Prime Number Analysis Using Analytic and Probabilistic Approaches

Ethan Ashbrook, Collin Jung, Akash Pardeshi, James Schmidt, Faculty Adviser

Introduction

Goal

Learn about prime numbers via methods regarding calculus and probability.

A prime number is any number whose only divisors are one and itself. In number theory, many interesting facets of primes have been discovered and explored. In certain respects, however, primes remain mysterious and understanding them remains a challenge. We wanted to find ways that we could research primes that were more advanced then just finding the next largest prime number.

Two Approaches:

- Analytically: use the Prime Number Theorem to approximate primes with a continuous function.
- Probabilistically: treat primes as random variables and employ probabilistic functions.

1. Analytic Approach

The analytic approach is centered around the Prime Number Theorem, a statement relating the prime numbers to another function:

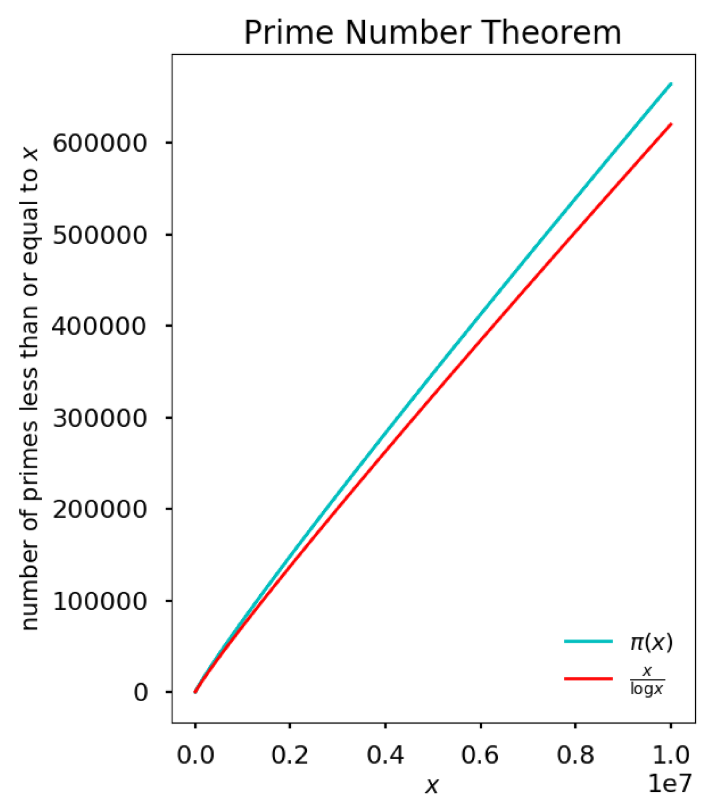
Theorem 1

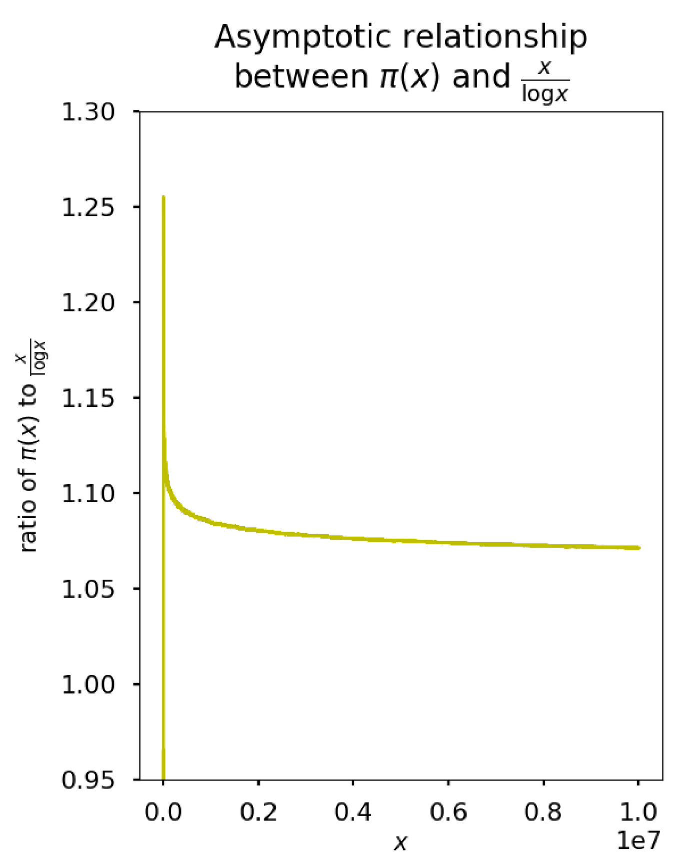
Let $\pi(x)$ denote a function that gives the number of primes below x. The Prime Number Theorem (PNT) states that

