Definitive Proof of Legendre's Conjecture

Kenneth A. Watanabe, PhD

June 12, 2019

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

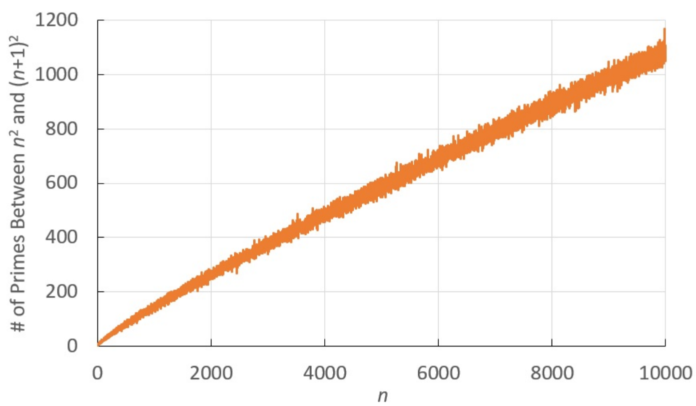

Figure 1: The number of primes between n^2 and $(n+1)^2$ steadily increases with increasing n.

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 odd numbers less than n. Then all the composite numbers in the set that are evenly divisible by 3 are identified. Then all the composite numbers evenly divisible by 5, 7, 11 ... $\lambda(\sqrt{n})$ are identified 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 us start with the set of all odd integers less than or equal to integer n excluding 1 as shown below.

$\mathbb{O} = \{3,5,7,9,11,13,15,17,19,21,23,25,27,29,31,33,35,37,\ldots \text{n}\}$

If n is odd, there are $\frac{(n-1)}{2}$ elements in the list. If n is even, there are $\frac{(n-2)}{2}$ elements in the list. In either case, as $n \to \infty$, the number of elements in the list approaches $\frac{n}{2}$.

Looking at those elements in $\mathbb{O}$ that are evenly divisible by 3 but not including 3, we notice that every third element after 3 (highlighted in yellow) beginning with 9, is divisible by 3.

{3,5,7, 9 ,11,13, 15 ,17,19, 21 ,23,25, 27 ,29,31, 33 ,35,37,... n}

Let the function $z_3(n)$ equal the number of odd integers less than or equal to n that are evenly divisible by 3 excluding 3. As $n \to \infty$, $z_3(n)$ approaches the following equation:

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

Looking at those elements in $\mathbb{O}$ that are evenly divisible by 5 but not including 5, we notice that every fifth element after 5 (highlighted in yellow) beginning with 15, is divisible by 5.

{3,5,7,9,11,13, 15 ,17,19,21,23, 25 ,27,29,31,33, 35 ,37,... ,n}

But notice that, of the set of elements divisible by 5, every third element is also divisible by 3.

{ 15 ,25,35, 45 ,55,65, 75 ,85,95, 105 ,... ,n}

So to avoid double counting, we must multiply the number of elements evenly divisible by 5 by $(\frac{2}{3})$. 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. As $n \to \infty$, $z_5(n)$ approaches the following equation:

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

Looking at those elements in $\mathbb{O}$ that are evenly divisible by 7, we notice that every seventh element after 7 beginning with 21, is divisible by 7.

But notice that every 3rd element (yellow) is also divisible by 3 and every 5th element (green) is divisible by 5.

{21,35,49,63,77,91,105,119,133,147,161,175…n}

{21,35,49,63,77,91,105,119,133,147,161,175…n}

So to avoid double counting, we must multiply the number of elements evenly divisible by 7 by (2/3) and (4/5). 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 5 or 3. As $n \to \infty$, $z_7(n)$ approaches the following equation:

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

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

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

or

$$z_p(n) \lim_{n\to\infty} = (\frac{n}{2})(\frac{1}{p}) \prod_{\substack{q=3 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q}$$

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:

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

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

$$k(n) = (\frac{n}{2}) \sum_{\substack{p=3 \\ p \text{ prime}}}^{\lambda(\sqrt{n})} \left((\frac{1}{p}) \prod_{\substack{q=3 \\ q \text{ prime}}}^{l(p)} \frac{(q-1)}{q} \right)$$

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

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

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

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

$$k(n) = (\tfrac{n}{2})W(\lambda(\sqrt{n}))$$

Thus, the number of primes less than or equal to $n \lim_{n\to\infty}$ is the total number of odd numbers less than n minus $k(n)$:

