

# CS 2051: Patterns in Primes

Anthony Hong  
Georgia Institute of Technology

Yongyu Qiang  
Georgia Institute of Technology

Aravinth Venkatesh Natarajan  
Georgia Institute of Technology

March 12th, 2023

## 1 Background

(hook to be added here)

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  $p_1, p_2, \dots, p_n$ , some arbitrary finite collection of prime numbers. Let  $N = p_1 p_2 \dots p_n$ , 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  $g \notin \{p_1, p_2, \dots, p_n\}$ . To show this, suppose that  $g \in \{p_1, p_2, \dots, p_n\}$ . Then, since  $p_1, p_2, \dots, p_n$  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,  $g \notin \{p_1, p_2, \dots, p_n\}$ , and we have found a new prime, as required.  $\square$

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.  $\square$

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  $n - 1$ , so we're done.  $\square$

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

(add more stuff here)

## 2 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  $\text{Li}(n)$  as

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

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  $\text{Li}(n)$  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  $\text{Li}(n)$  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:

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

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

### 3 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  $p$

$$i \leq p \leq 2i$$

The proof for it is as follows:

We’ll start by proving Lemma 1:

$$\begin{aligned} \frac{4^n}{2^n} &\leq \binom{2n}{n} \\ 4^n = (1+1)^{2n} &= \sum k = 0^{2n} \binom{2n}{k} \end{aligned}$$

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,

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

Therefore,

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

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  $r$ ,

$$p^r \leq 2n$$

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  $\frac{(2n)!}{n! * n!}$

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:

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

This is equivalent to:

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

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  $\left(\frac{2n}{p^i} \bmod 1\right) \geq 0.5$  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,

$$\begin{aligned} v_p\left(\binom{2n}{n}\right) &= r \leq L \\ &\equiv r \leq \log_p 2n \end{aligned}$$

Therefore, it follows that

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

We are able to prove our initial Lemma 2,

$$p^r \leq 2n$$

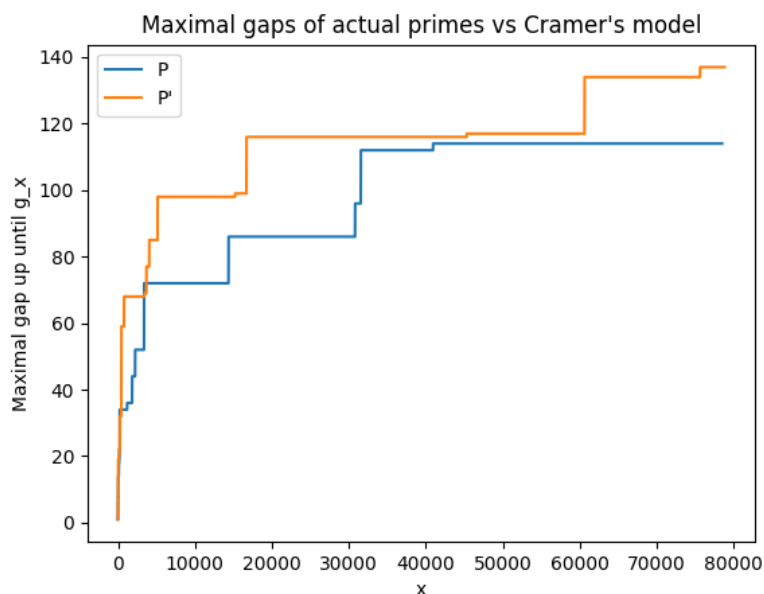
(We will proceed to prove Lemma 3 and 4, which will, in combination prove Bertrand's Postulate)

## 4 Extending the Model

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

## 5 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

```

## 6 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.

## 7 References

- Terence Tao’s Blog
- Wikipedia