Prime Gaps and Asymptotic Behavior of Primes: A Hypothetical Approach

Hyeon Jun Ahn

March 11, 2025

Abstract: Based on heuristics related to Cramér's conjecture, this paper proposes a suitable hypothesis and investigates its implications. The study encompasses prime gaps, Andrica's conjecture, the mean of consecutive prime numbers, and a detailed analysis of Oppermann's conjecture.

1. Introduction

According to the prime number theorem, the number of primes less than n is asymptotic to $n/nlogn$, and the average gap between primes less than n is $logn$. Therefore, nth prime is asymptotic to $nlogn$; that is

$$\lim_{n\to\infty} \frac{p_n}{nlogn} = 1$$

which can be recast as $p_n \sim nlogn$. In other words, $nlogn$ approximates p_n in the sense that the relative error of this approximation approaches 0 as n approaches infinity. So, we have

$$p_{n+1} + p_n \sim (n+1)log(n+1) + nlogn$$

because

$$\begin{aligned}
&\lim_{n\to\infty} \frac{p_{n+1} + p_n}{(n+1)log(n+1) + nlogn} \\
&= \lim_{n\to\infty} \left(\frac{1}{\frac{(n+1)log(n+1)}{p_{n+1}} + \frac{nlogn}{p_{n+1}}} + \frac{1}{\frac{(n+1)log(n+1)}{p_n} + \frac{nlogn}{p_n}} \right) \\
&= \lim_{n\to\infty} \left(\frac{1}{1+1} + \frac{1}{1+1} \right) = 1
\end{aligned}$$

This result shows it is possible to add $p_n \sim nlogn$ and $p_{n+1} \sim (n+1)log(n+1)$. However, subtraction is not possible; that is,

$$p_n - p_{n+1} \nsim (n+1)log(n+1) - nlogn \tag{1}$$

Rather, it holds that

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{(n+1)log(n+1)-nlogn} = \infty$$
$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{(n+1)log(n+1)-nlogn} = 0$$

proof. Note that

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{logp_n} = \infty, \quad \liminf_{n\to\infty} \frac{p_{n+1}-p_n}{logp_n} = 0 \tag{2}$$

E. Westzynthius proved the former in 1931 [1,2], Daniel Goldston, János Pintz and Cem Yıldırım proved the latter in 2005 [3]. First, we claim

$$\lim_{n\to\infty} \frac{log(nlogn)}{logp_n} = 1 \tag{3}$$

For every $2 \leq n \in \mathbb{N}$,

$$\frac{log(nlogn)}{logp_n} = log_{p_n}(nlogn)$$

put $k(n) = log_{p_n}(nlogn)$. Then, we obtain $p_n^{k(n)} = nlogn$ which yields

$$\frac{nlogn}{p_n^{k(n)}} = 1$$

Knowing that $p_n \sim nlogn$, we consider the limit of $p_n^{1-k(n)}$;

$$\lim_{n\to\infty} p_n^{1-k(n)} = \lim_{n\to\infty} \frac{p_n}{p_n^{k(n)}} = \lim_{n\to\infty} \frac{p_n}{nlogn}\frac{nlogn}{p_n^{k(n)}} = 1 \times 1 = 1$$

Therefore, $\lim_{n\to\infty} k(n) = 1$ as claimed. We also claim that

$$\lim_{n\to\infty} \frac{log(nlogn)}{(n+1)log(n+1)-nlogn} = 1 \tag{4}$$

By the L'Hôpital's rule,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{log(nlogn)}{(n+1)log(n+1)-nlogn} \\
&= \lim_{n\to\infty} \frac{logn + log(logn)}{(n+1)log(n+1)-nlogn} \\
&\overset{\text{L'H}}{=} \lim_{n\to\infty} \frac{1/n + 1/nlogn}{log(n+1)-logn} \\
&= \lim_{n\to\infty} \frac{logn+1}{nlogn(log(n+1)-logn)} \\
&= \lim_{n\to\infty} \frac{logn+1}{lognlog(1+1/n)^n} \\
&= \lim_{n\to\infty} \frac{logn+1}{logn} \\
&= 1
\end{aligned}$$