$\pi^*(n) = (\frac{n}{2}) - k(n)$
$\pi^*(n) = (\frac{n}{2}) - (\frac{n}{2})W(\lambda(\sqrt{n}))$
$\pi^*(n) = (\frac{n}{2})(1 - W(\lambda(\sqrt{n})))$
where $\pi^*(n)$ is the predicted number of prime numbers less than n.
The equation for the number of primes less than n as $n \to \infty$ is:

$$\pi^*(n) = (\frac{n}{2})(1 - W(\lambda(\sqrt{n}))) \tag{1}$$

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

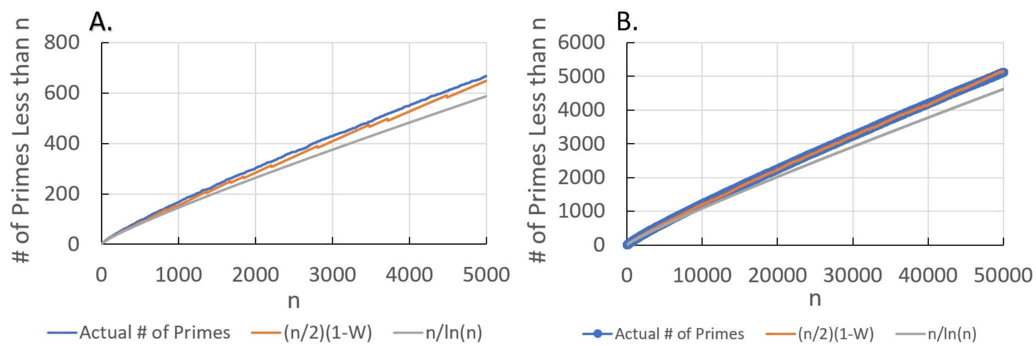

Figure 2: The actual number of primes less than n (blue) is slightly underestimated by equation 1 (orange) for values of n up to 5,000 (A). But for values of n up to 50,000, (B) the curves are virtually indistinguishable. The curve for $n/\ln(n)$ (gray) was also included for comparison.

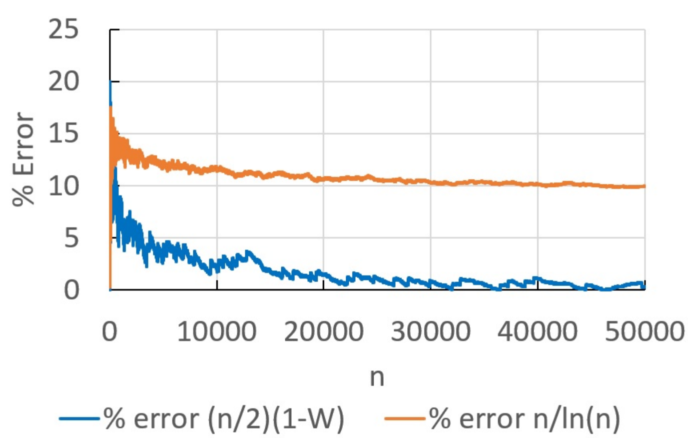

Figure 3: Comparison of equation 1 and $n/\ln(n)$ to the actual number of primes less than n. As n increases, the difference between equation 1 and the actual number of primes rapidly decreases (blue line). The difference between $n/\ln(n)$ and the actual number of primes decreases at a much slower rate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

$1 - W(3) = 1 - (\frac{1}{3}) = \frac{2}{3}$
$1 - W(5) = 1 - (\frac{1}{3}) - (\frac{2}{3})(\frac{1}{5}) = (\frac{2}{3})(\frac{4}{5})$
$1 - W(7) = 1 - (\frac{1}{3}) - (\frac{2}{3})(\frac{1}{5}) - (\frac{2}{3})(\frac{4}{5})(\frac{1}{7}) = (\frac{2}{3})(\frac{4}{5})(\frac{6}{7})$
$1 - W(11) = 1 - (\frac{1}{3}) - (\frac{2}{3})(\frac{1}{5}) - (\frac{2}{3})(\frac{4}{5})(\frac{1}{7}) - (\frac{2}{3})(\frac{4}{5})(\frac{6}{7})(\frac{1}{11}) = (\frac{2}{3})(\frac{4}{5})(\frac{6}{7})(\frac{10}{11})$

Notice the value of $1 - W(p_i)$ (yellow) can be substituted into the green part of $1 - W(p_{i+1})$. Therefore, these equations for $1 - W(p_i)$ can recursively defined as:

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

Using equation 1 to determine the number of primes less than n, we can calculate the number of primes between n^2 and $(n+1)^2$.

