SMALL GAPS BETWEEN PRIMES

SARAH FUJIMORI

1. Introduction

A major question of analytic number thoery is whether there exist infinitely many pairs of primes with a bounded difference. The twin prime conjecture is still an open problem:

Conjecture 1.1. *(Twin Prime Conjecture) There exist infinitely many pairs of primes* (p, q) *such that* $p - q = 2$.

In recent years, there has been significant progress towards solving this problem. In this paper, we state the Bombieri-Vinogradov Theorem and use it to prove a theorem of Goldston, Pintz, and Yildirim from 2005. We also introduce the Elliott-Halberstam conjecture and how proving it helps reduce the proven bound for gaps between primes and describe the recent work of Terence Tao, Maynard, and others.

2. The Bombieri-Vinogradov theorem

The Bombieri-Vinogradov theorem can help us prove Goldston, Pintz, and Yildirim's result. We state it below:

Theorem 2.1. *(Bombieri-Vinogradov) Let* A *be a fixed, positive real number. For all* $x \geq 2$ *and* Q *satisfying* $Q \in [\sqrt{x}(\log x)^{-A}, \sqrt{x}]$. *Additionally, let*

$$\psi(x; q, a) = \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \Lambda(n)$$

Then we have

$$\sum_{q \leq Q} \max_{y < x} \max_{\substack{1 \leq a \leq q \\ \gcd(a,q)=1}} \left| \psi(y; q, a) - \frac{y}{\phi(q)} \right| = O(Q(\log x)^5 \sqrt{x})$$

3. The Work of Goldston, Pintz, and Yildirim

Theorem 3.1. *(Goldston-Pintz-Yildirim) Let* $f(p)$ *be the smallest prime greater than* p. *Then*

$$\liminf_{p \to \infty} \frac{f(p) - p}{\log p} = 0$$

Proof. Let H, N, and R be real numbers satisfying $H = O(\log N)$ and $H < N$, $\log N = O(\log R)$, and $\log R \leq \log N$. Let k and l be arbitrary positive integers.

Define $\mathcal{H}$ as a tuple $\{h_1, h_2, \cdots h_k\} \subseteq [1, H] \cap \mathbb{Z}$, and for some prime p, define $\Omega(p) = \{a : \exists h \in \mathcal{H}, a \equiv -h \pmod p\}$. More generally, for some squarefree integer d, define

Date: June 2019.

$\Omega(d) = \{a : \forall p|d, a \in \Omega(p)\}$. Let $\mathcal{H}$ be a tuple such that $|\Omega(p)| < p$ for all primes p, and enumerate its elements as $h_1, h_2, \cdots h_k$.

Let $H = 6$, and let $\mathcal{H} = \{3, 5\}$. Then, $\Omega(2) = \{1\}$ and $\Omega(3) = \{0, 1\}$, so $\Omega(6) = \{1\}$.

Define two functions λ_R and Λ_R as

$$\lambda_R(d; a) = \begin{cases} 0 & d < R \\ \left(\frac{1}{a!}\right) \mu(d) \left(\log \frac{R}{d}\right)^a & d \geq R, \end{cases}$$

$$\begin{aligned} \Lambda_R(n; \mathcal{H}, a) &= \sum_{d|P(n;\mathcal{H})} \lambda_R(d; a) \\ &= \frac{1}{a!} \sum_{\substack{d|P(n;\mathcal{H}) \\ d \leq R}} \mu(d) \left(\log \frac{R}{d}\right)^a \end{aligned}$$

where $P(n; \mathcal{H}) = \prod_{i=1}^{k} (n + h_i)$. Let $H = 6$, and let $\mathcal{H} = \{3, 5\}$ as in the previous example. Let $a = 3$, $R = 9$, $n = 4$. Then, $P(n, \mathcal{H}) = (4 + 3)(4 + 5) = 63$.

