On the sum of the first n primes

Javier Cilleruelo
Departamento de Matemáticas
Facultad de Ciencias
Universidad Autónoma de Madrid
28049, Madrid, España
franciscojavier.cilleruelo@uam.es

Florian Luca
Instituto de Matemáticas
Universidad Nacional Autonoma de México
C.P. 58089, Morelia, Michoacán, México
fluca@matmor.unam.mx

November 9, 2007

Abstract

In this note, we show that the set of n such that the arithmetic mean of the first n primes is an integer is of asymptotic density zero. We use the same method to show that the set of n such the sum of the first n primes is a square is also of asymptotic density zero. We also prove that both the arithmetic mean of the first n primes as well as the square root of the sum of the first n primes are well distributed modulo 1.

1 The Main Results

Let p_n be the nth prime. It is clear that if $n > 1$, then the geometric mean of the first n primes, namely the number $(p_1 \dots p_n)^{1/n}$, is not an integer.

However, it happens sometimes that the arithmetic mean of the first n primes is an integer. In fact, putting

$$s_n = \sum_{i=1}^{n} p_i,$$

and

$$\mathcal{A} = \{n : s_n/n \in \mathbb{Z}\},$$

then one checks that

$$\mathcal{A} = \{1, 23, 53, 853, 11869, 117267, 339615, 3600489, \ldots\}.$$

This appears as sequence $A045345$ in [3], where the next three larger members of $\mathcal{A}$ are shown. Regular heuristics seem to suggest that $\mathcal{A}$ should be an infinite set. Indeed, assuming that s_n is uniformly distributed in arithmetic progressions of modulus n, it would follow that $s_n \equiv 0 \pmod{n}$ with a probability of $1/n$. Hence, putting $\mathcal{A}(x) = \mathcal{A} \cap [1, x]$, the above heuristics suggest that

$$\#\mathcal{A}(x) \sim \sum_{n \le x} \frac{1}{n} = \log x + O(1), \tag{1}$$

and, in particular, $\mathcal{A}$ should be an infinite set, albeit not a very dense one.

While we can neither show that $\mathcal{A}$ is infinite, nor can we show an upper bound on $\#\mathcal{A}(x)$ comparable to the one predicted by heuristics (1), we can at least show that $\mathcal{A}$ is of asymptotic density zero.

Theorem 1. *There exists a positive constant c_0 such that the inequality*

$$\#\mathcal{A}(x) < x \exp\left(-c_0 (\log x)^{3/5} (\log\log x)^{-1/5}\right) \tag{2}$$

holds for all $x \ge e$.

Our method is elementary and uses only the known bounds for the difference $|\pi(x) - \mathrm{li}(x)|$ (see, for example, Chapter 5 in [4]). In particular, under the Riemann hypothesis, our argument shows that

$$\#\mathcal{A}(x) \ll (x \log x)^{5/6}.$$

We also put $\mathcal{B} = \{n : \ s_n \text{ is an square}\}$. The sequence

$$\mathcal{B} = \{9, 2474, 6694, 7785, 709838, 126789311423, \ldots\}$$

appears as sequence $A003397$ in [3]. In [1], it was shown that $\mathcal{B}$ is a set of asymptotic density zero but no effective upper bound on $\#\mathcal{B}(x)$ was given. The proof from [1] uses sieves. Heuristic arguments show that $\mathcal{B}(x) \sim \sqrt{8\log x}$ as $x \to \infty$. Here, we use the same method as for the proof of Theorem 1 to get the following upper bound.

Theorem 2. *There exists a positive constant c_1 such that the inequality*

$$\#\mathcal{B}(x) < x\exp(-c_1(\log x)^{3/5}(\log\log x)^{-1/5}) \tag{3}$$

holds for all $x \geq e$.

A problem with a similar flavor was studied in [2] where it was shown that the set of n such that the sum $\phi(1) + \cdots + \phi(n)$ is a square is of asymptotic density zero, where for a positive integer m we write $\phi(m)$ for the Euler function of m. That proof also uses sieve methods. Our proofs, however, use an argument completely different which can perhaps be applied to strengthen the result from [2]. We leave this as a challenge to the reader.

