Definitive Proof of Legendre's Conjecture

Kenneth A. Watanabe, PhD

January 23, 2020

1 Abstract

Legendre's conjecture, states that there is a prime number between n^2 and $(n+1)^2$ for every positive integer n. In this paper, an equation was derived that accurately determines the number of prime numbers less than n for large values of n. Then, using this equation, it was proven by induction that there is at least one prime number between n^2 and $(n+1)^2$ for all positive integers n thus proving Legendre's conjecture for sufficiently large values n. The error between the derived equation and the actual number of prime numbers less than n was empirically proven to be very small (0.291% at n = 50,000), and it was proven that the size of the error declines as n increases, thus validating the proof.

2 Functions

Before we get into the proof, let me define a few functions that are necessary.

Let the function $l(x)$ represent the largest prime number less than x. For example, $l(10.5) = 7, l(20) = 19$ and $l(19) = 17$.

Let the function $\lambda(x)$ represent the largest prime number less than or equal to x. For example, $\lambda(10.5) = 7, \lambda(20) = 19$ and $\lambda(23) = 23$.

Let the function $z_p(n)$ equal the number of odd integers less than or equal to n that are evenly divisible by p and not equal to p, and not evenly divisible by another prime number less than p. For example $z_5(25) = 1$ since, excluding 5, there are only 2 odd integers {15, 25} less than or equal to 25 that are evenly divisible by 5 and only one of them {25} is not divisible by a prime lower than 5.

Let the function $k(n)$ represent the number of composite numbers in the set of odd integers less than or equal to n excluding 1. For example, $k(15) = 2$ since there are two composite numbers 9 and 15 that are less than or equal to 15.

Therefore, if there are x elements in the set of odd integers less than n, then $\pi(n) = x - k(n)$ where $\pi(n)$ is the number of prime numbers less than n, the prime counting function.

3 Introduction

Legendre's conjecture, proposed by Adrien-Marie Legendre (1752-1833), states that there is a prime number between n^2 and $(n+1)^2$ for every positive integer n. The conjecture is one of Landau's four problems (1912) on prime numbers [1]. The Legendre conjecture is the simplest of the Landau problems, and because all the Landau problems are related, a proof of Legendre's conjecture may lead to proofs of the other problems. As of this paper, all of Landau's problems are unproven.

A graph of the number of primes between n^2 and $(n+1)^2$ (Figure 1) for all n from 2 to 10,000 shows that the number of primes steadily increase with increasing n. This is an indication that Legendre's conjecture is likely true.

In order for Legendre's conjecture to be false, there must be a prime gap g larger than $2n+1$, the difference between n^2 and $(n+1)^2$. The gap must start at prime p, such that $p < n^2$ and $p + g > (n+1)^2$. For example, if $n = 100$, the distance between n^2 and $(n+1)^2$ is 201. The first prime gap over 201 occurs at $p = 20,831,323$ [2] which is well beyond n^2 or 10,000. For $n = 500$, the distance is 1001, and the first prime gap greater than 1001 occurs at $p = 1,693,182,318,746,371$ [2] which is even further beyond n^2 or 250,000. The prime gaps of size $2n+1$ start at a $p >> n^2$, another indication that Legendre's conjecture is very likely true.

A heuristic proof can be performed using the prime number theorem which states that $\frac{n}{\ln(n)} \lim_{n\to\infty} = \pi(n)$. It can easily be proven that $\frac{(n+1)^2}{\ln((n+1)^2)} - \frac{n^2}{\ln(n^2)} > 1$ for all $n > 2$. Therefore at a sufficiently large value of n, Legendre's Conjecture is true. However, the error between $\frac{n}{\ln(n)}$ and $\pi(n)$ is quite large ($>$10% error for $n = 50,000$) . So the question arises, what value of n is sufficiently large? Also, for a given value of n with a small % error, it is difficult to prove that the error will not spike to $>$100% at some greater

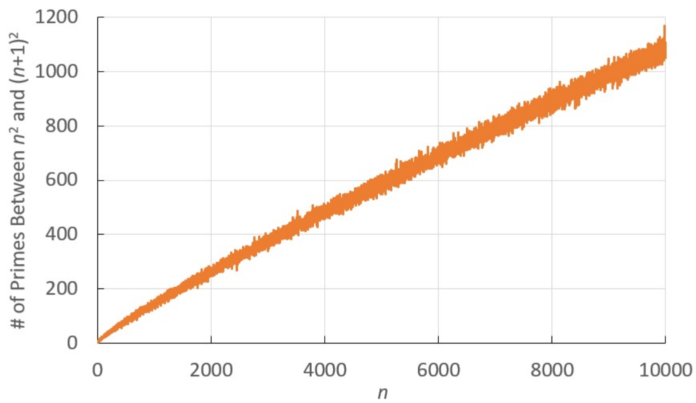

Figure 1: The number of primes between n^2 and $(n+1)^2$ steadily increases with increasing n.

value of n. These reasons make it difficult to accept a proof of Legendre's conjecture based on the prime number theorem.

