#A50 INTEGERS 23 (2023)

ON RATIOS OF CONSECUTIVE PRIME GAPS

William D. Banks
Department of Mathematics, University of Missouri, Columbia, Missouri
bankswd@missouri.edu

Received: 5/15/22, Revised: 3/4/23, Accepted: 7/7/23, Published: 7/21/23

Abstract

Let p_n be the nth smallest prime and $d_n := p_{n+1} - p_n$ the gap between p_n and p_{n+1}. For any fixed $c \geq 0$, we conjecture that the estimate

$$\bigl|\{p_n \leq x : d_{n+1}/d_n \geq c\}\bigr| = (c+1)^{-1}\pi(x) + O\bigl(x(\log x)^{-3/2+\varepsilon}\bigr)$$

holds for any $\varepsilon > 0$, and we give a heuristic argument to support this conjecture which is based on a strong form of the Hardy-Littlewood conjectures.

1. Introduction

Let $p_1 := 2 < p_2 := 3 < p_3 := 5 < \cdots$ be the sequence of prime numbers. The Prime Number Theorem implies that the nth prime gap

$$d_n := p_{n+1} - p_n$$

has length $\log p_n$ on average; in other words, the nth normalized prime gap

$$\widehat{d}_n := d_n / \log p_n$$

takes the value one on average. For any fixed number $c \geq 0$, heuristics based on Cramér's probabilistic model of the primes lead to the conjecture that

$$\lim_{N\to\infty} N^{-1}\bigl|\{n \leq N : \widehat{d}_n \geq c\}\bigr| = e^{-c}. \tag{1}$$

Thus, we expect that the normalized prime gaps are distributed according to a Poisson process. We refer the reader to the expository article [7] of Soundararajan for an excellent account of these intriguing statistics.

The conjectural relation (1) also leads to a natural conjecture concerning ratios of consecutive prime gaps. More specifically, it seems likely that for any fixed $c \geq 0$ the following relation holds:

$$\lim_{N\to\infty} N^{-1}\bigl|\{n \leq N : d_{n+1}/d_n \geq c\}\bigr| = (c+1)^{-1}. \tag{2}$$

DOI: 10.5281/zenodo.8174512

Indeed, (1) suggests that for any fixed $x \geq 0$ the normalized prime gap $\widehat{d}_n$ lies in the infinitesimal interval $(x, x + dx)$ with probability $e^{-x}\,dx$. The probability that both events $\widehat{d}_n \in (x, x + dx)$ and $\widehat{d}_{n+1} \in (y, y + dy)$ happen simultaneously is therefore $e^{-x-y}\,dx\,dy$, assuming these two events are independent. Integrating over all pairs (x, y) with $y \geq cx$, we expect that

$$\lim_{N\to\infty} N^{-1}\big|\big\{n \leq N : \widehat{d}_{n+1}/\widehat{d}_n \geq c\big\}\big| = \int_0^\infty \int_{cx}^\infty e^{-x-y}\,dy\,dx = (c+1)^{-1}.$$

Since $\log p_{n+1} = \log p_n + o(1)$ as $n \to \infty$, it follows that $\widehat{d}_{n+1}/\widehat{d}_n \to d_{n+1}/d_n$, and in this way we arrive at the conjectural relation (2).

Note that (2) can be reformulated as follows. Let $\pi(x)$ denote the prime counting function, and for any fixed $c \geq 0$ let $\pi_c(x)$ be the function given by

$$\pi_c(x) := \big|\big\{p_n \leq x : d_{n+1}/d_n \geq c\big\}\big|.$$

Then (2) is equivalent to the conjectural relation

$$\pi_c(x) \sim (c+1)^{-1}\pi(x) \qquad (x \to \infty). \tag{3}$$

In this note, we present a heuristic argument, based on a quantitative form of the Hardy-Littlewood prime k-tuple conjecture, to support the following stronger form of the conjecture (3).

Conjecture 1. For any $c \geq 0$ and $\varepsilon > 0$, one has the estimate

$$\pi_c(x) = (c+1)^{-1}\pi(x) + O\big(x(\log x)^{-3/2+\varepsilon}\big),$$

where the implied constant depends only on c and ε.

The results of the present paper are inspired by a celebrated work of Lemke Oliver and Soundararajan [4] that studies the surprisingly erratic distribution of pairs of consecutive primes among the $\phi(q)^2$ permissible reduced residue classes modulo q. In [4] a conjectural explanation for this phenomenon is offered, which is based on a strong form of the Hardy-Littlewood conjecture (see [4, Equation (2.4)]).

2. Preliminaries

2.1 Notation

Let $\mathbb{P}$ denote the set of primes in $\mathbb{N}$.