Now, put

$$F(n) = \frac{logp_n}{log(nlogn)} \frac{log(nlogn)}{(n+1)log(n+1) - nlogn} = \frac{logp_n}{(n+1)log(n+1) - nlogn}$$

then, because of (3),(4), we have $\lim_{n\to\infty} F(n) = 1$ and thus, for sufficiently large M,

$$\frac{1}{2} < F(n) < \frac{3}{2}$$

where $n > M$. Multipling by $(p_{n+1} - p_n)/logp_n$ leads us to

$$\frac{1}{2} \frac{p_{n+1} - p_n}{logp_n} < \frac{p_{n+1} - p_n}{logp_n} F(n) < \frac{3}{2} \frac{p_{n+1} - p_n}{logp_n}$$

By the Squeeze Theorem and (2), we obtain

$$\begin{aligned} \limsup_{n\to\infty} \frac{p_{n+1} - p_n}{logp_n} F(n) = \infty \\ \liminf_{n\to\infty} \frac{p_{n+1} - p_n}{logp_n} F(n) = 0 \end{aligned} \tag{5}$$

Knowing that

$$\frac{p_{n+1} - p_n}{logp_n} F(n) = \frac{p_{n+1} - p_n}{(n+1)log(n+1) - nlogn}$$

so, we can represent (5) as

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{(n+1)log(n+1) - nlogn} = \infty$$
$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{(n+1)log(n+1) - nlogn} = 0$$

Therefore, we need another method to find the approximate expression of $p_{n+1} - p_n$. $(n+1)log(n+1) - nlogn$ is not appropriate although $p_n \sim nlogn$. In this paper, instead of finding a solution, We will approach this problem in a different way. ■

Cramer conjecture is a conjecture regerding the gaps between prime numbers. The conjecture states that

$$g_n := p_{n+1} - p_n = O((logp_n)^2)$$

holds where O is a big O notation. And sometimes the following formulation is called Cramer's conjecture;

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

which is stronger than former. This conjecture is based on the Cramér random model, a model for the distribution of primes. In this model, the probability that a positive integer $n \geq 3$ is a prime is approximately $1/logn$.

But Maier's theorem shows that the Cramér random model does not adequately describe the distribution of primes on short intervals, and a refinement of Cramér's model taking into account divisibility by small primes suggests that

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{(logp_n)^2} \geq 2\exp(-\gamma) \approx 1.1229\cdots$$

These conjecture say that the limit superior of $g_n/(logp_n)^2$ converges. (But János Pintz suggested that it may diverge[4].) It is supported that there exists m such that the superior of $g_n/(logp_n)^m$ converges by the preceding several heuristics. So, Let μ be the smallest m that satisfies the following conditions:

$$m \in \mathbb{N}, \ \lim_{n\to\infty} \frac{g_n}{(logp_n)^m} = 0$$

which implies that

$$\lim_{n\to\infty} \frac{g_n}{(log(nlogn))^\mu} = 0 \tag{6}$$

because

$$\lim_{n\to\infty} \frac{g_n}{(log(nlogn))^\mu} = \lim_{n\to\infty} \frac{g_n}{(logp_n)^\mu}(\frac{logp_n}{log(nlogn)})^\mu = 0 \times 1 = 0$$

(See (3)). To avoid the possibility of such m not having a minimum, μ is assumed to be a natural number for convenience. Nevertheless, μ may not exist as such m doesn't exist, but in this paper, it is assumed to exist, and we will examine what conclusion we can reach.

