# The Distribution of Prime Numbers and the Gaps Between Primes

The prime numbers are a random sequence of natural numbers, which have been studied as far back as the ancient Greeks, with even Euclid constructing proofs in his 'Elements'. As they are a random sequence of numbers, it is impossible to find a formula or iterative process to accurately and efficiently locate primes at a macroscopic scale.

The largest prime to have ever been found is $23,249,425$ digits long and was found using computational brute force in the Great Internet Prime Search after almost 2 years of searching (for which the previous was $22,338,618$ digits long); which, with the computation power at our disposal in the modern day, shows the difficulty of finding large primes.

The mystery that the primes wield reaches across all of number theory is vast, and due to their random nature, many seemingly trivial results and conjectures, such as Goldbach's conjecture, appear very difficult to conceive and construct proofs for, even when their implications can be relatively simple.

In my essay, I would like to focus primarily on the gaps between the prime numbers, both in an arbitrary sense, and have a brief discussion on more specific gaps between subsequent prime numbers, namely the twin prime conjecture; and also wish to talk about the general density of the primes on a macroscopic scale.

## 1 The Endless Sequence

In this chapter I want to focus on arguably the most important property of the prime numbers: that there are infinitely many of them. If there were to be finitely many, we would be able to construct a complete set of the prime

numbers and analyse all their properties with ease. However, as we shall see soon, this is not the case with the prime numbers, which leads rise to the study of primes being far more difficult.

Before we prove the existence of infinitely many primes, we must first introduce our first theorem, The Fundamental Theorem of Arithmetic, which shall allow us to prove the main result in this chapter, and will prove key in proving results later on:

**Theorem 1.1 - The Fundamental Theorem of Arithmetic:**

1. Every positive integer greater than 1 may be written as the product of prime numbers
2. This product is unique up to permutations

A proof for this theorem may be found in MA132 Foundations.

The principle behind the existence of this theorem is very intuitive. Prime numbers, by definition, may not be factored further than themselves, so therefore it makes sense that they can be viewed as the 'building blocks' of the natural numbers. The uniqueness part of this theorem is more subtle, but also follows from the composition of natural numbers, and the concept of equivalence between numbers.

The Fundamental Theorem of Arithmetic is very important to the study of the prime numbers, as allows us to form a connection between the primes; which are very unpredictable and rare as they get large; and the natural numbers, for which we have a vast understanding, and can easily construct proofs with.

Now on to the main theorem of this section, the proof that there are infinitely many prime numbers. This has been proved using various different methods; but our proof, originally discovered by Euler, gives rise to a result which will be useful later:

**Theorem 1.2:** There are infinitely many prime numbers

**Note:** Throughout this essay, we will denote the set of prime numbers by $\mathbb{P}$

**Proof:** First we consider the following:

$$\begin{aligned}\prod_{p\in\mathbb{P}}\left(\sum_{k\geq 0}\frac{1}{p^k}\right) &= \left(\sum_{k\geq 0}\frac{1}{2^k}\right)\left(\sum_{k\geq 0}\frac{1}{3^k}\right)\left(\sum_{k\geq 0}\frac{1}{5^k}\right)\dots \\ &= \sum_{k,l,m,\dots\geq 0}\frac{1}{2^k\cdot 3^l\cdot 5^m\cdot\dots} \\ &= \frac{1}{2^0\cdot 3^0\cdot 5^0\cdot\dots}+\frac{1}{2^1\cdot 3^0\cdot 5^0\cdot\dots}+\frac{1}{2^0\cdot 3^1\cdot 5^0\cdot\dots}+\dots\end{aligned}$$

By the Fundamental Theorem of Arithmetic, $\forall n\in\mathbb{N}_{>1}\exists!$ prime factorisation s.t. $n=2^{k_2}\cdot 3^{k_3}\cdot 5^{k_5}\cdot\dots$ for $k_i\geq 0$

$$\begin{aligned}\Rightarrow\prod_{p\in\mathbb{P}}\left(\sum_{k\geq 0}\frac{1}{p^k}\right) &= \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\dots \\ &= \sum_{n=1}^{\infty}\frac{1}{n}\end{aligned}$$

As $p$ is prime, we know that $\frac{1}{p}<1$, meaning $\sum_{k\geq 0}\frac{1}{p^k}=\frac{1}{1-\frac{1}{p}}$ by the formula for infinite geometric series.