4 Methodology

To calculate the number of primes between n^2 and $(n+1)^2$, we need a function that accurately predicts the number of primes less than n. Although the prime number theorem states that $\frac{n}{\ln(n)} \lim_{n\to\infty} = \pi(n)$, this equation differs significantly from $\pi(n)$ even for very large values of n. At $n = 1,000,000$, the value of $\frac{n}{\ln(n)}$ underestimates $\pi(n)$ by 7.8%. Even at $n = 100,000,000$, the value of $\frac{n}{\ln(n)}$ underestimates $\pi(n)$ by 5.8%. Because the error is so large and it is difficult to calculate the precise error for a given value of n, a better equation for $\pi(n)$ is necessary.

In this paper, an equation is derived that more precisely determines the number of prime numbers less than n, and as n increases, the accuracy of the equation increases very rapidly. Then, using this equation, it is proven by induction that there is at least one prime number between n^2 and $(n+1)^2$ thus proving the Legendre conjecture is true.

To derive an equation to determine the number of prime numbers less than n, we start with the set of all integers less than n excluding 1. Then we remove all the even integers from the set. Then we remove all the integers evenly divisible by 3 from the set. Then we remove all the integers evenly divisible by 5, 7, 11, 13 ... $\lambda(\sqrt{n})$ where $\lambda(\sqrt{n})$ is the largest prime number less than or equal to $\sqrt{n}$. We only have to go up to $\lambda(\sqrt{n})$ because there are no prime numbers greater than $\sqrt{n}$ that evenly divide n that are not evenly divisible by a lower prime number. By summing up the number of composite numbers in the set of odd numbers less than n and subtracting this from the total number of odd numbers less than n, gives us the number of prime numbers less than n.

Let $\mathbb{I}_n$ represent the set of all integers less than or equal to integer n excluding 1 as shown below.

$\mathbb{I}_n = \{2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,\ldots n\}$

Let the function $z_2(n)$ equal the number of integers in $\mathbb{I}_n$ that are evenly divisible by 2 excluding 2. Notice that every other element beginning with 4 (highlighted in yellow), is divisible by 2. Thus, the number of elements evenly divisible by 2, excluding 2 is defined as follows:

$$z_2(n) = \lfloor \tfrac{n}{2} \rfloor - 1$$

Notice that 1 is subtracted from $\lfloor \frac{n}{2} \rfloor$ since we are excluding 2 from the set of integers evenly divisible by 2.

As $n \to \infty$, the number of even integers in $\mathbb{I}_n$ approaches $n/2$. This gives us the following equation:

$$z_2(n) \lim_{n\to\infty} = \tfrac{n}{2}.$$

In the set of integers $\mathbb{I}_n$, every third element starting with 6 (highlighted in yellow), is evenly divisible by 3.

$\mathbb{I}_n = \{2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,\ldots n\}$

However, the integers 6, 12, 18, etc. are even, so to avoid double counting, we have to subtract these values out. Let the function $z_3(n)$ equal the number of integers in $\mathbb{I}_n$ that are evenly divisible by 3 excluding 3, and not even. This gives us the following equation:

$$z_3(n) = \lfloor \tfrac{n}{3} \rfloor - \lfloor \tfrac{n}{6} \rfloor - 1.$$

Notice that 1 is subtracted since we are excluding 3 from the set of integers evenly divisible by 3.

As $n \to \infty$, $z_3(n)$ approaches the following equation:

$$z_3(n) \lim_{n\to\infty} = (\tfrac{1}{2})(\tfrac{n}{3})$$

This equation states that as n gets large, the number of odd integers approaches $n/2$, and one third of them are evenly divisible by 3.

Looking at those elements in $\mathbb{I}_n$ that are evenly divisible by 5 but not including 5, we notice that every fifth element after 5 (highlighted in yellow) beginning with 10, is divisible by 5.

{2,3,4,5,6,7,8,9, 10 ,11,12,13,14, 15 ,16,17,18,19, 20 ,21,22,23,24, 25 ,26,27, 28,29, 30 ,31,32,33,34, 35 ,36,37,. . . ,n}

But notice that, of the set of elements divisible by 5, every other element is evenly divisble by 2 and every third element is evenly divisible by 3. There are some elements that are divisible by both 2 and 3. So to avoid double counting, we have to subtract the elements evenly divisible by 2 and 3 without double counting the elements divisible by both 2 and 3. Let the function $z_5(n)$ equal the number of odd integers less than or equal to n that are evenly divisible by 5 excluding 5, but not evenly divisible by 3 or 2. Using the principle of inclusion/exclusion [3], we get the following equation for $z_5(n)$:

$$z_5(n) = \lfloor \frac{n}{5} \rfloor - (\lfloor \frac{n}{10} \rfloor + \lfloor \frac{n}{15} \rfloor) + \lfloor \frac{n}{30} \rfloor - 1$$

As $n \to \infty$, $z_5(n)$ approaches the following equation:

$$z_5(n) \lim_{n \to \infty} = (\tfrac{1}{2})(\tfrac{2}{3})(\tfrac{n}{5})$$

This equation states that as n gets large, of the odd integers that are not evenly divisible by 3, one fifth of them are evenly divisible by 5.

Looking at those elements in $\mathbb{I}_n$ that are evenly divisible by 7, we notice that every seventh element after 7 beginning with 14, is divisible by 7.

But notice that every other element is divisible by 2, and every third element (yellow) is divisible by 3 and every fifth element (green) is divisible by 5.