2. Prime gap

Remark 1. For every $k > 0$,

$$\lim_{n\to\infty} \frac{p_n^k}{(nlogn)^k} = \lim_{n\to\infty} (\frac{p_n}{nlogn})^k = 1 \tag{7}$$

Lemma 1. For every $k > 0$,

$$\lim_{n\to\infty} \frac{(log(nlogn))^\mu}{(nlogn)^k} = 0 \tag{8}$$

proof. Let $x = nlogn$, then by L'Hôpital's rule,

$$\lim_{n\to\infty} \frac{(log(nlogn))^\mu}{(nlogn)^k} = \lim_{x\to\infty} \frac{(logx)^\mu}{x^k} \overset{\text{L'H}}{=} \lim_{x\to\infty} \frac{\mu(logx)^{\mu-1}}{kx^k} \overset{\text{L'H}}{=} \cdots$$

$$\overset{\text{L'H}}{=} \lim_{x\to\infty} \frac{\mu!}{k^\mu x^k} = 0 \ \blacksquare$$

(6) and (7) allow us to conclude that for every $k > 0$,

$$\lim_{n\to\infty} \frac{p_{n+1}-p_n}{p_n^k} \frac{(nlogn)^k}{(log(nlogn))^\mu} = \lim_{n\to\infty} \frac{p_{n+1}-p_n}{(log(nlogn))^\mu} \frac{(nlogn)^k}{p_n^k} = 0 \times 1 = 0 \quad (9)$$

Hence, we have

$$\lim_{n\to\infty} \frac{p_{n+1}-p_n}{p_n^k} = (8) \times (9) = 0 \tag{10}$$

or

$$\lim_{n\to\infty} \frac{p_n^k}{p_{n+1}-p_n} = \infty$$

By epsilon-delta argument, we now obtain

$$\begin{aligned} \forall k > 0,\ \exists N \in \mathbb{N} \quad s.t.\ \ n \geq N \ &\Rightarrow\ g_n := p_{n+1} - p_n < p_n^k \\ &\Rightarrow\ p_n < p_{n+1} < p_n + p_n^k \end{aligned} \tag{11}$$

3. About Andrica's conjecture

Andrica's conjecture is a conjecture regarding the gaps between prime numbers. The conjecture states that the inequality

$$\sqrt{p_{n+1}} - \sqrt{p_n} < 1$$

holds for all $n \in \mathbb{N}$. And a strong version of Andrica conjecture is as follows; Except for $p_n \in \{3, 7, 13, 23, 31, 113\}$, that is $n \in \{2, 4, 6, 9, 11, 30\}$, one has

$$\sqrt{p_{n+1}} - \sqrt{p_n} < \frac{1}{2}; \qquad \textit{equivalently} \qquad g_n := p_{n+1} - p_n < p_n^{1/2} + \frac{1}{4}$$

In this chapter, we prove that

$$\lim_{n\to\infty} (\sqrt{p_{n+1}} - \sqrt{p_n}) = 0$$

proof. Let $\epsilon > 0$, $k \in (0, \frac{1}{2})$, Then, clearly

$$\lim_{n\to\infty} \frac{p_n^k}{(\sqrt{p_n}+\epsilon)^2 - p_n} = \lim_{n\to\infty} \frac{p_n^k}{2\epsilon\sqrt{p_n}+\epsilon^2} = 0$$

Thus,

$$\begin{aligned} \forall \epsilon > 0,\ \forall k \in (0, \tfrac{1}{2}),\ \exists N_1 \in \mathbb{N} \quad s.t.\ \ n > N_1 \ &\Rightarrow\ p_n^k < (\sqrt{p_n}+\epsilon)^2 - p_n \\ &\Rightarrow\ p_n + p_n^k < (\sqrt{p_n}+\epsilon)^2 \end{aligned}$$

Meanwhile,

$$\forall k \in (0, \frac{1}{2}),\ \exists N_2 \in \mathbb{N},\quad s.t.\ \ n > N_2 \ \Rightarrow\ p_{n+1} < p_n + p_n^k \qquad (\because (11))$$

Put N=max(N_1, N_2), Then we obtain

$$n > N \ \Rightarrow\ p_{n+1} < (\sqrt{p_n} + \epsilon)^2$$

which can be represented as, for $n > N$,

$$\sqrt{p_{n+1}} - \sqrt{p_n} < \epsilon$$

Since this inequation holds for every $\epsilon > 0$, the epsilon-delta argument allow us to conclude

$$\lim_{n\to\infty} (\sqrt{p_{n+1}} - \sqrt{p_n}) = 0 \ \blacksquare \tag{12}$$

Furthermore, let $y > 1,\ x < \frac{y-1}{y}$, then, since $\forall L > 0,\ \exists M \in \mathbb{N}\ s.t.\ n > M \Rightarrow p_n^{1/y} > L$, the generalized binomial theorem allow us to obatain

