Patterns in Primes

*Note: Sub-titles are not captured in Xplore and should not be used

1st Yongyu Qiang
Georgia Institute of Technology

2nd Aravinth Venkatesh Natarajan
Georgia Institute of Technology

3rd Anthony Hong
Georgia Institute of Technology

***Abstract*—**
***Index Terms*—**

I. Introduction

II. Background

A common result many students learn early on in number theory is about the infinitude of prime numbers. This fact is also known as Euclid's theorem, and we include a short summary of Euclid's original proof.

Euclid's Theorem. *The set of all prime numbers is larger in cardinality than any finite collection of prime numbers.*

Proof. Consider

$$\{p_1, p_2, \ldots, p_n\},$$

some arbitrary finite collection of prime numbers. Let

$$N = p_1 p_2 \ldots p_n,$$

and consider $P = N + 1$. P is either prime or not prime.

First, let P be prime. Then, we have constructed a new prime number and we are done.

Now, let P not be prime. Let g be a prime factor of P. We propose that $g \notin \{p_1, p_2, \ldots, p_n\}$. To show this, suppose for contradiction that $g \in \{p_1, p_2, \ldots, p_n\}$. Then, since $p_1, p_2, \ldots, p_n$ are all factors of N, we have $g|N$. $g|P$ and $g|N$, so we must also have $g|P - N$, i.e. $g|1$. But $g > 1$ (g is prime), so g cannot possibly divide 1. Therefore, $g \notin \{p_1, p_2, \ldots, p_n\}$, and we have found a new prime, as required. □

A natural next step from here is to explore how prime numbers are distributed. For now, we'll focus particularly on prime gaps and how small or large they can be. A bit of thinking leads to the observation that there are certain restrictions on what prime gaps can look like. First, we can see that prime gaps can be odd only finitely many times.

Lemma. *There exist only finitely many odd prime gaps.*

Proof. Notice that all primes $p > 2$ are odd. Then $p_{n+1} - p_n$ is even for all $n > 1$, so there exists only finitely many n such that $p_{n+1} - p_n$ is odd. □

In fact, $n = 1$ yields the only odd prime gap, namely $(p_1, p_2) = (2, 3)$ with difference 1. On the other hand, we have a much more promising observation for large prime gaps, namely that we can make them arbitrarily large.

Lemma. *There exist prime gaps of arbitrarily large size.*

Proof. We'll show that given $n \in \mathbb{Z}^+$, we can construct an interval of size at least $n-1$ of only composite numbers. Then the first primes immediately before and after this interval will have gap of at least n.

Let $n \in \mathbb{Z}^+$. Now consider the interval

$$[n! + 2, n! + n].$$

By definition of the factorial, we have $i|n!$ for all $i \in [2, n]$. We also trivially have $i|i$. Therefore, we have $i|(n!+i)$, and so i is a divisor of $n!+i$ for all $i \in [2, n]$. Then all of $[n!+2, n!+n]$ is composite, and this interval has size $n - 1$, as desired. □

Although this result is nice, we soon realize that it does not give us a very strong bound, in the sense that it is rather wasteful. To find a prime gap of size n by this method, we must consider numbers of order $n!$. By Stirling's approximation, we have

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

asymptotically, which is worse than exponential growth in n. Put in context, our current method suggests that finding a prime gap of size 10 requires us to find numbers of magnitude about 3 million. In reality, we can find such a gap of size 10 at $(p_{30}, p_{31}) = (113, 127)$, which is much smaller than 3 million, so certainly we can do better.

For that, we'll need better tools. Let us first define the prime counting function $\pi(n)$.

Definition. $\pi(n) :=$ number of primes $\leq n$.

Now we can introduce the prime number theorem, which characterizes the growth of $\pi(n)$ as n gets large. Note that we will use this result without proof in this paper, as even relatively simpler proofs rely on tools from analysis.

Prime Number Theorem (PMT). $\frac{n}{\ln(n)}$ asymptotically approximates $\pi(n)$. Put more formally,

$$\lim_{n\to\infty} \frac{\pi(n)}{\frac{n}{\ln(n)}} = 1.$$

From this result, we can quickly argue by the pigeonhole principle that there should exist a prime gap of size at least

Identify applicable funding agency here. If none, delete this.

$\ln(n)$ in the interval $[2, n]$. If we take buckets to be the $\frac{n}{\ln(n)}$ prime numbers less than or equal to n and our pigeons to be the integers in $[2, n]$, then by the pigeonhole principle, at least one prime number will correspond to $\ln(n)$ or more integers, making a gap of size at least $\ln(n)$. Note that this pigeonhole argument already gives an asymptotically better bound for large gaps than our previous result, albeit just barely, as we only now only need to go to e^x for a gap of size x rather than $x!$.

We can also apply the same argument in reverse to get a small gap of size at most $\ln(n)$, but getting any better bounds generally requires the use of analysis. In an effort to avoid that, we can instead explore the Cramér random model, which considers randomness as a method to approximate $\pi(n)$.