Theorems 1 and 2 show that the sequence of averages of the first n primes, as well as the sequence of square-roots of the sums of the first primes are, in general, not integers. We also prove more, namely that the fractional parts of both these sequences are well distributed in $[0, 1)$.

Theorem 3. *The sequence $\left\{\left(\frac{s_n}{n}\right)\right\}_{n\geq 1}$ is well distributed in $[0, 1)$.*

Theorem 4. *The sequence $\{(s_n^{1/2})\}_{n\geq 1}$ is well distributed in $[0, 1)$.*

Obviously, Theorems 3 and 4 already imply that both $\mathcal{A}$ and $\mathcal{B}$ have asymptotic densities zero, but Theorems 1 and 2 give us effective upper bounds on their counting functions.

Before proceeding to the proofs, we give a brief outline of the technique used to prove Theorem 1. We need to prove that if s_n denotes the sum of the first n primes, then s_n/n is an integer for a zero proportion of all positive integers n. Suppose that $\pi(x) \sim \text{Li}(x)$ were an exact formula. Then s_n/n would be an integer extremely rarely for the simple reason that $s_{n+m}/(n+m)$ - s_n/n could not be an integer for n large and $m \leq T(n)$, where $T(n)$ is a suitably chosen increasing function of n. Indeed, this is so essentially because $1/(n + m) - 1/n = -m/(n(n + m))$ is tiny for m much smaller than n.

Now, $\pi(x) \sim \text{Li}(x)$ is not actually an exact formula. Still, the error is small enough that $s_{n+m}/(n+m) - s_n/n$ is very rarely an integer for n large and m running through an interval $[0, T(n)]$, with our suitable function $T(n)$. Then the fact that s_n/n is an integer only for a zero proportion of all n follows almost immediately upon an application of Cauchy's inequality. The proof of Theorem 2 follows a similar plan of attack.

In what follows, we use p and q with or without subscripts for prime numbers, and the Landau symbols O and o and the Vinogradov symbols $\gg$, $\ll$ and $\asymp$ with their usual meanings. The constants implied by these symbols are absolute. We write $c_0, c_1, \ldots$ for positive computable constants which are labeled increasingly throughout the paper.

Acknowledgements. We thank the referees for comments which improved the quality of this paper. This work was done in July of 2007 when the second the author visited the UAM in Madrid, Spain. He thanks the host institution for its hospitality. The first author was supported in part by Grant MTM 2005-04730 of MYCIT (Spain), the second author was supported in part by project SEP-CONACyT 46755 (Mexico), and both authors were also supported by the project *Red Iberoamericana de Teoría de Números* of the UAM-Grupo Santander (Spain).

2 Preliminary Results

We recall that

$$\text{Li}(x) = \int_2^x \frac{dt}{\log t}$$

is the logarithmic integral of x. We put $\pi(x) = \#\{p \le x\}$ and write

$$E(x) = \max\{|\pi(y) - \text{Li}(y)| : 2 \le y \le x\}.$$

The following estimate for $E(x)$ is well-known (see Chapter 5 of [4]).

Lemma 1. *There exists a constant $c_2 > 0$ such that*

$$|E(x)| \le x \exp\left(-c_2 (\log x)^{3/5} (\log\log x)^{-1/5}\right)$$

holds for all $x > e$.

Lemma 1 above and some straightforward algebraic manipulations yield the following estimates.

Lemma 2. *The estimates*

$$s_m = \int_2^{\mathrm{Li}^{-1}(m)} \frac{t}{\log t} dt + O(m(\log m)E(p_m)), \tag{4}$$

and

$$s_{m+k} - s_m = k\mathrm{Li}^{-1}(m) + O\left(k \log(m+k)(E(p_{m+k}) + k)\right) \tag{5}$$

hold, where Li^{-1} *is the inverse function of the logarithmic integral function* $\mathrm{Li}(x)$*.*

