A simple framework for improving the Prime Number Theorem regarding estimating the nth prime number

Shreyansh Jaiswal
Atomic Energy Central School - 6 Mumbai, India

1 August 2024

Abstract

The Prime Number Theorem (PNT) offers a foundational approximation for the distribution of prime numbers and aids in estimating the nth prime number $p(n)$ through $p(n) \sim n \log n$. This paper proposes enhancements to this approximation by introducing a correction factor $C(k)$, refining the estimate to $p(n) \approx C(k) \cdot n \cdot \log n$. The derivation of $C(k)$ is explored, alongside its asymptotic behavior and empirical analysis. A generalized formula for $p(n)$ is also derived, eliminating variables other than n and e (Euler's number). Empirical comparisons with traditional methods demonstrate the accuracy and computational efficiency of these new approaches. Ideal conditions for optimal performance of $C(k)$ are examined. Graphical representations and statistical analyses support the validity of the proposed refinements. The paper concludes with a discussion on the implications of these findings and potential areas for future research.

1 Prime Number Theorem

The Prime Number Theorem (PNT) describes the asymptotic distribution of prime numbers. It provides a profound insight into how primes are distributed among the integers. Specifically, the PNT states that the number of primes less than or equal to n, denoted $\pi(n)$, is asymptotically equal to $\frac{n}{\log n}$[2]. In mathematical terms, this can be expressed as:

$$\lim_{n \to \infty} \frac{\pi(n)}{\frac{n}{\log n}} = 1[2]$$

which can be written as,

$$\pi(n) \sim \frac{n}{\log n}[2]$$

where the symbol $\sim$ means that the ratio of $\pi(n)$ to $\frac{n}{\log n}$ approaches 1 as n approaches infinity. This theorem was first conjectured by Gauss and Legendre in the late 18th century and was later proved independently by Hadamard and de la Vallée Poussin in 1896.[2]

The PNT also aids in estimating the nth prime number, $p(n)$. An important corollary of the PNT is that the nth prime can be approximated by:

$$p(n) \sim n \log n[2]$$

This approximation becomes more accurate as n increases.

The significance of the Prime Number Theorem lies not only in its ability to approximate the number of primes up to a given number but also in its implications for the overall understanding of number theory. The distribution of primes influences various areas of mathematics and has applications in fields such as cryptography, where large prime numbers are crucial.

This work explores slight improvements to the PNT, through intuition, logic, assumption, and more.

2 The Correction Factor $C(k)$

We can begin our work with presuming the exsistence of a function $C(k)$, which adjusts the estimation of $p(n)$. Let $C(k)$ be a function such that:

$$p(n) = C(k) \cdot n \cdot \log n$$

Accounting for the complex and unpredictable distribution of primes, we can safely assume that $C(k)$ is not a constant, but a function (of k, which is related to n). However, as we will discuss later in the paper, it isn't technically ideal for us to be able to obtain the "absolute" $C(k)$ for any n in practical uses. Meaning, we can say that, for any $C(k)$

$$C(k-\delta) > C(k) > C(k+\delta),\ |p_n - p(n)| \to 0$$

Where p_n represents any actual prime. The "range" of C(k) explains the variability of $C(k)$
(The reason $C(k-\delta) > C(k)$ is disscussed later in the paper, for now we can just understand it as a representation.)

3 Finding $C(k)$

In order to find $C(k)$, we can rearrange our equation to be written as:

$$C(k) = \frac{p(n)}{n \log n}$$

As we are deriving a formula for $C(k)$, we first assume that

$$p = n \log n$$

Hence, rearranging, we get:

$$n = \frac{p}{\log n}$$

For simplifying this equation, we can assume that $n \approx \frac{p}{\log p}$. If we plug this n value into $p(n) \approx C(k) \cdot n \cdot \log n$, we can simplify as follows:

$$p(n) \approx C(k) \cdot \left(\frac{p}{\log p}\right) \cdot \log\left(\frac{p}{\log p}\right)$$

This further simplifies to:

$$C(k) \approx \frac{\log p}{\log\left(\frac{p}{\log p}\right)}$$

3.1 Justifying taking $n = \frac{p}{\log p}$ instead of $n = \frac{p}{\log n}$

To simplify our expressions and reduce computational complexity, we use the approximation $n = \frac{p}{\log p}$, assuming that n tends to infinity. This approach is justified under the asymptotic analysis where n is large.

We aim to show that changing the value of n does not significantly alter $C(k)$. For this, we will examine the ratio of the actual value $C(k)_r$ to the assumed value $C(k)_a$ of $C(k)$.

Define:

$$C(k)_a = \frac{\log p}{\log\left(\frac{p}{\log p}\right)}$$

and

$$C(k)_r = \frac{p}{n \log n}.$$

First, we expand $C(k)_a$:

$$C(k)_a = \frac{\log p}{\log p - \log(\log p)}.$$

Next, we compute the ratio $\frac{C(k)_a}{C(k)_r}$:

$$\frac{C(k)_a}{C(k)_r} = \left(\frac{\log p}{\log p - \log(\log p)}\right) \cdot \left(\frac{n \log n}{p}\right).$$

Substitute $n = \frac{p}{\log p}$ into the expression:

$$\frac{C(k)_a}{C(k)_r} = \left(\frac{\log p}{\log p - \log(\log p)}\right) \cdot \left(\frac{\frac{p}{\log p} \cdot \log n}{p}\right).$$

Simplify the expression:

$$\frac{C(k)_a}{C(k)_r} = \frac{\log p \cdot \frac{p \log n}{\log p}}{p \cdot (\log p - \log(\log p))} = \frac{\log n}{\log p - \log(\log p)}.$$

According to the Prime Number Theorem, $n \log n \approx p$, so $\log n \approx \log p - \log(\log p)$. Hence:

$$\frac{\log n}{\log p - \log(\log p)} \approx \frac{\log p - \log(\log p)}{\log p - \log(\log p)} \to 1.$$

As $k \to \infty$, $\log(\log p)$ grows much more slowly compared to $\log p$, allowing us to approximate:

$$\frac{C(k)_a}{C(k)_r} \approx \frac{\log p}{\log p} = 1.$$

Thus, we can conclude:

$$\lim_{k\to\infty} \frac{C(k)_a}{C(k)_r} = 1$$

which implies that $C(k)_a$ and $C(k)_r$ are asymptotically equivalent. Therefore:

$$C(k)_a \sim C(k)_r.$$

4 Understanding p

In our analysis, we initially approximated n using $\frac{p}{\log n}$ and later refined it to $\frac{p}{\log p}$ for computational convenience. However, using p directly as the exact value of the prime does not provide sufficient accuracy. Thus, we need to modify our approach to account for this discrepancy.

Empirical calculations suggest that, within a range of primes from the first prime $P(1)$ to the Nth prime $P(N)$, the value of p can be approximated by a function involving both $P(N)$ and $P\left(\frac{N}{2}\right)$. Specifically, we approximate p as:

$$p \approx P(N) \cdot P\left(\frac{N}{2}\right)$$

where $P(N)$ denotes the N-th prime. This approximation helps in refining our model to better match empirical data and improves the accuracy of our calculations.