For an arbitrary set $\mathcal{S}$, we use $\mathbf{1}_\mathcal{S}$ to denote its indicator function:

$$\mathbf{1}_\mathcal{S}(n) := \begin{cases} 1 & \text{if } n \in \mathcal{S}, \\ 0 & \text{if } n \notin \mathcal{S}. \end{cases}$$

Throughout the paper, implied constants in symbols O, $\ll$ and $\gg$ may depend (where obvious) on the parameters c and ε but are independent of other variables except where indicated. For given functions F and G, the notations $F \ll G$, $G \gg F$ and $F = O(G)$ are all equivalent to the statement that the inequality $|F| \le k|G|$ holds with some constant $k > 0$.

2.2 The Modified Singular Series

Gallagher [2] has shown that the relation (1) is a consequence of Hardy and Littlewood's [3, p. 61] quantitative version of the prime k-tuple conjecture, which asserts that for every finite subset $\mathcal{H}$ of $\mathbb{Z}$ one has

$$\sum_{n\le x}\prod_{h\in\mathcal{H}}\Lambda(n+h) = (\mathfrak{S}(\mathcal{H})+o(1))x \qquad (x\to\infty). \tag{4}$$

Here, Λ is the von Mangoldt function, and $\mathfrak{S}(\mathcal{H})$ is the singular series defined by

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1-\frac{|(\mathcal{H} \bmod p)|}{p}\right)\left(1-\frac{1}{p}\right)^{-|\mathcal{H}|}.$$

To prove (1), Gallagher showed that for any fixed $\ell \ge 1$ one has

$$\sum_{\substack{\mathcal{H}\subseteq[0,n]\\ |\mathcal{H}|=\ell}} \mathfrak{S}(\mathcal{H}) \sim \binom{n+1}{\ell} \qquad (n\to\infty). \tag{5}$$

In other words, the singular series has an average value of one.

In their study of the distribution of primes in longer intervals, Montgomery and Soundararajan [6] employ a more precise form of the Hardy-Littlewood conjecture (4), which is supported by results in their earlier paper [5]: If $\mathcal{H}$ is any finite set of integers, then

$$\sum_{n\le x}\prod_{h\in\mathcal{H}}\mathbf{1}_{\mathbb{P}}(n+h) = \mathfrak{S}(\mathcal{H})\int_2^x \frac{du}{(\log u)^{|\mathcal{H}|}} + O(x^{1/2+\varepsilon}).$$

In [6] the authors also introduce the modified singular series

$$\mathfrak{S}_0(\mathcal{H}) := \sum_{\mathcal{H}'\subseteq\mathcal{H}} (-1)^{|\mathcal{H}\setminus\mathcal{H}'|}\mathfrak{S}(\mathcal{H}'),$$

which satisfies

$$\mathfrak{S}(\mathcal{H}) = \sum_{\mathcal{H}'\subseteq\mathcal{H}} \mathfrak{S}_0(\mathcal{H}'),$$

with $\mathfrak{S}(\varnothing) = \mathfrak{S}_0(\varnothing) = 1$. The modified singular series arises naturally in the following formulation of the Hardy-Littlewood conjecture (regarding the elements of $\mathcal{H}$ as being small relative to x): If $\mathcal{H}$ is any finite set of integers, then

$$\sum_{n\le x}\prod_{h\in\mathcal{H}}\left(\mathbf{1}_{\mathbb{P}}(n+h)-\frac{1}{\log n}\right)=\mathfrak{S}_0(\mathcal{H})\int_2^x\frac{du}{(\log u)^{|\mathcal{H}|}}+O(x^{1/2+\varepsilon}). \tag{6}$$

Here, the term $1/\log n$ being subtracted from $\mathbf{1}_{\mathbb{P}}(n+h)$ represents the probability that the "random number" $n+h$ is a prime number.

2.3 Essential Estimates

Montgomery and Soundararajan [6, Theorem 2] gave the following refinement of Gallagher's estimate (5):

$$\sum_{\substack{\mathcal{H}\subseteq[0,n]\\|\mathcal{H}|=\ell}}\mathfrak{S}_0(\mathcal{H})=\frac{\mu_\ell}{\ell!}(-n\log n+An)^{\ell/2}+O_\ell\big(n^{\ell/2-1/(7\ell)+\varepsilon}\big),$$

showing that the modified singular series exhibits square-root cancellation in each variable. Here, μ_ℓ is the ℓth moment of the standard Gaussian:

$$\mu_\ell:=\begin{cases}1\cdot3\cdots(\ell-1) & \text{if } \ell \text{ is even,}\\ 0 & \text{if } \ell \text{ is odd,}\end{cases}$$

and A is given by

$$A:=2-C_0-\log 2\pi \tag{7}$$

with C_0 the Euler-Mascheroni constant. For small values of ℓ, one can be more precise; in particular, [6, Equation (16)] implies for the case $\ell=2$:

$$\sum_{\substack{\mathcal{H}\subseteq[0,n]\\|\mathcal{H}|=2}}\mathfrak{S}_0(\mathcal{H})=-\tfrac12 n\log n+\tfrac12 An+O(n^{1/2+\varepsilon}). \tag{8}$$

