A note on gaps

Hisanobu Shinya

January 7, 2009

Abstract

Let $p_k$ denote the $k$-th prime and $d(p_k) = p_k - p_{k-1}$, the difference between consecutive primes. We denote by $N_\epsilon(x)$ the number of primes $\leq x$ which satisfy the inequality $d(p_k) \leq (\log p_k)^{2+\epsilon}$, where $\epsilon > 0$ is arbitrary and fixed, and by $\pi(x)$ the number of primes less than or equal to $x$. In this paper, we first prove a theorem that $\lim_{x \to \infty} N_\epsilon(x)/\pi(x) = 1$. A corollary to the proof of the theorem concerning gaps between consecutive squarefree numbers is stated.

Email address: shinyah18@yahoo.co.jp

2000 MSC: 11B05.

1 Introduction

Let $p_k$ denote the $k$-th prime and for $k > 1$,

$$d(p_k) := p_k - p_{k-1}. $$

Concerning $d(p_k)$, Harald Cramér conjectured that there exists a positive real number $M$ such that

$$d(p_{k+1}) = p_{k+1} - p_k \leq M(\log p_k)^2$$

for all $k \geq 1$. 

Cramér himself showed \[2\] that
\[ p_{k+1} - p_k = o((\log p_k)^3) \]
for all but at most \(o(x/(\log x)^4)\) primes \(\leq x\).

We denote the number of primes less than or equal to a positive real number \(x\) with \(\pi(x)\). In this paper, we prove the following theorem, which supports Cramér’s conjecture.

**Theorem 1.** Let \(x\) be any positive real number and \(N_\epsilon(x)\) the number of primes \(\leq x\) which satisfy the inequality
\[ d(p_k) \leq (\log p_k)^{2+\epsilon}, \]
where \(\epsilon > 0\) is arbitrary and fixed. Then we have
\[ \lim_{x \to \infty} \frac{N_\epsilon(x)}{\pi(x)} = 1. \]

Note that the function for the upper bound of \(d(p_k)\) in Theorem 1 is \((\log p_k)^2\), while that in Cramér’s conjecture is \((\log p_{k-1})^2\); the function is evaluated at the smaller prime of the gap in Cramér’s conjecture. Nevertheless, the prime number theorem implies that
\[ \lim_{k \to \infty} \frac{p_k}{p_{k+1}} = 1, \]
so replacing the function for the upper bound in Cramér’s conjecture by \((\log p_k)^2\) does not change the statement of the conjecture essentially.

In proving Theorem 1 the following lemmas are used.

**Lemma 1.** \([1\ pp. 77]\) For any arithmetical function \(a(n)\) let
\[ A(x) = \sum_{n \leq x} a(n), \]
where \(A(x) = 0\) if \(x < 1\). Assume that \(f\) has a continuous derivative on the interval \([y, x]\), where \(0 < y < x\). Then we have
\[ \sum_{y < n \leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)dt. \]

**Lemma 2.** \([2]\) For all \(k \geq 1\), we have
\[ d(p_{k+1}) = O\left(p_k^{\frac{3}{2} + \frac{7}{200}}\right). \]
2 Proof of Theorem 1

In this section, we give a proof of Theorem 1.

Define

\[ d(p_1) := p_1. \]

We begin by an elementary analysis of the partial sums

\[ Q(t) := \sum_{p_k \leq t} d(p_k) - \sum_{n \leq t} 1. \]

**Lemma 3.** For all positive real numbers \( t \geq 2 \), if \( p_{k-1} \leq t \leq p_k \) \((k \geq 2)\) then we have

\[ -(p_k - p_{k-1}) \leq Q(t) \leq 0. \]

**Proof.** It is easy to see that for \( N = 2, 3, \ldots \), we have

\[
\sum_{p_k \leq p_N} d(p_k) = d(p_1) + d(p_2) + d(p_3) + \cdots + d(p_N) = p_1 + p_2 - p_1 + p_3 - p_2 + \cdots + p_N - p_{N-1} = p_N.
\]

Hence, for \( N \geq 2 \) we have

\[ Q(p_N) = 0. \quad (1) \]

If \( t \) satisfies \( p_{N-1} < t < p_N \) \((N \geq 2)\), then the first partial sums of \( Q \) is constant, and the second ones decrease by 1 as \( t \) increases by 1. The lemma now follows from (1).