5 Making the Approximation of $C(k)$ Calculable Without Knowledge of Specific Primes

For extremely large primes, we aim to simplify the approximation of $C(k)$ without requiring detailed knowledge of specific primes. We use the following approximations based on the Prime Number Theorem (PNT) and empirical observations:

5.1 Assumption for $P\left(\frac{N}{2}\right)$

We first assume the nth prime, $P(n) = e^{k+1}$, for our model, we need to justify, $P\left(\frac{N}{2}\right)$ as e^k, To justify the approximation of $P\left(\frac{N}{2}\right)$ as e^k, we need to consider the relationship between $P(N)$ and $P\left(\frac{N}{2}\right)$ carefully, using the Prime Number Theorem (PNT) and empirical observations.

1. Prime Number Theorem and Prime Approximation:

   According to the Prime Number Theorem, the n-th prime $P(N)$ is approximated asymptotically by:

   $$P(N) \approx n \log n$$

   we use the exponential approximation, where:

   $$P(N) \approx e^{k+1}$$

2. Relating $P\left(\frac{N}{2}\right)$ to $P(N)$:

   To estimate $P\left(\frac{N}{2}\right)$, we need to understand the growth rate of primes. A common heuristic is that the n-th prime is not simply twice the value of the $\frac{n}{2}$-th prime. Instead, we use a logarithmic approximation based on empirical observations.

   Given $P(N) \approx e^{k+1}$, we find N such that:

   $$\pi(P(N)) \approx \frac{P(N)}{\log P(N)}$$

Thus:

$$\pi(P(N)) \approx \frac{e^{k+1}}{\log e^{k+1}} = \frac{e^{k+1}}{k+1}$$

For $\frac{N}{2}$:

$$\pi\left(P\left(\frac{N}{2}\right)\right) \approx \frac{P\left(\frac{N}{2}\right)}{\log P\left(\frac{N}{2}\right)}$$

We approximate $P\left(\frac{N}{2}\right)$ using:

$$\pi\left(P\left(\frac{N}{2}\right)\right) \approx \frac{1}{2} \cdot \pi(P(N))$$

Substituting $\pi(P(N))$:

$$\pi\left(P\left(\frac{N}{2}\right)\right) \approx \frac{1}{2} \cdot \frac{e^{k+1}}{k+1}$$

This implies:

$$P\left(\frac{N}{2}\right) \approx e^k$$

3. Rationale for Approximation:

   Empirical observations and numerical experiments often show that $P\left(\frac{N}{2}\right) \approx e^k$ is a reasonable approximation. The exact growth rate of primes means $P\left(\frac{N}{2}\right)$ is generally smaller than $P(N)$ but not simply half. Instead, the exponential approximation e^k provides a tolerable approximation.

Now that we've justified the approximation of $P(N/2) \approx e^k$, we use these values, to obtain an approximation for $C(k)$
Putting values into $p \approx P(N) \cdot P\left(\frac{N}{2}\right)$, we can simplify $C(k)$,

$$C(k) = \frac{2k+1}{(2k+1) - \log(2k+1)}$$

6 Using $p(n)$ for Approximating $\pi(n)$

Just like Gauss's approximation of the prime counting function $\pi(n) \sim \frac{n}{\log n}$, the logarithmic integral method $Li(n) = \int_2^n \frac{dt}{\log t}$, and our new $p(n)$ can be interchanged to approximate $\pi(n)$. Hence, we can derive:

$$\pi(n) \approx \frac{C(k) \cdot n}{\log n}$$

This simplifies to:

$$\pi(n) \sim \frac{\frac{2k+1}{(2k+1)-\log(2k+1)} \cdot n}{\log n}$$

The exact range of k for which this new approximation of $\pi(n)$ would optimally work has not yet been determined by empirical analysis. Further research and data analysis would be required to precisely define this range of k. However, we can adjust the value of k nicely, even if we only have some few values of real $\pi(n)$.

7 Asymptotic Analysis of $C(k)$

Given the function:

$$C(k) \approx \frac{2k+1}{(2k+1) - \log(2k+1)}$$

we want to determine the limit as $k \to \infty$.

First, rewrite $C(k)$:

$$C(k) = \frac{2k+1}{(2k+1) - \log(2k+1)}$$

$$C(k) \approx \frac{2k+1}{(2k+1) - \log(2k+1)} = \frac{2k+1}{2k+1\left(1 - \frac{\log(2k+1)}{2k+1}\right)}$$

Simplify the fraction:

$$C(k) \approx \frac{1}{1 - \frac{\log(2k+1)}{2k+1}}$$

As $k \to \infty$, $\frac{\log(2k+1)}{2k+1} \to 0$. Therefore, the denominator approaches 1:

$$C(k) \approx \frac{1}{1-0} = 1$$

Hence, the limit of $C(k)$ as $k \to \infty$ is:

$$\lim_{k\to\infty} C(k) = 1$$

This limit implies that our function for primes converges to that of the PNT's, thus implying that our equation is asymptotically valid.

8 Empirical Analysis

Please note that this section (Empirical Analysis) uses the approximation of $C(k) \approx \frac{\log p}{\log\left(\frac{p}{\log p}\right)} \approx \frac{2k+1}{(2k+1)-\log(2k+1)}$

8.1 Real $C(k)$ and Concept of e^k

For a range from the 1st prime to the 10^6th prime, the real $C(k)$ is calculated to be approximately 1.12035490203611. Using the concept of e^k, $C(k)$ is estimated as $\frac{32}{32-\ln(32)} \approx 1.12458745$. This shows that our concept for estimating large primes could be valid.

8.2 Comparison with Traditional Formula

Comparing the absolute differences between the traditional formula $n \cdot \ln(n)$ and the new formula $C(k) \cdot n \cdot \ln(n)$ for primes up to 10^6 gives promising results for the new equation. (See graph in Figure 1).