$$\therefore\prod_{p\in\mathbb{P}}\left(\frac{1}{1-\frac{1}{p}}\right)=\sum_{n=1}^{\infty}\frac{1}{n}$$

As we know the RHS diverges, and that every term in the LHS is finite, this means that the LHS must be an infinite product. This in turn means that there are infinitely many prime numbers. $\square$ (Dunham, 1999)

Unsurprisingly, this theorem holds. We shall discuss the connotations of there being infinitely many primes later, but for now, we shall prove and discuss as different result which will allow us an insight into the density of the primes:

**Theorem 1.3:** $\sum_{p\in\mathbb{P}}\frac{1}{p}$ diverges

**Proof:** We have already shown that:

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p \in \mathbb{P}} \left( \frac{1}{1 - \frac{1}{p}} \right)$$

$$\begin{aligned}
\Rightarrow \log \left[ \sum_{n=1}^{\infty} \frac{1}{n} \right] &= \log \prod_{p \in \mathbb{P}} \left( \frac{1}{1 - \frac{1}{p}} \right) \\
&= \sum_{p \in \mathbb{P}} \log \left( \frac{1}{1 - \frac{1}{p}} \right) \\
&= \sum_{p \in \mathbb{P}} - \log \left( 1 - \frac{1}{p} \right) \\
&= \sum_{p \in \mathbb{P}} \left( \frac{1}{p} + \frac{1}{2p^2} + \frac{1}{3p^3} + \dots \right) \\
&= \sum_{p \in \mathbb{P}} \left( \frac{1}{p} \right) + \frac{1}{2} \sum_{p \in \mathbb{P}} \left( \frac{1}{p^2} \right) + \frac{1}{3} \sum_{p \in \mathbb{P}} \left( \frac{1}{p^3} \right) + \dots
\end{aligned}$$

We know that $\sum_{n=1}^{\infty} \frac{1}{n^k}$ converges $\forall k > 1$, meaning that $\sum_{p \in \mathbb{P}} \left( \frac{1}{p^k} \right)$ converges $\forall k > 1$.

$$\therefore \log \left[ \sum_{n=1}^{\infty} \frac{1}{n} \right] = \sum_{p \in \mathbb{P}} + N(*)$$

where

$$N = \sum_{i=2}^{\infty} \left[ \frac{1}{i} \sum_{p \in \mathbb{P}} \left( \frac{1}{p^i} \right) \right] \in \mathbb{R}$$

This means that because the LHS of $(*)$ is infinite, the RHS is also infinite. So as $N \in \mathbb{R} \Rightarrow \sum_{p \in \mathbb{P}} \frac{1}{p}$ diverges. $\square$ (Dunham, 1999)

This is quite an interesting result when we analyse it further. We know from MA131 that the sums of the reciprocals of the square numbers converges to $\frac{\pi^2}{6}$ (otherwise known as the Basel Problem). Therefore, if we compare the sequences of the partial sums of each series, we can compare their term by term progressions. We know that by the comparison test, if the sequence of partial sums is eventually less than that of a convergent

series, than that series itself converges. This means that because the infinite series of reciprocals prime numbers diverges, there reaches a point where its sequence of partial sums is always greater than that of the squares. This in turn means that the prime numbers are denser than the squares.

This is an interesting result as it is known that the $n^{th}$ and $(n-1)^{th}$ square numbers differ by the $n^{th}$ odd number, so it is well understood how the squares are distributed in the natural numbers, and they appear to be seemingly frequently, especially considering that there is a 1-million-digit difference between the largest two known prime numbers as mentioned before.

This result may be applied to other series of natural numbers and their reciprocals, such as the cubes, to gage how the primes are distributed in on the macrocosmic scale.

## 2 Arbitrary Gaps

We have gauged the density of the primes in comparison with other sequences of natural numbers, but here we shall explore the possibility of having 'arbitrary' lower and upper bounds between subsequent prime numbers.

To begin, we shall investigate constructing lower bounds for gaps between prime numbers, of any size. Naturally, due to the random and infinite nature of the prime numbers, one would expect any gap between consecutive prime numbers to occur, and nevertheless infinitely often.