Using all the preliminary lemmas above, Theorem 1 is proved as follows. Let \( \epsilon > 0 \) be arbitrary.

We define a function \( \delta_\epsilon(x) \) such that

\[ N_\epsilon(x) = (1 - \delta_\epsilon(x))\pi(x). \quad (2) \]

It is plain that \( 0 < \delta_\epsilon(x) < 1 \). We show that

\[ \lim_{x \to \infty} \delta_\epsilon(x) = 0, \quad (3) \]
thereby proving Theorem 1.

By the definition of $N_\varepsilon(x)$, there are $(1 - \delta_\varepsilon(x))\pi(x)$ primes which are $\leq x$ and satisfy the inequality

$$d(p_k) \leq (\log p_k)^{2+\varepsilon}.$$  

This is equivalent to stating the following:

There are $\delta_\varepsilon(x)\pi(x)$ primes which are $\leq x$ and satisfy

$$d(p_k) > (\log p_k)^{2+\varepsilon}. \quad (4)$$

At this point, we pay our attention to the partial sums

$$\sum_{p \leq x} \frac{d(p)}{p} - \sum_{n \leq x} \frac{1}{n}.$$  

Define

$$p(n) := \begin{cases} 1 & : n \text{ is prime} \\ 0 & : \text{otherwise}. \end{cases}$$

In Lemma 1, we choose

$$f(t) = \frac{1}{t}, \quad a(n) = d(n)p(n) - 1, \quad \text{and} \quad y = 1/2,$$

and obtain

$$\sum_{p \leq x} \frac{d(p)}{p} - \sum_{n \leq x} \frac{1}{n} = \frac{Q(x)}{x} + \int_1^x \frac{Q(t)dt}{t^2}. \quad (5)$$

By Lemmas 2 and 3, it is plain that the limit of the right side of (5) as $x \to \infty$ exists. Since

$$\sum_{n \leq x} \frac{1}{n} \sim \log x, \quad \text{as } x \to \infty,$$

it follows from (5) that

$$\sum_{p \leq x} \frac{d(p)}{p} \sim \log x, \quad \text{as } x \to \infty. \quad (6)$$
Now, given $\epsilon$ and $x > 0$, let $S_\epsilon(x)$ be the set of all primes $\leq x$ which satisfy the last inequality in (4), $|S_\epsilon(x)|$ the number of elements in $S_\epsilon(x)$, and

$$\chi_{\epsilon,x}(n) := \begin{cases} 1 & : n \in S_\epsilon(x) \\ 0 & : \text{otherwise.} \end{cases}$$

Choosing $f(t) = q_{\epsilon}(t) := \frac{(\log t)^{2+\epsilon}}{t}$, $a(n) = \chi_{\epsilon,x}(n)$, and $y = 1/2$ in Lemma (1) we have for each $w \leq x$

$$\sum_{p \in S_\epsilon(x), p \leq w} \frac{d(p)}{p} \geq \sum_{n \leq w} \frac{(\log n)^{2+\epsilon}\chi_{\epsilon,x}(n)}{n} = \frac{\xi_{\epsilon,x}(w)(\log w)^{2+\epsilon}}{w} - \int_{2}^{w} \xi_{\epsilon,x}(t)q'_{\epsilon}(t)dt,$$

where

$$\xi_{\epsilon,x}(t) := \sum_{n \leq t} \chi_{\epsilon,x}(n)$$

and $0 \leq \xi_{\epsilon,x}(w) \leq |S_\epsilon(x)|$ for $w \leq x$. In particular, when $w = x$, we have

$$\sum_{p \in S_\epsilon(x), p \leq x} \frac{d(p)}{p} \geq \frac{|S_\epsilon(x)|(\log x)^{2+\epsilon}}{x} - \int_{2}^{x} \xi_{\epsilon,x}(t)q'_{\epsilon}(t)dt. \quad (7)$$

But since

$$q'_{\epsilon}(t) = \frac{(2 + \epsilon)(\log t)^{1+\epsilon} - (\log t)^{2+\epsilon}}{t^2},$$

it is plain that for each arbitrary $\epsilon > 0$, there exists $t_\epsilon > 0$ such that $q'(\epsilon)(t) < 0$ for all $t \geq t_\epsilon$. This in turn implies that for each arbitrary $\epsilon > 0$, with

$$\xi_{\epsilon,x}(t) \geq 0 \quad \text{for all } t \geq 0,$$

there exists a positive real number $M_\epsilon$, which depends only on $\epsilon$, such that the integral in (7) satisfies

$$- \int_{2}^{x} \xi_{\epsilon,x}(t)q'_{\epsilon}(t)dt > -M_\epsilon \quad (8)$$
for all $x \geq 2$.