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

There are two cases. The first case is where $p_i \leq n < p_{i+1} - 1$ 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 $p_i \leq n < p_{i+1} - 1$.

Let us prove for all $p_i \leq n < p_{i+1} - 1$, 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 1.

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

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

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

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

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

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

$\Delta\pi(n^2) = (\frac{(2n+1)}{2})(1 - W(\lambda(n)))$
$\Delta\pi(3^2) = (\frac{(2\times3+1)}{2})(1 - W(\lambda(3)))$
$\Delta\pi(3^2) = (\frac{7}{2})(1 - (\frac{1}{3}))$
$\Delta\pi(3^2) = (\frac{7}{2})(\frac{2}{3})$
$\Delta\pi(3^2) = (\frac{7}{3}) > 1$

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

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

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

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

This proves that for all $p_i \leq n < p_{i+1} - 1$ where p is a prime number, 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 $n = p - 1$.

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

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:

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

The difference between $\pi^*(n^2)$ and $\pi^*((n+1)^2)$ gives:

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

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

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

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

To prove $\Delta\pi(n^2) > 1$ for all $n = p_{i+1} - 1$, we will use induction.
Base case $p_{i+1} = 5, p_i = 3$ and $n = p_{i+1} - 1 = 4$.
Plugging 4 for n, and 5 for p_{i+1} and 3 for p_i into equation 4 gives:

$\Delta\pi(4^2) = (5 - 1)(1 - W(3))/2$
$\Delta\pi(4^2) = 4(1 - (\frac{1}{3}))/2$
$\Delta\pi(4^2) = 4(\frac{2}{3})/2$
$\Delta\pi(4^2) = \frac{4}{3} > 1$

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

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

Since we know $\frac{(p_{i+2}-1)}{p_{i+1}} > 1$ and we assumed $(p_{i+1} - 1)(1 - W(p_i))/2 > 1$, 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.

We start by calculating the errors associated with the derivation of the $W(x)$ function. Expanding the W(x) function of equation 1, we get the following equation:

$$W(\lambda(\sqrt{n})) = \frac{1}{3}+(\frac{1}{5})(\frac{2}{3})+(\frac{1}{7})(\frac{2}{3})(\frac{4}{5})+...+(\frac{1}{\lambda(\sqrt{n})})(\frac{2}{3})(\frac{4}{5})(\frac{6}{7})(\frac{10}{11})...(\frac{l(\lambda(\sqrt{n}))-1)}{l(\lambda(\sqrt{n}))}).$$

The first fraction of the W(x) function is $\frac{1}{3}$. This is an estimate for the number of elements in the set of odd integers less than or equal to n that are evenly divisible by 3. A graph of difference between the actual fraction of elements evenly divisible by 3 excluding 3, versus $\frac{1}{3}$ (Figure 4A) shows that the difference decreases as n gets large. Only odd values of n were plotted since even values of n have the same number of odd integers as $n-1$ and does not add additional information. The graph starts at $n = 9$ since $W(\lambda(\sqrt{n}))$ is not defined for values of n less than 9.

For example, for $n = 9$, there are 4 odd integers less than or equal to n, 1 of which {9} is evenly divisible by 3. So the difference is $(\frac{1}{3})-(\frac{1}{4}) = 0.08333$.
For $n = 11$, there are 5 odd integers less than or equal to n, 1 of which {9} is evenly divisible by 3. So the difference is $(\frac{1}{3}) - (\frac{1}{5}) = 0.13333$.
For $n = 13$, there are 6 odd integers less than or equal to n, 1 of which {9} is evenly divisible by 3. So the difference is $(\frac{1}{3}) - (\frac{1}{6})) = 0.16667$.
For $n = 15$, there are 7 odd integers less than or equal to n, 2 of which {9,15} are evenly divisible by 3. So the difference is $(\frac{1}{3}) - (\frac{2}{7}) = 0.04762$.