Now let us compose one of these intervals:
Let $n \in \mathbb{N}$
Consider $n! = 1 \cdot 2 \cdot \ldots (n-1) \cdot n$
$\Rightarrow \forall i = 1, 2, \ldots, n : n \mid n!$
$\Rightarrow \forall i = 2, \ldots, n : n + i \nmid n! + i$
$\Rightarrow \forall m \in [n! + 2, n! + n], m \notin \mathbb{P}$

Therefore; there are gaps between prime numbers of size at least n-1.

Obviously for this construction, as n gets very large, these gaps appear several orders of magnitude apart from each other, which seemingly becomes obsolete, but it does however prove the existence of gaps between prime numbers that get arbitrarily large.

This result does have some sense to it. The conditions for a number to be prime become increasingly difficult to satisfy as they become large; as for a prime number p, all numbers up to $\sqrt{p}$ inclusive must be verified to not factor into p.

However, this does not tell the whole story. In a similar way to a lower bound for gaps between prime numbers, we are able to construct a maximum interval between two consecutive primes. However, as we shall see, this gap does indeed become arbitrarily large.

We shall now investigate Bertrand's Postulate, a statement that gives us such a bound between prime numbers. However, we must prove a series of lemmas which will ultimately allow us to prove this theorem.

Firstly we shall define:

$$C_n := \binom{2n}{n}$$

We shall be using a proof by contradiction for our proof of Bertrand's Postulate, where we create a false inequality using $C_n$ and its prime factors. This brings us to our first lemma:

**Lemma 2.1:** $\forall n \in \mathbb{N}$:

$$C_n \geq \frac{4^n}{2n}$$

**Proof:**

$$\begin{aligned}
4^n &= 2^{2n} \\
&= (1+1)^{2n} \\
&= \sum_{k=0}^{2n} \binom{2k}{k} \\
&= 2 + 2\sum_{k=0}^{n-1} \binom{2k}{k} \\
&\leq 2 + (2n-1)\binom{2n}{n} \\
&\leq 2n\binom{2n}{n} \\
&= 2nC_n \quad \square
\end{aligned}$$

Now that we have produced a lower bound that we may use, we must now search for an upper bound. This is a multi-step process and we shall begin with our next lemma:

**Lemma 2.2:** $\forall n \in \mathbb{N}$, none of the prime powers of $C_n$, that is to say the greatest power of prime factors that divide $C_n$, exceed $2n$

**Example:** Let $n = 4$: $C_n = 70 = 2 \cdot 5 \cdot 7$. $2^1$,$5^1$,$7^1 < 2(4) = 8$.

**Proof:** Let $n \in \mathbb{N}$, $p \in \mathbb{P}$

We shall denote the highest power $k$ of $p$ s.t. $p^k \mid n$ by $v_p(n)$
Notice that $v_p(n) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \ldots$; and if $p^k > n$ then $\left\lfloor \frac{n}{p^k} \right\rfloor = 0$

$$\begin{aligned}
C_n = \binom{2n}{n} &= \frac{(2n)!}{2(n!)} \\
\Rightarrow v_p(C_n) &= v_p((2n)!) - 2v_p(n!) \\
&= \left( \left\lfloor \frac{2n}{p} \right\rfloor - 2\left\lfloor \frac{n}{p} \right\rfloor \right) + \left( \left\lfloor \frac{2n}{p^2} \right\rfloor - 2\left\lfloor \frac{n}{p^2} \right\rfloor \right) + \ldots
\end{aligned}$$

If $p^k > 2n$ $(> n)$, then $\left( \left\lfloor \frac{2n}{p^k} \right\rfloor - 2\left\lfloor \frac{n}{p^k} \right\rfloor \right) = 0 - 2(0) = 0$

$\forall a, b > 0$, $\lfloor a+b \rfloor - \lfloor a \rfloor - \lfloor b \rfloor = 0$ or $1 \Rightarrow \lfloor 2a \rfloor - 2\lfloor a \rfloor = 0$ or $1$
$\therefore$ For $p^k \leq 2n$: $\left( \left\lfloor \frac{2n}{p^k} \right\rfloor - 2 \left\lfloor \frac{n}{p^k} \right\rfloor \right) = 0$ or $1$
$\Rightarrow v_p(C_n) \leq k$ for the largest integer $k$ s.t. $p^k \leq 2n$
$\Rightarrow p^{v_p(C_n)} \leq 2n \quad \square$

We now have the first component of our upper bound, but we still need more information to complete our proof.