Proof. Since $\mathrm{Li}(x) = (1+o(1))x/\log x$ as $x \to \infty$, we have that $\mathrm{Li}^{-1}(x) = (1+o(1))x \log x$ as $x \to \infty$. Furthermore, since

$$(\mathrm{Li}^{-1})'(\mathrm{Li}(x)) = \frac{1}{\mathrm{Li}'(x)} = \log x,$$

we get that

$$(\mathrm{Li}^{-1})'(x) = \log(\mathrm{Li}^{-1}(x)) = (1+o(1)) \log x \qquad \text{as } x \to \infty.$$

We can write

$$m = \pi(p_m) = \mathrm{Li}(p_m)(1+\varepsilon_m),$$

with $|\varepsilon_m| \le E(p_m)/\mathrm{Li}(p_m) = o(1)$ as $m \to \infty$. Therefore $p_m = \mathrm{Li}^{-1}(m/(1+\varepsilon_m))$ and then

$$|p_m - \mathrm{Li}^{-1}(m)| = |\mathrm{Li}^{-1}(m/(1+\varepsilon_m)) - \mathrm{Li}^{-1}(m)| \ll \varepsilon_m m \log m,$$

Thus,

$$p_m = \mathrm{Li}^{-1}(m) + O((\log m)E(p_m)).$$

Then,

$$s_n = \sum_{1 \le m \le n} p_m = \sum_{1 \le m \le n} \mathrm{Li}^{-1}(m) + O(n(\log n)E(p_n)).$$

Finally we can write

$$\sum_{1 \le m \le n} \mathrm{Li}^{-1}(m) = \int_0^n \mathrm{Li}^{-1}(t)dt + \sum_{1 \le m \le n} \int_{m-1}^m \left(\mathrm{Li}^{-1}(m) - \mathrm{Li}^{-1}(t)\right) dt =$$

$$= \int_2^{\mathrm{Li}^{-1}(n)} \frac{t}{\log t} dt + O\left(\sum_{1 \le m \le n} \log m\right) = \int_2^{\mathrm{Li}^{-1}(n)} \frac{t}{\log t} dt + O(n \log n).$$

For the second one, we certainly have that

$$\begin{aligned} p_{m+j} &= \mathrm{Li}^{-1}(m+j) + O((\log(m+k))E(p_{m+k})) \\ &= \mathrm{Li}^{-1}(m) + (\mathrm{Li}^{-1}(m+j) - \mathrm{Li}^{-1}(m)) + O((\log(m+k))E(p_{m+k})) \end{aligned}$$

for all $j = 1, \dots, k$. Since

$$\mathrm{Li}^{-1}(m+j) - \mathrm{Li}^{-1}(m) = O(j(\mathrm{Li}^{-1})'(m+j)) \ll k \log(m+k),$$

when $j = 1, \dots, k$, we get that

$$p_{m+j} = \mathrm{Li}^{-1}(m) + O\left(\log(m+k)(E(p_{m+k}) + k)\right)$$

for all $j = 1, \dots, k$. Summing up these estimates for $j = 1, \dots, k$ we get

$$s_{m+k} - s_m = \sum_{j=1}^{k} p_{m+j} = k\mathrm{Li}^{-1}(m) + O\left(k \log(m+k)(E(p_{m+k}) + k)\right).$$

□

In particular, we have the estimates

$$s_m = (1+o(1))\frac{m^2 \log m}{2} \qquad \text{and} \qquad s_{m+k} - s_m = (1+o(1))km\log m \quad (6)$$

as $m \to \infty$, assuming that $k = o(m)$.

Lemma 3. *Let g, h denote the functions*

$$g(x) = \frac{\mathrm{Li}^{-1}(x)}{x} - \frac{\int_2^{\mathrm{Li}^{-1}(x)} \frac{s}{\log s} ds}{x^2}, \quad (7)$$

and

$$h(x) = \frac{\mathrm{Li}^{-1}(x)}{2\left(\int_2^{\mathrm{Li}^{-1}(x)} \frac{s}{\log s} ds\right)^{1/2}}. \quad (8)$$

Then the estimates

$$\begin{aligned} g(x) &= \frac{\log x}{2}(1+o(1)), & g'(x) &= \frac{1}{2x}(1+o(1)), \\ h(x) &= \left(\frac{\log x}{2}\right)^{1/2}(1+o(1)), & h'(x) &= \frac{1}{2(2x\log x)^{1/2}}(1+o(1)) \end{aligned}$$

hold when $x \to \infty$.

