 2, the double sum on the right-hand-side of (3) can be written as follows

$$\sum_{n\leq x} \sum_{d|n} \mu(d) log^2 \left(\frac{n}{d}\right) = \sum_{n\leq x} \Lambda(n) logn + \sum_{mn\leq x} \Lambda(m)\Lambda(n)$$

Now, we observe that

$$\sum_{n\leq x} \mu(n) logn = \sum_{p\leq x} log^2 p + O(\sqrt{x} log^2 x)$$

It is due to the prime powers on $\Lambda(n)$ have made negligible contribution to the sum. We also claim that

$$\sum_{mn\leq x}\Lambda(m)\Lambda(n)=\sum_{pq\leq x}logplopq+O(x)$$

Similarly, the right-hand-side of (3) does give rise to the left-hand-side of (2). Now, to prove (2), it suffices to show that the right-hand-side of (3) can be estimated as follows:

$$\sum_{n\leq x}\sum_{d|n}\mu(d)log^2\left(\frac{n}{d}\right)=2xlogx+O(x)$$

To obtain the above equation, we give the following estimates without proof which can be found in [2] and [4].

$$\sum_{n\leq x}\frac{\mu(n)}{n}=O(1)$$

$$\sum_{n\leq x}\frac{\mu(n)}{n}log\left(\frac{x}{n}\right)=O(1)$$

$$\sum_{n\leq x}\frac{\mu(n)}{n}log^2\left(\frac{x}{n}\right)=2xlogx+O(1)$$

We thus obtain our desired equality

$$\sum_{n\leq x}\sum_{d|n}\mu(d)log^2\left(\frac{n}{d}\right)=2xlogx+O(x)$$

This formula is also widely used in many proofs of analytic number theory, in addition to PNT.

3 The Proof of the PNT

We will now present the elementary proof due to Selberg and Erdos. We start by studying the proposition given by Erdos (from [3]) and deduce the PNT with the help of Selberg Symmetry formula. As is observed by Erdos (in [3]), we do not need the full symmetry formula bu rather in this form

$$\sum_{p\leq x}log^2p+\sum_{pq\leq x}logplogq=2xlogx+o(xlogx)$$

Define $\vartheta(x)$ as

$$\vartheta(x) = \sum_{p \leq x} logp$$

and let a and A be defined as

$$a = \lim_{x \to \infty} inf \frac{\vartheta(x)}{x} \text{ and } A = \lim_{x \to \infty} sup \frac{\vartheta(x)}{x}$$

We now prove the proposition of Erdos:

Proposition(Erdos). *For every $c > 0$, there exists $\delta(c) > 0$ such that*

$$\pi(x(1+c)) \text{ - } \pi(x) > \delta(c)\frac{x}{logx}$$

for sufficiently large x

Proof of Proposition (from [3]) : Recall that $\pi(x) \sim \frac{x}{logx} \Leftrightarrow \vartheta(x) \sim x$. It therefore suffices to show that for every $c > 0$, there exists $\delta(c) > 0$ such that

$$\vartheta(x(1+c)) \text{ - } \vartheta(x) > \delta(c)x$$

for sufficiently large x.

We proceed by contradiction. If the proposition is false, then there exists a constant $c > 0$ such that

$$\vartheta(x(1+c)) \text{ - } \vartheta(x) > o(c)x$$

for arbitrarily large values of x. Consider the set S of all constants c, and let $C = sup(S)$. Since $0 < a \leq A < \infty$, we have $C < \infty$. We will show that

$$\vartheta(x(1+c)) \text{ - } \vartheta(x) > o(c)x$$

for arbitrarily large values of x. We require the following lemma:

Lemma 3 : *Whenever $x < y$, we have that $\vartheta(y) - \vartheta(x) = 2(y-x) + o(x)$*

Proof : By the Selberg Symmetry formula, we have that

$$\sum_{x<p \leq y} (logp)^2 \leq 2(y-x)logy + o(ylogy)$$

First suppose that $x \geq \frac{y}{log^2 y}$. In this case, we have that $logx = (1+o(1))logy$, so for all primes p such that $x < p \leq y$, we have $logp = (1+o(1))logy$.