$$\begin{aligned}
&\lim_{n\to\infty} \frac{p_n^x}{(p_n^{1/y} + \epsilon)^y - p_n} \\
&= \lim_{n\to\infty} \frac{p_n^x}{(p_n + \binom{y}{1} p_n^{(y-1)/y}\epsilon + \binom{y}{2} p_n^{(y-2)/y}\epsilon^2 + \cdots) - p_n} \\
&= \lim_{n\to\infty} \frac{p_n^x}{(\binom{y}{1} p_n^{(y-1)/y}\epsilon + \binom{y}{2} p_n^{(y-2)/y}\epsilon^2 + \cdots)} = 0 \ \ (\because x < \frac{y-1}{y})
\end{aligned}$$

In the same method as the proof of (12),

$$\forall y > 1, \quad \lim_{n\to\infty} (p_{n+1}^{1/y} - p_n^{1/y}) = 0$$

3-1. The arithmetic mean, the geometric mean and the harmonic mean of primes

The relation between the arithmetic mean and the geometric mean of nth prime and $(n+1)$th prime is as follows:

$$\frac{p_{n+1} + p_n}{2} \sim \sqrt{p_{n+1} p_n}$$

proof.

$$\begin{aligned}
&\lim_{n\to\infty} (\sqrt{p_{n+1}} - \sqrt{p_n}) = 0 \\
&\Rightarrow \lim_{n\to\infty} (\sqrt{p_{n+1}} - \sqrt{p_n})^2 = 0 \\
&\Rightarrow \lim_{n\to\infty} (p_{n+1} + p_n - 2\sqrt{p_{n+1} p_n}) = 0
\end{aligned} \tag{13}$$

Thus,

$$\lim_{n\to\infty} \frac{p_{n+1}+p_n}{2\sqrt{p_{n+1}p_n}} = \lim_{n\to\infty} (\frac{p_{n+1}+p_n-2\sqrt{p_{n+1}p_n}}{2\sqrt{p_{n+1}p_n}}+1) = 1 \blacksquare \tag{14}$$

Furthermore,

$$\lim_{n\to\infty} (\frac{p_{n+1}+p_n}{2} - \sqrt{p_{n+1}p_n}) = 0$$

trivially holds by (13). And similarly, the relation between the arithmetic mean and the harmonic mean of nth prime and $(n+1)$th prime is as follows:

$$\frac{p_{n+1}+p_n}{2} \sim \frac{2p_{n+1}p_n}{p_{n+1}+p_n}$$

proof. By (14)

$$\lim_{n\to\infty} \frac{2p_{n+1}p_n}{p_{n+1}+p_n}\frac{2}{p_{n+1}+p_n} = \lim_{n\to\infty} (\frac{2\sqrt{p_{n+1}p_n}}{p_{n+1}+p_n})^2 = 1 \blacksquare$$

In a similar manner to before, it is also true that

$$\lim_{n\to\infty} (\frac{p_{n+1}+p_n}{2} - \frac{2p_{n+1}p_n}{p_{n+1}+p_n}) = 0$$

proof.

$$\begin{aligned}
&\lim_{n\to\infty} (\frac{p_{n+1}+p_n}{2} - \frac{2p_{n+1}p_n}{p_{n+1}+p_n}) \\
&= \lim_{n\to\infty} \frac{(p_{n+1}+p_n)^2-4p_{n+1}p_n}{2(p_{n+1}+p_n)} = \lim_{n\to\infty} \frac{(p_{n+1}-p_n)^2}{2(p_{n+1}+p_n)} \\
&\leq \lim_{n\to\infty} \frac{(p_{n+1}-p_n)^2}{4p_n} = \lim_{n\to\infty} (\frac{p_{n+1}-p_n}{2\sqrt{p_n}})^2 = 0 \ (\because (10))
\end{aligned}$$

By the relation between the arithmetic mean and the harmonic mean,

$$\lim_{n\to\infty} (\frac{p_{n+1}+p_n}{2} - \frac{2p_{n+1}p_n}{p_{n+1}+p_n}) = 0 \blacksquare$$

Hence,

$$\frac{p_{n+1}+p_n}{2} \sim \sqrt{p_{n+1}p_n} \sim \frac{2p_{n+1}p_n}{p_{n+1}+p_n}$$

Therefore, the arithmetic mean, geometric mean, and harmonic mean of nth and $(n+1)$th primes are asymptotically equal as n approaches infinity.