Proof. It is easy to check that $g(x) \sim (\log x)/2$. For the asymptotic behavior of $g'(x)$ it suffices to prove that $g'(\mathrm{Li}(x))\mathrm{Li}(x) \sim \frac{1}{2}$. We write

$$g(\mathrm{Li}(x)) = \frac{x}{\mathrm{Li}(x)} - \frac{\int_2^x \frac{s}{\log s} ds}{\mathrm{Li}^2(x)}.$$

Since $\mathrm{Li}'(x) = 1/\log x$, we have

$$\begin{aligned} g'(\mathrm{Li}(x))\mathrm{Li}(x) &= \frac{1}{\mathrm{Li}^2(x)} \left((\log x)\mathrm{Li}^2(x) - 2x\mathrm{Li}(x) + 2\int_2^x \frac{s}{\log s} ds \right) \\ &= \frac{1}{\mathrm{Li}^2(x)} \left(\log x \left(\frac{x}{\log x} + \frac{(1+o(1))x}{\log^2 x} \right)^2 \right. \\ &- 2x \left(\frac{x}{\log x} + \frac{(1+o(1))x}{\log^2 x} \right) + 2 \left(\frac{x^2}{2\log x} + \frac{x^2}{4\log^2 x} \right. \\ &+ \left. \left. \frac{(1+o(1))x^2}{8\log^3 x} \right) \right), \end{aligned}$$

which tends to $1/2$ when $x \to \infty$.

For the second function h, it is also easy to check that

$$h(x) \sim ((\log x)/2)^{1/2} \qquad \text{as } x \to \infty.$$

To show the asymptotic behavior of $h'(x)$, it suffices to prove that

$$h'(\mathrm{Li}(x))\mathrm{Li}(x)(\log \mathrm{Li}(x))^{1/2} \to \frac{1}{2^{3/2}} \qquad \text{as } x \to \infty.$$

We have

$$\begin{aligned} \left(h^2(\mathrm{Li}(x))\right)' &= \left(\frac{x^2}{4\int_2^x \frac{sds}{\log s}} \right)' = \frac{1}{4} \left(2x \int_2^x \frac{sds}{\log s} - \frac{x^3}{\log x} \right) \left(\int_2^x \frac{sds}{\log s} \right)^{-2} = \\ &\frac{1}{4} \left(2x \left(\frac{x^2}{2\log x} + \frac{x^2(1+o(1))}{4\log^2 x} \right) - \frac{x^3}{\log x} \right) \left(\int_2^x \frac{sds}{\log s} \right)^{-2} \sim \frac{1}{2x}, \end{aligned} \tag{9}$$

as $x \to \infty$. We can then write

$$\begin{aligned} h'(\mathrm{Li}(x))\mathrm{Li}(x)(\log \mathrm{Li}(x))^{1/2} &= \left(h^2(\mathrm{Li}(x))\right)' \frac{\log x}{2h(\mathrm{Li}(x))} \mathrm{Li}(x) \ (\log \mathrm{Li}(x))^{1/2} \sim \\ &\sim \frac{1}{2x} \frac{(\log x)\mathrm{Li}(x)}{2} \frac{(\mathrm{Li}(x))^{1/2}}{h(\mathrm{Li}(x))} \sim \frac{1}{2x} \frac{x}{2} \sqrt{2} = \frac{1}{2\sqrt{2}}. \end{aligned}$$