Throughout the sequel, we denote (as in [4])

$$\alpha(u):=1-\frac{1}{\log u}\qquad(u>1).$$

Taking into account that $|\alpha(u)|<\frac12$ for $u\in[2,3]$ and $0<\alpha(u)<1$ for all $u\ge3$, the following is a straightforward variant of Banks and Guo [1, Lemma 2.3].

Lemma 1. *Let f be an arithmetic function such that*

$$\|f\|_\infty:=\sup\{|f(n)|:n\ge1\}<\infty.$$

Uniformly for $2 \le u \le x$ and $y \ge (\log x)^3$ we have

$$\sum_{\substack{n\le y\\ 2|n}} f(n)\alpha(u)^n = \sum_{\substack{n\ge 1\\ 2|n}} f(n)\alpha(u)^n + O\big(x^{-1}\|f\|_\infty\big),$$

where the implied constant is absolute.

Lemma 2. *Let $c \ge 0$ be fixed. Uniformly for $2 \le u \le x$ and $y \ge (\log x)^2$ we have*

$$\sum_{\substack{m,n\le y\\ 2|m,\,2|n\\ n\ge cm}} \alpha(u)^{m+n} = (4c+4)^{-1}\big((\log u)^2 + O(\log u)\big),$$

where the implied constant depends only on c.

Proof. Writing $\alpha \coloneqq \alpha(u)$ we have

$$\sum_{\substack{n\ge 1\\ 2|n,\, n\ge cm}} \alpha^n = \frac{\alpha^{2\lceil cm/2\rceil}}{1-\alpha^2}.$$

Since

$$|\alpha|^{cm} \le |\alpha|^{2\lceil cm/2\rceil} < |\alpha|^{cm-2}, \qquad \alpha^{-2} = 1 + O((\log u)^{-1}),$$

and

$$1-\alpha^2 = 2(\log u)^{-1} + O((\log u)^{-2}),$$

it follows that

$$\sum_{\substack{n\ge 1\\ 2|n,\, n\ge cm}} \alpha^n = \tfrac{1}{2}\alpha^{cm}(\log u + O(1)). \tag{9}$$

Using Lemma 1 with $f \coloneqq \mathbf{1}_{\mathbb{N}}$ we see that

$$g(m) \coloneqq \sum_{\substack{n\le y\\ 2|n,\, n\ge cm}} \alpha^n = \sum_{\substack{n\ge 1\\ 2|n,\, n\ge cm}} \alpha^n + O(x^{-1}) = \tfrac{1}{2}\alpha^{cm}\log u + O(\alpha^{cm} + x^{-1});$$

in particular, $\|g\|_\infty \le \log u$. A second application of Lemma 1 with $f \coloneqq g$ gives

$$\begin{aligned}
\sum_{\substack{m,n\le y\\ 2|m,\,2|n\\ n\ge cm}} \alpha^{m+n} &= \sum_{\substack{m\le y\\ 2|m}} g(m)\alpha^m = \sum_{\substack{m\ge 1\\ 2|m}} g(m)\alpha^m + O(x^{-1}\log u)\\
&= \sum_{\substack{m\ge 1\\ 2|m}} \big(\tfrac{1}{2}\alpha^{(c+1)m}\log u + O(\alpha^{(c+1)m} + \alpha^m x^{-1})\big) + O(x^{-1}\log u)\\
&= \frac{\tfrac{1}{2}\alpha^{2c+2}}{1-\alpha^{2c+2}}(\log u + O(1)) + O(x^{-1}\log u).
\end{aligned}$$

Since

$$\frac{\frac{1}{2}\alpha^{2c+2}}{1-\alpha^{2c+2}} = \frac{\frac{1}{2}(\log u - 1)^{2c+2}}{(\log u)^{2c+2} - (\log u - 1)^{2c+2}} = (4c+4)^{-1}\log u + O(1),$$

the result follows. □

The following statement is a straightforward extension of [1, Lemma 2.4].