Though it is obvious that Figure 4A is a declining curve, to be rigorous, we must prove that the curve declines for all n. Notice that in Figure 4A, the local maxima occur at $n_i = 7 + 6i$ where i is an integer greater than or equal to 0. The value of i also corresponds to the number of composite integers less than n that are evenly divisible by 3. Let $\epsilon_3(n)$ represent the error between $\frac{1}{3}$ and the actual fraction of odd integers less than n that are divisible by 3. Examining the values of $\epsilon_3(n)$ at the local maxima gives the following:

$\epsilon_3(13) = \frac{1}{3} - \frac{1}{6}$
$\epsilon_3(19) = \frac{1}{3} - \frac{2}{9}$

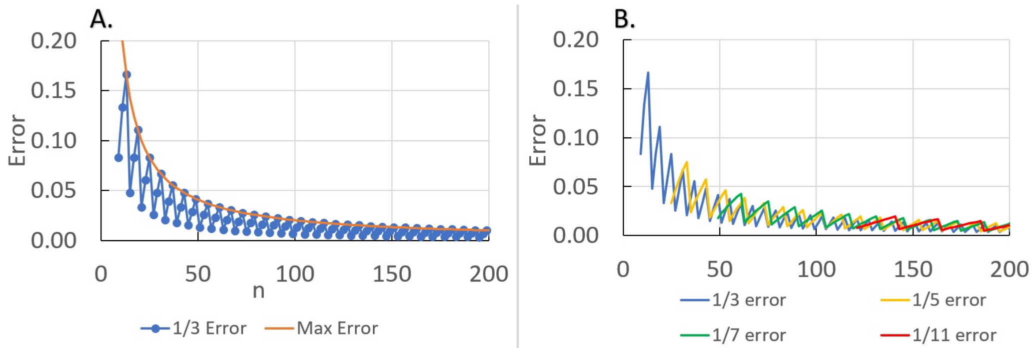

Figure 4: Graph of the error from estimating the fraction of elements evenly divisible by 3 as 1/3 (A) and the fraction of elements evenly divisible by 3, 5, 7 and 11 as 1/3, 1/5, 1/7 and 1/11 respectively (B).

$\epsilon_3(25) = \frac{1}{3} - \frac{3}{12}$
$\epsilon_3(31) = \frac{1}{3} - \frac{4}{15}$

Let $\epsilon_3^*(n)$ represent the upper bound on the value of $\epsilon_3(n)$. In other words, for all values of n, $\epsilon_3(n) <= \epsilon_3^*(n)$. By fitting a curve through the local maxima, we can derive $\epsilon_3^*(n)$ as follows:

$$\begin{aligned}
\epsilon_3^*(n) &= \frac{1}{3} - \frac{(n-7)}{3(n-1)} \\
\epsilon_3^*(n) &= \frac{(n-1)}{3(n-1)} - \frac{(n-7)}{3(n-1)} \\
\epsilon_3^*(n) &= \left(\frac{1}{3}\right)\left(\frac{6}{(n-1)}\right)
\end{aligned}$$

Since n is in the denominator and a constant is in the numerator, this proves that $\epsilon_3^*(n)$ approaches 0 as $n \lim \to \infty$.

The next set of fractions in the W(x) function is $(\frac{1}{5})(\frac{2}{3})$. The fraction $\frac{1}{5}$, is an estimate for the number of elements in the set of odd integers less than or equal to n that are evenly divisible by 5. As can be seen in Figure 4B, this curve also appears to be declining with local maxima at $n_i = 13 + 10i$. Fitting a curve to the local maxima gives us the following equation:

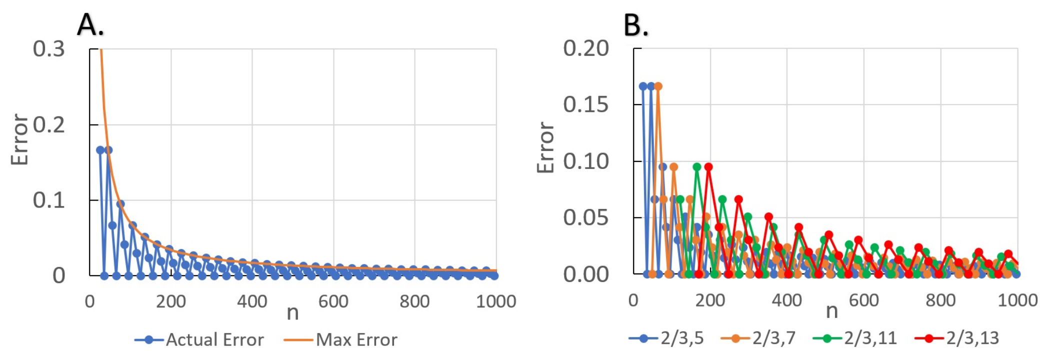

Figure 5: Graph of the error from estimating the fraction of elements divisible by 5 that are not evenly divisible by 3 as 2/3 (A) and the fraction of elements divisible by 5, 7, 11 and 13 that are not evenly divisible by 3 (B).

$$\epsilon_5^*(n) = \frac{1}{5} - \frac{(n-13)}{5(n-1)}$$
$$\epsilon_5^*(n) = \frac{(n-1)}{5(n-1)} - \frac{(n-13)}{5(n-1)}$$
$$\epsilon_5^*(n) = \left(\frac{1}{5}\right)\left(\frac{12}{(n-1)}\right)$$

Since n is in the denominator and a constant is in the numerator, this proves that $\epsilon_5^*(n)$ also approaches 0 as n increases.

The general formula for the maximum error for all prime numbers p less than $\lambda(\sqrt{n})$ is

$$\epsilon_p^*(n) = \left(\frac{1}{p}\right)\left(\frac{(3p-3)}{(n-1)}\right)$$

Also note that in Figure 4B, every successive prime number, the initial error is decreasing.
$\epsilon_3^*(9) > \epsilon_5^*(25) > \epsilon_7^*(49) > \epsilon_{11}^*(121)... > \epsilon_p^*(p^2)$
The fraction $\frac{2}{3}$ in the term $(\frac{1}{5})(\frac{2}{3})$ represents the number of elements in the set of odd integers less than or equal to n that are divisible by 5 but not

evenly divisible by 3. Let $\epsilon_{2/3,5}(n)$ represent the difference between $\frac{2}{3}$ and the fraction of elements less than n that are evenly divisible by 5 and not evenly divisible by 3 (Figure 5A). The local maxima occur at n= 45, 75, 105, ... 45+30i ...

$\epsilon_{2/3,5}(45) = \frac{2}{3} - \frac{2}{4}$
$\epsilon_{2/3,5}(75) = \frac{2}{3} - \frac{4}{7}$
$\epsilon_{2/3,5}(105) = \frac{2}{3} - \frac{6}{10}$
$\epsilon_{2/3,5}(135) = \frac{2}{3} - \frac{8}{13}$

Fitting a curve through the local maxima gives the following equation:

$\epsilon^*_{2/3,5}(n) = (\frac{2}{3})(\frac{10}{(n-5)})$

Since n is in the denominator and a constant is in the numerator, this proves that $\epsilon^*_{2/3,5}(n)$ also approaches 0 as n increases.

By fitting a curve through the local maxima of all the errors for fractions of the form $\frac{(q-1)}{q}$ in the $W(x)$ function, we get the general formula for maximal error as follows:

$$\epsilon^*_{(q-1)/q,p}(n) = \frac{(q+1)}{q}\frac{p}{(n-p)}$$

Since n is in the denominator for all sources of error, all sources of error approach 0 as n increases. The local maxima may align in some areas and not align in other areas resulting in peaks in Figure 3. However, since local maxima of all the curves in Figure 4B decline with increasing n, subsequent alignments of local maxima will result in peaks with a lower magnitude.

7 Maximum Error

Since the curve in Figure 3 is far from smooth, and even though we know that the error decreases as n increases, this raises a question. What if all the peaks in all the curves in Figure 4B and 5B happen to align at some very large value of n, is it possible that we encounter a very large error $>$ 100%? Since we know the upper limits on the errors for each of the fractions in W(x), we can combine all the maximal errors to determine the maximum possible error for all values of n.

Let $z^*_5(n)$ be the maximum possible fraction of odd numbers less than n that are evenly divisible by 5 and not evenly divisible by 3.