□

3 Proof of Theorem 1

It clearly suffices to prove inequality (2) when the left hand side of it is replaced by $\#(\mathcal{A}\cap(x/2,x])$. We subdivide the interval $(x/2,x]$ in intervals $\mathcal{E}_j$ of length T each, $j=1,\ldots,[x/2T]+1$, and split the set of index j in two sets J_1 and J_2 according to whether $|\mathcal{A}\cap\mathcal{E}_j|\le 1$ or not. We note that $|\mathcal{A}\cap\mathcal{E}_j|^2\le 4\binom{|\mathcal{A}\cap\mathcal{E}_j|}{2}$ when $j\in J_2$. Thus, by the Cauchy-Schwartz inequality,

$$\begin{aligned}\#(\mathcal{A}\cap(x/2,x]) &= \sum_{j\in J_1}|\mathcal{A}\cap\mathcal{E}_j|+\sum_{j\in J_2}|\mathcal{A}\cap\mathcal{E}_j|\\ &\le |J_1|+|J_2|^{1/2}\left(\sum_{j\in J_2}|\mathcal{A}\cap\mathcal{E}_j|^2\right)^{1/2}\\ &\le \frac{x}{T}+2\left(\frac{x}{T}\right)^{1/2}\left(\sum_{j\in J_2}\binom{|\mathcal{A}\cap\mathcal{E}_j|}{2}\right)^{1/2}. \qquad (10)\end{aligned}$$

The pairs $(m,m')\in\mathcal{A}^2$ with $m<m'$ counted by the second sum above satisfy that $m'=m+k$ for some k, $1\le k\le T$. Thus,

$$\begin{aligned}\sum_{j\in J_2}\binom{|\mathcal{A}\cap\mathcal{E}_j|}{2} &\le \sum_{1\le k\le T}\#\{m:\ m\in(x/2,x-k],\ m,m+k\in\mathcal{A}\}\\ &\le \sum_{1\le k\le T}\#\left\{m:\ m\in(x/2,x-k],\ \frac{s_{m+k}}{m+k}-\frac{s_m}{m}\in\mathbb{Z}\right\}. \qquad (11)\end{aligned}$$

For any $m\in(x/2,x-k]$ and $k\le T$ such that $\frac{s_{m+k}}{m+k}-\frac{s_m}{m}\in\mathbb{Z}$, we write

$$\frac{s_{m+k}}{m+k}-\frac{s_m}{m}=\frac{s_{m+k}-s_m}{m}-\frac{ks_m}{m^2}-\frac{k(s_{m+k}-s_m)}{m(m+k)}+\frac{k^2s_m}{m^2(m+k)}. \qquad (12)$$

Since $m+k\le x$, we use Lemma 2 to obtain that

$$\frac{s_{m+k}-s_m}{m}=k\frac{\mathrm{Li}^{-1}(m)}{m}+O\left(\frac{k(\log m)(E(p_{\lfloor x\rfloor})+k)}{m}\right),$$

$$\frac{ks_m}{m^2}=k\frac{\int_2^{\mathrm{Li}^{-1}(m)}\frac{sds}{\log s}}{m^2}+O\left(\frac{k(\log m)E(p_{\lfloor x\rfloor})}{m}\right),$$

and

$$\frac{k^2 s_m}{m^2(m+k)} = O\left(\frac{k^2 \log m}{m}\right),$$

therefore

$$\frac{s_{m+k}}{m+k} - \frac{s_m}{m} = kg(m) + O\left(\frac{k(\log m)(E(p_{\lfloor x \rfloor}) + k)}{m}\right), \tag{13}$$

where $g(t)$ is the function defined in Lemma 3.

Using the fact that the left hand side of formula (13) is an integer, we have proved that for all m counted in (11) we have

$$\|kg(m)\| \ll \varepsilon(x), \tag{14}$$

where $\varepsilon(x) = T(\log x)(E(p_{\lfloor x \rfloor}) + T)x^{-1}$ and $\| \cdot \|$ denotes the distance to the closest integer. Then, if we write $g_k(y) = kg(y)$ and $I_l = [l - \varepsilon(x), l + \varepsilon(x)]$, by (11) and (14) we have

$$\begin{aligned} \sum_{j \in J_2} \binom{|\mathcal{A} \cap \mathcal{E}_j|}{2} &\le \sum_{k \le T} \#\{m :\ m \in (x/2, x-k],\ \|g_k(m)\| \le \varepsilon(x)\} \\ \le \sum_{k \le T} \#\{m :\ &m \in (x/2, x-k],\ \exists l \in \mathbb{Z},\ l - \varepsilon(x) \le g_k(m) \le l + \varepsilon(x)\} \\ &\le \sum_{k \le T} \sum_{g_k(x/2) \le l \le g_k(x)} \#\{m :\ m \in [x/2, x] \cap g_k^{-1}(I_l)\}. \end{aligned}$$