$$\Lambda_R(n; \mathcal{H}, a) = \frac{1}{a!} \sum_{\substack{d|P(n;\mathcal{H}) \\ d \leq R}} \mu(d) \left(\log \frac{R}{d}\right)^a = \frac{1}{6} \sum_{\substack{d|63 \\ d \leq 9}} \mu(d) \left(\log \frac{9}{d}\right)^3$$

$$= \frac{1}{6}(\mu(1) \log 9 + \mu(3) \log 3 + \mu(7) \log \frac{9}{7} + \mu(9) \log(1)) = \frac{1}{6}(\log \frac{7}{3})$$

The motivation for these choices of functions comes from the following identity:

Proposition 3.2. *Suppose m is a positive integer, and n has more than m distinct prime factors. Then, we have*

$$\sum_{d|n} \mu(d) \left(\log \frac{n}{d}\right)^m = 0$$

.

Proof. We prove this by strong induction on m. We first prove a slightly different identity, that

$$\sum_{d|n} \mu(d) (\log d)^m = 0$$

For our base case, we have $m = 1$, so the sum is $\sum_{d|n} \mu(d) (d)$. Recall that this is $-\Lambda$ where Λ is the von Mangoldt function, so since n has more than 1 factors, the sum is 0 and the identity holds.

Next, assume that our claim holds for all k satisfying $1 \leq k \leq m$. We show that

$$\sum_{d|n} \mu(d) (\log d)^{m+1} = 0,$$

where n has more than $m + 1$ prime factors. Factor out a prime from n, so that $n = p^a b$ for positive integers a and b. We separate the sum based on $\gcd(d, p)$:

$$\sum_{d|n} \mu(d) (\log d)^{m+1} = \sum_{\substack{d|n \\ \gcd(d,p)=1}} \mu(d) (\log d)^{m+1} + \sum_{\substack{d|n \\ \gcd(d,p)>1}} \mu(d) (\log d)^{m+1}$$

$$= \sum_{d|b} \mu(d) (\log d)^{m+1} + \sum_{d|p^{a-1}b} \mu(pd) (\log pd)^{m+1}$$

In the sum on the right, if d is a multiple of p, then we have $\mu(d) = 0$, so we only sum over $d|b$:

$$= \sum_{d|b} \mu(d) (\log d)^{m+1} + \sum_{d|b} \mu(pd) (\log pd)^{m+1}$$

Since $\gcd(d, p) = 1$ now in the sum on the right, we can factor $\mu(pd)$ into $\mu(p)\mu(d)$:

$$= \sum_{d|b} \mu(d) (\log d)^{m+1} - \sum_{d|b} \mu(d) (\log d - \log p)^{m+1}$$

$$= \sum_{d|b} \mu(d) \left((\log d)^{m+1} - (\log d - \log p)^{m+1}\right)$$

We expand out the sum using the Binomial Theorem:

$$= \sum_{d|b} \mu(d) \left((\log d)^{m+1} - \sum_{i=0}^{m+1} \binom{m+1}{i} (\log d)^{m+1-i} (-\log p)^{i} \right)$$

$$= \sum_{d|b} \mu(d) \left(- \sum_{i=1}^{m+1} \binom{m+1}{i} (\log d)^{m+1-i} (-\log p)^{i} \right)$$

Shifting the indices of the second sum down by 1,

$$= \sum_{d|b} \mu(d) \left(- \sum_{i=0}^{m} \binom{m+1}{i+1} (\log d)^{m-i} (-\log p)^{i+1} \right)$$

Switching the sums yields

$$= -\sum_{i=0}^{m} \binom{m+1}{i+1} (-\log p)^{i+1} \sum_{d|b} \mu(d) (\log d)^{m-i}$$

Since b has more than m factors and the $m-i$ ranges from 0 to m, the sum $\sum_{d|b} \mu(d) (\log d)^{m-i}$ is always 0 by the induction hypothesis. Thus, the general identity holds.