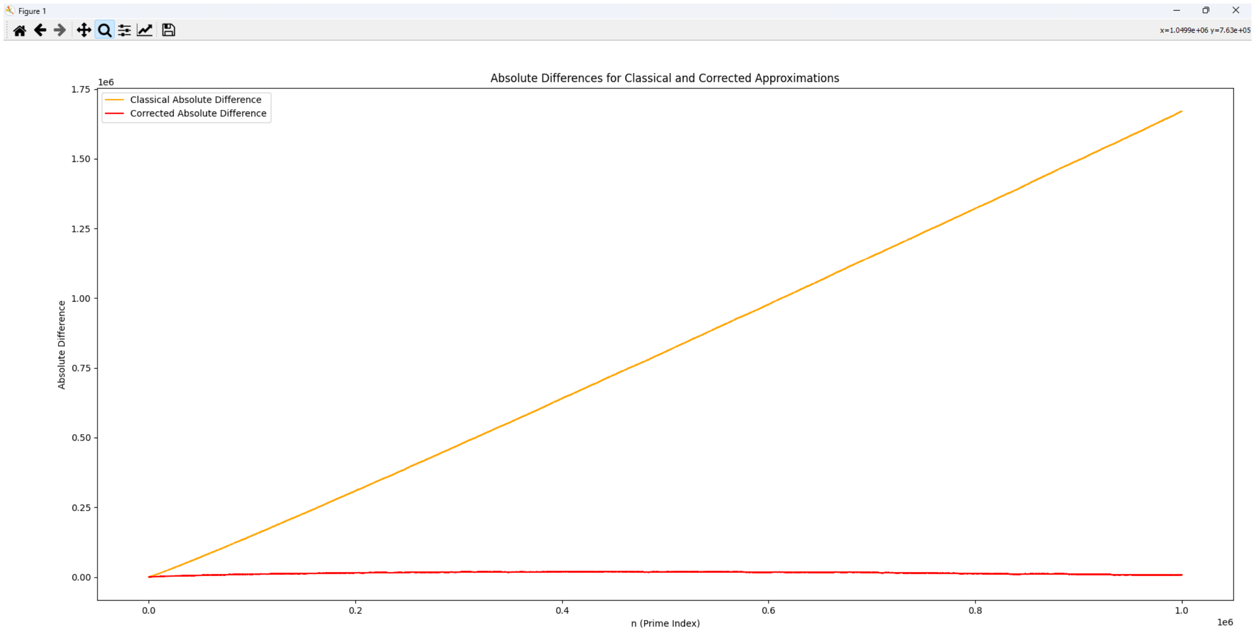

Figure 1: Comparison of Prime Estimation Formulas

Observing the graph (Figure 1), one can clearly deduce that the new equation performs significantly better.

8.3 Accuracy Comparison with Prime Counting Algorithms

The graph (Figure 2) compares three methods for approximating $\pi(n)$:

- Gauss's method $\pi(n) \approx \frac{n}{\log n}$,
- logarithmic integral method $Li(n) = \int_2^n \frac{dt}{\log t}$,
- and our new equation $\pi(n) \sim \frac{\frac{2k+1}{(2k+1)-\log(2k+1)} \cdot n}{\log n}$.

(Note - k is chosen based on empirical values).

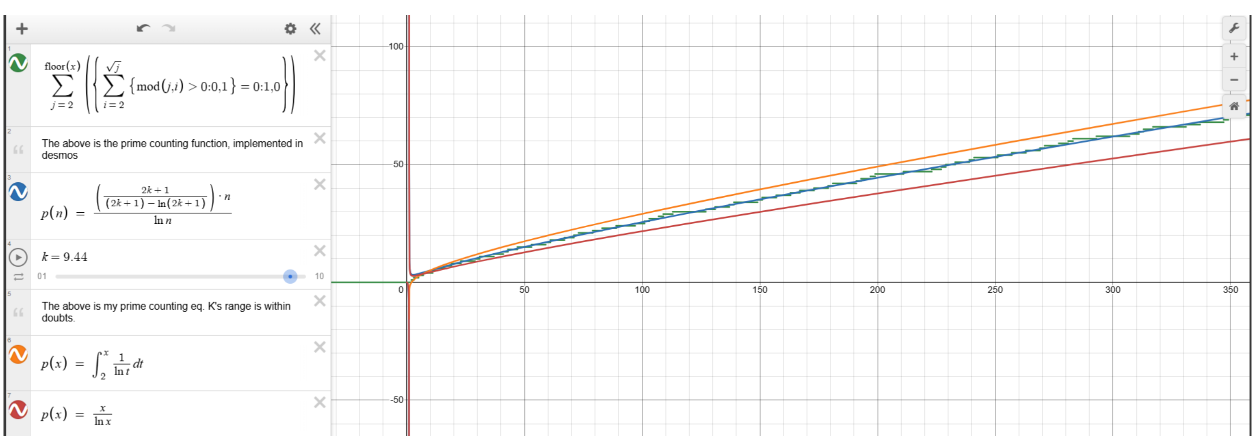

Figure 2: Accuracy of Prime Counting Algorithms

We can see that our equation has the potential to be a better approximation for $\pi(n)$, provided we determine the optimal ranges for k. Additionally, the new equation promises less complexity and shorter computing times compared to the logarithmic integral method.

8.4 Other comparison data of classical and new prime approximation

This section contains various data comparing both the equations (PNT's equation and our new equation). Please note that all these comparisons have been done for up to the 10^6th prime.

1. Absolute difference vs nth prime for our new equation (see Figure 3). 2. Percentage errors vs nth prime (see Figure 4). 3. Histogram analysis of Frequency vs Magnitude of errors for our new equation (see Figure 5). 4. Compared to PNT's approximation (see Figure 6).