Since g_k is an increasing function, $g_k^{-1}(I_l)$ is also an interval, and we have that $\frac{|I_l|}{|g_k^{-1}(I_l)|} = g_k'(\xi)$ for some $\xi \in (x/2, x]$. Lemma 3 says that $g'(y) \sim 1/2y$, then we have that $|g_k^{-1}(I_l)| = |I_l|/g_k'(\xi) \ll \varepsilon(x)/(k/x)$. So we have that

$$\#\{m :\ m \in [x/2, x] \cap g_k^{-1}(I_l)\} \ll \frac{x\varepsilon(x)}{k} + 1.$$

On the other hand we have

$$g_k(x) - g_k(x/2) = k\int_{x/2}^{x} g'(t)dt \ll k\int_{x/2}^{x} \frac{dt}{t} \ll k.$$

Thus,

$$\sum_{j \in J_2} \binom{|\mathcal{A} \cap \mathcal{E}_j|}{2} \ll \sum_{k \le T} k\left(\frac{x\varepsilon(x)}{k} + 1\right) \ll T^2(\log x)(E(p_{\lfloor x \rfloor}) + T). \tag{15}$$

We substitute the last inequality (15) in (11) and (10) and we get

$$\#(\mathcal{A} \cap (x/2, x]) \ll x/T + \left(xT(\log x)(E(p_{\lfloor x \rfloor}) + T)\right)^{1/2}.$$

We now take $T = \lfloor (x/((\log x)E(p_{\lfloor x \rfloor})))^{1/3} \rfloor$ and get

$$\begin{aligned} \#(\mathcal{A} \cap (x/2, x]) &\ll (x^2(\log x)E(p_{\lfloor x \rfloor}))^{1/3} + x^{5/6}(\log x)^{1/6}/E^{1/3}(p_{\lfloor x \rfloor}) \\ &\ll (x^2(\log x)E(2x \log x))^{1/3} + x^{5/6}(\log x)^{1/6}. \end{aligned} \tag{16}$$

Lemma 1 leads to the desired conclusion. Assuming the Riemann Hypothesis, we have that $E(y) \ll y^{1/2} \log y$ for all y, which via estimate (16) gives

$$\#(\mathcal{A} \cap (x/2, x]) \ll (x \log x)^{5/6}.$$

4 Proof of Theorem 2

We put $b_n = s_n^{1/2}$ and let $\mathcal{B} = \{n : \ b_n \in \mathbb{Z}\}$. The proof is similar to the previous one. We proceed as before to obtain

$$\#(\mathcal{B} \cap (x/2, x]) \le x/T + 2(x/T)^{1/2} \left(\sum_{j \in J_2} \binom{|\mathcal{B} \cap \mathcal{E}_j|}{2} \right)^{1/2}, \tag{17}$$

where

$$\sum_{j \in J_2} \binom{|\mathcal{B} \cap \mathcal{E}_j|}{2} \le \sum_{1 \le k \le T} \# \left\{ m, \ m \in (x/2, x-k], \ s_{m+k}^{1/2} - {s_m}^{1/2} \in \mathbb{Z} \right\}. \tag{18}$$

For any $m \in (x/2, x-k], \ k \le T$ such that $b_{m+k} - b_m \in \mathbb{Z}$, we use estimate (6) to get

$$b_{m+k} - b_m = \frac{s_{m+k} - s_m}{b_{m+k} + b_m} \ll k(\log m)^{1/2}.$$

We assume that $k = o(x)$ as $x \to \infty$ and apply Lemma 2 to write

$$\begin{aligned} b_{m+k} - b_m &= \frac{s_{m+k} - s_m}{2 s_m^{1/2}} - \frac{(s_{m+k} - s_m)(s_{m+k}^{1/2} - s_m^{1/2})}{2 s_m^{1/2}(s_{m+k}^{1/2} + s_m^{1/2})} \\ = \frac{k \mathrm{Li}^{-1} m + O(k \log(m+k)(E(p_m) + k))}{2\left(\int_2^{\mathrm{Li}^{-1}(m)} \frac{s}{\log s} ds + O(m(\log m)E(p_m))\right)^{1/2}} &+ O\left(\frac{k^2 (\log m)^{1/2}}{m}\right) \\ &= k h(m) + O\left(\frac{k(\log m)^{1/2}(E(p_m) + k)}{m}\right), \end{aligned} \tag{19}$$