Our next proposition will narrow down the interval in which we find the prime factors of $C_n$, allowing our proof to develop more easily.

**Lemma 2.3:** Let $n \in \mathbb{N}$, $p \in \mathbb{P} \backslash \{2\}$ with $\frac{2n}{3} < p < n$. Then $p \nmid C_n$.

**Proof:**
Case 1: $n \leq 4$
2,3 are the only primes less than or equal to 4
$\Rightarrow p = 3$
This means that we only need to consider $C_4 = 8$: $3 \nmid 8$
$\therefore$ True for $n \leq 4$
Case 2: $n > 4$
$\frac{2n}{3} < p < n$
$\Rightarrow \frac{1}{n} < \frac{1}{p} < \frac{3}{2n}$
$\Rightarrow 1 < \frac{n}{p} < \frac{3}{2}$
$\Rightarrow 2 \leq \left\lfloor \frac{2n}{p} \right\rfloor < 3$ and $2 \leq 2 \left\lfloor \frac{n}{p} \right\rfloor < 3$
$\Rightarrow \left\lfloor \frac{2n}{p} \right\rfloor = 2 \left\lfloor \frac{n}{p} \right\rfloor = 2$
$\Rightarrow \left\lfloor \frac{2n}{p} \right\rfloor - 2 \left\lfloor \frac{n}{p} \right\rfloor = 0$

For $k \geq 2$: $n^{k-1} > 3 \left(\frac{3}{2}\right)^{k-1}$
$\Rightarrow \left(2^{k+1}\right)\left(n^{k-1}\right) > 3^k \cdot 2n$
$\Rightarrow \left(\frac{2n}{3}\right)^k > 2n$
$\Rightarrow p^k > 2n$
$\therefore$ For $k \geq 2$, $\left\lfloor \frac{2n}{p^k} \right\rfloor - 2 \left\lfloor \frac{n}{p^k} \right\rfloor = 0$
Recall that: $v_p(C_n) = \sum_k \left\lfloor \frac{2n}{p^k} \right\rfloor - 2 \left\lfloor \frac{n}{p^k} \right\rfloor$
$\Rightarrow v_p(C_n) = 0 for \frac{1}{n} < \frac{1}{p} < \frac{3}{2n}$
$\Rightarrow p \nmid C_n$ for $n > 4$

$\therefore \forall n \in \mathbb{N}, n \in \mathbb{P}$ s.t. $\frac{2n}{3} < p < n$, $p \nmid C_n \quad \square$

**Definition:** We shall define the *Primorial Function*:

$$x\# := \prod_{\substack{p \in \mathbb{P} \\ p \leq x}} p$$

with

$$1\# = 1$$

**Lemma 2.4:** Let $n \in \mathbb{N}$, $n\# < 4^n$

**Proof:**

$n = 1$, $1\# < 4^1$
$n = 2$, $n\# < 4^2$
Suppose $k\# < 4^k \ \forall k < n \in \mathbb{N}$

If $n \notin \mathbb{P}$, $n\# = (n-1)\# \leq 4^n < 4^{n-1}$
If $n \in \mathbb{P}_{>2}$, $n = 2m + 1$ for $m \in \mathbb{N}$

$$\binom{2m+1}{m} = \frac{(2m+1)!}{(m!)((m+1)!)}$$

$\Rightarrow \binom{2m+1}{m}$ is divisible by all primes $p$ where $m + 1 < p < 2m + 1$

$$\left( \prod_{\substack{p \in \mathbb{P} \\ m+1 < p \leq x}} p \right) \Bigg| \binom{2m+1}{m}$$
$$\Rightarrow \frac{(2m+1)\#}{(m+1)\#} \Bigg| \binom{2m+1}{m}$$
$$\Rightarrow \frac{(2m+1)\#}{(m+1)\#} \leq \binom{2m+1}{m}$$

Now:

$$\begin{aligned}\binom{2m+1}{m} &< \binom{2m+1}{0} + \binom{2m+1}{1} + \ldots + \binom{2m+1}{m} \\ &= \frac{1}{2}\left[\binom{2m+1}{0} + \binom{2m+1}{1} + \ldots + \binom{2m+1}{2m+1}\right] \\ &= \frac{1}{2}\left(2^{2m+1}\right) \\ &= 2^{2m} \\ &= 4^m \\ \Rightarrow (2m+1)\# &< 4^m (m+1)\#\end{aligned}$$