{14, 21 ,28, 35 , 42 ,49,56, 63 , 70 ,77, 84 ,91,98, 105 ,112,119, 126 ,133, 140 , 147 ,154,161, 168 , 175 ,182, 189 ,196. . . n}

So to avoid double counting, we have to subtract the elements evenly divisible by 2, 3 or 5 without double counting the elements. Let the function $z_7(n)$ equal the number of odd integers less than or equal to n that are evenly divisible by 7 excluding 7, but not evenly divisible by 2,3 or 5. Using the principle of inclusion/exclusion, we get the following equation for $z_7(n)$:

$z_7(n) = \lfloor \frac{n}{7} \rfloor - (\lfloor \frac{n}{14} \rfloor + \lfloor \frac{n}{21} \rfloor + \lfloor \frac{n}{35} \rfloor) + (\lfloor \frac{n}{42} \rfloor + \lfloor \frac{n}{70} \rfloor + \lfloor \frac{n}{105} \rfloor) - \lfloor \frac{n}{210} \rfloor - 1$

As $n \to \infty$, $z_7(n)$ approaches the following equation:

$$z_7(n) \lim_{n\to\infty} = (\tfrac{1}{2})(\tfrac{2}{3})(\tfrac{4}{5})(\tfrac{n}{7})$$

This equation states that as n gets large, of the odd integers that are not evenly divisible by 3 or 5, one seventh of them are evenly divisible by 7.

The general formula for the number of elements in $\mathbb{I}_n$ that are evenly divisible by prime number p excluding p, and not evenly divisible by a prime number less than p is as follows:

$$z_p(n) \lim_{n\to\infty} = (\tfrac{1}{2})(\tfrac{2}{3})(\tfrac{4}{5})(\tfrac{6}{7})(\tfrac{10}{11})\dots(\tfrac{(l(p)-1)}{l(p)})(\tfrac{n}{p})$$

or

$$z_p(n) \lim_{n\to\infty} = (\frac{n}{p}) \prod_{\substack{q=2 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q}$$

The total number of composite numbers in the set of odd numbers less than or equal to n, defined as $k(n)$, is thus defined as follows:

$$k(n) \lim_{n\to\infty} = z_2(n) + z_3(n) + z_5(n) + z_7(n) + z_{11}(n) + ... + z_{\lambda(\sqrt{n})}(n)$$

Plugging in the values of $z_p(n)$ gives:

$$k(n) = n \sum_{\substack{p=2 \\ p \text{ prime}}}^{\lambda(\sqrt{n})} \left((\frac{1}{p}) \prod_{\substack{q=2 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q} \right)$$

Let us define the function $W(x)$, which represents the fraction of the odd numbers less than n that are composite numbers:

$$W(x) = \sum_{\substack{p=2 \\ p \text{ prime}}}^{x} \left((\frac{1}{p}) \prod_{\substack{q=2 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q} \right)$$

where $x = \lambda(\sqrt{n})$ and the sum and products are over prime numbers.

Then the equation for $k(n)$ simplifies to the following:

$$k(n) = nW(\lambda(\sqrt{n}))$$

Let $\pi^*(n)$ be the predicted number of prime numbers less than n for large values of n. The number of primes less than n is the number of elements in $\mathbb{I}_n$ minus $k(n)$:

$\pi^*(n) = |\mathbb{I}_n| - k(n)$

As $n \to \infty$, $|\mathbb{I}_n|$ approaches n, therefore

$\pi^*(n) = n - k(n)$

$\pi^*(n) = n - nW(\lambda(\sqrt{n}))$

$\pi^*(n) = n(1 - W(\lambda(\sqrt{n})))$

The equation for the number of primes less than n as $n \to \infty$ is:

$$\pi^*(n) = n(1 - W(\lambda(\sqrt{n}))) \tag{1}$$

To verify that no mistakes were made in the derivation of equation 1 and to determine at what point the equation converges to the actual number of prime numbers less than n, the actual number of primes less than n (blue) was plotted against equation 1 (orange) in Figure 2A. Equation 1 slightly underestimated the actual number of primes for $n <= 5,000$, but for $n <= 50,000$ in Figure 2B, the curves were virtually indistinguishable. The curve for the actual number of primes less than n (blue) was made thicker so it can be viewed since it was completely obscured by the number of primes predicted by equation 1 (orange). The curve for the prime number theorem $\frac{n}{\ln(n)}$ (gray) was also included for comparison and grossly underestimates the actual number of prime numbers less than n.

A graph of the absolute difference between equation 1 and the actual number of primes less than n for $n = 20$ to 50,000, shows that as n increases, the error decreases (Figure 3). As n increases, the difference between equation 1 and the actual number of primes decreases down to 0.291% at $n = 50,000$ (blue line). The difference between the prime number theorem $\frac{n}{\ln(n)}$ and the actual number of primes decreases at a much slower rate and at $n = 50,000$, the percent difference is 10% (orange line). More will be discussed about the error later in this paper.