4. About Oppermann conjecture

Oppermann conjecture is a conjecture regarding the distribution of prime numbers. It is closely related to but stronger than Legendre conjecture, Andrica conjecture, and Brocard conjecture. The conjecture states that for every integer $n \geq 1$,

$$\pi(n^2 - n) < \pi(n^2) < \pi(n^2 + n)$$

Notation 1. Let $\hat{p}(x)$ is the largest prime number less than x, $\hat{P}(x)$ is the smallest prime number greater than x.

$$e.g.\ \ \hat{p}(10) = 7,\ \ \hat{P}(10) = 11$$

Lemma 2. Let $f\colon \mathbb{R} \to \mathbb{R}$ is an increasing function and m is constant, then

$$\forall n \geq M,\ p_n < p_{n+1} < f(p_n) \quad \Rightarrow \quad \forall x \geq p_M,\ \exists p \in \mathbb{P}\ \ s.t.\ \ x < p < f(x)$$

proof. Suppose for contradiction that there exists an $x \geq p_M$ such that an open inteval $(x, f(x))$ doesn't contain any prime number. Then we have $\hat{P}(x) > f(x)$. Knowing that, by the definition, $\hat{p}(x) \leq x$ and $\hat{P}(x)$ is the next prime number after $\hat{p}(x)$, we can conclude

$$\hat{p}(x) < \hat{P}(x) < f(\hat{p}(x))$$

But, because f is an increasing function, $\hat{p}(x) \leq x$ implies $f(\hat{p}(x)) \leq f(x) < \hat{P}(x)$. It's a contradiction. ∎

Lemma 3. By **Lemma 2** and (11) implies that

$$\forall k > 0,\ \exists M_1 \in \mathbb{R}\ \ s.t.\ \ \exists p \in \mathbb{P}\ \ with\ \ x < p < x + x^k\ \ for\ x \geq M_1 \tag{15}$$

Lemma 4.

$$\forall k > 0,\ \exists M_2 \in \mathbb{R},\ \ s.t.\ \ \exists p \in \mathbb{P}\ \ with\ \ x - x^k < p < x\ \ for\ x \geq M_2 \tag{16}$$

proof. In the **Lemma 3**, let $x = m + m^k$ and $x \geq M_2$ where $M_2 = M_1 + M_1^k$, then there is a prime number in the open interval (m, x). Also, since $x > m$, we have $(m, x) \subset (x - x^k, x)$. Hence, there is a prime in the open interval $(x - x^k, x)$. ∎

We now prove that for every $k > 0$, there exists $M \in \mathbb{R}$ such that

$$x \geq M \ \Rightarrow\ \pi(x^k - x) < \pi(x) < \pi(x^k + x) \tag{17}$$

proof. By (15) and (16),

$$\forall k > 0,\ \exists M_2 \in \mathbb{R}\ \ s.t.\ \ \exists p, q \in \mathbb{P}\ \ with\ \ x - x^k < p < x < q < x + x^k\ \ for\ x \geq M_2$$

Substitute $x = t^m$ where $m = \dfrac{1}{k}$, then

$$\forall m > 0,\ \exists M' \in \mathbb{R}\ \ s.t.\ \ \exists p, q \in \mathbb{P}\ \ with\ \ t^m - t < p < t^m < q < t^m + t\ \ for\ t \geq M'$$

($c.f.$ $x = t^m$ yields $M_2 = (M')^m$) which implies that

$$\forall m > 0,\ \exists M' \in \mathbb{R}\ \ s.t.\ \ t \geq M' \ \Rightarrow\ \pi(t^m - t) < \pi(t^m) < \pi(t^m + t)\ ∎$$

Furthermore, how many primes exist in $(x^k, x^k + x)$? In other words, what is the result of $\lim_{x\to\infty}(\pi(x^k + x) - \pi(x^k))$?