With (8), the inequality (7) becomes

$$\sum_{p \in S_\varepsilon(x), p \leq x} \frac{d(p)}{p} = \frac{|S_\varepsilon(x)|(\log x)^{2+\varepsilon}}{x} - \int_2^x \xi_\varepsilon(t)q_\varepsilon(t)dt$$

$$\geq \frac{|S_\varepsilon(x)|(\log x)^{2+\varepsilon}}{x} - M_\varepsilon. \quad (9)$$

Furthermore, by (4), we have $|S_\varepsilon(x)| = \delta_\varepsilon(x)\pi(x)$, and so (9) gives

$$\sum_{p \in S_\varepsilon(x), p \leq x} \frac{d(p)}{p} \geq \frac{|S_\varepsilon(x)|(\log x)^{2+\varepsilon}}{x} - M_\varepsilon$$

$$= \frac{\delta_\varepsilon(x)\pi(x)(\log x)^{2+\varepsilon}}{x} - M_\varepsilon. \quad (10)$$

Finally, by (6) and the prime number theorem

$$\pi(x) \sim \frac{x}{\log x}, \quad \text{as } x \to \infty,$$

letting $x \to \infty$, the inequality (10) becomes

$$\log x \sim \sum_{p \leq x} \frac{d(p)}{p} \geq \sum_{p \in S_\varepsilon(x), p \leq x} \frac{d(p)}{p} \geq \frac{\delta_\varepsilon(x)\pi(x)(\log x)^{2+\varepsilon}}{x} - M_\varepsilon,$$

by which (3) follows immediately. This completes the proof of Theorem 1.

3 Distribution of squarefree numbers

We note the following theorem, which is a corollary to the proof of Theorem 1 in the previous section.

**Theorem 2.** Let $\{a_k\}$ be the sequence of squarefree numbers,

$$s(n) := \begin{cases} 1 & : n = a_k \\ 0 & : \text{otherwise}, \end{cases}$$

$$d(a_k) := a_k - a_{k-1}.$$
and
\[ S(x) := \sum_{n \leq x} s(n). \]

Let \( \epsilon > 0 \) be arbitrary and \( N_\epsilon(x) \) the number of squarefree numbers \( \leq x \) which satisfy the inequality
\[ d(a_k) \leq M(\log a_k)^{1+\epsilon} \]
for some positive constant \( M \). Then we have
\[ \lim_{x \to \infty} \frac{N_\epsilon(x)}{S(x)} = 1. \]

**Proof.** We recall that [3]
\[ S(x) = Ax + O(x^{\frac{1}{2}}), \quad (11) \]
for some constant \( A \). By (11), it is easy to see that
\[ d(a_k) = O(a_k^{\frac{1}{2}}). \]

If we recall how Lemma [3] was derived with Lemma [2], it is plain that
\[ R(t) := \sum_{a_k \leq t} d(a_k) - \sum_{n \leq t} 1 = O(t^{\frac{1}{2}}). \]

The rest of the proof follows a similar pattern. In particular, we note the following analogues:

1. the analogue of [6] is
\[ \sum_{a_k \leq x} \frac{d(a_k)}{a_k} \sim \log x, \quad \text{as } x \to \infty; \]

2. the analogue of \( Q(t) \) is \( R(t) \);

3. the analogue of \( q_\epsilon(t) \) is the function \( \frac{(\log t)^{1+\epsilon}}{t} \).

\[ \square \]

From Theorem [2] one may wonder if the exact order of \( d(a_k) \) is \( O(\log a_k) \).
References

[1] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, New York, 1976.

[2] A. Granville, Harald Cramér and the distribution of prime numbers, Scandinavian Actuarial J. 1 (1995), 12-28

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th ed., Oxford, New York, 1960.