$z^*_5(n) = (\frac{1}{5} + \epsilon^*_5)(\frac{2}{3} + \epsilon^*_{2/3,5})$

$z_5^*(n) = (\frac{1}{5})(\frac{2}{3}) + (\frac{2}{3})\epsilon_5^* + (\frac{1}{5})\epsilon_{2/3,5}^* + \epsilon_5^*\epsilon_{2/3,5}^*$
$z_5^*(n) = (\frac{1}{5})(\frac{2}{3}) + \epsilon_5^{max}$
where ϵ_5^{max} is the maximal error between $z_5^*(n)$ and $(\frac{1}{5})(\frac{2}{3})$. Solving for ϵ_5^{max} gives:
$\epsilon_5^{max} = z_5^*(n) - (\frac{1}{5})(\frac{2}{3})$
$\epsilon_5^{max} = (\frac{1}{5} + \epsilon_5^*)(\frac{2}{3} + \epsilon_{2/3,5}^*) - (\frac{1}{5})(\frac{2}{3})$

Likewise, ϵ_7^{max} and ϵ_{11}^{max} would be:
$\epsilon_7^{max} = (\frac{1}{7} + \epsilon_7^*)(\frac{2}{3} + \epsilon_{2/3,7}^*)(\frac{4}{5} + \epsilon_{4/5,7}^*) - (\frac{1}{7})(\frac{2}{3})(\frac{4}{5})$
$\epsilon_{11}^{max} = (\frac{1}{11} + \epsilon_{11}^*)(\frac{2}{3} + \epsilon_{2/3,11}^*)(\frac{4}{5} + \epsilon_{4/5,11}^*) + (\frac{6}{7} + \epsilon_{6/7,11}^*) - (\frac{1}{11})(\frac{2}{3})(\frac{4}{5})(\frac{6}{7})$

Using this technique, we can calculate the maximum possible error between $W(\lambda(\sqrt{n}))$ and the fraction of composite numbers less than n. Let $\epsilon_{max}(n)$ represent the maximum error of the combination of all the $\epsilon_p^*(n)$ and $\epsilon_{(q-1)/q,p}^*(n)$.

$\epsilon_{max}(n) = \epsilon_3^* + \epsilon_5^{max} + \epsilon_7^{max} + \epsilon_{11}^{max} + ... + \epsilon_r^{max}$

$\epsilon_{max}(n) = \epsilon_3^* + (\frac{1}{5} + \epsilon_5^*)(\frac{2}{3} + \epsilon_{2/3,5}^*) - (\frac{1}{5})(\frac{2}{3}) + (\frac{1}{7} + \epsilon_7^*)(\frac{2}{3} + \epsilon_{2/3,7}^*)(\frac{4}{5} + \epsilon_{4/5,7}^*) - (\frac{1}{7})(\frac{2}{3})(\frac{4}{5}) + (\frac{1}{11} + \epsilon_{11}^*)(\frac{2}{3} + \epsilon_{2/3,11}^*)(\frac{4}{5} + \epsilon_{4/5,11}^*)(\frac{6}{7} + \epsilon_{6/7,11}^*) - (\frac{1}{11})(\frac{2}{3})(\frac{4}{5})(\frac{6}{7}) + ... + (\frac{1}{r} + \epsilon_r^*)\prod_{q=3}^{l(r)}((\frac{q-1}{q}) + \epsilon_{q-1/q,p}^*) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q})$
where r is $\lambda(\sqrt{n})$.

Substituting for ϵ_p^* and $\epsilon_{q-1/q,p}^*$ gives the following:
$\epsilon_{max}(n) = \frac{9-3}{3(n-1)} + (\frac{1}{5} + \frac{15-3}{5(n-1)})(\frac{2}{3} + (\frac{4}{3})(\frac{5}{n-5})) - (\frac{1}{5})(\frac{2}{3}) + (\frac{1}{7} + \frac{21-3}{7(n-1)})(\frac{2}{3} + (\frac{4}{3})(\frac{7}{n-7}))(\frac{4}{5} + (\frac{6}{5})(\frac{7}{n-7})) - (\frac{1}{7})(\frac{2}{3})(\frac{4}{5}) + (\frac{1}{11} + \frac{33-3}{11(n-1)})(\frac{2}{3} + (\frac{4}{3})(\frac{11}{n-11}))(\frac{4}{5} + (\frac{6}{5})(\frac{11}{n-11}))(\frac{6}{7} + (\frac{8}{7})(\frac{11}{n-11})) - (\frac{1}{11})(\frac{2}{3})(\frac{4}{5})(\frac{6}{7}) + ... + (\frac{1}{r} + \frac{3r-3}{r(n-1)})\prod_{q=3}^{l(r)}((\frac{q-1}{q}) + (\frac{q+1}{q})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q})$
where r is $\lambda(\sqrt{n})$.