where h is the function defined in lemma 3. Thus, we have proved that if $b_{m+k} - b_m \in \mathbb{Z},\ x/2 < m \le m-k,\ k \le T$, then we have

$$\|kh(m)\| \ll \varepsilon(x), \tag{20}$$

where $\varepsilon(x) = T(\log x)^{1/2}(E(p_{\lfloor x \rfloor}) + T)x^{-1}$.

Since the following argument is similar to the proof of Theorem 1, we omit some details. We write $h_k(y) = kh(y)$ and $I_l = [l - \varepsilon(x), l + \varepsilon(x)]$ to obtain

$$\sum_{j \in J_2} \binom{|\mathcal{B} \cap \mathcal{E}_j|}{2} \le \sum_{k \le T} \sum_{h_k(x/2) \le l \le h_k(x)} \#\{m:\ m \in [x/2, x] \cap h_k^{-1}(I_l)\}.$$

As before, we can see that $|h_k^{-1}(I_l)| \ll |I_l|/h_k'(\xi) \ll \varepsilon(x)x(\log x)^{1/2}/k$ and also that $h_k(x) - h_k(x/2) \ll k/(\log x)^{1/2}$. Then

$$\begin{aligned} \sum_{j \in J_2} \binom{|\mathcal{B} \cap \mathcal{E}_j|}{2} &\ll \sum_{k \le T} \frac{k}{(\log x)^{1/2}} \left(\frac{\varepsilon(x)x(\log x)^{1/2}}{k} + 1 \right) \\ &\ll T^2(\log x)^{1/2}(E(p_{\lfloor x \rfloor}) + T). \end{aligned} \tag{21}$$

Substituting the above inequality (21) in (11) and (10), we get

$$\#(\mathcal{B} \cap (x/2, x]) \ll x/T + (xT(\log x)^{1/2}(E(p_{\lfloor x \rfloor}) + T))^{1/2}.$$

We take $T = \lfloor (x/((\log x)^{1/2} E(p_{\lfloor x \rfloor})))^{1/3} \rfloor$ and finally we obtain

$$\begin{aligned} \#(\mathcal{B} \cap (x/2, x]) &\ll (x^2(\log x)^{1/2} E(p_{\lfloor x \rfloor}))^{1/3} + x^{5/6}(\log x)^{1/12}/E^{1/3}(p_{\lfloor x \rfloor}) \\ &\ll (x^2(\log x)^{1/2} E(2x \log x))^{1/3} + x^{5/6}(\log x)^{1/12}. \end{aligned} \tag{22}$$

Again Lemma 1 leads to the desired conclusion.

5 Proofs of Theorems 3 and 4

The Weil criterion for the uniform distribution says that a sequence $\{a_n\}_{n \ge 1}$ is well distributed modulo 1 if and only if for any integer $m \ne 0$ we have that

$$\sum_{n \le x} \exp(2\pi i m a_n) = o(x) \qquad \text{as } x \to \infty. \tag{23}$$

We will use this criterion for the sequences $a_n = s_n/n$ and $b_n = {s_n}^{1/2}$. To prove estimate (23), it suffices to prove that

$$\sum_{x/2<n\le x} \exp(2\pi i m a_n) = o(x) \qquad \text{as } x \to \infty. \tag{24}$$

Writing

$$\sum_{x/2<n\le x} \exp(2\pi i m a_n) = \frac{1}{T} \sum_{x/2<n\le x-T} \sum_{0\le k<T} \exp(2\pi i m a_{n+k}) + O(T),$$

we get

$$\left| \sum_{x/2<n\le x} \exp(2\pi i m a_n) \right| \le \frac{1}{T} \sum_{x/2<n\le x-T} \left| \sum_{0\le k<T} \exp(2\pi i m (a_{n+k} - a_n)) \right| + O(T).$$