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

In other words, $\pi(x)$ begins to behave like $\frac{x}{log(x)}$ as $x \to \infty$

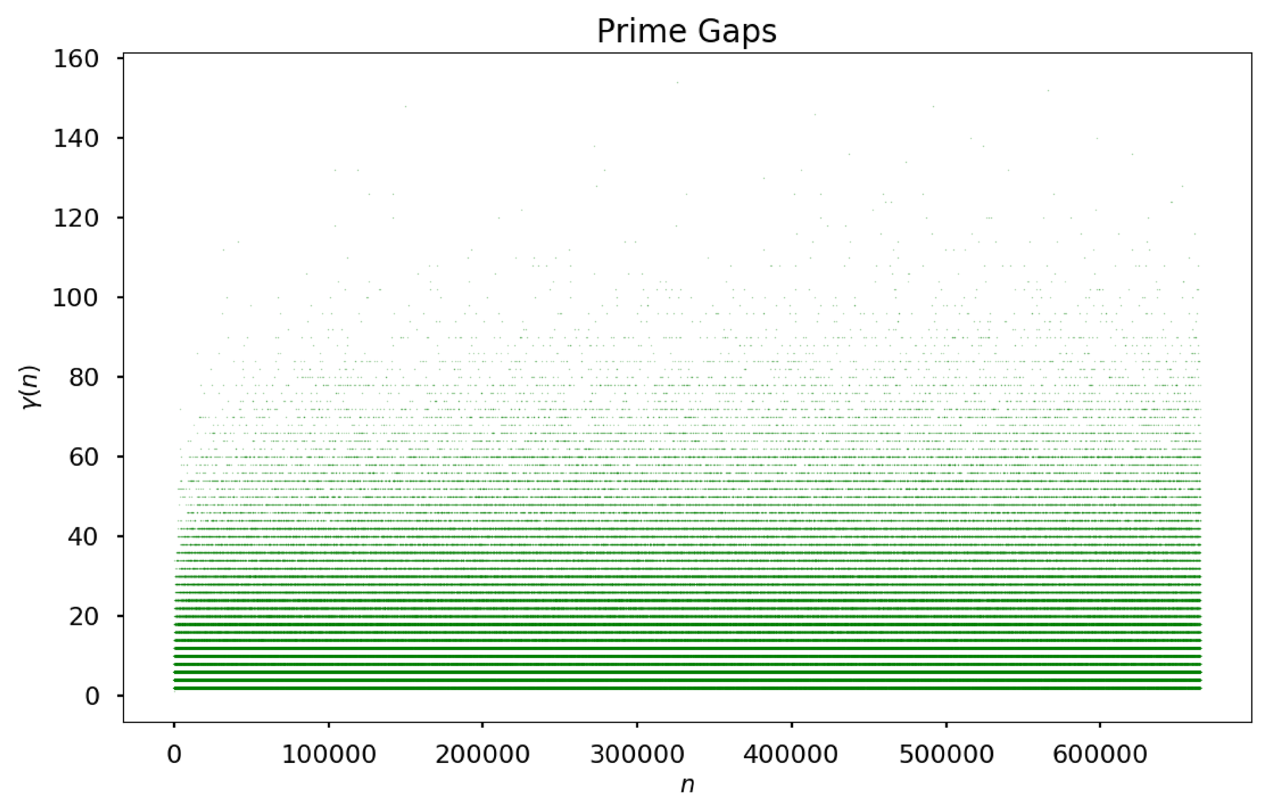
The ratio of the two functions approaches 1 as $x \to \infty$, however that doesn't necessarily mean that the two functions are equal. As shown below, they appear to diverge as x gets bigger, however their ratio converges to 1.





2. Probabilistic Method

If we treat the primes in a probabilistic manner, then we can use ideas from probability theory to understand the primes. Define a function $\gamma(n)$ that outputs the nth prime gap: $p_{n+1} - p_n$. By treating $\gamma(n)$ as a random variable, we may apply theorems from probability to it.



The prime gaps up to $n = 10^7$.