III. Cramér's Random Model

Recall that the PMT tells us that for large n,

$$\pi(n) \approx \frac{n}{\ln(n)}.$$

But since we also have n total numbers less than or equal to n, we can divide by this size to get a rough prime density $\delta(n)$:

$$\delta(n) = \frac{\pi(n)}{n} = \frac{\frac{n}{\ln n}}{n} = \frac{1}{\ln(n)}.$$

So for a random number x in the interval $[n, n+kn]$ for large n and fixed k, the probability that x is prime is approximately $1/n \approx 1/x$. (As a side note, this density $\delta(n)$ is also the rationale behind an alternate approximation for $\pi(n)$ with the logarithmic integral $\mathrm{Li}(n)$ as

$$\pi(n) \approx \int_2^n \delta(x)\,dx = \int_2^n \frac{1}{\ln(x)}\,dx = \mathrm{Li}(n),$$

which actually turns out to be a much better approximation to $\pi(n)$ than the traditional PMT. In fact, if we assume the Riemann hypothesis, we have that the error of $\mathrm{Li}(n)$ from $\pi(n)$ is bounded by $O(n^{1/2+\epsilon})$ for any $\epsilon > 0$, meaning that roughly the first half of the digits of $\mathrm{Li}(n)$ will be correct, but that is outside the scope of this paper.)

Cramér's random model uses this idea as a naive approach to emulate the distribution of prime numbers. Consider a random subset of the natural numbers, where the independent probability that a number n is chosen is $1/\ln(n)$. Let's call this random set P', where P is the set of actual prime numbers. Cramér conjectured that P', which consists of our "fake primes," accurately models the distribution of P.

According to this heuristic, we have the resulting claim, which is known as Cramér's conjecture:

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{(\ln p_n)^2} = 1$$

where p_n denotes the n-th prime.

(Some directions to take these ideas):

- Problems with Cramér's naive model and ways we can improve it (with modern results)
- How Cramér's model fares depending on the size and location of the interval, calculating asymptomatic statistics

IV. Bertrand's Postulate

Another problem in number theory is finding the bounds in which you would find a prime number. Of these, one of the paramount significance is **Bertrand's Postulate**:
Bertrand's postulate states that for an integer i > 1, there is at least one prime number p

$$i \le n \le 2i$$

The proof for it is as follows:
We'll start by proving Lemma 1:

$$\frac{4^n}{2n} \le \binom{2n}{n}$$

$$4^n = (1+1)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k}$$

Since, $\binom{2n}{0}$ is 1 and $\binom{2n}{2n}$ is 1, this is the same as

$$\equiv 2 + \sum_{k=1}^{2n-1} \binom{2n}{k}$$

Since the largest term in this summation is $\binom{2n}{n}$ (since for $\binom{n}{k}$, $k = n/2$ will give the largest term) and there are 2n terms,

$$(2 + \sum_{k=1}^{2n-1} \binom{2n}{k}) \le (2n * \binom{2n}{n})$$

Therefore,

$$\frac{4^n}{2n} \le \binom{2n}{n}$$

Let's now prove Lemma 2:
For a given prime p, let's define r as the greatest number for which $p^r | \binom{2n}{n}$. Lemma 2 is as follows, for such an r,

$$p^r \le 2n$$

Firstly, we have to introduce Legendre's Formula:
Legendre's Formula states that for any prime number p, and any integer n, let's define the function $v_p(n)$ as the exponent of the largest power of p that divides n. Let L = $\lfloor \log_p n \rfloor$
Legendre's Formula is:

$$v_p(n!) = \sum_{i=1}^{L} \left\lfloor \frac{n}{p^i} \right\rfloor$$

$\binom{2n}{n}$ can also be written as $\frac{(2n)!}{n!*n!}$)
F Finding the largest exponent of p, r, that divides $\frac{(2n)!}{n!*n!}$ is the same as finding the largest exponent of p, r, that divides each component of $\frac{(2n)!}{n!*n!}$, i.e. $(2n)!$, $n!$

In this case, L = $\lfloor \log_p 2n \rfloor$
Writing this in terms Legendre's Formula, we get that:

$$v_p(\binom{2n}{n}) = \sum_{i=1}^{L} \left\lfloor \frac{2n}{p^i} \right\rfloor - 2\sum_{i=1}^{L} \left\lfloor \frac{n}{p^i} \right\rfloor$$

This is equivalent to:

$$v_p(\binom{2n}{n}) = \sum_{i=1}^{L} \left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor$$