By Supposition, as $m+1 < n$, $(m+1)\# < 4^{m+1}$
$\Rightarrow (2m+1)\# < 4^m \cdot 4^{m+1} = 4^{2m+1}$
$\Rightarrow n\# < 4^n$
$\therefore$ If true for $k < n$, it is true for $n$.
$\therefore$ By the Principle of Induction, $n\# < 4^n \forall n \in \mathbb{N}$ $\square$

We now must construct a more general form of this lemma:

**Corollary 2.5:** Let $n \in \mathbb{R}_+$, $n\# < 4^n$

**Proof:** We have already proved this for a natural number $n$, so all that is left is to prove it for $n \in \mathbb{R}_+ \backslash \mathbb{N}$
Suppose $n \in \mathbb{R}_+ \backslash \mathbb{N}$
$\Rightarrow n \notin P$
$\Rightarrow n\# = \lfloor n \rfloor\# < 4^{\lfloor n \rfloor} \leq 4^n$ $\square$

Now we have all of the ingredients we need for our proof of Bertrand's Postulate, but first let's formally state it:

**Theorem 2.6 - Bertand's Postulate:** Let $n \in \mathbb{N}$. Then $\exists p \in \mathbb{P}$ s.t. $n \leq p \leq 2n$

An equivalent statement would be to say that $p_{n+1} < 2p_n$, where $p_n$ denotes the $n^{th}$ prime number.

**Proof:** Suppose $\exists n \in \mathbb{N}$ s.t. $\nexists p \in \mathbb{P}$ s.t. $n \leq p \leq 2n$

Case 1: $n \leq 4$

We shall do this by counter example, as some of our inequalities do not hold for this case.

$n = 1$: $p = 2 \in [1, 2]$
$n = 2$: $p = 3 \in [2, 4]$
$n = 3$: $p = 5 \in [3, 6]$
$n = 4$: $p = 7 \in [4, 8]$
$\therefore$ by counter examples, we have shown that it is true for $n \leq 4$.

Case 2: $n > 4$

All factors of $(2n)!$ are less than $2n$, so by the definition of $C_n$, all factors of $C_n$ (and thus all of its prime factors) are less than $2n$.

$\therefore$ by our supposition, all the prime factors of $C_n$ are less than $n$.
$\therefore$ by Lemma 2.3, none of the prime factors are greater than $\frac{2n}{3}$
By FTA, we can write $C_n$ as the product of its prime factors:

$$C_n = p_1^{a_k} \cdot p_2^{a_2} \cdot \ldots p_k^{a_k}$$

where $p_i$ are prime numbers and $a_i > 0$.
As $n > 4 \Rightarrow \sqrt{2n} < \frac{2n}{3}$
Consider the following form of $C_n = P_1 \cdot P_2$, where

$$P_1 = \prod_{p_i \leq \sqrt{2n}} p_i^{a_i}, P_2 = \prod_{\sqrt{2n} < p_i \leq \frac{2n}{3}} p_i^{a_i}$$

By Lemma 2.2, no terms in $P_1$ exceed 2n

$$\Rightarrow P_1 = \prod_{p_i \leq \sqrt{2n}} p_i^{a_i} \leq \prod_{k=1}^{\lfloor \sqrt{2n} \rfloor} 2k \leq (2n)^{\sqrt{2n}} (*)$$

For $p_i > \sqrt{2n}$, $p_i^k > 2n$ for $k > 2$

$$\Rightarrow P_2 = \prod_{\sqrt{2n} < p_i \leq \frac{2n}{3}} p_i^{a_i} = \prod_{\sqrt{2n} < p_i \leq \frac{2n}{3}} p_i < \prod_{p_i \leq \frac{2n}{3}} p_i = \left(\frac{2n}{3}\right) \#$$

$\Rightarrow$By Corollary 2.5: $P_2 < 4^{\left(\frac{2n}{3}\right)} (**)$
$(*)$ and $(**) \Rightarrow C_n = P_1 P_2 \leq (2n)^{\sqrt{2n}} \cdot 4^{\left(\frac{2n}{3}\right)}$
By Lemma 2.1: $C_n \geq \frac{4^n}{2n}$

$$\therefore \frac{4^n}{2n} \leq C_n \leq (2n)^{\sqrt{2n}} \cdot 4^{\left(\frac{2n}{3}\right)}$$