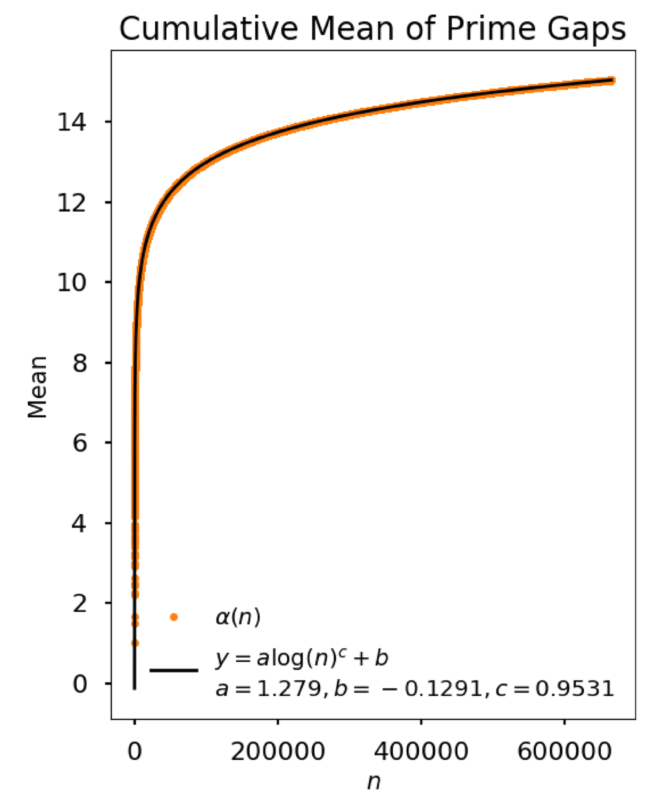
In searching for regularity in $\gamma(n)$, we defined a data set $\mathcal{D}_n$ which we could analyze.

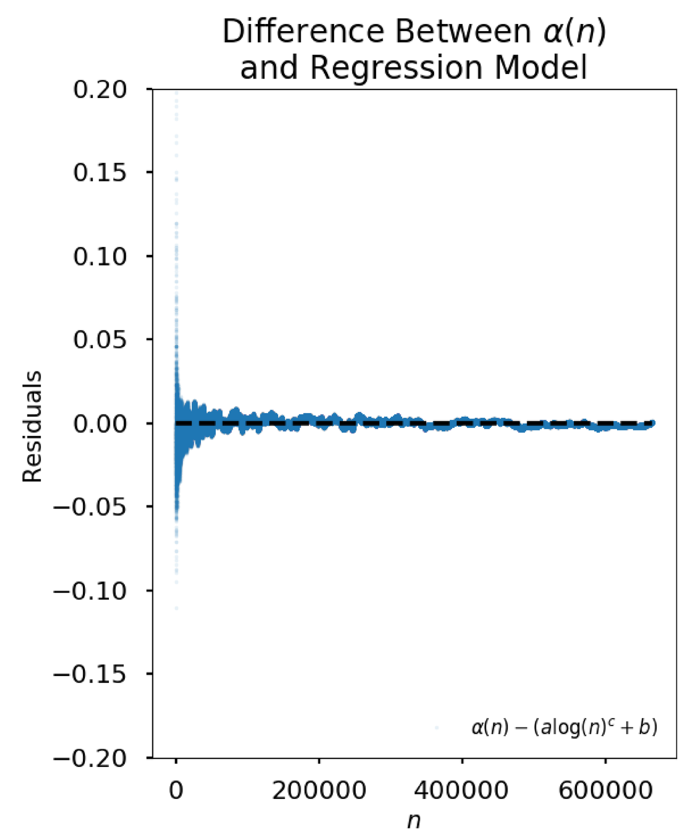
$$\mathcal{D}_n = \{\gamma(1), \gamma(2), \ldots, \gamma(n)\}$$

We defined functions $\alpha(n)$ and $\beta(n)$ that output the mean and variance of $\mathcal{D}_n$, respectively.

$$\alpha(n) = \frac{1}{n}\sum_{i=1}^{n}\gamma(i) \qquad \beta(n) = \frac{1}{n}\sum_{i=1}^{n}(\alpha(n) - \gamma(i))^2$$

We plotted $\alpha(n)$ and $\beta(n)$ and discovered that the data seemed to follow a pattern. Given the apparent regularity, we performed regression with the model $a\log(n)^c + b$.