Lemma 3. *Fix $\theta \geq 0$, $\xi = 0$ or 1, and $\lambda > 0$. For all $u \geq 2$ the sums*

$$F(\theta,\xi,\lambda;u) := \sum_{\substack{n\geq 1\\ 2|n}} n^\theta (\log n)^\xi \alpha(u)^{\lambda n}$$

and

$$G(\xi,\lambda;u) := \sum_{\substack{n\geq 1\\ 2|n}} \mathfrak{S}_0(\{0,n\}) n^\xi \alpha(u)^{\lambda n}$$

satisfy the estimates

$$F(\theta,0,\lambda;u) = \tfrac{1}{2}\lambda^{-(1+\theta)}\Gamma(1+\theta)(\log u)^{1+\theta} + O((\log u)^\theta), \tag{10}$$

$$F(\theta,1,\lambda;u) = \tfrac{1}{2}\lambda^{-(1+\theta)}(\log 2)\Gamma(1+\theta)(\log u)^{1+\theta} + O((\log u)^\theta), \tag{11}$$

$$G(0,\lambda;u) = \tfrac{1}{2}\lambda^{-1}\log u - \tfrac{1}{2}\log\log u + O(1), \tag{12}$$

$$G(1,\lambda;u) = \tfrac{1}{2}\lambda^{-2}(\log u)^2 + O(\log u), \tag{13}$$

where the implied constants depend only on θ, ξ and λ.

Lemma 4. *For any set $\mathcal{Z} \subseteq [0,n]$ of cardinality $k := |\mathcal{Z}|$ we have*

$$\sum_{\substack{\mathcal{A}\subseteq\mathcal{Z}\\ |\mathcal{A}|=1}} \sum_{\substack{\mathcal{B}\subseteq[0,n]\setminus\mathcal{Z}\\ |\mathcal{B}|=1}} \mathfrak{S}_0(\mathcal{A}\cup\mathcal{B}) = O_k(n^{1/2+\varepsilon}) \tag{14}$$

and

$$\sum_{\substack{\mathcal{B}\subseteq[0,n]\setminus\mathcal{Z}\\ |\mathcal{B}|=2}} \mathfrak{S}_0(\mathcal{B}) = -\tfrac{1}{2}n\log n + \tfrac{1}{2}An + O_k(n^{1/2+\varepsilon}), \tag{15}$$

where A is defined by (7).

Proof. Let s,u be arbitrary integers such that $0 \leq s < u \leq n$. Using the translation-invariance of the singular series and applying [1, Lemma 2.2], we see that

$$\begin{aligned}\sum_{s<t<u} \mathfrak{S}_0(\{s,t\}) &= \sum_{0<t<u-s} \mathfrak{S}_0(\{0,t\}) \ll (u-s)^{1/2+\varepsilon} \ll n^{1/2+\varepsilon},\\ \sum_{s<t<u} \mathfrak{S}_0(\{t,u\}) &= \sum_{0<t<u-s} \mathfrak{S}_0(\{t,u-s\}) \ll (u-s)^{1/2+\varepsilon} \ll n^{1/2+\varepsilon}.\end{aligned} \tag{16}$$

Next, let x, y be arbitrary integers such that $0 \le x < y \le n$. Suppose that $a \in \mathcal{Z}$ and $a \notin (x, y)$. Then

$$\sum_{b\in(x,y)} \mathfrak{S}_0(\{a,b\}) = \begin{cases} \sum_{a<t<y} \mathfrak{S}_0(\{a,t\}) - \sum_{a<t<x} \mathfrak{S}_0(\{a,t\}) & \text{if } a \le x; \\ \sum_{x<t<a} \mathfrak{S}_0(\{t,a\}) - \sum_{y<t<a} \mathfrak{S}_0(\{t,a\}) & \text{if } a \ge y, \end{cases}$$

hence from (16) it follows that

$$\sum_{b\in(x,y)} \mathfrak{S}_0(\{a,b\}) \ll n^{1/2+\varepsilon}. \tag{17}$$

Now suppose $\mathcal{Z} = \{z_1, \ldots, z_k\}$ with

$$z_0 := -1 < z_1 < \cdots < z_k < z_{k+1} := n + 1.$$

For $j = 1, \ldots, k$ let (x_j, y_j) be the open interval with $x_j := z_j$ and $y_j := z_{j+1}$. Using (17) we have for each $a \in \mathcal{Z}$:

$$\sum_{b\in[0,n]\setminus\mathcal{Z}} \mathfrak{S}_0(\{a,b\}) = \sum_{j=1}^{k} \sum_{b\in(x_j,y_j)} \mathfrak{S}_0(\{a,b\}) = O_k(n^{1/2+\varepsilon}).$$

Summing this bound over all $a \in \mathcal{Z}$, we obtain (14).