Dividing the above equality by $logy$ yields the lemma. Now suppose, we have $x \leq \frac{y}{log^2 y}$. In this case, we have the following results:

$$\vartheta(y) - \vartheta(x) = \vartheta(y) - \vartheta\left(\frac{y}{log^2 y}\right) + \vartheta\left(\frac{y}{log^2 y}\right) - \vartheta(x)$$

which gives us

$$\vartheta(y) - \vartheta(x) = 2(y - x) + o(y)$$

where we used $x = o(y)$ to apply in the first case and deduce the equality $\square$

Let $\varepsilon > 0$, and take $c > C - \frac{\varepsilon}{2}$. As $x \to \infty$, running through the values where

$$\vartheta(x(1+C)) - \vartheta(x) = o(x)$$

we have by Lemma 3 that

$$\vartheta(x(1+C)) - \vartheta(x) = \varepsilon x + o(x)$$

and since our $\varepsilon > 0$ was arbitrary, we have the equality

$$\vartheta(x(1+C)) - \vartheta(x) = o(x)$$

as desired. Now, from the Selberg Symmetry formula, we have that

$$\textstyle\sum_{p \leq x(1+C)} log^2 p + \sum_{pq \leq x(1+C)} logplogq = 2Cxlogx + o(xlogx)$$

Now, since for all primes $p \in (x, x(1+C)]$ we have that $logp < log(x(1+C))$, we have

$$\textstyle\sum_{x<p\leq x(1+C)} log^2 p \leq (\vartheta(x(1+C)) - \vartheta(x))\, log(x(1+C)) = o(xlogx)$$

By combining the previous results, we obtain

$$\textstyle\sum_{x<p\leq x(1+C)} \left(\vartheta(\frac{x}{p}(1+C)) - \vartheta\left(\frac{x}{p}\right)\right) logp = 2Cxlogx + o(xlogx)$$

(4)

We will now use the following lemma

Lemma 4: *Consider the limit as $x \to \infty$, running through the values where*

$$\vartheta(x(1+c)) - \vartheta(x) = o(x)$$

Then, we have that

$$\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right) = 2C\frac{x}{p} + o\left(\frac{x}{p}\right)$$

for all the primes $p \leq x(1+C)$ outside a set P of primes $p \leq x(1+C)$ satisfying

$$\sum_{p\in P} \frac{logp}{p} = o(logx)$$

Proof: We proceed by contradiction. If the lemma is false, then there exists $b_1, b_2 > 0$ such that

$$\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right) < (2C - b_1)\frac{x}{p} \text{ for all } p \in P \text{ and } \sum_{p\in P} \frac{logp}{p} \sim b_2 logx$$

Using the elementary estimate,

$$\sum_{p\leq P} \frac{logp}{p} = (1 + o(1)))logx$$

Using the result of Lemma 3, we can find that

$$\begin{gathered}\sum_{p\leq x(1+C)} \left(\vartheta(\frac{x}{p}(1+C)) - \vartheta\left(\frac{x}{p}\right)\right) logp = \\ \sum_{p\in P} \left(\vartheta(\frac{x}{p}(1+C)) - \vartheta\left(\frac{x}{p}\right)\right) logp + \\ \sum_{p\leq x(1+C), p\notin P} \left(\vartheta(\frac{x}{p}(1+C)) - \vartheta\left(\frac{x}{p}\right)\right) logp \leq \\ b_2(2C - b_1)xlogx + 2C(1 - b_2)xlogx + o(xlogx) = (2C - b_{1b_2})xlogx + o(xlogx)\end{gathered}$$

But the above result is the contradiction to (4), hence the lemma is true. □

Using the two lemmas mentioned above, we can construct the proof of Proposition by Erdos, which can be found in [3].□