5 The Proof of Legendre's Conjecture

Now that we have an equation that accurately determines the number of primes less than n for large values of n, we can prove Legendre's conjecture

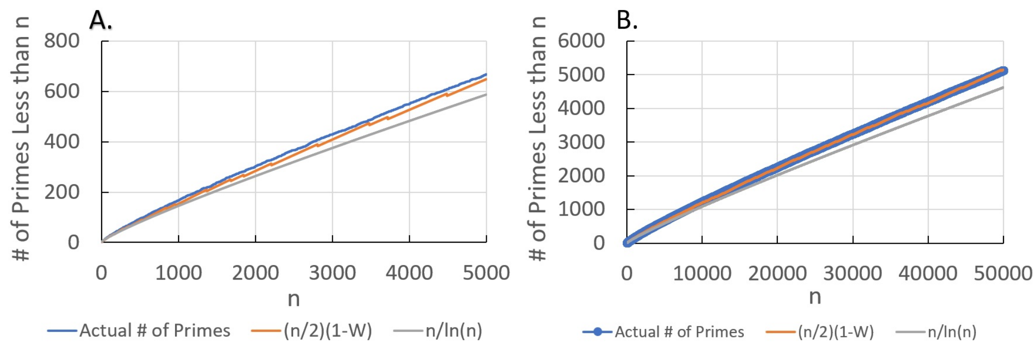

Figure 2: The actual number of primes less than n (blue) is slightly underestimated by equation 1 (orange) for values of n up to 5,000 (A). But for values of n up to 50,000, (B) the curves are virtually indistinguishable. The curve for $n/\ln(n)$ (gray) was also included for comparison.

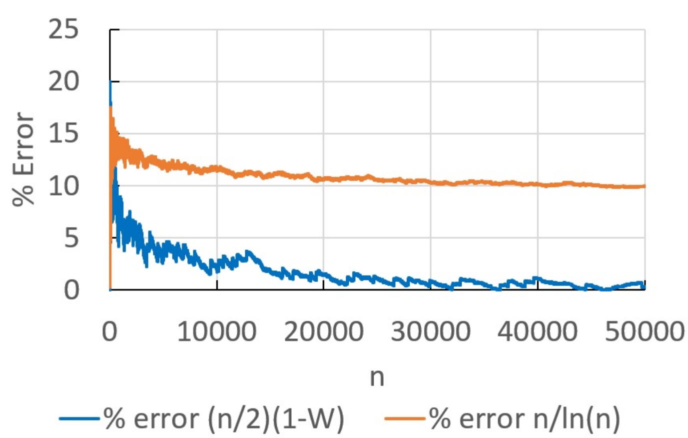

Figure 3: Comparison of equation 1 and $n/\ln(n)$ to the actual number of primes less than n. As n increases, the difference between equation 1 and the actual number of primes rapidly decreases (blue line). The difference between $n/\ln(n)$ and the actual number of primes decreases at a much slower rate (orange line).

by induction. However, to perform proof by induction, we must first get $(1 - W(p_{i+1}))$ in terms of $W(p_i)$. To do this, we must look at the actual values of $(1 - W(p_i))$.

$1 - W(2) = 1 - (\frac{1}{2}) = \frac{1}{2}$

$1 - W(3) = 1 - (\frac{1}{2}) - (\frac{1}{2})(\frac{1}{3}) = (\frac{1}{2})(\frac{2}{3})$

$1 - W(5) = 1 - (\frac{1}{2}) - (\frac{1}{2})(\frac{1}{3}) - (\frac{1}{2})(\frac{2}{3})(\frac{1}{5}) = (\frac{1}{2})(\frac{2}{3})(\frac{4}{5})$

$1 - W(7) = 1 - (\frac{1}{2}) - (\frac{1}{2})(\frac{1}{3}) - (\frac{1}{2})(\frac{2}{3})(\frac{1}{5}) - (\frac{1}{2})(\frac{2}{3})(\frac{4}{5})(\frac{1}{7}) = (\frac{1}{2})(\frac{2}{3})(\frac{4}{5})(\frac{6}{7})$

Notice the value of $1 - W(p_{i+1})$ is the same as $1 - W(p_i)$ minus $(1 - W(p_i))(\frac{1}{p_{i+1}})$. Therefore, these equations for $1 - W(p_i)$ can recursively defined as:

$$\begin{aligned} 1 - W(p_{i+1}) &= (1 - W(p_i)) - (1 - W(p_i))(\frac{1}{p_{i+1}}) \\ 1 - W(p_{i+1}) &= (1 - W(p_i))\left(1 - (\frac{1}{p_{i+1}})\right) \end{aligned}$$

$$1 - W(p_{i+1}) = (1 - W(p_i))\left(\frac{(p_{i+1} - 1)}{p_{i+1}}\right) \tag{2}$$

or

$$1 - W(p) = \prod_{\substack{q=2 \\ q \text{ prime}}}^{p} \frac{(q-1)}{q}$$

Using equation 1 to determine the number of primes less than n, we can calculate the number of primes between n^2 and $(n+1)^2$. If this number is greater than or equal to 1 for all n, then we have proven Legendre's Conjecture.

$\pi^*(n^2) = (n^2)(1 - W(\lambda(n)))$

$\pi^*((n+1)^2) = ((n+1)^2)(1 - W(\lambda(n+1)))$

There are two cases. The first case is where $p_i \le n < p_{i+1} - 1$ in which case $\lambda(n) = \lambda(n+1) = p_i$. The second case is where $n = p_i - 1$ in which case $\lambda(n) = p_{i-1}$ and $\lambda(n+1) = p_i$.

Case 1: Let us look at the case where $p_i \le n < p_{i+1} - 1$.
Let us prove for all $p_i \le n < p_{i+1} - 1$, there is at least 1 prime number between n^2 and $(n+1)^2$. That means the difference between $\pi^*((n+1)^2)$ and $\pi^*(n^2)$ must be greater than or equal to 1.