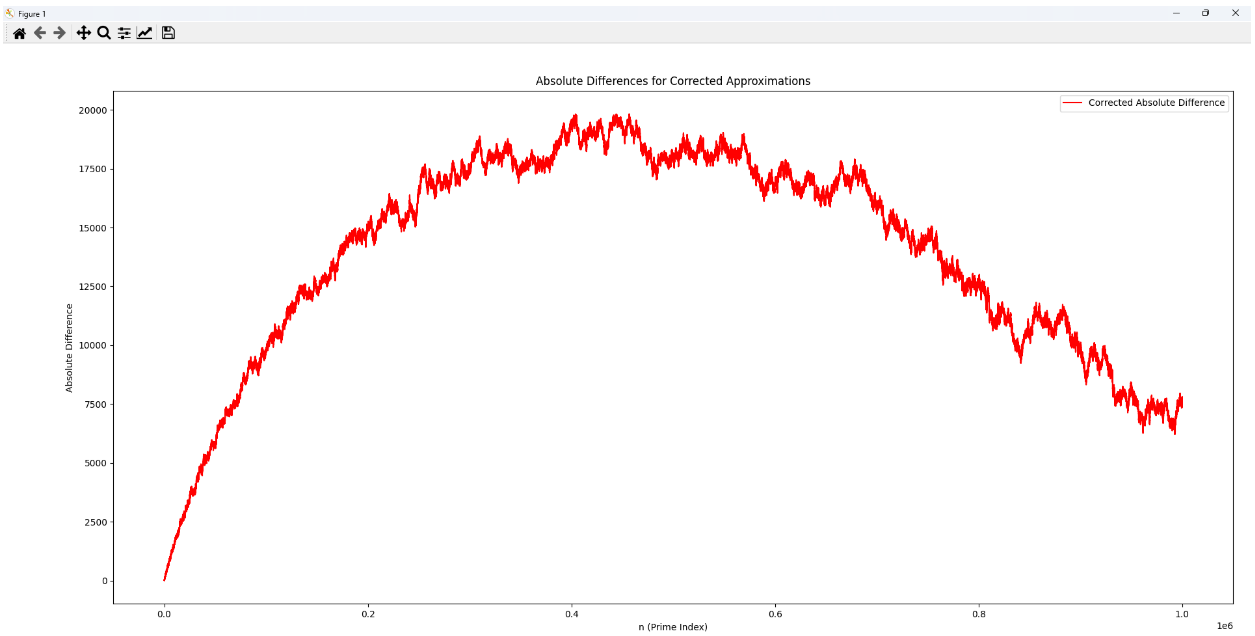

Figure 3: Absolute Difference vs nth Prime for Corrected Approximations

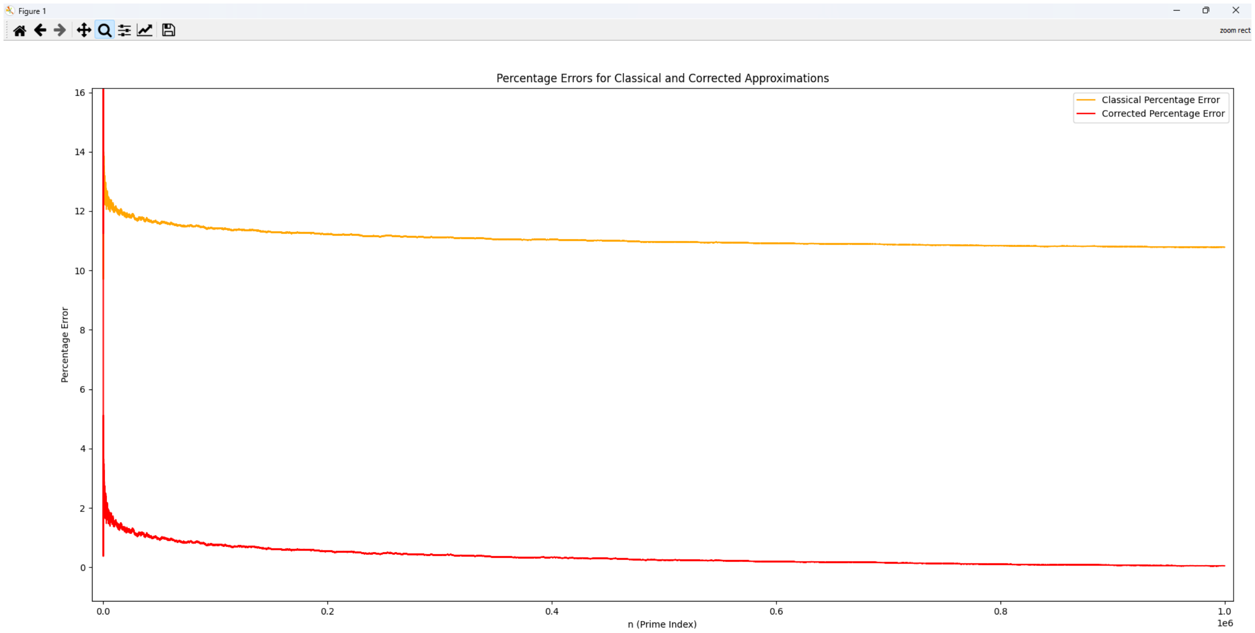

Figure 4: Percentage Errors vs nth Prime

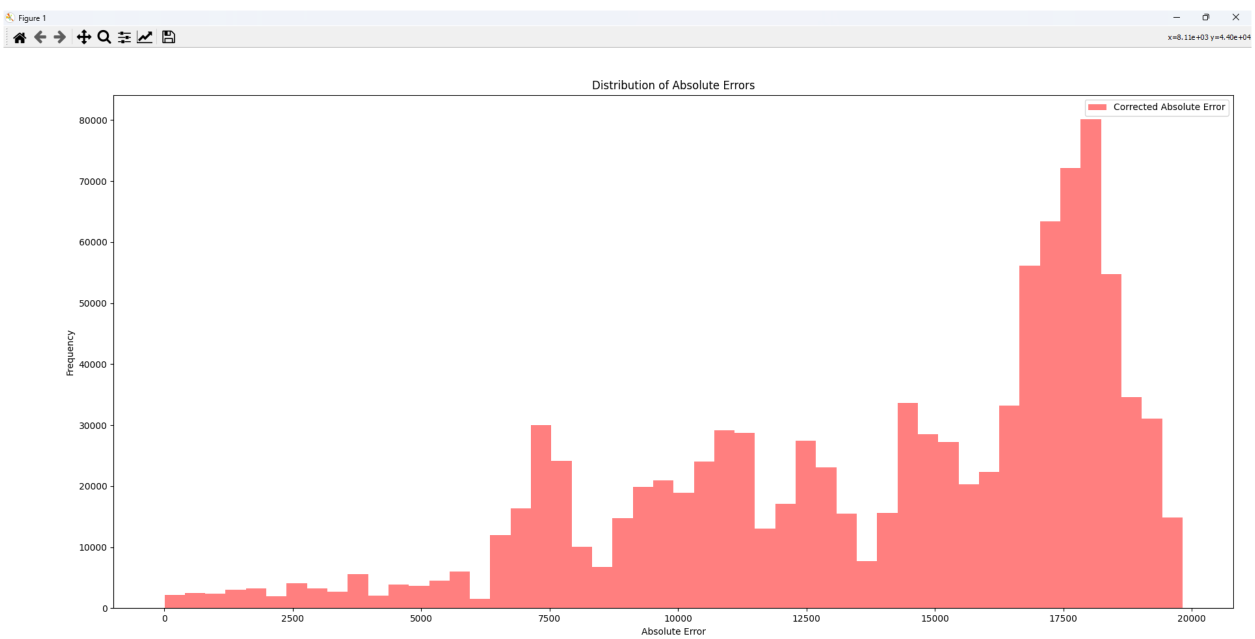

Figure 5: Histogram Analysis of Frequency vs Magnitude of Errors for New Equation