After learning the Erdos's result of the above proposition, Selberg proved the PNT. His proof is as follows:

Proofs of the PNT (from [3]): We require three lemmas:

Lemma 5: *We have* $a + A = 2$

Proof: By definition of A, we can choose a large x, such that $\vartheta(x) = Ax + o(x)$. It easily follows that

$$\sum_{p \leq x} log^2 p \leq Axlogx + o(xlogx)$$

Now, subtracting the above equation from the symmetry formula, we get

$$\sum_{pq \leq x} logplogq = \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) logp = (2 - A)xlogx + o(xlogx)$$

We now recall the following well-known fact

$$\sum_{p \leq P} \frac{logp}{p} = (1 + o(1)))logx$$

Combining the above equation with definition of a, we get

$$\sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) logp \geq \sum_{p \leq x} \left(\frac{ax}{p}\right) logp = axlogx + o(xlogx)$$

We then have that

$$0 = \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) logp - \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) logp = (2 - a - A)xlogx + o(xlogx)$$

from which we can deduce that $2 - a - A \geq 0$, so $a + A \leq 2$. We repeat the aobe argument with large x chosen so that $\vartheta(x) = ax + o(x)$ we find that $a + A \geq 2$. Thus, it follows that $a + A = 2$. $\square$

Lemma 6: *Choose large* x *such that* $\vartheta(x) = Ax + o(x)$. *Then*

$$\vartheta\left(\frac{x}{p'}\right) = a\left(\frac{x}{p'}\right) + o\left(\frac{x}{p'}\right)$$

for all primes $p^{'}$ *outside of set* P *of primes* $p \leq x$ *satisfying*

$$\sum_{p \in P} \frac{logp}{p} = o(logx)$$

Proof: We proceed by contradiction. If the lemma is false, then there exist $b_1, b_2 > 0$ and a set P of primes such that for all $p \in P$, we have

$$\vartheta\left(\frac{x}{p}\right) > (a+b_1)\frac{x}{p} \text{ and } \sum_{p \in P} \frac{logp}{p} > b_2 logx$$

As in the proof of Lemma 5, we recall

$$\sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) logp = (2-A)xlogx + o(xlogx) = axlogx + o(xlogx)$$

where the last step is the application of Lemma 6. Combining those results, we obtain the following:

$$axlogx + o(xlogx) = \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right)$$

which then gives us

$$\sum_{p \in P} \vartheta\left(\frac{x}{p}\right) + \sum_{p \notin P, p \leq x} \vartheta\left(\frac{x}{p}\right) >$$
$$b_2(a+b_1)xlogx + (1-b_2)axlogx + o(xlogx) = axlogx + b_1 b_2 xlogx + o(xlogx)$$

We know have that

$$axlogx + o(xlogx) > axlogx + b_1 b_2 xlogx + o(xlogx)$$

It is a contradiction. Thus the lemma is true. □

Lemma 7: *Consider the set of primes p' such that*

$$\vartheta\left(\frac{x}{p'}\right) = a\left(\frac{x}{p'}\right) + o\left(\frac{x}{p'}\right)$$

let p_1 be the smallest member of this set. Then $p_1 < x^{\varepsilon}$ for every $\varepsilon > 0$, and

$$\vartheta\left(\frac{x}{p_1 p'}\right) = a\left(\frac{x}{p_1 p'}\right) + o\left(\frac{x}{p_1 p'}\right)$$

for all primes p' outside a set P of primes $p \leq x$ satisfying

$$\sum_{p \leq x} \frac{logp}{p} = o(logx)$$

Proof: Trivial using Lemma 6. □

Having proven Lemma 5,6 and 7, we can now prove the PNT with the Proposition(Erdos). Take $p^{'}$ as Lemma 7, and let $p^{'}$ be any prime such that

$$\vartheta\left(\frac{x}{p'}\right) = a\left(\frac{x}{p'}\right) + o\left(\frac{x}{p'}\right)$$

and let $p^{"} < x/p_1$. Assume, for the sake of contradiction, that $\frac{x}{p_1 p^{"}} < \frac{x}{p'}$. Furthermore, let $\delta \in (a, \frac{A}{a-1})$, and let the closed interval I be defined by

$$I = \left[\frac{p'}{p_1}, \frac{p'}{p_1}\left(\frac{A}{a} - \delta\right)\right]$$