$\pi^*(n^2) = (n^2)(1 - W(\lambda(n)))$

$\pi^*((n+1)^2) = ((n+1)^2)(1 - W(\lambda(n+1))) = ((n+1)^2)(1 - W(\lambda(n)))$

Let $\Delta\pi(n^2)$ be the difference between $\pi((n+1)^2)$ and $\pi(n^2)$.

$\Delta\pi(n^2) = \pi^*((n+1)^2) - \pi^*(n^2)$

$\Delta\pi(n^2) = ((n+1)^2)(1 - W(\lambda(n))) - (n^2)(1 - W(\lambda(n)))$

$\Delta\pi(n^2) = ((n+1)^2 - n^2)(1 - W(\lambda(n)))$

$\Delta\pi(n^2) = ((n^2 + 2n + 1) - n^2)(1 - W(\lambda(n)))$

$$\Delta\pi(n^2) = (2n+1)(1 - W(\lambda(n))) \tag{3}$$

To prove $\Delta\pi(n^2) \ge 1$ for all $p_i \le n < p_{i+1} - 1$, we will use induction.
Base case $n = 3$. Plugging 3 for n into equation 3 gives us the following:

$\Delta\pi(n^2) = (2n+1)(1 - W(\lambda(n)))$

$\Delta\pi(2^2) = (2 \times 3 + 1)(1 - W(\lambda(3)))$

$\Delta\pi(2^2) = (7)(1 - (\frac{1}{2}) - (\frac{1}{2})(\frac{1}{3}))$

$\Delta\pi(2^2) = (\frac{7}{3}) > 1$

Assuming $\Delta\pi(n^2) > 1$ for all $p_i \le n < p_{i+1} - 1$, we must prove that $\Delta\pi((n+1)^2) > 1$.
Plugging $n+1$ for n in equation 3 gives the following:

$\Delta\pi(n^2) = (2n+1)(1 - W(\lambda(n)))$

$\Delta\pi((n+1)^2) = (2(n+1)+1)(1 - W(\lambda(n+1)))$

$\Delta\pi((n+1)^2) = (2n+3)(1 - W(\lambda(n)))$

Taking the ratio of $\Delta\pi((n+1)^2)/\Delta\pi(n^2)$ gives

$\Delta\pi((n+1)^2)/\Delta\pi(n^2) = (2n+3)(1 - W(\lambda(n)))/(2n+1)(1 - W(\lambda(n)))$

$\Delta\pi((n+1)^2)/\Delta\pi(n^2) = \frac{(2n+3)}{(2n+1)} > 1$

This proves that for all $p_i \le n < p_{i+1} - 1$ where p, there is at least 1 prime number between n^2 and $(n+1)^2$. In fact, since $\Delta\pi((n+1)^2) > \Delta\pi(n^2)$, this proves that the number of primes between n^2 and $(n+1)^2$ increases with increasing n, which is corroborated by the data in Figure 1.

Case 2: Let us look at the case where $n = p - 1$.

$\pi^*(n^2) = (n^2)(1 - W(\lambda(n)))$

$\pi^*((n+1)^2) = ((n+1)^2)(1 - W(\lambda(n+1)))$

Suppose $n = p_{i+1} - 1$, then $\lambda(n) = p_i$ and $\lambda(n+1) = p_{i+1}$.
Substituting p_i for $\lambda(n)$ and substituting p_{i+1} for $\lambda(n+1)$ gives the following:

$\pi^*(n^2) = (n^2)(1 - W(p_i))$

$\pi^*((n+1)^2) = ((n+1)^2)(1 - W(p_{i+1}))$

$\pi^*((n+1)^2) = ((n+1)^2)(\frac{(p_{i+1}-1)}{p_{i+1}})(1 - W(p_i))$ using equation 2

Let $\Delta\pi(n^2)$ be the difference between $\pi^*(n^2)$ and $\pi^*((n+1)^2)$.

$\Delta\pi(n^2) = \pi^*((n+1)^2) - \pi^*(n^2)$

$\Delta\pi(n^2) = ((n+1)^2)(\frac{(p_{i+1}-1)}{p_{i+1}})(1 - W(p_i)) - (n^2)(1 - W(p_i))$

$\Delta\pi(n^2) = (\frac{(n+1)^2(p_{i+1}-1)}{p_{i+1}} - n^2)(1 - W(p_i))$

Substituting n with $p_{i+1} - 1$ gives the following:

$\Delta\pi(n^2) = (\frac{p_{i+1}^2(p_{i+1}-1)}{p_{i+1}} - (p_{i+1} - 1)^2)(1 - W(p_i))$

$\Delta\pi(n^2) = (p_{i+1}^2 - p_{i+1} - (p_{i+1}^2 - 2p_{i+1} + 1))(1 - W(p_i))$

$\Delta\pi(n^2) = (p_{i+1}^2 - p_{i+1} - p_{i+1}^2 + 2p_{i+1} - 1))(1 - W(p_i))$

$\Delta\pi(n^2) = (p_{i+1} - 1)(1 - W(p_i))$

$$\Delta\pi(n^2) = (p_{i+1} - 1)(1 - W(p_i)) \tag{4}$$

To prove $\Delta\pi(n^2) \geq 1$ for all $n = p_{i+1} - 1$, we will use induction.
Base case $p_{i+1} = 3, p_i = 2$ and $n = p_{i+1} - 1 = 2$.
Plugging 2 for n, and 3 for p_{i+1} and 2 for p_i into equation 4 gives:

$\Delta\pi(2^2) = (3 - 1)(1 - W(2))$

$\Delta\pi(2^2) = 2(1 - (\frac{1}{2}))$

$\Delta\pi(2^2) = 1$

Assuming $\Delta\pi(n^2) > 1$ for all $n = p_{i+1} - 1$
we must prove $\Delta\pi(n^2) > 1$ for all $n = p_{i+2} - 1$

$\Delta\pi((p_{i+2} - 1)^2) = (p_{i+2} - 1)(1 - W(p_{i+1}))$

$\Delta\pi((p_{i+2} - 1)^2) = (p_{i+2} - 1)(\frac{(p_{i+1}-1)}{p_{i+1}})(1 - W(p_i))$ using equation 2

$\Delta\pi((p_{i+2} - 1)^2) = \frac{(p_{i+2}-1)}{p_{i+1}}(p_{i+1} - 1)(1 - W(p_i))$

Since we know $\frac{(p_{i+2}-1)}{p_{i+1}} > 1$ and we assumed $(p_{i+1} - 1)(1 - W(p_i)) > 1$, the product must be greater than 1. This proves that for all $n = p - 1$ where p is a prime number, there is at least 1 prime number between n^2 and $(n+1)^2$ and that the number of prime numbers between n^2 and $(n+1)^2$ also increases with increasing n.

6 Error Analysis

Unlike the prime number theorem, equation 1 is very accurate (0.291% error at $n = 50,000$) and the limits on the error can be precisely determined. Figure 3 shows that the relative difference between the actual number of primes and the number of primes predicted by equation 1, decreases as n increases. This is expected since the limit $n \to \infty$ was used to estimate number of composite numbers less than n. However, a figure does not make a proof. To prove the error does decreases as n increases, we have to look at each source of error in the derivation of equation 1.

The $W(\sqrt{n})$ function estimates the fraction of composite integers less than n. Determining the difference between $W(\sqrt{n})$ and the actual number of composite integers less than n will determine the error in the $\pi^*(n)$ function. Then proving that this error declines with increasing n will confirm the proof of Legendre's conjecture. Expanding the W(x) function, gives the following equation:

$$W(\lambda(\sqrt{n})) = (\tfrac{1}{2})+(\tfrac{1}{2})(\tfrac{1}{3})+(\tfrac{1}{2})(\tfrac{2}{3})(\tfrac{1}{5})+(\tfrac{1}{2})(\tfrac{2}{3})(\tfrac{4}{5})(\tfrac{1}{7})+...+\tfrac{1}{\lambda(\sqrt{n})}\prod_{q=2}^{l(\lambda(\sqrt{n}))}(\tfrac{q-1}{q}).$$

where the product is over prime numbers.

The first fraction $(\frac{1}{2})$, is an estimate for the fraction of elements in $\mathbb{I}_n$ that are evenly divisible by 2, excluding 2. This means that $(\frac{1}{2})$ is an estimate for $z_2(n)/n$, or $(\frac{n}{2})$ is an estimate for $z_2(n)$. The difference between $(\frac{n}{2})$ and $z_2(n)$ is the error. A graph of difference between $(\frac{n}{2})$ and $z_2(n)$ (Figure 4A) shows that the difference is either 1 or 1.5 depending on whether n is even or odd. This difference occurs because 2 is excluded in $z_2(n)$ giving a difference of 1, and if n is odd and addition 0.5 is added to the error. Though there will always be an absolute error of 1 or 1.5, as n gets large, the relative error (Figure 4B) becomes insignificant.

The next pair of fractions is $(\frac{1}{2})(\frac{1}{3})$ or $(\frac{1}{6})$. This is an estimate for the number of elements in $\mathbb{I}_n$ that are evenly divisible by 3 and not evenly divisible by 2, excluding 3. This means that $(\frac{1}{6})$ is an estimate for $z_3(n)/n$, or $(\frac{n}{6})$ is an estimate for $z_3(n)$. A graph of difference between $z_3(n)$ and $(\frac{n}{6})$ (Figure 4C) shows that the difference ranges from $(1/2)$ to $(4/3)$. The average difference is $11/12$ and the curve is cyclical with a period of 6. That means the difference between $z_3(n)$ and $(\frac{n}{6})$ is the same as the difference between $z_3(n+6)$ and $(\frac{n+6}{6})$. Though there will always be an absolute error of at least $(1/2)$, as n gets large, the relative error (Figure 4D) becomes insignificant.