To prove (15), we observe that

$$\begin{aligned} \sum_{\substack{\mathcal{B}\subseteq[0,n]\setminus\mathcal{Z} \\ |\mathcal{B}|=2}} \mathfrak{S}_0(\mathcal{B}) &= \sum_{\substack{\mathcal{B}\subseteq[0,n] \\ |\mathcal{B}|=2}} \mathfrak{S}_0(\mathcal{B}) - \sum_{\substack{\mathcal{A}\subseteq\mathcal{Z} \\ |\mathcal{A}|=1}} \sum_{\substack{\mathcal{B}\subseteq[0,n]\setminus\mathcal{Z} \\ |\mathcal{B}|=1}} \mathfrak{S}_0(\mathcal{A}\cup\mathcal{B}) - \sum_{\substack{\mathcal{A}\subseteq\mathcal{Z} \\ |\mathcal{A}|=2}} \mathfrak{S}_0(\mathcal{A}) \\ &= S_1 - S_2 - S_3 \quad \text{(say)}. \end{aligned}$$

By (8) we have

$$S_1 = -\tfrac{1}{2} n \log n + \tfrac{1}{2} A n + O(n^{1/2+\varepsilon}).$$

By (14) we also have $S_2 = O_k(n^{1/2+\varepsilon})$. Finally, $S_3 = O_k(\log\log n)$ since the trivial bound $\mathfrak{S}_0(\mathcal{H}) \ll \log\log n$ holds for any $\mathcal{H} \subseteq [0, n]$ with $|\mathcal{H}| = 2$. Combining all of these estimates, we derive (15). □

3. The Heuristic Argument

We denote

$$g_{h,k}(n) := \mathbf{1}_{\mathbb{P}}(n)\mathbf{1}_{\mathbb{P}}(n+h)\mathbf{1}_{\mathbb{P}}(n+h+k) \prod_{\substack{0<t<h+k \\ t\neq h}} \big(1 - \mathbf{1}_{\mathbb{P}}(n+t)\big),$$

so that

$$g_{h,k}(n) = \begin{cases} 1 & \text{if } n = p_m \in \mathbb{P},\ d_m = h \text{ and } d_{m+1} = k; \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\pi_c(x) = \sum_{n \le x} \sum_{\substack{h,k \ge 1 \\ k \ge ch}} g_{h,k}(n).$$

Taking into account the trivial bound

$$\big|\{p_n \le x : \max\{\delta_n, \delta_{n+1}\} > (\log x)^3\}\big| \ll x(\log x)^{-3},$$

it follows that

$$\begin{aligned} \pi_c(x) &= \big|\{p_n \le x : \max\{\delta_n, \delta_{n+1}\} \le (\log x)^3,\ \delta_{p'} \ge c\,\delta_p\}\big| + O\big(x(\log x)^{-3}\big) \\ &= \sum_{\substack{h,k \le (\log x)^3 \\ 2|h,\, 2|k \\ k \ge ch}} S_{h,k}(x) + O\big(x(\log x)^{-3}\big), \end{aligned}$$

where

$$S_{h,k}(x) \coloneqq \sum_{n \le x} g_{h,k}(n).$$

For now, fix two even integers $h, k \in [1, (\log x)^3]$. Put $\mathcal{Z} = \mathcal{Z}_{h,k} \coloneqq \{0, h, h+k\}$ and write $\widetilde{\mathbf{1}}_{\mathbb{P}}(n) \coloneqq \mathbf{1}_{\mathbb{P}}(n) - 1/\log n$. Then

$$\begin{aligned} S_{h,k}(x) &= \sum_{n \le x} \prod_{l \in \mathcal{Z}} \mathbf{1}_{\mathbb{P}}(n+l) \prod_{m \in [0,h+k] \setminus \mathcal{Z}} \big(1 - \mathbf{1}_{\mathbb{P}}(n+m)\big) \\ &= \sum_{n \le x} \prod_{l \in \mathcal{Z}} \left(\frac{1}{\log n} + \widetilde{\mathbf{1}}_{\mathbb{P}}(n+l)\right) \prod_{m \in [0,h+k] \setminus \mathcal{Z}} \left(1 - \frac{1}{\log n} - \widetilde{\mathbf{1}}_{\mathbb{P}}(n+m)\right) \\ &= \sum_{\mathcal{A} \subseteq \mathcal{Z}} \sum_{\mathcal{B} \subseteq [0,h+k] \setminus \mathcal{Z}} (-1)^{|\mathcal{B}|} \sum_{n \le x} \left(\frac{1}{\log n}\right)^{3-|\mathcal{A}|} \left(1 - \frac{1}{\log n}\right)^{h+k-2-|\mathcal{B}|} \prod_{t \in \mathcal{A} \cup \mathcal{B}} \widetilde{\mathbf{1}}_{\mathbb{P}}(n+t) \end{aligned}$$