Notice that if $p^{"} \in I$ then we have

$$a\frac{x}{p'} + o(\frac{x}{p'}) \leq A\frac{x}{p_1 p'} + o(\frac{x}{p_1 p'})$$

But for any $p^{"}$ satisfying the equality

$$\vartheta\left(\frac{x}{p'}\right) = a\left(\frac{x}{p'}\right) + o\left(\frac{x}{p'}\right)$$

we have the monotonicity of ϑ that

$$a\frac{x}{p'} + o(\frac{x}{p'}) = \vartheta\frac{x}{p'} \geq \vartheta\frac{x}{p_1 p'} = A\left(\frac{x}{p_1 p'}\right) + o\left(\frac{x}{p_1 p'}\right)$$

Now, observe that we have the following inequalities:

$$\sum_{p \in I} \frac{logp}{p} > \eta \cdot \frac{\left(\frac{p'}{p_1}\right)}{log\left(\frac{p'}{p_1}\right)} \cdot \frac{log\left(\frac{p'}{p_1}\right)}{\left(\frac{p'}{p_1}\right)} = \eta$$

where the first inequality is a trivial bound and the second inequality follows from Proposition(Erdos). Similar to proof of Proposition(Erdos), it is possible to construct $(clogx)$ many such disjoint intervals I, and if P denotes the set of primes in the union of this interval. We find that

$$\sum_{p \in P} \frac{logp}{p} > c'(logx)$$

for some constant c'. This contradicts Lemma 7, so we must have $\frac{x}{p_1 p^{"}} \geq \frac{x}{p'}$, it is possible to choose $p' = p_1$. from which we can find $p^{"} \leq 1$. It follows from the second part of Lemma 7 that

$$\vartheta(\frac{x}{p_1}) = A(\frac{x}{p_1}) + o(\frac{x}{p_1})$$

and by assumption, we have that

$$\vartheta(\frac{x}{p'}) = \vartheta(\frac{x}{p_1}) = a(\frac{x}{p_1}) + o(\frac{x}{p_1})$$

Adding the above two equalities, we find that

$$2\vartheta(\frac{x}{p_1}) = (a + A)(\frac{x}{p_1}) + o(\frac{x}{p_1}) = 2(\frac{x}{p_1}) + o(\frac{x}{p_1})$$

and dividing by 2 on both sides yields that

$$\vartheta(\frac{x}{p_1}) = (\frac{x}{p_1}) + o(\frac{x}{p_1})$$

By the first part of Lemma 7, we have $p_1 < x^{\varepsilon}$ for every $\varepsilon > 0$ so from the above equality, we deduce that

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} \Leftrightarrow \vartheta(x) \sim x$$

We recall that $\pi(x) \sim (\frac{x}{logx}) \Leftrightarrow \vartheta(x) \sim x$. Thus we have proved the PNT.$\square$

The above proof of the PNT is the first elementary proof, but it is not the easiest elementary proof. Just days after the above proof was discovered, Selberg and Erdős jointly managed to simplify the above arguments to a large extent.

4 References

[1] Steve Balady. Annotation: A discussion of the fundamental ideas behind Selberg's "An elementary proof of the prime-number theorem". August 2006.

[2] Atle Selberg. An elementary proof of the prime-number theorem. Annals of Mathemat- ics, Second Series, 50(2):305–313, April 1949.

[3] Paul Erdős. On a new method in elementary number theory which leads to an ele- mentary proof of the prime number theorem. Proceedings of the National Academy of Sciences, 35(7):374–384, July 1949.

[4] Ashvin Swaminathan. On the Selberg-Erdos proof of the prime number theorem.