For $n > 467$: $(2n)^{\sqrt{2n}} \cdot 4^{\left(\frac{2n}{3}\right)} \leq \frac{4^n}{2n} \Rightarrow\Leftarrow$

Now all that remains for $n \in (4, 467]$, is to give a sequence of prime numbers, ending greater than 467, where each prime is no greater than double the last:

$$5, 7, 13, 23, 43, 83, 163, 317, 631$$

$\therefore$ Supposition false $\forall n \in \mathbb{N}$

$\therefore$ Bertrand's Postulate Holds $\square$ (*Proof of Bertrands Postulate*, n.d.)

We have indeed already proved that there are infinitely many primes, and that they are relatively dense compared to other sequences of natural numbers, but Bertrands postulate brings us one step closer to realising the true nature of the distribution of prime numbers, and it implies that irrespective of how large you may choose a prime number to be, there will always be another within a known distance. However, as with the argument with arbitrarily large gaps between prime numbers, this bound becomes incredible large, especially when considering prime numbers which are millions of digits long.

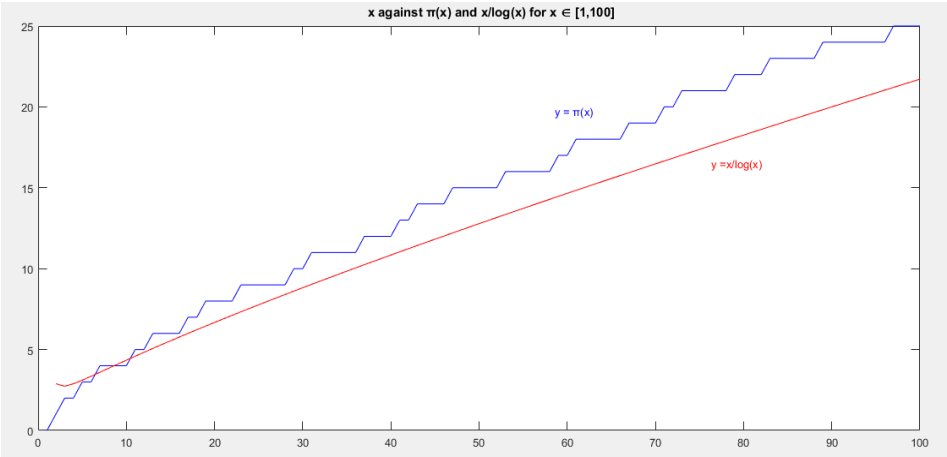
# 3 The Prime Number Theorem

When discussing the density of the prime numbers, it would be ideal for us to have some sort of formula or iterative process such that we can calculate the number of primes either less than a given integer, or between two given integers. However, due to the random nature of the prime numbers, no such a formula/such process is not known, and likely does not exist.

This then breeds the question as to whether there is a way of estimating the general vicinity of primes numbers and thus their general density, especially when considering large integers, where we know that large prime numbers are very difficult to locate.

We shall define $\pi(x) :=$ "The number of primes less than x"

When Gauss was young, he conjectured that $\pi(x) \sim \frac{x}{\log x}$ (shock Gauss was clever), which was later proved by Jacques Hadamard and Charles Jean de la Valle-Poussin in the 19th century using the Riemann Zeta Function, and has been coined the Prime number Theorem. (Selberg, 1949)




Firstly, notice that this is a strictly increasing function (as $\forall x, x > \log x$) which coincides with our theorem that there are infinitely many prime numbers.

Now, we shall take a look at the growth of this function, as we have previously mentioned that the primes become scarcer as they become large. Taking the first derivative of $\frac{x}{\log x}$ we can look at its rate of growth at $x$:

We shall by first defining:

$$f(x) = \frac{x}{\log x}$$

$$\Rightarrow f'(x) = \frac{\log x - 1}{(\log x)^2} = \frac{1}{logx} - \frac{1}{(logx)^2}$$

$\frac{1}{(logx)^n} > 0$ is a decreasing function for $n \geq 1$, so as $\frac{1}{logx} - \frac{1}{(logx)^2}$, we have that $f'(x)$ is a strictly positive, decreasing function. This then implies that $f(x)$ is always increasing, but the rate at which it grows decreases logarithmically as x gets large.