We now prove our original claim; that

$$\sum_{d|n} \mu(d) \left(\log \frac{n}{d}\right)^{m} = 0$$

. By the Binomial Theorem,

$$\sum_{d|n} \mu(d) \left(\log \frac{n}{d}\right)^{m} = \sum_{d|n} \mu(d) \sum_{k=0}^{m} \binom{m}{k} (\log n)^{k} (-\log d)^{m-k}$$

Switching the order of the sums,

$$= \sum_{k=0}^{m} \binom{m}{k} (\log n)^{k} \sum_{d|n} \mu(d) (-\log d)^{m-k}$$

We proved that $\sum_{d|n} \mu(d)(-\log d)^a$ for any integer a, so the sum is 0, as desired. ☐

If for all i with $1 \leq i \leq k$, $n + h_i$ is prime, then $P(n; \mathcal{H})$ has exactly k prime factors, so Λ_R detects this property, but with the sum truncated at R.

We aim to approximate the sum

$$\sum_{N<n\leq 2N} \left(\Lambda_R(n; \mathcal{H}, k+l)\right)^2.$$

Proposition 3.3.

$$\sum_{N<n\leq 2N} (\Lambda_R(n; \mathcal{H}, k+l))^2 = NT + O\left(\left(\sum_{d\leq R} |\Omega(d)||\lambda_R(d; k+l)\right)^2\right)$$

where

$$T = \sum_{d_1,d_2} \frac{|\Omega(\operatorname{lcm}(d_1,d_2))|}{\operatorname{lcm}(d_1,d_2)} \lambda_R(d_1; k+l)\lambda_R(d_2; k+l)$$

Proof. By the definition of Λ_R, we have

$$\sum_{N<n\leq 2N} \left(\sum_{d|P(n;\mathcal{H})} (\lambda_R(d; k+l))^2\right)$$

$$= \sum_{N<n\leq 2N} \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} \lambda_R(d_1; k+l)\lambda_R(d_2; k+l)$$

Suppose $d_1|P(n; \mathcal{H})$. Then, for every prime p dividing d_1, $P(n; \mathcal{H}) = (n+h_1)(n+h_2)\cdots(n+h_k) \equiv 0 \pmod{p}$, so there exists i satisfying $n \equiv -h_i \pmod{p}$. Thus, since $n \in \Omega(p)$ for every prime $p|d_1$, we conclude that $d_1|P(n; \mathcal{H})$ is equivalent to $n \in \Omega(d_1)$, and similarly for d_2:

$$= \sum_{d_1,d_2\leq R} \lambda_R(d_1; k+l)\lambda_R(d_2; k+l) \sum_{\substack{N<n\leq 2N \\ n\in\Omega(d_1),\Omega(d_2)}} 1$$

Since $n \in \Omega(d_1), \Omega(d_2) \iff n \in \Omega(\operatorname{lcm}(d_1, d_2))$, we have

$$\sum_{\substack{N<n\leq 2N \\ n\in\Omega(\operatorname{lcm}(d_1,d_2))}} 1 = \frac{N|\Omega(\operatorname{lcm}(d_1,d_2))|}{\operatorname{lcm}(d_1,d_2)} + O(|\Omega(d_1)||\Omega(d_2)|)$$

Recall that we set

$$T = \sum_{d_1,d_2} \frac{|\Omega(\operatorname{lcm}(d_1,d_2))|}{\operatorname{lcm}(d_1,d_2)} \lambda_R(d_1; k+l)\lambda_R(d_2; k+l)$$

Then, we have

$$\sum_{N<n\leq 2N} (\Lambda_R(n; \mathcal{H}, k+l))^2 = NT + O\left(\sum_{d_1,d_2\leq R} |\Omega(d_1)||\Omega(d_2)|(\lambda_R(d_1; k+l)\lambda_R(d_2; k+l))\right)$$

$$= NT + O\left(\left(\sum_{d\le R} |\Omega(d)||\lambda_R(d;k+l)\right)^2\right)$$

as desired. □

Evaluating T uses complex analysis, so we will skip it; the result is the following:

Lemma 3.4.

$$\sum_{N<n\le 2N} (\Lambda_R(n;\mathcal{H},k+l))^2 = \frac{G(\mathcal{H})}{(k+2l)^2}\binom{2l}{l} N(\log R)^{k+2l} + O\left(N(\log N)^{k+2l-1}(\log\log N)^c\right)$$

where

$$G(\mathcal{H}) = \prod_{p\,prime}\left(1 - \frac{|\Omega(p)|}{p}\right)\left(1-\frac{1}{p}\right)^{-k},$$

and c is a constant.

Next, we analyze the behavior of another function. Let

$$\varpi(n) = \begin{cases} \log n & n \text{ prime} \\ 0 & \text{otherwise} \end{cases}$$

Note that $\sum_{n\le x}\varpi(n) = \theta(x)$.
We look at the expression

$$\sum_{N<n\le 2N} \varpi(n+h)(\Lambda_R(n;\mathcal{H},k+l))^2$$

where h is an integer satisfying $h \le H$. We claim the following:

Lemma 3.5.

$$\sum_{N<n\le 2N} \varpi(n+h)(\Lambda_R(n;\mathcal{H},k+l))^2 = NT' + O\left(\frac{N}{(\log N)^{\frac{A}{3}}}\right)$$

where

$$T' = \sum_{d_1,d_2\le R} \frac{\lambda_R(d_1;k+l)\lambda_R(d_2;k+l)}{\phi(\mathrm{lcm}(d_1,d_2))} \sum_{b\in\Omega(\mathrm{lcm}(d_1,d_2))} \delta((b+h,\mathrm{lcm}(d_1,d_2)))$$

with $\delta((a,b)) = 0$ if $a=b$ and 1 otherwise.

Proof. If $R < N$, then by the definition of $\varpi(n+h)$, the sum is equal to

$$\sum_{N<n\le 2N} \varpi(n+h)(\Lambda_R(n;\mathcal{H}\backslash\{h\},k+l))^2$$

Let $\vartheta(y;a,q) = \sum_{\substack{y<n\le 2y \\ n\equiv a \pmod q}} \varpi(n)$. An equivalent form of the Bombieri-Vinogradov Theorem says that for fixed $A>0$, there exists $C>0$ such that when $Q \le \frac{x^{\frac{1}{2}}}{(\log x)^C}$,

$$\sum_{q\le Q} \max_{y\le x} \max_{\substack{1\le a\le q-1 \\ \gcd(a,q)=1}} \left|\vartheta(y;a,q) - \frac{y}{\phi(q)}\right| = O\left(\frac{x}{(\log x)^A}\right).$$

Thus, we assume that $R \leq \frac{N^{\frac{\theta}{2}}}{(\log N)^C}$ for some constant C.

Again, we expand the square:

$$\sum_{N<n\leq 2N} \varpi(n+h)(\Lambda_R(n;\mathcal{H},k+l))^2 = \sum_{N<n\leq 2N} \varpi(n+h) \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} \lambda_R(d_1;\mathcal{H},k+l)\lambda_R(d_2;\mathcal{H},k+l)$$

$$= \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} \lambda_R(d_1;\mathcal{H},k+l)\lambda_R(d_2;\mathcal{H},k+l) \sum_{N<n\leq 2N} \varpi(n+h)$$

$$= \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} \lambda_R(d_1;\mathcal{H},k+l)\lambda_R(d_2;\mathcal{H},k+l) \sum_{b\in\Omega(\mathrm{lcm}(d_1,d_2))} \delta((b+h,\mathrm{lcm}(d_1,d_2)))\vartheta(N;b+h,\mathrm{lcm}(d_1,d_2))$$