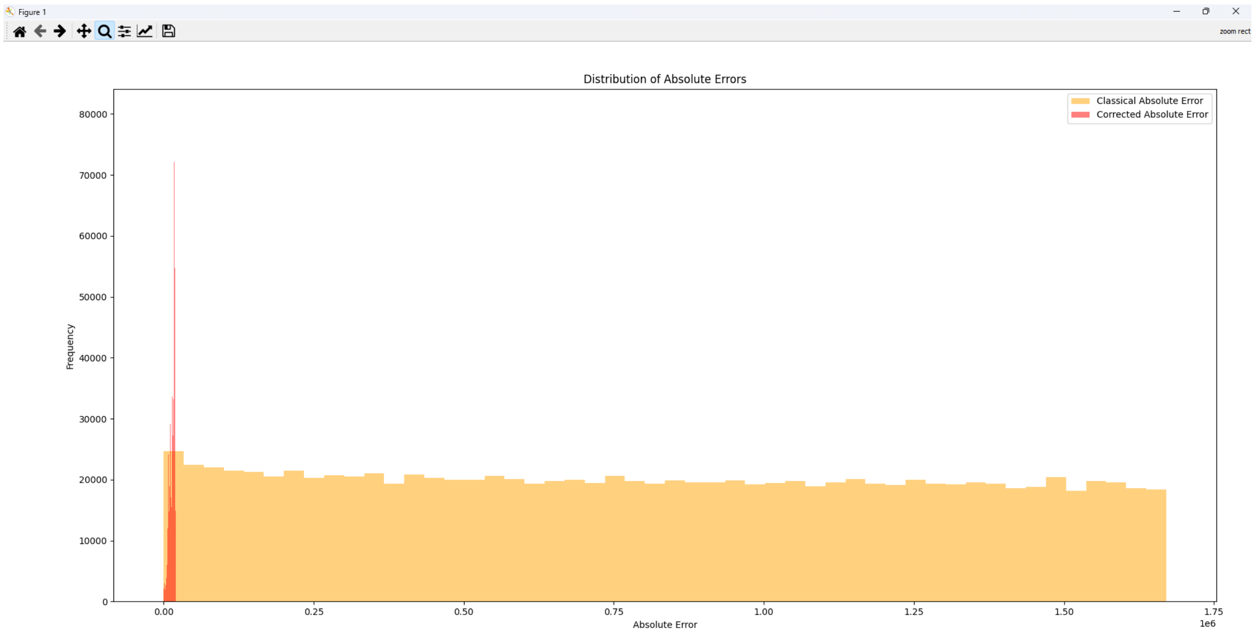

Figure 6: Comparison of Histogram Analysis of our equation with PNT's analysis

9 Ideal Conditions Give Ideal Outputs

This subsection discusses the ideal conditions under which this system would perform the best. Most of the work done regarding $C(k)$ in the paper is attempting to make $C(k)$ easier to calculate (i.e., making it calculable without knowing any primes). However, in this subsection, we are going to assume the ideal conditions. As ideally, we are having the nth prime, we can directly use the formula $C(k) = \frac{p(n)}{n \log n}$ instead of any approximations of $C(k)$. This value of $C(k)$ works best for the range of 1 to $p(n)$. The figures comparing $C(k)$ approximations vs. real $C(k)$ in the matter of predicting primes (for the ranges of 1- 10^6th and 10^7th prime) demonstrate this fact. (See Figure 7 and Figure 8). However, it is important to also remember the discussions of subsection 3.1 (Asymptotic analysis of real and approximate $C(k)$).

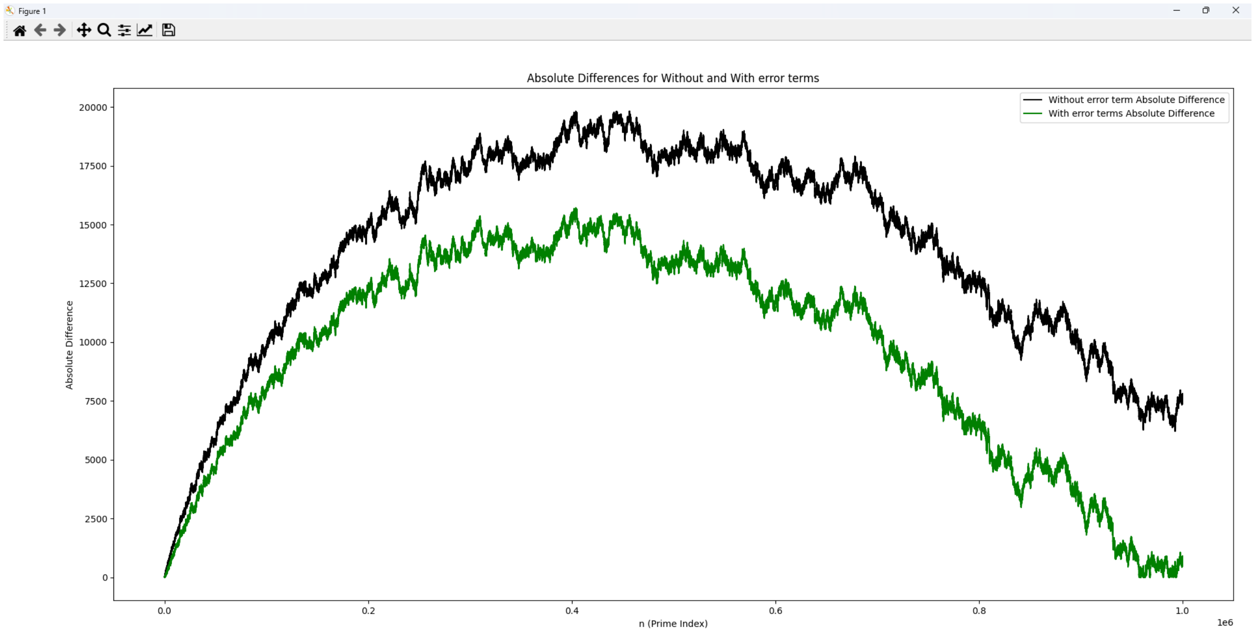

Figure 7: Comparison of $C(k)$ Approximations vs. Real $C(k)$ for $1 - 10^6$th Prime

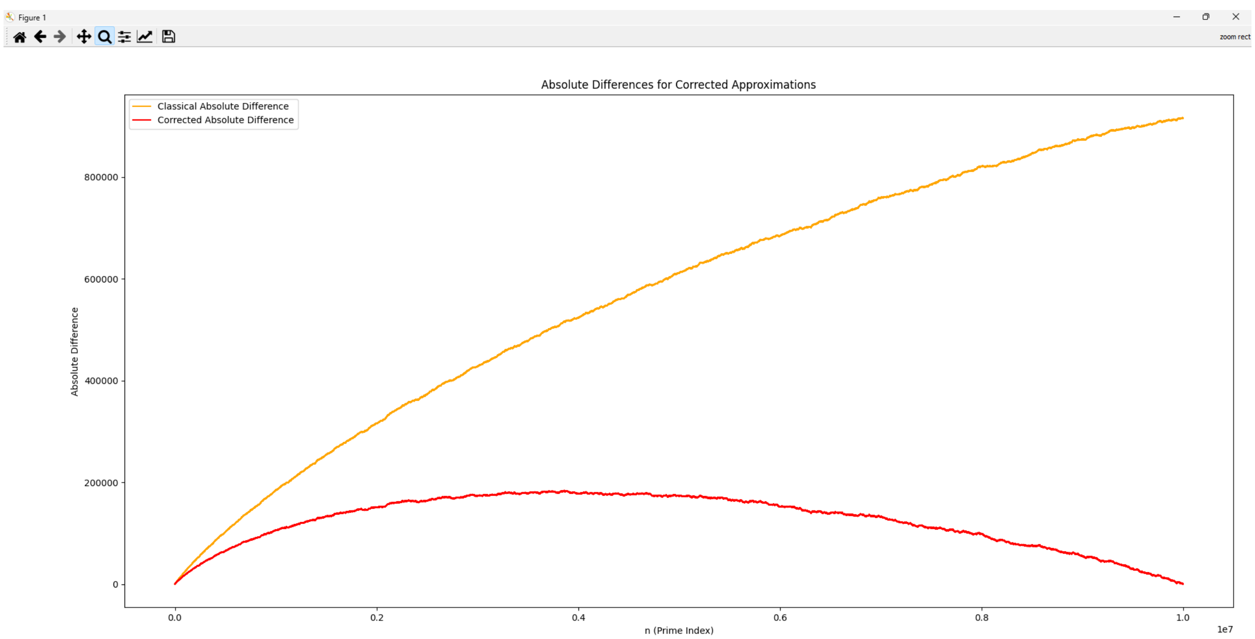

Figure 8: Comparison of $C(k)$ Approximations vs. Real $C(k)$ for $1 - 10^7$th Prime

10 Developing a piece wise system of equations for potential better approximations

In order to improve the prime prediction capability of our equation, we can convert the equation $p(n) \approx C(k) \cdot n \cdot \ln(n)$ into a piecewise function. The conditions in the piecewise function would help change the value of $C(k)$ relative to $p(n)$. Here is how this piecewise function can be represented:

$$p(n) = \begin{cases} C(k)_x \cdot n \cdot \ln(n) & \text{if } p(n)_x > p(n) > 1 \\ C(k)_y \cdot n \cdot \ln(n) & \text{if } p(n)_y > p(n) > p(n)_x \\ C(k)_z \cdot n \cdot \ln(n) & \text{if } p(n)_z > p(n) > p(n)_y \\ \ldots \end{cases}$$

The reader must note that $p(n)_x$, $p(n)_y$, and $p(n)_z$ in the piece wise functions need not be exact primes in practical algorithms. They have been represented as such for theoretical purposes. This piece wise approach allows for the variability in the number of conditions used, providing a more refined and accurate prediction of primes based on different ranges of $p(n)$. The reader must remember still, that this approach requires some clever way of implementation, in order to stay away from minute errors that may add up. The author, has yet to find an algorithm, which can help to use this approach correctly.