(cf. [4, Equations (2.5) and (2.6)]). Using the Hardy-Littlewood conjecture (6) and partial summation, we expect that the estimate

$$\sum_{n \le x} (\log n)^{-C} \prod_{t \in \mathcal{H}} \widetilde{\mathbf{1}}_{\mathbb{P}}(n+t) = \mathfrak{S}_0(\mathcal{H}) \int_2^x (\log u)^{-|\mathcal{H}|-C}\, du + O(x^{1/2+\varepsilon})$$

holds for any fixed $C \ge 0$ and any set $\mathcal{H}$ of nonnegative integers bounded above by $x^{o(1)}$ as $x \to \infty$; in particular, we expect that $S_{h,k}(x)$ is approximately

$$\sum_{\mathcal{A} \subseteq \mathcal{Z}} \sum_{\mathcal{B} \subseteq [0,h+k] \setminus \mathcal{Z}} (-1)^{|\mathcal{B}|} \mathfrak{S}_0(\mathcal{A} \cup \mathcal{B}) \int_2^x (\log u)^{-3-|\mathcal{B}|} \alpha(u)^{h+k-2-|\mathcal{B}|}\, du,$$

with an error not exceeding $O(x^{1/2+\varepsilon})$. For every integer $L \geq 0$ we now denote

$$\mathcal{D}_{h,k,L}(u) := \sum_{\substack{\mathcal{A}\subseteq\mathcal{Z}\ \mathcal{B}\subseteq[0,h+k]\setminus\mathcal{Z} \\ (|\mathcal{A}|+|\mathcal{B}|=L)}} (-1)^{|\mathcal{B}|}\mathfrak{S}_0(\mathcal{A}\cup\mathcal{B})(\alpha(u)\log u)^{-|\mathcal{B}|}\alpha(u)^{h+k},$$

so that

$$S_{h,k}(x) = \sum_{L=0}^{h+k+1}\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{D}_{h,k,L}(u)\,du + O(x^{1/2+\varepsilon}).$$

Next, arguing as in [4] (see also [1]) we conjecture that terms with $L \geq 3$ contribute no more than $O(x(\log x)^{-5/2})$ to the quantity $\pi_c(x)$. Noting that $\mathcal{D}_{h,k,1}$ is identically zero (since $\mathfrak{S}_0(\mathcal{H}) = 0$ for any singleton set $\mathcal{H}$), this leaves only terms with $L = 0$ or $L = 2$. Collecting terms according to the values $|\mathcal{A}|$ and $|\mathcal{B}|$, and summing over the variables h and k, we arrive at the estimate

$$\pi_c(x) = \sum_{j=1}^{4}\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{F}_j(u)\,du + O(x^{1/2+\varepsilon}), \tag{18}$$

where

$$\mathcal{F}_1(u) := \sum_{\substack{h,k\leq(\log x)^3 \\ 2|h,\,2|k \\ k\geq ch}} \alpha(u)^{h+k},$$

$$\mathcal{F}_2(u) := \sum_{\substack{h,k\leq(\log x)^3 \\ 2|h,\,2|k \\ k\geq ch}} \sum_{\substack{a_1,a_2\in\mathcal{Z}_{h,k} \\ a_1\neq a_2}} \mathfrak{S}_0(\{a_1,a_2\})\alpha(u)^{h+k},$$

$$\mathcal{F}_3(u) := -(\alpha(u)\log u)^{-1} \sum_{\substack{h,k\leq(\log x)^3 \\ 2|h,\,2|k \\ k\geq ch}} \sum_{a\in\mathcal{Z}_{h,k}} \sum_{b\in[0,h+k]\setminus\mathcal{Z}_{h,k}} \mathfrak{S}_0(\{a,b\})\alpha(u)^{h+k},$$

$$\mathcal{F}_4(u) := (\alpha(u)\log u)^{-2} \sum_{\substack{h,k\leq(\log x)^3 \\ 2|h,\,2|k \\ k\geq ch}} \sum_{\substack{b_1,b_2\in[0,h+k]\setminus\mathcal{Z}_{h,k} \\ b_1\neq b_2}} \mathfrak{S}_0(\{b_1,b_2\})\alpha(u)^{h+k}.$$

According to Lemma 2 we have

$$\mathcal{F}_1(u) = (4c+4)^{-1}((\log u)^2 + O(\log u)),$$

and thus the corresponding contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{F}_1(u)\,du = (4c+4)^{-1}\pi(x) + O\left(\frac{x}{(\log x)^2}\right). \tag{19}$$