This notion further exemplifies that the primes do, in general, get further apart as they get larger, as we have $\pi(x) \sim f(x)$.

Furthermore, as $\frac{1}{logx} - \frac{1}{(logx)^2}$ decreases logarithmically, this means that the

sparseness between primes begins to even out, as the rate at which $f'(x)$ decreases, also decreases. This gives rise to an apparent homogeneity in the prime numbers over large values of x (because when we take very large $a, b$ with $a < b$, we see that $\pi(b) - \pi(a) = 0$ for more values of $a$ and $b$ as they get large).

Rearranging the relation for $\pi(x)$, we obtain $p_n \sim n \log n$, where $p_n$ denotes the $n^{th}$ prime number. On the search for a formula to locate prime numbers, this proves a valuable estimate, especially for large values of n.

For example, if we consider the $100^{th}$ prime number, 541; using the relation above we obtain $p_{100} \approx 461$ (a 15% error), but if we consider the $1000000^{th}$ prime number 15,485,863 with corresponding estimate $p_{1000000} \approx$ 13,815,511; we have a 10% error - which displays an increase in precision for this relation as n gets larger.

When considering prime numbers of millions of orders of magnitude, even a 0.0001% error interval too large to find primes, even with the powers of modern computing, making this relation very useful when estimating large prime numbers.

Having such a tool for locating prime numbers is important in areas such a cryptography, where the fundamental principles rely on having strong, hard to locate prime numbers.

## 4 The Twin Prime Conjecture

So far, we have discussed the general density of the primes over the macroscopic scale, and arbitrary bounds which can be constructed which allow us to limit where we might find primes relative to each other.

Here, we shall discuss the possibility of a specific lower bound for the gaps between consecutive prime numbers which occurs infinitely often, and the reasons for the existence of such a lower bound, if it does indeed exist.

As there is only one even prime number, it is evident that there is no possibility of infinitely many intervals of size 1 between subsequent prime numbers, so we shall discuss the next smallest option: Twin Primes.

**Definition:** A pair of prime numbers $x, y$ are called *Twin Primes* if they have a difference of 2

As we have proven that there are indeed infinitely many prime numbers, one would intuitively think that there would be infinitely many cases of these twin primes existing. If this was not the case, this would imply that there would be a maximal case, after which there would be no further examples. If such a maximal case did exist, this would raise the question as to why such a point in the natural numbers occurs and could potentially give great insight into the prime numbers.

**Conjecture 4.1 - Twin Prime Conjecture:** There are infinitely many cases of twin primes

As mentioned before, all the prime numbers (excluding 2) are odd, which in turn means they all take the form of either $4k+1$ or $4k+3$ for some natural number $k$, which gives us a sensible place to start. If such a point where twin primes cease to occur were to exist, one reason might be that there are only finitely many examples of primes being of the form $4k+1$ or $4k+3$. So now we shall examine this with the help of the following theorem:

**Theorem 4.2 - Fermat's Little Theorem:** Let $p \in \mathbb{P}$, $a \in \mathbb{Z}$. Then $a^{p-1} \equiv 1(mod\ p)$

This theorem was proved in MA136 Introduction to Abstract Algebra, and will allows us the prove the following two lemmas:

**Lemma 4.3:** There are infinitely many primes of the form $4k+1$ for $k \in \mathbb{N}$

**Proof:** Let $N \in \mathbb{N}$, $N \geq 2$

Consider $M = (N!)^2 + 1$
$\forall x$ s.t. $1 < x \leq N$, $x \nmid M$ (as $x \mid N! \Rightarrow x \mid (N!)^2$)

By FTA, $M$ may be written as a product of its prime factors
$\Rightarrow \exists p \in \mathbb{P}$ s.t. $p \mid M$
As $p \in \mathbb{P}, p > N$

$M \equiv 0(mod\ p)$

$\Rightarrow (N!)^2 \equiv -1(mod\ p)$
$\Rightarrow [(N!)^2]^{\frac{p-1}{2}} \equiv [-1(mod\ p)]^{\frac{p-1}{2}}$
$\Rightarrow (N!)^{p-1} \equiv (-1)^{\frac{p-1}{2}}(mod\ p)$