Let $L = \lambda_R(d_1;\mathcal{H},k+l)\lambda_R(d_2;\mathcal{H},k+l)$, and assume that $R \leq \frac{N^{\frac{\theta}{2}}}{(\log N)^C}$ for some constant C. Because of the Bombieri-Vinogradov Theorem, we split the sum up based on the value of $|\Omega(\mathrm{lcm}(d_1,d_2))|$:

$$= \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} L \sum_{\substack{b\in\Omega(\mathrm{lcm}(d_1,d_2)) \\ |\Omega(\mathrm{lcm}(d_1,d_2)|\leq(\log N)^{\frac{A}{2}}}} \delta((b+h,\mathrm{lcm}(d_1,d_2)))\vartheta(N;b+h,\mathrm{lcm}(d_1,d_2))$$

$$+ \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} L \sum_{\substack{b\in\Omega(\mathrm{lcm}(d_1,d_2)) \\ |\Omega(\mathrm{lcm}(d_1,d_2)|>(\log N)^{\frac{A}{2}}}} \delta((b+h,\mathrm{lcm}(d_1,d_2)))\vartheta(N;b+h,\mathrm{lcm}(d_1,d_2))$$

As in the statement of our lemma, set

$$T' = \sum_{\substack{d_1|P(n;\mathcal{H}) \\ d_2|P(n;\mathcal{H})}} \frac{L}{\phi(\mathrm{lcm}(d_1,d_2))}$$

Let $\tau_k(n) = \sum_{d|n} d^k$. We can approximate the second sum as

$$N(\log R)^{2k+2l}(\log N) \sum_{d_1,d_2\leq R} \frac{(\tau_k(\mathrm{lcm}(d_1,d_2)))(\Omega(\mathrm{lcm}(d_1,d_2)))}{\left((\log N)^{\frac{A}{2}}\right)(\mathrm{lcm}(d_1,d_2))}$$

$$= O\left(\frac{N}{(\log N)^{\frac{A}{3}}}\right)$$

We apply the Bombieri-Vinogradov Theorem to the first part of the sum, since we have a bound on the number of terms, so we can approximate the first sum as NT', with a smaller error. Combining these two approximations gives the desired result,

$$\sum_{N<n\leq 2N} \varpi(n+h)(\Lambda_R(n;\mathcal{H},k+l))^2 = NT' + O\left(\frac{N}{(\log N)^{\frac{A}{3}}}\right)$$

□

Again, evaluating T' uses complex analysis, so we skip it and state the result:

Lemma 3.6. *Assuming that* $R \leq \frac{N^{\frac{\theta}{2}}}{(\log N)^C}$ *for a sufficiently large* $C > 0$ *depending on* k *and* l,

$$\sum_{N<n\leq 2N} \varpi(n+h)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

$$= \begin{cases} \frac{G(\mathcal{H}\cup\{h\})}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l} + O\left(N(\log N)^{k+2l-1}(\log\log N)^c\right) & h \notin \mathcal{H} \\ \frac{G(\mathcal{H})}{(k+2l+1)!}\binom{2(l+1)}{l+1}N(\log R)^{k+2l+1} + O\left(N(\log N)^{k+2l}(\log\log N)^c\right) & h \in \mathcal{H} \end{cases}$$

(Note that these are very similar; the expression for the second case is the expression for the first case with k *shifted down 1 and* l *shifted up by 1.)*

We now use these lemmas for the main proof. Consider the expression

$$\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{h\leq\mathcal{H}}\varpi(n+h)-\log 3N\right)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

If this expression is positive, then one of the terms must be greater than 0, there exists some n satisfying $N < n \leq 2N$ such that $\sum_{h\leq H}\varpi(n+h) - \log 3N > 0$. Recall that $H < N$. If only one of $n+h$ for $1 \leq h \leq H$ is prime, then the sum is $\log(n+h_1)$ for some h_1 between 1 and H, and we have $\log(n+h_1)-\log(3N) \leq \log(2N+H)-\log(3N) < \log(3N)-\log(3N) = 0$. Thus, there must exist two values of h such that $n+h$ is prime; we can confirm that two primes is enough because the sum will be at least $\log(n+1)^2 > \log(N+1)^2 > \log 3N$.

