1 Introduction

The Prime Number Theorem (PNT) describes the asymptotic distribution of prime numbers. It states that the number of primes less than or equal to \( n \), denoted \( \pi(n) \), is approximately \( \frac{n}{\log n} \). This provides a useful approximation for the number of primes up to \( n \).

The PNT also aids in estimating the \( n \)th prime number, \( p(n) \), using \( p(n) \approx n \log n \). This work explores slight improvements to the PNT, through intuition, logic, assumption, and more.

2 The Correction Factor \( C(k) \)

I began my thinking by assuming the existence of a function \( C(k) \), which adjusts the estimation of \( p(n) \). Let \( C(k) \) be a function such that:

\[
p(n) \approx C(k) \cdot n \cdot \log n
\]

The function \( C(k) \) aims to account for the complex and unpredictable distribution of primes, and I assumed that this \( C(k) \) only works for a specific range of primes.

3 Finding \( C(k) \)

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

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

Accounting for the concept of \( C(k) \) working most effectively for a "defined" set of ranges, I had the idea of finding the "best \( n \)" for any true prime \( p \), and then using that best \( n \) to find \( C(k) \). Mathematically, this means:

\[
p - n \log n \rightarrow 0
\]

Or, we can assume that:

\[
p \approx n \log n
\]

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

The main idea for using $n = \frac{p}{\log p}$ is to simplify our expression and to reduce calculation load. This justification is done assuming that $n$ tends to infinity (Asymptotic value of $n$). As we have to prove that changing the value of $n$ doesn’t alter $C(k)$ that much, we will take the ratio of $C(k)_r$ (actual value of $C(k)$) and $C(k)_a$ (assumed value of $C(k)$). Hence, we have to prove that $\frac{C(k)_a}{C(k)_r} \to 1$. $C(k)_a = \frac{\log p}{\log \left( \frac{p}{\log p} \right)}$ and $C(k)_r = \frac{\log p}{\log p - \log(\log p)}$. If we expand $C(k)_a$, we get $\frac{\log p}{\log p - \log(\log p)}$.

Taking the ratio of $C(k)_a$ and $C(k)_r$:

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

According to PNT, $n \log n \approx p$ (primes). Hence, we get:

$$\frac{\log p \cdot p}{p \cdot (\log p - \log(\log p))}$$

As we are doing asymptotic analysis, we can ignore $\log(\log p)$ as it grows slowly compared to other terms. Hence, we get the ratio as:

$$\frac{p \log p}{p \log p} \to 1$$

Hence, we can say that $\frac{C(k)_a}{C(k)_r} \to 1$, and $C(k)_a \sim C(k)_r$.

### 4 What’s $p$?

As we have approximated $n$ from $\frac{p}{\log n}$ to $\frac{p}{\log p}$ for easier calculations, directly using $p$ as the exact value of the prime won’t yield accurate results. We have to modify the value of $p$. According to empirical calculations, for a range of primes from the 1st prime $P(1)$ to the $N$th prime $P(N)$:

$$p \approx P(N) \cdot P \left( \frac{N}{2} \right)$$
5 Can this equation be modified further for huge primes?

For extremely large primes, we can approximate $P(N)$ as $e^{k+1}$, and assuming $P(N)$ is approximately twice $P(N/2)$, we can approximate $P(N/2)$ as $e^k$. Substituting these values into the formula for $C(k)$:

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

Hence, for very large primes, we approximate:

$$p(n) \approx \frac{2k + 1}{(2k + 1) - \log(2k + 1)} \cdot n \cdot \log n$$

where $C(k) = \frac{2k+1}{(2k+1) - \log(2k+1)}$ is optimized for a range of primes from 2 to $e^{k+1}$.

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:

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

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 Behavior of $C(k)$

As $k \rightarrow \infty$, $\log(2k + 1)$ grows less significantly compared to the other terms in $C(k)$. Therefore, as $k \rightarrow \infty$, $C(k) \rightarrow 1$. This implies that our new equation for $p(n)$ converges to the traditional approximation $n \log n$ as $k$ approaches infinity.

8 Empirical Analysis

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

For a range from the 1st prime to the 100th 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 Counting Formulas](image)

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 $\pi(n) \sim \frac{2k+1}{\log n} \left( \frac{\log(2k+1)}{\log n} \right)^n$.

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

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 $10^6$th prime.
1. Absolute difference vs \(n\)th prime for our new equation (see Figure 3). 2. Percentage errors vs \(n\)th 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 2: Accuracy of Prime Counting Algorithms

Figure 3: Absolute Difference vs \(n\)th Prime for Corrected Approximations
Figure 4: Percentage Errors vs $n$th 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