The second function $\mathcal{F}_2(u)$ splits naturally as a sum

$$\mathcal{F}_2(u) = \mathcal{G}_1(u) + \mathcal{G}_2(u) + \mathcal{G}_3(u),$$

where

$$\begin{aligned}
\mathcal{G}_1(u) &:= \sum_{\substack{h,k \le (\log x)^3 \\ 2|h,\, 2|k \\ k \ge ch}} \mathfrak{S}_0(\{0,h\})\alpha(u)^{h+k},\\
\mathcal{G}_2(u) &:= \sum_{\substack{h,k \le (\log x)^3 \\ 2|h,\, 2|k \\ k \ge ch}} \mathfrak{S}_0(\{0,h+k\})\alpha(u)^{h+k},\\
\mathcal{G}_3(u) &:= \sum_{\substack{h,k \le (\log x)^3 \\ 2|h,\, 2|k \\ k \ge ch}} \mathfrak{S}_0(\{h,h+k\})\alpha(u)^{h+k}.
\end{aligned}$$

To estimate $\mathcal{G}_1(u)$ we apply Lemma 1 together with (9) to the inner sum over k, deriving that

$$\mathcal{G}_1(u) = \left(\tfrac{1}{2}\log u + O(1)\right) \sum_{\substack{h \le (\log x)^3 \\ 2|h}} \mathfrak{S}_0(\{0,h\})\alpha(u)^{(c+1)h}.$$

Using Lemma 1 again followed by (12) (with $\lambda := c+1$), we have

$$\mathcal{G}_1(u) = (4c+4)^{-1}((\log u)^2 + O(\log u \log\log u)).$$

Hence, the corresponding contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{G}_1(u)\,du = (4c+4)^{-1}\pi(x) + O\left(\frac{x\log\log x}{(\log x)^2}\right). \tag{20}$$

To estimate $\mathcal{G}_2(u)$, we write $m := h+k$ and use Lemma 1 along with (10), (13) and the trivial bound $\mathfrak{S}_0(\{0,m\}) \ll \log\log m$:

$$\begin{aligned}
\mathcal{G}_2(u) &= \sum_{\substack{m \le 2(\log x)^3 \\ 2|m}} \mathfrak{S}_0(\{0,m\})\alpha(u)^m \sum_{\substack{h \le (\log x)^3 \\ 2|h \\ m \ge (c+1)h}} 1\\
&= (2c+2)^{-1} \sum_{\substack{m \le 2(\log x)^3 \\ 2|m}} \mathfrak{S}_0(\{0,m\})\alpha(u)^m(m+O(1))\\
&= (4c+4)^{-1}(\log u)^2 + O(\log u \log\log\log x).
\end{aligned}$$

The contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{G}_2(u)\,du = (4c+4)^{-1}\pi(x) + O\bigg(\frac{x\log\log\log x}{(\log x)^2}\bigg). \tag{21}$$

By the translation-invariance of the singular series we have

$$\mathcal{G}_3(u) = \sum_{\substack{k\le(\log x)^3\\ 2|k}} \mathfrak{S}_0(\{0,k\})\alpha(u)^k \sum_{\substack{h\le(\log x)^3\\ 2|h\\ h\le k/c}} \alpha(u)^h.$$

Using Lemma 1 together with (9) (with $1/c$ in place of c) it follows that

$$\begin{aligned}\sum_{\substack{h\le(\log x)^3\\ 2|h\\ h\le k/c}} \alpha(u)^h &= \sum_{\substack{h\ge1\\ 2|h}} \alpha(u)^h - \sum_{\substack{h\ge1\\ 2|h,\, h>k/c}} \alpha(u)^h + O(x^{-1})\\ &= \frac{\alpha(u)^2}{1-\alpha(u)^2} - \tfrac12\alpha(u)^{k/c}(\log u + O(1)) + O(x^{-1})\\ &= \tfrac12\log u - \tfrac12\alpha(u)^{k/c}\log u + O(1)\end{aligned}$$

uniformly for u in the range $2\le u\le x$; hence, taking into account the trivial bound $\mathfrak{S}_0(\{0,k\}) \ll \log\log k$ and applying Lemma 1 again, we see that $\mathcal{G}_3(u)$ is equal to

$$(\tfrac12\log u)\bigg(\sum_{\substack{k\ge1\\ 2|k}} \mathfrak{S}_0(\{0,k\})\alpha(u)^k - \sum_{\substack{k\ge1\\ 2|k}} \mathfrak{S}_0(\{0,k\})\alpha(u)^{(1+1/c)k}\bigg) + O\bigg(\frac{\log\log x}{x}\bigg).$$