We conclude that since there is a interval of length H in $(N, 2N+H]$, we have

$$\min_{N<p_r\leq 2N+H} p_{r+1}-p_r \leq H$$

By Lemma 3.4, we have

$$\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{h\leq\mathcal{H}}\varpi(n+h)-\log 3N\right)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

$$= \sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{h\leq\mathcal{H}}\varpi(n+h)\right)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2 - \sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\log 3N\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

$$= \sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{h\leq\mathcal{H}}\varpi(n+h)\right)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

$$-\log 3N\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}} G(H)\left(\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l} + O\left(N(\log N)^{k+2l-1}(\log\log N)^c\right)\right)$$

$$= \sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}\varpi(n+h)+\sum_{\substack{h\leq\mathcal{H}\\h\notin\mathcal{H}}}\varpi(n+h)\right)\left(\Lambda_R(n;\mathcal{H},k+l)\right)^2$$

$$-\log 3N\left(\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}+O\left(N(\log N)^{k+2l-1}(\log\log N)^c\right)\right)\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}G(H)$$

For the sum $\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}G(H)$, we state a theorem of Gallagher:

Theorem 3.7.

$$\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}G(H)\sim H^k$$

as $H\to\infty$

We omit the proof, as it is beyond the scope of this paper.

Using Gallagher's theorem, the expression above is equal to

$$=\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{N<n\leq 2N}\left(\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}\varpi(n+h)+\sum_{\substack{h\leq\mathcal{H}\\h\notin\mathcal{H}}}\varpi(n+h)\right)(\Lambda_R(n;\mathcal{H},k+l))^2$$

$$-H^k\log N\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}+o\left(NH^k(\log N)^{k+2l+1}\right)$$

Using Lemma 3.6, we can simplify the first sum:

$$=\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}\frac{G(\mathcal{H}\cup\{h\})}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}+O\left(N(\log N)^{k+2l-1}(\log\log N)^c\right)$$

$$+\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}\frac{G(H)}{(k+2l+1)!}\binom{2(l+1)}{l+1}N(\log R)^{k+2l-1}+O\left(N(\log N)^{k+2l}(\log\log N)^c\right)$$

$$-H^k\log N\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}+o\left(NH^k(\log N)^{k+2l+1}\right)$$

$$=\left(\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}\right)\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}G(\mathcal{H}\cup\{h\})$$

$$+\left(\frac{1}{(k+2l+1)!}\binom{2(l+1)}{l+1}N(\log R)^{k+2l-1}\right)\sum_{\substack{\mathcal{H}\subseteq[1,H]\\|\mathcal{H}|=k}}\sum_{\substack{h\leq\mathcal{H}\\h\in\mathcal{H}}}G(H)$$

$$-H^k\log N\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}+o\left(NH^k(\log N)^{k+2l+1}\right)$$

where we include the error bounds on the sums in the last error bound.

Using Gallagher's theorem, the first sum is just the sum over all tuples of length $k+1$, so this is equal to

$$=\frac{1}{(k+2l)!}\binom{2l}{l}N(\log R)^{k+2l}H^{k+1}+\frac{1}{(k+2l+1)!}\binom{2(l+1)}{l+1}N(\log R)^{k+2l-1}H^k$$

$$-H^k \log N \frac{1}{(k+2l)!}\binom{2l}{l} N(\log R)^{k+2l} + o\left(NH^k(\log N)^{k+2l+1}\right)$$

Factoring out $\frac{1}{(k+2l)!}\binom{2l}{l} NH^k(\log R)^{k+2l}$ yields

$$= \frac{1}{(k+2l)!}\binom{2l}{l} NH^k(\log R)^{k+2l}\left(H + \frac{2k(2l+1)}{(k+2l+1)(l+1)}(\log R) - \log N + \varepsilon(\log N)\right)$$

where the ε comes from the error term and is greater than 0.