11 Validating the Concept of $C(k)$ Using Rosser's Theorem's Bounds

Rosser's Theorem provides important bounds for the n-th prime number, $p(n)$, which can be directly related to the correction factor $C(k)$ introduced in our work. This subsection explores how Rosser's Theorem ensures that $C(k)$ remains within well-defined boundaries.

11.1 Rosser's Theorem

Rosser's Theorem states that for $n \geq 1$:

$$n \log n < p(n) < n(\log n + \log \log n) \tag{1}$$

11.2 Correction Factor $C(k)$

The correction factor $C(k)$ is proposed to improve the approximation of $p(n)$ using:

$$p(n) \approx C(k) \cdot n \log n \tag{2}$$

11.3 Derivation of Bounds for $C(k)$

To show that $C(k)$ is valid within Rosser's bounds, we derive the range for $C(k)$.

11.3.1 Lower Bound

From Rosser's Theorem:

$$n \log n < p(n) \tag{3}$$

Substituting $p(n) \approx C(k) \cdot n \log n$:

$$n \log n < C(k) \cdot n \log n \tag{4}$$

Dividing both sides by $n \log n$:

$$1 < C(k) \tag{5}$$

11.3.2 Upper Bound

From Rosser's Theorem:

$$p(n) < n(\log n + \log \log n) \tag{6}$$

Substituting $p(n) \approx C(k) \cdot n \log n$:

$$C(k) \cdot n \log n < n(\log n + \log \log n) \tag{7}$$

Dividing both sides by $n \log n$:

$$C(k) < 1 + \frac{\log \log n}{\log n} \tag{8}$$

11.4 Combined Bound for $C(k)$

Combining both the lower and upper bounds:

$$1 < C(k) < 1 + \frac{\log \log n}{\log n} \tag{9}$$

11.5 Verification with Graphs

Empirical analysis using graphs shows that for $k > 5$, the bounds $1 < C(k) < 1 + \frac{\log \log n}{\log n}$ are satisfied. As k approaches infinity, the term $\frac{\log \log n}{\log n}$ diminishes, leading $C(k)$ to approach 1, which aligns with the asymptotic behavior of prime distribution.

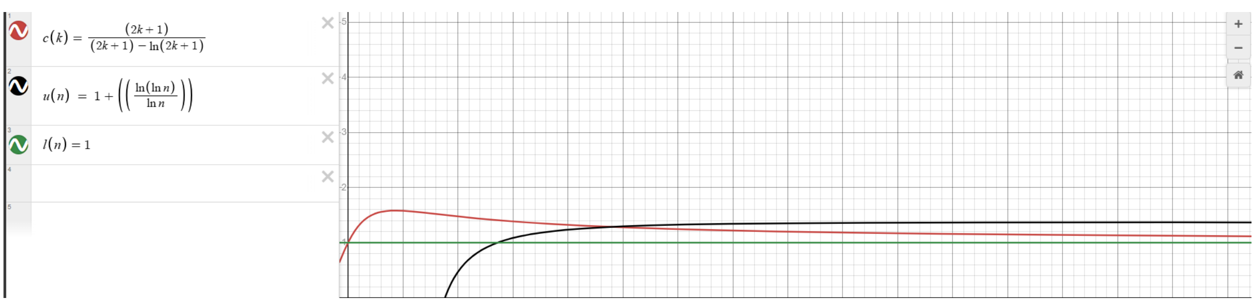

Figure 9: Graph shows that for $k > 5$, the bounds $1 < C(k) < 1 + \frac{\log \log n}{\log n}$ are satisfied.

Rosser's Theorem effectively supports the validity of the correction factor $C(k)$ by providing theoretical bounds within which $C(k)$ must lie. The bounds $1 < C(k) < 1 + \frac{\log \log n}{\log n}$ are consistent with our empirical findings and ensure that the refined approximation for $p(n)$ is both accurate and theoretically sound.

12 Deriving the Simplest Generalized $p(n)$ Formula on the Basis of $C(k)$

This subsection focuses on deriving the simplest generalized function for estimating the n-th prime number based on $C(k)$, which should only involve the variable n (and Euler's number e). While $C(k)$ works well for ranges between 2 and e^{k+1}, and can be used to approximate the n-th prime by setting $p \approx e^{k+1}$ and finding $C(k)$, it is not yet a generalized equation. The goal of this subsection is to address this and derive a more universal formula.

Based on the Prime Number Theorem, we know that:

$$p(n) \sim n \ln(n)$$

In our context, we use $p \approx e^{k+1}$. Therefore, we can equate:

$$e^{k+1} \sim n \ln(n)$$

Rearranging this, we have:

$$e^{k+1} = n \ln(n)$$

Taking the natural logarithm of both sides yields:

$$k + 1 = \ln(n \ln(n))$$

Thus:

$$k = \ln(n \ln(n)) - 1$$

Since $\ln(e) = 1$, we substitute:

$$k = \ln(n \ln(n)) - \ln(e)$$

Applying logarithm properties:

$$k = \ln\left(\frac{n \ln(n)}{e}\right)$$

Substituting this into $2k + 1$, we get:

$$2k + 1 = \ln\left(\frac{(n \ln(n))^2}{e}\right)$$

Substituting $2k + 1$ into the formula for $C(k)$, we have:

$$C(k) = \frac{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)}{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\right)}$$

Finally, substituting $C(k)$ into the formula for $p(n)$, we obtain:

$$p(n) = \frac{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) \cdot n \cdot \ln n}{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\right)}$$

If we do asymptotic analysis of the formula, we can obtain ourselves a rough formula,

$$p(n) \sim n \cdot \ln n \cdot \left(1 + \frac{\ln \ln n}{\ln n} - \frac{1}{2 \ln n}\right)$$

(Proofs and derivation of this formula done in Section 14, (Asymptotic Expansion of the Generalised Formula)

13 Limit of $p(n)$ as $n \to \infty$

Given the function:

$$p(n) = \frac{\left(\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) \cdot n \cdot \ln(n)\right)}{\left(\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right)\right)\right)}$$

we want to determine if it approaches $n \ln(n)$ as $n \to \infty$.