If we focus on the last term:
$(\frac{1}{r} + \frac{3r-3}{r(n-1)})\prod_{q=3}^{l(r)}((\frac{q-1}{q}) + (\frac{q+1}{q})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q}) =$

$\frac{1}{r}(1 + \frac{3r-3}{(n-1)})\prod_{q=3}^{l(r)}((\frac{q-1}{q}) + (\frac{q+1}{q})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q}) =$

$\frac{1}{r}(1 + \frac{3r-3}{(n-1)})\prod_{q=3}^{l(r)}(\frac{q-1}{q})(1 + (\frac{q+1}{q-1})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q}) =$

$\frac{1}{r}(1 + \frac{3r-3}{(n-1)})\prod_{q=3}^{l(r)}(\frac{q-1}{q})\prod_{q=3}^{l(r)}(1 + (\frac{q+1}{q-1})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(r)}(\frac{q-1}{q}) =$

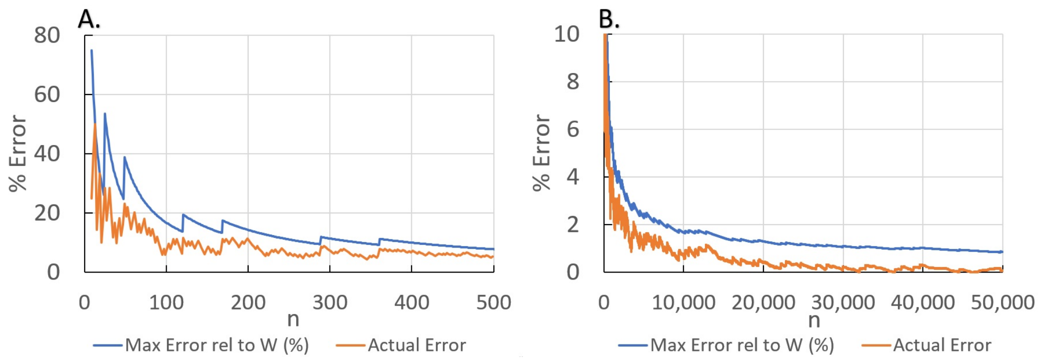

Figure 6: Maximum relative error between $W(\lambda(\sqrt{n}))$ and the fraction of odd composite numbers less than n. The maximum relative error of $W(\lambda(\sqrt{n}))$ (blue line) declines with increasing n but has local maxima at $n = p^2$. The maximum error is always greater than the actual error (orange line).

$$\frac{1}{r}\prod_{q=3}^{l(p)}(\frac{q-1}{q})(1+\frac{3r-3}{(n-1)})\prod_{q=3}^{l(r)}(1+(\frac{q+1}{q-1})(\frac{r}{n-r})) - \frac{1}{r}\prod_{q=3}^{q=l(p)}(\frac{q-1}{q}) =$$

$$\frac{1}{r}\prod_{q=3}^{l(p)}(\frac{q-1}{q})\{(1+\frac{3r-3}{(n-1)})\prod_{q=3}^{l(r)}(1+(\frac{q+1}{q-1})(\frac{r}{n-r})) - 1\}$$

Now we can define $\epsilon_{max}(n)$ as follows:
$$\epsilon_{max}(n) = \sum_{p=3}^{r}\frac{1}{p}\prod_{q=3}^{l(p)}(\frac{q-1}{q})\{(1+\frac{3p-3}{(n-1)})\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{p}{n-p})) - 1\}$$

Dividing $\epsilon_{max}(n)$ by $W(\lambda(\sqrt{n}))$ gives the maximum error relative to $W(\lambda(\sqrt{n}))$. A plot of the maximum error relative to $W(\lambda(\sqrt{n}))$ is shown in Figure 6. Notice that the error spikes at the values where $n = p^2$. When $n = p^2$, a new term is added to $\epsilon_{max}(n)$ and $W(\lambda(\sqrt{n}))$. The largest prime p, such that $p^2 < 50,000$ is $p = 223$ and $p^2 = 49,729$. The value of $\epsilon_{max}(49,729)/W(223) = 0.0087407$ or 0.87407%. If we can prove that $\epsilon_{max}(n)$ is a declining curve, then that means that the maximum error in $W(\lambda(\sqrt{n}))$ cannot exceed 0.87407% for $n > 50,000$.