Recall that $R \leq \frac{N^{\frac{1}{4}}}{(\log N)^C}$ for a constant C; then, taking the logarithm of both sides yields $\log R \leq \frac{1}{4}\log N - \log(C \log N) \leq \frac{1}{4}\log N$. Thus, since the part of the expression that we factored out is positive, for the expression to be positive we must have

$$H \geq (\log N)\left(1 + \varepsilon - \frac{2k(2l+1)}{4(k+2l+1)(l+1)}\right).$$

Recall that k and l are arbitrary, so if we set $l = \lfloor\sqrt{k}\rfloor$, then we have $\frac{H}{\log N} > 0$. We conclude that the original expression is positive, and so taking the limit as $N \to \infty$ of $\min_{N<p_r\leq 2N+H} \frac{p_{r+1}-p_r}{\log N} \leq \frac{H}{\log N}$ gives

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

as desired.

□

4. The Elliott-Halberstam Conjecture

Let $\pi(x)$ denote the number of primes less than or equal to x, and let $\pi(x;q,a)$ denote the number of primes p such that $p \leq x$ and $p \equiv a \pmod{q}$. We know that for a, b satisfying $\gcd(a,n) = 1$ and $\gcd(b,n) = 1$, we have $\lim_{x\to\infty} \frac{\pi(x;n,a)}{\pi(x;n,b} = 1$, which implies that

$$\pi(x;q,a) \sim \frac{\pi(x)}{\phi(q)}$$

The Elliott-Halberstam conjecture is a generalization of the Bombieri-Vinogradov Theorem; it is as follows:

Conjecture 4.1. *(Elliott-Halberstam) Define an error function for the Dirichlet approximation of $\pi(x;q,a)$: let*

$$E(q;x) = \max_{\gcd(a,q)=1}\left|\pi(x;q,a) - \frac{\pi(x)}{\phi(q)}\right|$$

.

Then for every $\theta < 1$ and $A > 0$, there exists a constant c such that for all $x > 2$,

$$\sum_{1\leq q\leq x^\theta} E(x;q) \leq \frac{cx}{(\log x)^A}$$

(Note that the conjecture fails when $\theta = 1$.

If we assume the Elliott-Halberstam conjecture on primes in arithmetic progression, then Terence Tao, James Maynard, and others proved that we have

Theorem 4.2.

$$\liminf_{p\to\infty} f(p) - p \leq 6$$

Without assuming the Elliott-Halberstam conjecture, the current best bound is also due to Terence Tao, James Maynard, and others:

Theorem 4.3.

$$\liminf_{p\to\infty} f(p) - p \leq 246$$

We omit these proofs, since they are beyond the scope of this paper.

References

[Gal76] P.X. Gallagher. On the distribution of primes in short intervals. *Mathematika*, 23(1):4–9, June 1976.

[GGPY09] D. Goldston, S.W. Graham, J. Pintz, and C. Yildirim. Small gaps between primes or almost primes. *Transactions of the American Mathematical Society*, 361(10):5285–5330, October 2009.

[GMPY06] D. Goldston, Y. Motohashi, J. Pintz, and C. Yildirim. Small gaps between primes exist. In *Proceedings of the Japan Academy*, number 4 in 82, April 2006.

[GPY09] D. Goldston, J. Pintz, and C. Yildirim. Primes in tuples i. *Annals of Mathematics*, 170(2):819–862, 2009.

[TM14a] T. Tao and J. Maynard. Bounded intervals with many primes, February 2014.

[TM14b] T. Tao and J. Maynard. Using the generalized elliott-halberstam conjecture to enlarge the sieve support yet further, January 2014.