First, we simplify the logarithmic expressions:

$$\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) = \ln\left((n \cdot \ln(n))^2\right) - \ln(e) = 2\ln(n \cdot \ln(n)) - 1 = 2(\ln(n) + \ln(\ln(n))) - 1$$

Thus, the numerator becomes:

$$[2(\ln(n) + \ln(\ln(n))) - 1] \cdot n \cdot \ln(n)$$

For the denominator, we have:

$$\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right)\right) = 2(\ln(n) + \ln(\ln(n))) - 1 - \ln\left(2(\ln(n) + \ln(\ln(n))) - 1\right)$$

As $n \to \infty$, $\ln(n) \to \infty$ and $\ln(\ln(n)) \to \infty$, we approximate:

$$\ln\left(2(\ln(n) + \ln(\ln(n))) - 1\right) \approx \ln(2\ln(n)) = \ln(2) + \ln(\ln(n))$$

Hence, the denominator becomes:

$$2(\ln(n) + \ln(\ln(n))) - 1 - (\ln(2) + \ln(\ln(n))) = 2\ln(n) + \ln(\ln(n)) - 1 - \ln(2)$$

Putting it all together:

$$p(n) = \frac{[2(\ln(n) + \ln(\ln(n))) - 1] \cdot n \cdot \ln(n)}{2\ln(n) + \ln(\ln(n)) - 1 - \ln(2)}$$

For large n:

$$p(n) \approx \frac{2\ln(n) \cdot n \cdot \ln(n)}{2\ln(n)} = n\ln(n)$$

Therefore, the limit of $p(n)$ as $n \to \infty$ is:

$$\lim_{n \to \infty} p(n) = n\ln(n)$$

This limit implies that our new generalised equation converges to the approximation of the PNT, for extremely large n, hence maintaining our new equation's asymptotic correctness.

14 Error Bounds in Prime Number Approximation

In this section, we analyze the error bounds associated with two prime number approximation formulas: the traditional approximation $p(n) = n \ln n$ and our generalized formula. We employ Rosser's Theorem to understand the error bounds and derive functions that define the upper and lower bounds of the error.

14.1 Error Bounds for the Traditional Approximation

The traditional approximation for the n-th prime number is given by:

$$p(n) = n \ln n$$

While this formula is derived from the Prime Number Theorem and is asymptotically correct, it does not provide an exact value for the n-th prime number. The error associated with this approximation can be quantified using Rosser's Theorem.

14.1.1 Rosser's Theorem

Rosser's Theorem provides explicit bounds for the n-th prime number p_n. The theorem states that:

$$n \ln n < p_n < n(\ln n + \ln \ln n)$$

This inequality indicates that the actual n-th prime number lies between $n \ln n$ and $n(\ln n + \ln \ln n)$.

14.1.2 Deriving the Error Bounds

Using the bounds from Rosser's Theorem, we can define the error associated with the traditional approximation as:

$$\text{Error} = |p_n - p(n)|$$

From the inequalities provided by Rosser's Theorem, we have:

$$0 < |p_n - n \ln n| < n \ln \ln n$$

This expression indicates that the error is strictly positive and bounded above by $n \ln \ln n$. Therefore, the error bounds can be expressed as:

$$0 < j(n) < |p_n - p(n)| < i(n) = n \ln \ln n$$

where $j(n) = 0$ represents the lower bound, and $i(n) = n \ln \ln n$ represents the upper bound.

14.2 Error Bounds for the Generalized Formula

The generalized formula for estimating the n-th prime number is given by:

$$p(n) = \frac{\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) \cdot n \cdot \ln n}{\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n \cdot \ln(n))^2}{e}\right)\right)}$$

This formula is a refinement of the traditional approximation, incorporating additional logarithmic terms to improve accuracy.

14.2.1 Asymptotic Expansion of the Generalized Formula

To derive error bounds for the generalized formula, it is essential to analyze its asymptotic behavior as n becomes large. The generalized formula for the n-th prime number is:

$$p(n) = \frac{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\cdot n\cdot\ln n}{\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\right)}$$

This expression can be divided into two main components: the numerator and the denominator. We will first simplify these components individually before considering the asymptotic expansion.

Numerator Expansion The numerator of the generalized formula is:

$$\text{Numerator} = \ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\cdot n\cdot\ln n$$

We can simplify the logarithmic expression as follows:

$$\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) = \ln\left((n\cdot\ln(n))^2\right) - \ln(e)$$

Since $\ln(e) = 1$, this reduces to:

$$\ln\left((n\cdot\ln(n))^2\right) - 1 = 2\ln(n\cdot\ln(n)) - 1$$

Further, expanding $\ln(n\cdot\ln(n))$ yields:

$$\ln(n\cdot\ln(n)) = \ln(n) + \ln(\ln(n))$$

Therefore, the numerator becomes:

$$\text{Numerator} = [2(\ln n + \ln\ln n) - 1]\cdot n\cdot\ln n$$

This expression suggests that the dominant term in the numerator for large n is:

$$2\ln n\cdot n\cdot\ln n$$

with the term involving $\ln\ln n$ contributing a smaller adjustment.

Denominator Expansion The denominator of the generalized formula is:

$$\text{Denominator} = \ln\left(\frac{(n\cdot\ln(n))^2}{e}\right) - \ln\left(\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)\right)$$

Using the earlier simplification for $\ln\left(\frac{(n\cdot\ln(n))^2}{e}\right)$, the denominator becomes:

$$2(\ln n + \ln\ln n) - 1 - \ln\left(2(\ln n + \ln\ln n) - 1\right)$$

For large n, the logarithmic subtraction term $\ln\left(2(\ln n + \ln\ln n) - 1\right)$ can be approximated as:

$$\ln(2\ln n + 2\ln\ln n - 1) \approx \ln(2\ln n) = \ln(2) + \ln(\ln n)$$

Substituting this into the denominator:

$$\text{Denominator} \approx 2\ln n + 2\ln\ln n - 1 - (\ln(2) + \ln(\ln n))$$

For large n, the dominant term is $2\ln n$, while the other terms provide a smaller correction.

Combining the Results Given the expansions of the numerator and denominator, we can approximate the generalized formula for large n as:

$$p(n) \sim \frac{[2(\ln n + \ln\ln n) - 1] \cdot n \cdot \ln n}{2\ln n + 2\ln\ln n - 1 - \ln(2) - \ln(\ln n)}$$

This can be simplified by canceling out the dominant terms $2\ln n$, leading to:

$$p(n) \sim n \cdot \ln n \cdot \left(1 + \frac{\ln\ln n}{\ln n} - \frac{1}{2\ln n}\right)$$

This expression reveals that the generalized formula corrects the traditional approximation by adding terms that account for the slower-growing $\ln\ln n$ and a small negative correction term $\frac{1}{2\ln n}$. These adjustments make the generalized formula more accurate for large n.

Interpretation of the Asymptotic Expansion The asymptotic expansion shows that the generalized formula provides a refined estimate of $p(n)$ by adjusting the traditional $n\ln n$ approximation. The additional terms $\frac{\ln\ln n}{\ln n}$ and $-\frac{1}{2\ln n}$ contribute to a more accurate estimation, particularly as n becomes large.