The next set of fractions is $(\frac{1}{2})(\frac{2}{3})(\frac{1}{5})$ or $(\frac{2}{30})$. This is an estimate for

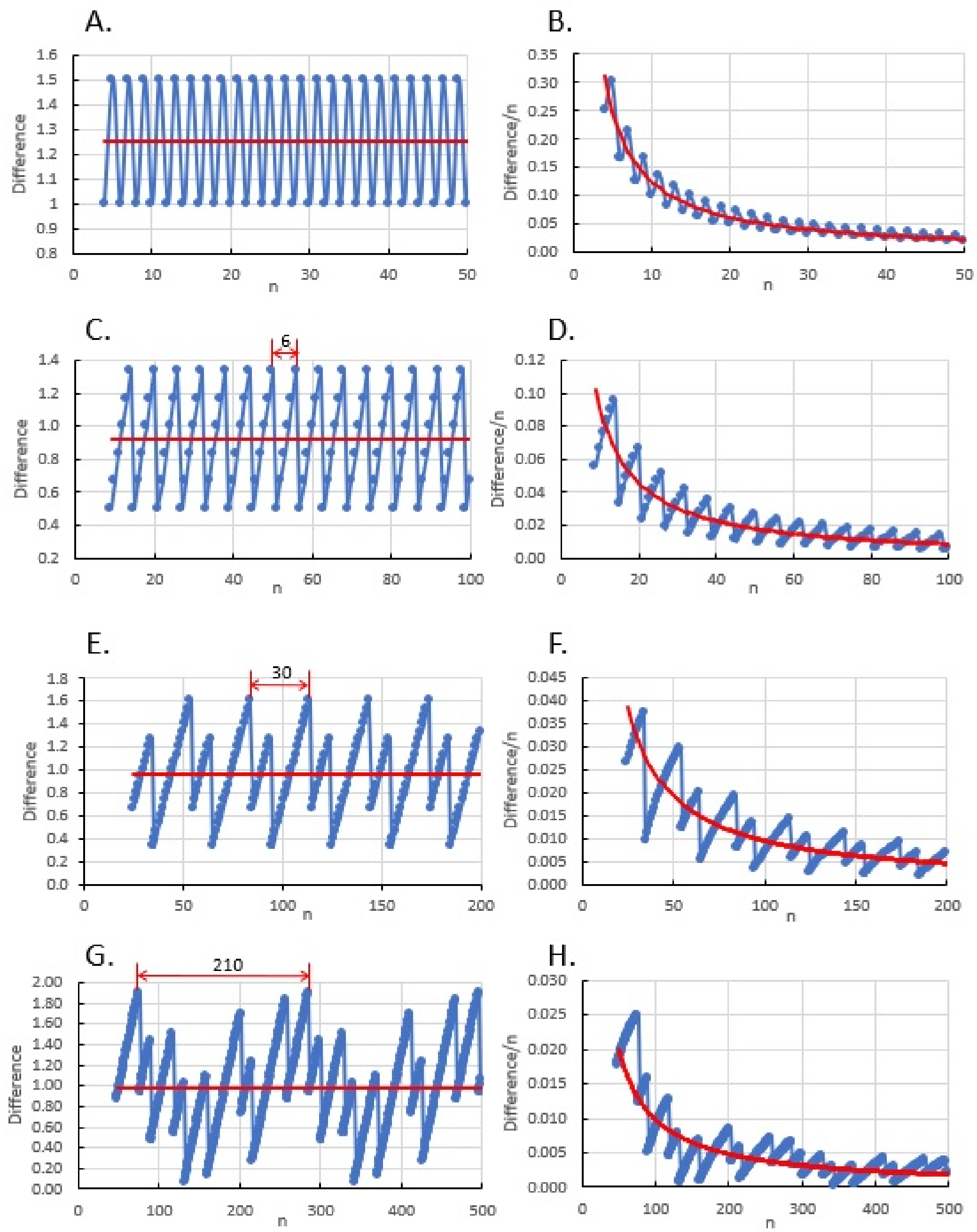

Figure 4: Graph of the error from estimating the fraction of elements evenly divisible by 2 as 1/2 (A and B) and the fraction of elements evenly divisible by 3, 5 and 7 as 1/3, 1/5 and 1/7 respectively (C through H). The red line denotes the average error.

the number of elements in $\mathbb{I}_n$ that are evenly divisible by 5 and not 3 or 2, excluding 5. This means that $(\frac{2}{30})$ is an estimate for $z_5(n)/n$, or $(\frac{n}{15})$ is an estimate for $z_5(n)$. A graph of difference between $z_5(n)$ and $(\frac{n}{15})$ (Figure 4E) shows that the difference ranges from (1/3) to (8/5). The average difference is 29/30 and the curve is cyclical with a period of 30. That means the difference between $z_5(n)$ and $(\frac{n}{15})$ is the same as the difference between $z_5(n+6)$ and $(\frac{n+6}{15})$. Though there will always be an absolute error of at least (1/3), as n gets large, the relative error (Figure 4F) becomes insignificant.

For the set of fractions $(\frac{1}{2})(\frac{2}{3})(\frac{4}{5})(\frac{1}{7})$ or $\frac{8}{210}$, the average difference between this and $z_7(n)$ is $1-(\frac{1}{2})(\frac{8}{210})$ or $(\frac{103}{105})$ with a period of 210.

For all prime numbers $p > 2$, the general formula for the average difference, $\overline{\epsilon_p(n)}$, between $\frac{1}{p}\prod_{q=2}^{l(p)}(\frac{q-1)}{q})$ and $z_p(n)$ is given by the equation below.

$$\overline{\epsilon_p(n)} = 1 - \frac{1}{2p}\prod_{q=2}^{l(p)}(\frac{q-1)}{q})$$

For $p = 2$, the average difference between $\frac{1}{2}$ and $z_2(n)$ is 1.25.