Remark 2. Note that

$$f_1 \sim g_1 \ \wedge \ f_2 \sim g_2 \ \to \ f_1 - f_2 \sim g_1 - g_2$$

doesn't always hold. (1) is a counterexample. So,

$$\lim_{x\to\infty} \frac{\pi(x^m + x) - \pi(x^m)}{(x^m + x)/log(x^m + x) - x^m/log(x^m)} = 1$$

may not hold. It is necessary for us to explore alternative methods.

Lemma 5. Let functions f and g be increasing and satisfy that $\forall x \in \mathbb{R},\ g(x) > f(x) > 0$. If $\lim_{x\to\infty}(g(x) - f(x)) = \infty$ and there exists $k \in (0, 1)$ such that $g(x)^k < g(x) - f(x)$ for sufficiently large x, then

$$\lim_{x\to\infty}(\pi(g(x)) - \pi(f(x))) = \infty$$

proof. Because of (16),

$$\begin{aligned} \forall j \in (0, k),\ \exists N \in \mathbb{R} \quad s.t.\ \ x \geq N \ &\Rightarrow\ \exists p \in \mathbb{P} \ \ with \ \ g(x) - g(x)^j < p < g(x) \\ &\Rightarrow\ \exists p \in \mathbb{P} \ \ with \ \ f(x) < p < g(x) \end{aligned}$$

Let a_n be a sequence defined by $a_1 = g(x)$ and $a_{n+1} = a_n - a_n^j$, then there exists a prime number in the open interval $(a_n - a_n^j, a_n) = (a_{n+1}, a_n)$ and for every $n \in \mathbb{N},\ a_1 \geq a_n$. We consider an m such that $f(x) < a_m,\ f(x) > a_{m+1}$ which forces $\pi(g(x)) - \pi(f(x)) \geq m - 1$. (Such m must exist since $a_n \to 0$ as $n \to \infty$, and depend on x.) Therefore, for sufficiently large x,

$$g(x) - f(x) < \sum_{n=1}^{m}(a_n - a_{n+1}) = \sum_{n=1}^{m} a_n^j \leq \sum_{n=1}^{m} a_1^j = m a_1^j$$

and thus, we obtain

$$m > \frac{g(x) - f(x)}{a_1^j} = \frac{g(x) - f(x)}{g(x)^j} > \frac{g(x)^k}{g(x)^j}$$

Note that

$$\lim_{x\to\infty} \frac{g(x)^k}{g(x)^j} = \infty \ (\because j \in (0, k))$$

Hence,

$$\lim_{x\to\infty}(\pi(g(x)) - \pi(f(x))) = \infty \ \blacksquare$$

Since $\forall x \in \mathbb{R},\ (x + x^m) > x^m > 0$ and for every $m > 0$, there exists $k \in (0, 1)$ such that $(x^m + x)^k < (x^m + x) - x^m = x$ for sufficiently large x,

$$\forall m > 0,\ \lim_{x\to\infty}(\pi(x^m + x) - \pi(x^m)) = \infty$$

References

[1] E. Westzynthius. Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind. Commentationes Helsingfors 5, Nr. 25, 37 S. (1931)., 1931.

[2] Erik Westzynthius. On the distribution of numbers that are not divisible by the first n primes. *Commentat. Phys.-Math.*, 5(25):1–37, 1931.

[3] Daniel A. Goldston, János Pintz, and Cem Y. Yildirim. Primes in tuples i. *Annals of Mathematics*, 170(2):819–862, 2009.

[4] János Pintz. Very large gaps between consecutive primes. *Journal of Number Theory*, 63(2):286–301, 1997.

Keyword: Prime; Prime gap; Andrica conjecture; Cramer conjecture; Oppermann conjecture; Arithmetic mean of primes; Geometric mean of primes; Harmonic mean of primes

E-mail: fghwer1051@gmail.com