14.2.2 Deriving the Error Bounds for the Generalized Formula

The generalized formula adjusts the estimate slightly upwards compared to the traditional approximation. This adjustment leads to a smaller error, particularly for large n.

Given the additional correction term $\frac{\ln\ln n}{\ln n}$, the error bounds can be defined as follows:

$$0 < j_{\text{generalized}}(n) < |p_n - p_{\text{generalized}}(n)| < i_{\text{generalized}}(n)$$

Here, the upper bound $i_{\text{generalized}}(n)$, is approximated using a wise guess, based on correction term $\frac{\ln\ln n}{\ln n}$:

$$i_{\text{generalized}}(n) \to n \cdot \frac{\ln\ln n}{\ln n}$$

All our analysis suggest that the upper bounds of the new generalised equation will be smaller (may be significantly smaller) than the upper bound for the traditional formula, reflecting the accuracy of our new formula.

14.3 Comparative Analysis of Error Bounds

The error bounds derived for both formulas highlight the advantages of the generalized formula over the traditional approximation. While Rosser's Theorem provides a strict upper bound for the traditional formula, the generalized formula achieves a tighter bound by incorporating additional logarithmic terms.

- **Traditional Approximation:** The error is bounded by $0 < |p_n - n\ln n| < n\ln\ln n$.

- **Generalized Formula:** The error is bounded by $0 < j_{\text{generalized}}(n) < |p_n - p_{\text{generalized}}(n)| < n \cdot \frac{\ln \ln n}{\ln n}$.

These bounds demonstrate the superiority of the generalized formula in providing a more accurate estimate of prime numbers, particularly for large n.

14.4 Conclusion

In conclusion, the generalized formula offers an improvement in the accuracy of prime number approximation by reducing the error bounds. This improvement is particularly evident when comparing the traditional formula's error bound of $n \ln \ln n$ to the generalized formula's tighter bound. The application of Rosser's Theorem provides a framework for understanding these error bounds.

15 Testing Our Generalized Equations

In this section, we test our generalized equations through empirical analysis.

15.1 Test #1

The following graph compares the absolute errors of two formulae: one, $C(k) \cdot n \ln(n)$ (optimized for the range of 2 to 10^6th prime) and second, our new generalized equation. By observing the graph, one can deduce that the new generalized equation still performs pretty well, and our hypothesis is that the slight decrease in performance is likely due to the fact that we assumed $e^{k+1} = n \ln(n)$ directly for simplicity.

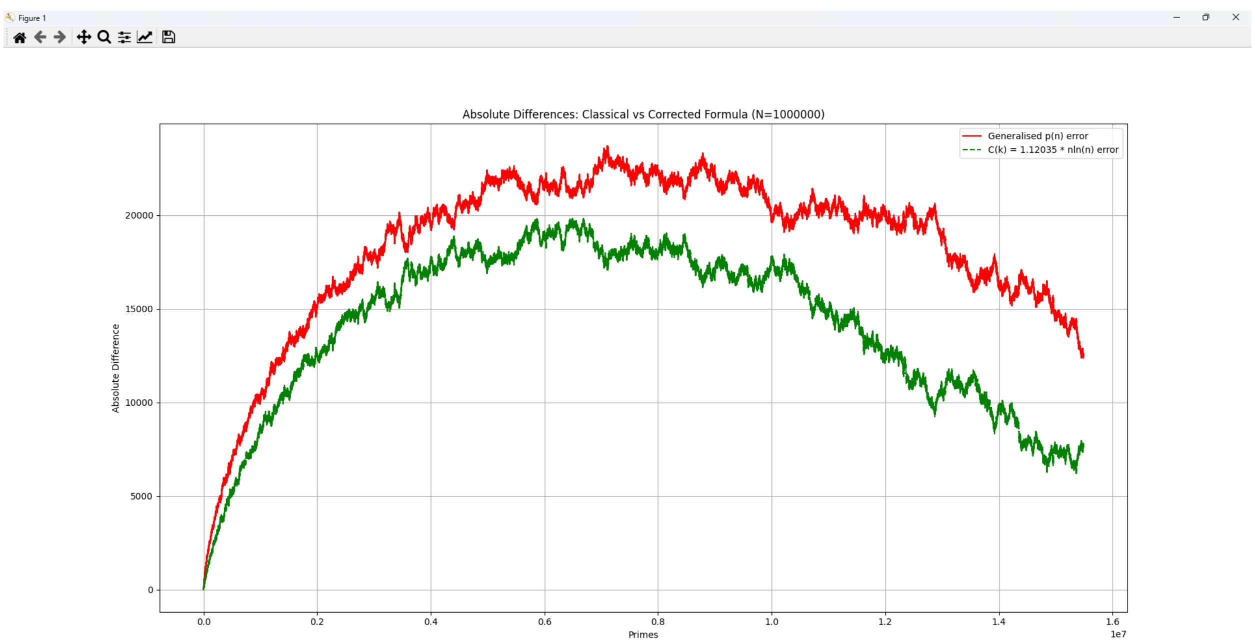

Figure 10: Comparison of Absolute Errors

15.2 Test #2

Though the graph of the analysis couldn't be unfortunately supplied. (Due to limitations in computational power), but we have observed that for the 10^7th prime, the error observed is 0.4%. For 2×10^{17} th prime, the error observed is 1.29% [2]

However, one thing interesting and important is that, comparatively, at least for upto the 10^7th prime, the $C(k)$ formula consistently shows low errors. (For the 10^7th prime, the error observed for the 10^7th prime is 0.25% (we have discussed about why this may be happening)

However, one can still argue that the generalised formula for $p(n)$ has it's own benefits. The major benefit is that, compared to the $C(k)$ $framework$, the generalised formula provides a more deterministic approach. Meaning , that the generalised formula will always produce the same results, while the $C(k)$ framework, doesn't guarantee that. Different users could use different $C(k)$ for the same range, or , their approximations of $C(k)$ used could be different. Some users, would be interested in applying the Ideal $C(k)$ framework, while others could use the different approximations of $C(k)$. In a sort of informal conclusion, we can conclude that the whole system of $C(k)$ framework and the generalised formula, are both useful for achieving different aims regarding prime number estimations.

16 Conclusion

This paper introduces a correction factor $C(k)$ to improve the Prime Number Theorem (PNT) for estimating the nth prime number. By refining the traditional approximation $p(n) \approx n \ln n$ to $p(n) \approx C(k) \cdot n \cdot \ln n$, we demonstrated enhanced accuracy in prime prediction. A generalized formula for $p(n)$ was derived, which eliminates the need for variables other than n and e. Empirical analyses show that this new approach performs well compared to traditional methods, although the simplified assumption $e^{k+1} = n \ln(n)$ introduces slight deviations. The findings suggest that these innovations can lead to more accurate and computationally efficient prime number approximations.

17 References

References

[1] Brilliant.org, "Distribution of Primes", https://brilliant.org/wiki/distribution-of-primes/, accessed July 21, 2024.

[2] Wikipedia, "Prime Number Theorem", https://en.wikipedia.org/wiki/Prime_number_theorem, accessed August 1, 2024.