The general formula for the period P_p of $\epsilon_p(n)$ is given by

$$P_p = \prod_{q=2}^{p} q$$

Since the average difference is always less than 1 for $p > 2$, estimating the average difference of 1, errs on the side of caution. Using an error of 1.25 for $p = 2$ and an error of 1 for $p > 2$, we can combine all the curves in Figure 4 to get a combined average error (Figure 5A) and the combined average relative error (Figure 5B).

The formula for the curve of the combined error $E(n)$ in Figure 5A is

$$E(n) = (1/4) + \pi(\lambda(\sqrt{n}))$$

where π is the prime counting function.

The formula for the curve of the combined relative eror $RE(n)$ in Figure 5B is

$$RE(n) = \frac{(1/4) + \pi(\lambda(\sqrt{n}))}{n}$$

Since we know that $\frac{\pi(n)}{n}$ goes to 0 as n increases, then the error $\frac{\pi(\lambda(\sqrt{n}))}{n}$ must also go to 0 as n increases. Therefore, the error declines as n increases.

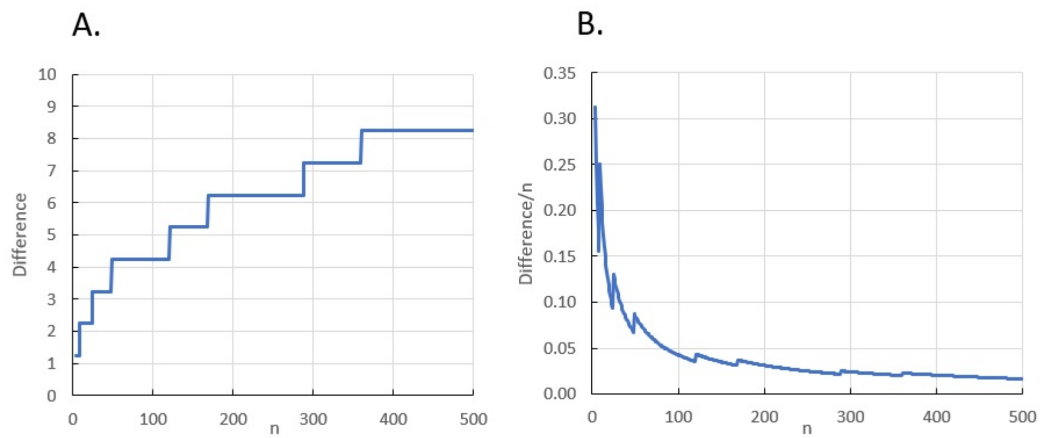

Figure 5: Graph of the combined errors in Figure 4. Graph of the absolute error 5A and graph of the relative error 5B.

7 Summary

In summary, the following equation was derived that accurately determines the number of prime numbers less than n for large values of n.

$$\pi^*(n) = n(1 - W(\lambda(\sqrt{n})))$$

where $\lambda(\sqrt{n})$ is the largest prime number less than or equal to $\sqrt{n}$ and $W(x)$ is defined as follows:

$$W(x) = \sum_{\substack{p=2 \\ p \text{ prime}}}^{x} \left((\frac{1}{p}) \prod_{\substack{q=2 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q} \right)$$

where x is a prime number, $l(p)$ is the largest prime number less than p, and the sum and products are over prime numbers.

It was then proven by induction, that the number of prime numbers between n^2 and $(n+1)^2$ is greater than 1 for all positive integers n, thus confirming the Legendre Conjecture.

It was empirically shown that the error between equation 1 and the actual number of primes less than n is very small ($\epsilon = 0.291\%$ for $n = 50,000$). It was proven that the relative error in the $W(\lambda(\sqrt{n}))$ approaches 0 as $n \to \infty$.

8 Future Directions

Future work will involve applying this technique to prove other prime number conjectures such as the Twin Prime Conjecture and Polignac's Conjecture [4]. Polignac's Conjecture states that there is an infinite number of prime pairs (p_1, p_2) such that $|p_2 - p_1| = 2i$ where i is an integer greater than 0. The Twin Prime Conjecture is the case where $i = 1$.

To prove the Twin Prime conjecture, we need to find the number of twin primes less than an integer n, $(\pi_2(n))$. To do this, we first pair odd numbers (x, y) such that $x+2 = y$ and $y <= n$. For example, (3,5),(5,7),(7,9),(9,11)...,(n-4,n-2),(n-2,n). Then by eliminating pairs that are divisible by 3, 5, 7, 11 etc, the remaining pairs are twin primes.
The number of twin primes less than n will approach the following equation as n gets large:

$$\pi_2(n) = P(1 - 2W(\lambda(\sqrt{n})))$$

where

$$W(x) = \sum_{\substack{p=3 \\ p \text{ prime}}}^{x} (1/p) \prod_{\substack{q=3 \\ q \text{ prime}}}^{l(p)} \frac{(q-2)}{q}.$$

Using proof by induction, it can be shown that the number of twin primes increases indefinitely as n increases.

References

[1] Janos Pintz. Landau's problems on primes. *J. Theory. Nombres Bordeaux*, 21(2):357–404, 2009.

[2] Thomas R. Nicely. First occurrence prime gaps.

[3] Stasys Jukna. *Extremal Combinatorics.* Springer, Berlin, Heidelberg, 2001.

[4] Alphonse de Polignac. *Recherches nouvelles sur les nombres premiers.* Comptes Rendus des Séances de l'Académie des Sciences, 1849.

9 Copyright Notice