$\alpha(n)$ is shown to $n = 10^7$ along with the regression model. The residual plot seems to indicate the differences converge to 0.

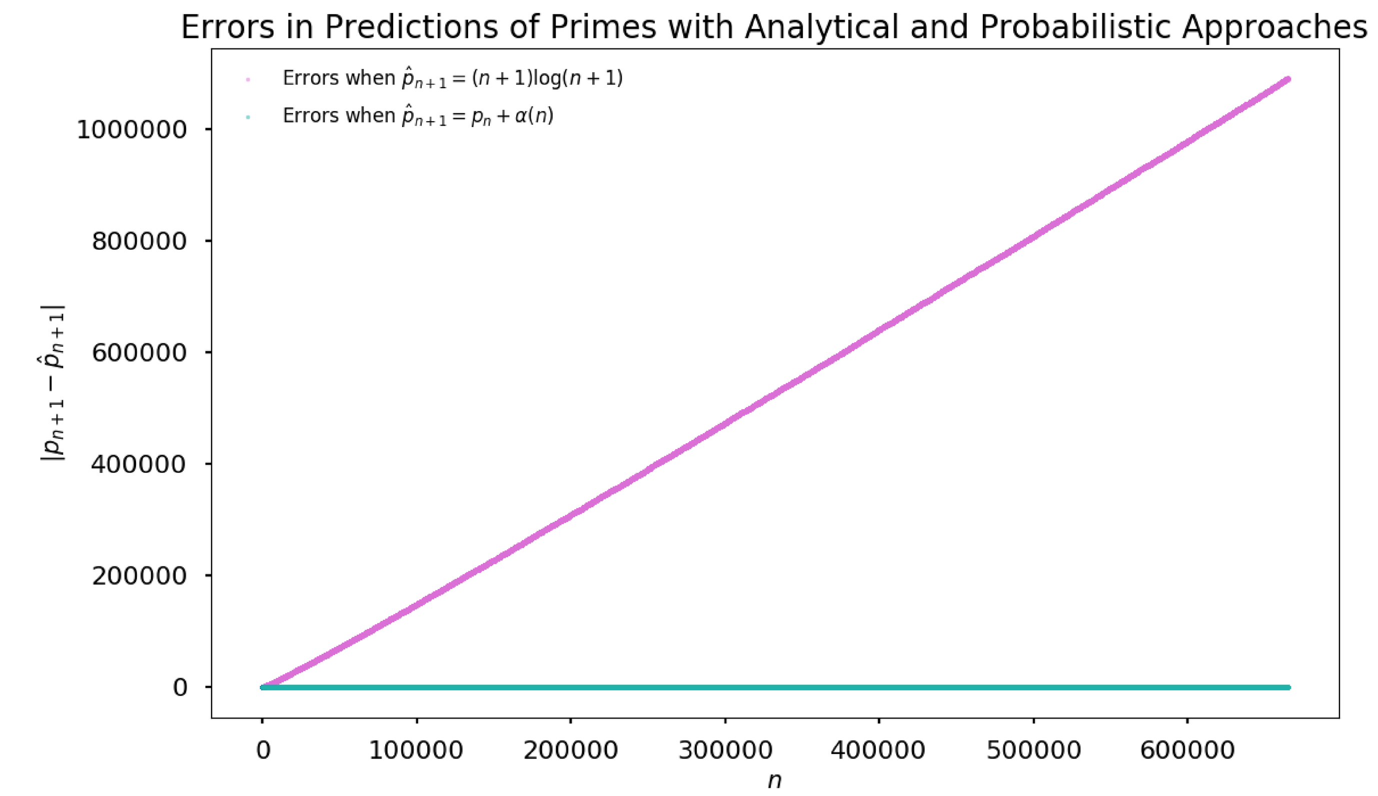
Comparisons

With two different ways of understanding primes, we decided to compare how we could use both models to potentially predict where primes would be.