Thinking intuitively, every term in $\sum_{i=1}^{L} \left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor$ must either be 0 or 1.
If $(\frac{2n}{p^i} \mod 1) \geq 0.5$ then the term would be 1, otherwise the term is 0. Therefore, the maximum value of this function would be if all the terms were equal to 1. Since we are only dealing with positive numbers, and the exponent is a monotonic function if both numbers are positive numbers,

$$v_p(\binom{2n}{n}) = r \leq L$$

$$\equiv r <= \log_p 2n$$

Therefore, it follows that

$$p^r <= p^{\log_p 2n} = 2n$$

We are able to prove our initial Lemma 2,

$$p^r <= 2n$$

Now, we move on to proving Lemma 3: Let's place some bounds on p for the lemma we're going to prove. If

$$\frac{2n}{3} < p \leq n, then R(n,p) = 0$$

$$\binom{2n}{n} = \frac{(2n)!}{n!^2}$$

Therefore, there are 2 factors of p in the numerator, p and 2p
In the denominator, there are 2 factors of p, each in n!
The conditions mean that 3p is too large to be in the numerator
Due to this, the factors of p in the numerator and denominator cancel out so the greatest power of p that divides this number is 1, i.e. p^0, giving R = 0
With this, we prove Lemma 3
Moving on to Lemma 4:
Let's define the primorial function:

$$x\# = \Pi_{p \leq x} x$$

I.e., x# is the product of all primes less than or equal to x
Lemma 4 states that for x $\geq$ 3, x# $\leq 4^x$
We proceed by induction:
This claim is easily verifiable for small odd numbers. Try one yourself!
For even numbers, x# = (x - 1)# as for x $\geq$ 3, there are no even prime numbers
Therefore, it is also bounded by x# $\leq 4^x$
Let's take a larger odd number, n = 2m + 1

$$\Pi_{p \leq n} p = \Pi_{p \leq m+1} * \Pi_{m+2 \leq p \leq 2m+1}$$

Since the Inductive Hypothesis is true for smaller odd numbers,

$$\Pi_{p \leq m+1} = 4^{m+1}$$

Now, let's take a detour, let's look at a binomial expansion

$$(1+1)^{2m+1} = \sum_{k=0}^{2m+1} \binom{2m+1}{k} 1^{2m+1-k} 1^k$$

Necessarily

$$\Pi_{m+2 \leq p \leq 2m+1} \leq \binom{2m+1}{m}$$

Since you can notice that all the primes on the left will divide $\binom{2m+1}{m}$

$$\binom{2m+1}{m \leq \sum_{k=0}^{2m+1} \binom{2m+1}{k} 1^{2m+1-k} 1^k}$$

$$\sum_{k=0}^{2m+1} \binom{2m+1}{k} 1^{2m+1-k} 1^k = 2^{2m+1}$$

Also, since

$$\binom{2m+1}{m} = \binom{2m+1}{m+1}$$

We can divide our bound by 2 since those numbers will be repeated
Therefore,

$$\Pi_{m+2 \leq p \leq 2m+1} \leq 2^{2m}$$

Consequently,

$$\Pi_{p \leq n} p = 4^{m+1} * 2^{2m}$$

Which is equivalent to

$$\Pi_{p \leq n} p \leq 4^n$$

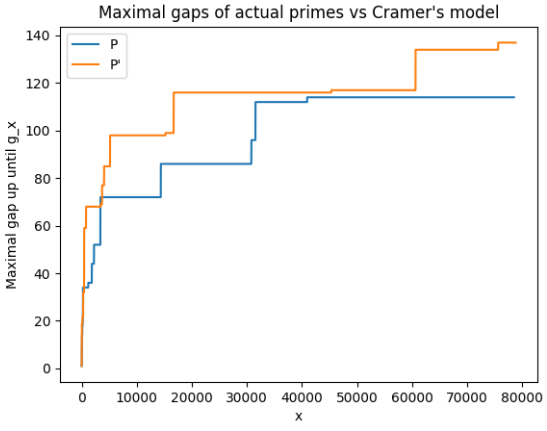
Now, we are finally ready to prove Bertrand's Postulate:

V. Extension/application/generalisation

- Connections from Cramér's conjecture to the Riemann hypothesis
- Other ways to use Cramér's technique of random modeling

VI. Preliminary Code and Illustrations

Cramér's random model allows us to heuristically test properties of primes. In this example, we graphically compare the maximal prime gap of the model and the actual primes.



```
def sieve(n):
    prime = [True for i in range(n + 1)]
    i = 2
    while i * i <= n:
        if prime[i]:
            for j in range(i * i, n + 1, i):
                prime[j] = False
        i += 1
    last_prime = -1
    gaps = []
    for i in range(n + 1):
        if prime[i]:
            if last_prime != -1:
                if len(gaps) == 0:
                    gaps.append(i - last_prime)
                else:
                    gaps.append(max(gaps[-1], i -
                                                last_prime
                                                )
                                                )
            last_prime = i
    return gaps

def cramer_model(n):
    # Assume 2 is in the model because 1 / ln(2) > 1
    gaps = []
    last_prime = 2
    for i in range(3, n + 1):
        if random.random() <= 1 / np.log(i):
            if len(gaps) == 0:
                gaps.append(i - last_prime)
            else:
                gaps.append(max(gaps[-1], i -
                                            last_prime
                                            )
                                            )
            last_prime = i
    return gaps
```

VII. Accuracy of the Model

Under heuristic testing with variations of Cramér's model, we should be able to support strong statements such as Bertrand's postulate and Legendre's conjecture. However, comparing with the actual primes, it should be clear that Cramér's model is inaccurate. Further work would involve the creation and tuning of other random models, to more closely emulate prime distributions.

References

[1] Terence Tao's Blog
[2] Wikipedia Article on Prime Gaps