If $n = p^2$, then the last term in $W(p)$ is:
$\frac{1}{p}\prod_{q=3}^{l(p)}(\frac{q-1}{q})$.

If $n = p^2$, then the last term in $\epsilon_{max}(n)$ is:

$\frac{1}{p}\prod_{q=3}^{l(p)}(\frac{q-1}{q})\{(1+\frac{3p-3}{(p^2-1)})\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{1}{p-1}))-1\}$

If we let $f(n)$ represent the ratio of the last term of $\epsilon_{max}(n)$ and the last term of $W(p)$, we have the following:
$f(n)=(1+\frac{3p-3}{(n-1)})\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{p}{n-p}))-1$

Plugging p^2 for n gives:
$f(p^2)=(1+\frac{3p-3}{(p^2-1)})\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{p}{p^2-p}))-1$
$f(p^2)=(1+\frac{3p-3}{(p^2-1)})\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{1}{p-1}))-1$

If we show that $f(n)\lim_{n\to\infty}=0$, then we know that the maximum possible relative error $\epsilon_{max}(n)/W(\lambda(\sqrt{n}))$ also goes to 0 as $n\to\infty$.

The fraction $\frac{3p-3}{(p^2-1)}$ approaches $3/p$ as p gets large and since p is in the denominator, the fraction will go to 0 as $p\to\infty$. Therefore $(1+\frac{3p-3}{(p^2-1)})\lim_{p\to\infty}=1$. Let $g(p)=\prod_{q=3}^{l(p)}(1+(\frac{q+1}{q-1})(\frac{1}{p-1}))$. We know that $g(p)>1$ since $g(p)$ is the product of numbers greater than 1. But if we can show that $g(p)$ approaches 1 as $p\to\infty$, then the product of $(1+\frac{3p-3}{(p^2-1)})$ and $g(p)$ will also go to 1, thus proving $f(n)$ goes to 0 as $n\to\infty$.

Since we know that $q\geq 3$, $\frac{q+1}{q-1}$ must be at less than or equal to 2. Therefore:
$g(p)\leq\prod_{q=3}^{l(p)}(1+\frac{3+1}{3-1}\frac{1}{p-1})$
$g(p)\leq\prod_{q=3}^{l(p)}(1+\frac{2}{p-1})$
$g(p)\leq(1+\frac{2}{p-1})^{\pi(p)-2}$
$g(p)\leq(1+\frac{2}{p-1})^{\pi(p)-2}$
$g(p)\leq((1+\frac{2}{p-1})^{\frac{(p-1)}{2}\frac{2(\pi(p)-2)}{(p-1)}}$
where π is the prime counting function. It is a well known fact that $(1+\frac{1}{n})^n$ approaches Euler's number e as $n\to\infty$. Therefore $(1+\frac{2}{p-1})^{(p-1)/2}$ also approaches e as $p\to\infty$. Therefore:
$g(p)\leq e^{\frac{2(\pi(p)-2)}{(p-1)}}$
It is also a well known fact that $\frac{\pi(p)}{p}\lim_{p\to\infty}=0$, therefore $e^{2(\pi(p)-2)/(p-1)}$ approaches e^0 which equals 1. Therefore $g(p)\lim_{p\to\infty}=1$ and the product of $(1+\frac{3p-3}{(p^2-1)})$ and $g(p)$ will also go to 1, thus proving $f(n)\lim_{n\to\infty}=0$. Since $f(n)$ goes to 0, that means that $\epsilon_{max}(n)$ must also go to zero as $n\to\infty$.

8 Summary

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

$$\pi^*(n) = (\tfrac{n}{2})(1 - W(\lambda(\sqrt{n})))$$

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

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

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 error in the $W(\lambda(\sqrt{n}))$ approaches 0 as $n \to \infty$ and cannot exceed 0.87407% for all $n > 50,000$.

9 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 [3]. 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] Alphonse de Polignac. *Recherches nouvelles sur les nombres premiers.* Comptes Rendus des Séances de l'Académie des Sciences, 1849.

10 Copyright Notice