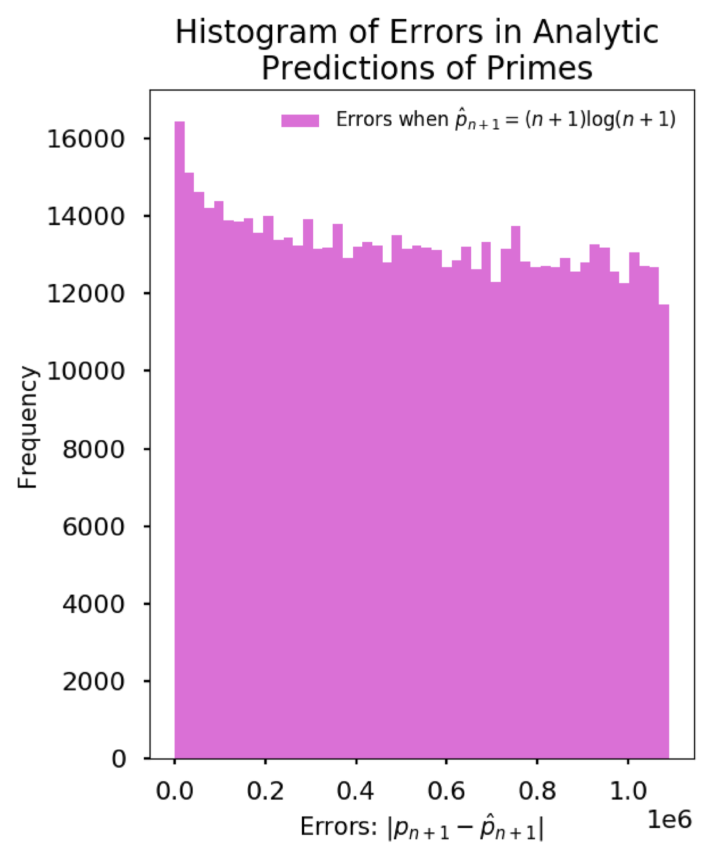
With an analytic approach, we can say $p_n \sim n\log(n)$, as a consequence of the PNT. Although this does not imply $p_n \to n\log(n)$, the best approximation we can make is $n\log(n)$.

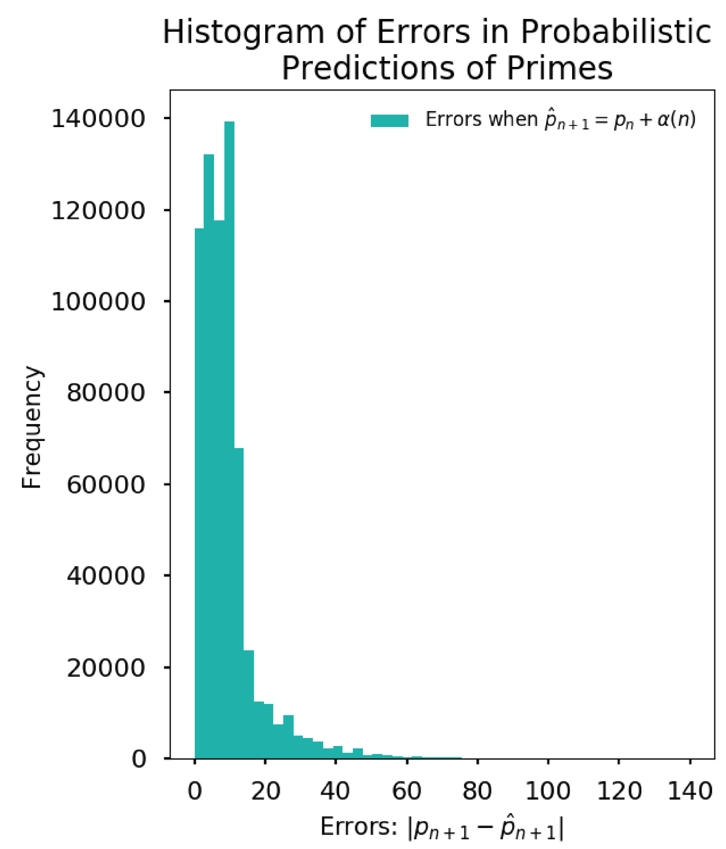
With the probabilistic approach, we take $p_{n-1} + \alpha(n-1)$ to be an approximation of the nth prime.

A natural measure of the accuracy of the predictions is to take the difference between the actual and predicted values, in absolute value: $Er(n) = |p_n - \hat{p}_n|$ where $\hat{p}_n$ is an approximation of the nth prime.



The first ten million errors for the two types of predictions.





Histograms of the first ten million errors for both methods

Conclusions

- Although the probabilistic method requires a stronger assumption, namely information about the previous prime, the conclusion is much stronger.
- Primes can be understood in different ways and each approach provides different insights into the nature of primes.

This IGL project was made possible by the generous support of the David Frankel ('76) Fund.