Using estimate (12) of Lemma 3 we have

$$\begin{aligned}\mathcal{G}_3(u) &= (\tfrac12\log u)\big(G(0,1;u) - G(0,1+1/c;u)\big) + O\bigg(\frac{\log\log x}{x}\bigg)\\ &= (4c+4)^{-1}(\log u)^2 + O(\log u),\end{aligned}$$

and thus the corresponding contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3}\alpha(u)^{-2}\mathcal{G}_3(u)\,du = (4c+4)^{-1}\pi(x) + O\bigg(\frac{x}{(\log x)^2}\bigg). \tag{22}$$

To bound $\mathcal{F}_3(u)$ we apply Lemma 4 (using (14) with $k := |\mathcal{Z}_{h,k}| = 3$) to deduce that

$$\mathcal{F}_3(u) \ll (\log u)^{-1} \sum_{\substack{h,k\le(\log x)^3\\ 2|h,\,2|k\\ k\ge ch}} (h+k)^{1/2+\varepsilon}\alpha(u)^{h+k}.$$

Setting $m := h + k$ and using Lemma 1 and the estimate (10), we have

$$\begin{aligned}\mathcal{F}_3(u) &\ll (\log u)^{-1} \sum_{\substack{m \le 2(\log x)^3 \\ 2 \mid m}} m^{1/2+\varepsilon} \alpha(u)^m \sum_{\substack{h \le (\log x)^3 \\ 2 \mid h \\ (c+1)h \le m}} 1 \\ &\ll (\log u)^{-1} \sum_{\substack{m \le 2(\log x)^3 \\ 2 \mid m}} m^{3/2+\varepsilon} \alpha(u)^m \ll (\log u)^{3/2+\varepsilon};\end{aligned}$$

hence the contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3} \alpha(u)^{-2} \mathcal{F}_3(u)\, du \ll \frac{x}{(\log x)^{3/2-\varepsilon}}. \tag{23}$$

Finally, to bound the quantity $\mathcal{F}_4(u)$ we apply Lemma 4 (using the estimate (15) with $k := |\mathcal{Z}_{h,k}| = 3$), which gives

$$\mathcal{F}_4(u) \ll (\log u)^{-2} \sum_{\substack{h,k \le (\log x)^3 \\ 2 \mid h,\, 2 \mid k \\ k \ge ch}} (h+k) \log(h+k) \alpha(u)^{h+k}.$$

Writing $m := h + k$ as before and using Lemma 1 and the estimate (11) it follows that

$$\begin{aligned}\mathcal{F}_4(u) &\ll (\log u)^{-2} \sum_{\substack{m \le 2(\log x)^3 \\ 2 \mid m}} m (\log m) \alpha(u)^m \sum_{\substack{h \le (\log x)^3 \\ 2 \mid h \\ (c+1)h \le m}} 1 \\ &\ll (\log u)^{-2} \sum_{\substack{m \le 2(\log x)^3 \\ 2 \mid m}} m^2 (\log m) \alpha(u)^m \ll \log u.\end{aligned}$$

Hence, the corresponding contribution to $\pi_c(x)$ is

$$\int_2^x (\log u)^{-3} \alpha(u)^{-2} \mathcal{F}_4(u)\, du \ll \frac{x}{(\log x)^2}. \tag{24}$$

Combining all of the estimates (18)–(24) above, we arrive at the statement of Conjecture 1.

Acknowledgements. The author thanks the anonymous referee for numerous helpful comments, corrections, and suggestions. Some of this work was done during a pleasant visit by the author to Duke University; the author thanks that institution for its kind hospitality. This research was supported in part by a grant from the University of Missouri Research Board.

References

[1] W. D. Banks and V. Z. Guo, Consecutive primes and Beatty sequences. *J. Number Theory* **191** (2018), 158–174.

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

[3] G. H. Hardy and J. E. Littlewood, Some problems of Partitio Numerorum (III): On the expression of a number as a sum of primes. *Acta Math.* **44** (1923), no. 1, 1–70.

[4] R. J. Lemke Oliver and K. Soundararajan, Unexpected biases in the distribution of consecutive primes. *Proc. Natl. Acad. Sci. USA* **113** (2016), no. 31, E4446–E4454.

[5] H. L. Montgomery and K. Soundararajan, Beyond pair correlation. pp. 507–514 in *Paul Erdős and his mathematics, I (Budapest, 1999).* Bolyai Soc. Math. Stud., 11, János Bolyai Math. Soc., Budapest, 2002.

[6] H. L. Montgomery and K. Soundararajan, Primes in short intervals. *Comm. Math. Phys.* **252** (2004), no. 1-3, 589–617.

[7] K. Soundararajan, The distribution of prime numbers, pp. 59–83 in *Equidistribution in number theory, an introduction.* NATO Sci. Ser. II Math. Phys. Chem. 237. Springer, Dordrech, 2007.