Estimate (12) shows that if $x/2 < n \le x - k$ and $k \le T$, then

$$a_{n+k} - a_n = kg(n) + O\left(\frac{T(\log x)(E(p_{\lfloor x \rfloor}) + T)}{x} \right).$$

We take $T = \lfloor (\log x)^2 \rfloor$ and use the estimate $E(p_{\lfloor x \rfloor}) \ll E(2x \log x) \ll x(\log x)^{-4}$. Then

$$a_{n+k} - a_n = kg(n) + O\left((\log x)^{-1} \right),$$

so we can write

$$\begin{aligned} \left| \sum_{0\le k<T} \exp(2\pi i m (a_{n+k} - a_n)) \right| &= \left| \sum_{0\le k<T} \exp(2\pi i m k g(n)) \left(1 + O\left(\frac{m}{\log x} \right) \right) \right| \\ &= \left| \sum_{0\le k<T} \exp(2\pi i m k g(n)) \right| + O(m \log x) \\ &= O\left(\min\left\{ T, \frac{1}{\|mg(n)\|} \right\} + m \log x \right). \end{aligned}$$

Then

$$\begin{aligned} \left| \sum_{x/2<n\le x} \exp(2\pi i m a_n) \right| &\ll \frac{1}{T} \sum_{x/2<n\le x} \min\left\{ T, \frac{1}{\|mg(n)\|} \right\} + \frac{mx}{\log x} \\ &\ll \#\left\{ n:\ x/2 < n \le x,\ \|mg(n)\| \le \frac{1}{T^{1/2}} \right\} + \frac{x}{T^{1/2}} + \frac{mx}{\log x}. \end{aligned} \tag{25}$$

If we write $g_m(y) = mg(y)$ and $I_l = [l - 1/T^{1/2}, l + 1/T^{1/2}]$ then

$$\begin{aligned} &\# \left\{ n : \ x/2 < n \le x, \ \|g_m(n)\| \le \frac{1}{T^{1/2}} \right\} \\ \le &\sum_{g_m(x/2) \le l \le g_m(x)} \#\{n : \ n \in g_m^{-1}(I_l) \cap (x/2, x]\}. \end{aligned} \tag{26}$$

Since g_m is an increasing function, we have that $|I_l|/|g_m^{-1}(I_l)| = g_m'(\xi)$ for some $\xi \in (x/2, x]$. Thus, by Lemma 3, we have

$$|g_m^{-1}(I_l)| \le \frac{|I_l|}{\min_{\xi \in (x/2,x]} g_m'(\xi)} \ll \frac{x}{mT^{1/2}}. \tag{27}$$

On the other hand, we have

$$g_m(x) - g_m(x/2) = m \int_{x/2}^{x} g'(t)dt \ll m. \tag{28}$$

Taking into account (25), (26), (27) and (28) we obtain

$$\left| \sum_{x/2 < n \le x} \exp(2\pi i m a_n) \right| \ll m \left(\frac{x}{T^{1/2} m} + 1 \right) + \frac{x}{T^{1/2}} + \frac{mx}{\log x} \ll \frac{mx}{\log x} = o(x)$$

as $x \to \infty$, and we finish the proof of Theorem 3.

The proof of Theorem 4 is similar but instead of estimate (12), we use estimate (19)

$$b_{n+k} - b_n = kh(n) + O\left(\frac{T(\log x)^{1/2}(E(p_{\lfloor x \rfloor}) + T)}{x} \right).$$

We give no further details.

References

[1] F. Luca, 'On the sum of the first n primes being a square', *Lithuanian Math. J.*, to appear.

[2] F. Luca and A. Sankaranarayanan, 'On positive integers n such that $\phi(1) + \phi(2) + \cdots + \phi(n)$ is a square', *Preprint*, 2007.

[3] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, http://www.research.att.com/~njas/sequences/

[4] A. Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie*, Berlin: Deutscher Verlag der Wissenschaften, 1963.