By Fermat's Little Theorem: $(N!)^{p-1} \equiv 1(mod\ p)$
$\Rightarrow 1(mod\ p) \equiv (-1)^{\frac{p-1}{2}}(mod\ p)$

As $N \geq 2 \Rightarrow p$ is odd
$\Rightarrow -1 \not\equiv 1(mod\ p)$
$\Rightarrow 1 \equiv (-1)^{\frac{p-1}{2}}$

This equation only holds when $\frac{p-1}{2}$ is even
$\Rightarrow \exists k \in \mathbb{N}$ s.t. $\frac{p-1}{2} = 2k$
$\Rightarrow p - 1 = 4k$
$\Rightarrow p = 4k + 1 > N$

This means that if there were to be a finite set of primes of the $4k + 1$, with the largest being $p_0$, we can take $N > p_0$ and construct a $p_1 \in \mathbb{P}$ with $p_1 > p_0$

$\therefore$ There are infinitely many primes of the form $4k+1$ $\square$ (Apostol, 2013)

Now that we have proved that there are infinitely many primes of the form $4k+1$, we must now check that there are infinitely many primes of the form $4k + 3$ to ensure that our suspicion was false.

**Lemma 4.4:** There are infinitely many primes of the form $4k + 3$, $k \in \mathbb{N}$

**Proof:** Our proof shall use contradiction

Suppose that there are finitely many primes of the form $4k + 3$
Let this set of primes be $\{p_1, p_2, \ldots, p_n\}$

Consider $N = (p_1 \cdot p_2 \cdot \ldots p_n)^2 + 2$ (notice that N is odd)
For $i = 1, \ldots, n$: $p_i \nmid N$

By FTA, $N$ may be written as the product of its prime factors.
As $p_i \nmid N\ \forall i$, they are all of the form $4k + 1$

Let $p \in \mathbb{P}$ s.t. $p \mid N$
As $p = 4k + 1 \Rightarrow p \equiv 1(mod\ 4)$

$N = (p_1 \cdot p_2 \cdot \ldots p_n)^2 + 2$
$\Rightarrow N - 2 = (p_1)^2 \cdot (p_2)^2 \cdot \ldots \cdot (p_n)^2$

For $i = 1, \ldots, n$ $p_i \equiv 3(mod\ 4)$
$\Rightarrow p_i^2 \equiv 1(mod\ 4)$
$\Rightarrow (p_1)^2 \cdot (p_2)^2 \cdot \ldots \cdot (p_n)^2 \equiv 1(mod\ 4)$
$\Rightarrow N - 2 \equiv 1(mod\ 4)$
$\Rightarrow N \equiv 3(mod\ 4)$
$\Rightarrow p \nmid N$ which contradicts that $p$ is a prime factor of $N$ $\square$ (*MT 430 Intro to Number Theory PROBLEM SET 2*, n.d.)

As we have shown, this suggestion for a counter proof for the Twin Prime conjecture did not yield any useful results about the existence of maximal twin primes. This indeed points us towards the conjecture being true.

Twin primes have been studied extensively, and a more general conjecture was posed, stating that all even gaps between prime numbers occur infinitely often, but neither this nor the twin prime conjecture have had proofs constructed for them to this day.

Recently, there have been some major breakthroughs in this the proving of this conjecture, most notably by James Maynard and Terrance Tao, which have led to the result:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \leq 600$$

(Maynard, 2013)
This huge discovery, coupled with an ongoing Polymath project, may one day uncover the proof of the Twin Prime Conjecture, or we may need to find a completely new way of analysing this unsolved mystery to finally tame this problem.

## References

Apostol, T. M. (2013). *Introduction to analytic number theory*. Springer Science & Business Media.

Dunham, W. (1999). Euler, the master of us all, volume 22 of the dolciani mathematical expositions. *The Mathematical Association of America, Washington, DC*.

Maynard, J. (2013). Small gaps between primes. *arXiv preprint arXiv:1311.4600*.

*Mt 430 intro to number theory problem set 2.* (n.d.). Retrieved from `https://www2.bc.edu/maksym-fedorchuk/430-pset-2-solutions.pdf`

*Proof of bertrands postulate.* (n.d.). Retrieved from `https://sites.math.washington.edu/ mathcircle/circle/2013-14/advanced/mc-13a-w10.pdf`

Selberg, A. (1949). An elementary proof of the prime-number theorem. *Annals of Mathematics*, 305–313.