ON THE INFINITUDE OF SOPHIE GERMAIN PRIMES

THEOPHILUS AGAMA

Abstract. In this paper we obtain the estimate

$$\#\left\{p \leq x \mid 2p+1, p \in \mathbb{P}\right\} \geq (1+o(1))\frac{\mathcal{D}}{(2+2\log 2)}\frac{x}{\log^2 x}$$

where $\mathbb{P}$ is the set of all prime numbers and $\mathcal{D} \geq 1$. This proves that there are infinitely many primes $p \in \mathbb{P}$ such that $2p+1 \in \mathbb{P}$ is also prime.

1. Introduction and statement

Let $\mathbb{P}$ denotes the set of all prime numbers, then we say a prime p is a Sophie Germain prime - named after the French mathematician Sophie Germain who encountered it in her investigations of Fermat's Last Theorem - if $2p+1$ is also a prime number. The motivation for the study of Sophie Germain primes is quite clear from a practical point of view (see [3]), as it owes it's application to cryptography and primality testing [2]. There has also been lot of computational work in verifying pushing the barrier of the largest known Sophie Germain prime, a worthwhile endeavor since the infinitude of such primes has been conjectured to hold. In the current paper we obtain a lower bound for the number of such primes less than a given threshold, thereby confirming the infinitude of such primes.
Let us denote $\vartheta : \mathbb{N} \longrightarrow \mathbb{C}$ to be function defined by

$$\vartheta(n) := \begin{cases} \log p & \textbf{if} \quad n = p \in \mathbb{P} \\ 0 & \textbf{otherwise} \end{cases}$$

then an natural step to take to obtain an estimate for the number of such primes is to obtain an estimate for the correlation

$$\sum_{n \leq x} \vartheta(n)\vartheta(2n+1)$$

or at the very least a non-trivial lower bound followed by a consequent appeal to partial summation to remove the weight ϑ. Analyzing such correlations is by no means an easy tussle but an appeal to the area method [1] provides with at least a non-trivial lower bound.

2. Preliminary results

In this section we restate and prove an earlier result which will certainly serve it's purpose and in many ways can be viewed as a black box to obtaining further results in the sequel. The proof of this result can be found in [1]. It could have been ignored and refereed but we deem it appropriate keeping in mind our intention to make the paper comprehensive.

Date: April 3, 2021.
Key words and phrases. Sophie Germain; prime.

Theorem 2.1. *Let* $\{r_j\}_{j=1}^n$ *and* $\{h_j\}_{j=1}^n$ *be any sequence of real numbers, and let* r *and* h *be any real numbers satisfying* $\sum\limits_{j=1}^n r_j = r$ and $\sum\limits_{j=1}^n h_j = h$, *and*

$$(r^2+h^2)^{1/2} = \sum_{j=1}^n (r_j^2+h_j^2)^{1/2},$$

then

$$\sum_{j=2}^n r_j h_j = \sum_{j=2}^n h_j\bigg(\sum_{i=1}^j r_i + \sum_{i=1}^{j-1} r_i\bigg) - 2\sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}.$$

Proof. Consider a right angled triangle, say ΔABC in a plane, with height h and base r. Next, let us partition the height of the triangle into n parts, not neccessarily equal. Now, we link those partitions along the height to the hypothenus, with the aid of a parallel line. At the point of contact of each line to the hypothenus, we drop down a vertical line to the next line connecting the last point of the previous partition, thereby forming another right-angled triangle, say $\Delta A_1B_1C_1$ with base and height r_1 and h_1 respectively. We remark that this triangle is covered by the triangle ΔABC, with hypothenus constituting a proportion of the hypothenus of triangle ΔABC. We continue this process until we obtain n right-angled triangles $\Delta A_jB_jC_j$, each with base and height r_j and h_j for $j=1,2,\dots n$. This construction satisfies

$$h = \sum_{j=1}^n h_j \text{ and } r = \sum_{j=1}^n r_j$$

and

$$(r^2+h^2)^{1/2} = \sum_{j=1}^n (r_j^2+h_j^2)^{1/2}.$$

Now, let us deform the original triangle ΔABC by removing the smaller triangles $\Delta A_jB_jC_j$ for $j=1,2,\dots n$. Essentially we are left with rectangles and squares piled on each other with each end poking out a bit further than the one just above, and we observe that the total area of this portrait is given by the relation

$$\begin{aligned}\mathcal{A}_1 &= r_1h_2 + (r_1+r_2)h_3 + \cdots (r_1+r_2+\cdots+r_{n-2})h_{n-1} + (r_1+r_2+\cdots+r_{n-1})h_n\\ &= r_1(h_2+h_3+\cdots h_n) + r_2(h_3+h_4+\cdots+h_n) + \cdots + r_{n-2}(h_{n-1}+h_n) + r_{n-1}h_n\\ &= \sum_{j=1}^{n-1} r_j \sum_{k=1}^{n-j} h_{j+k}.\end{aligned}$$

On the other hand, we observe that the area of this portrait is the same as the difference of the area of triangle ΔABC and the sum of the areas of triangles $\Delta A_jB_jC_j$ for $j=1,2,\dots,n$. That is

$$\mathcal{A}_1 = \frac{1}{2}rh - \frac{1}{2}\sum_{j=1}^n r_j h_j.$$

This completes the first part of the argument. For the second part, along the hypothenus, let us construct small pieces of triangle, each of base and height (r_i,h_i) $(i=1,2\dots,n)$ so that the trapezoid and the one triangle formed by partitioning becomes rectangles and squares. We observe also that this construction satisfies the relation

$$(r^2+h^2)^{1/2}=\sum_{i=1}^{n}(r_i^2+h_i^2)^{1/2},$$

Now, we compute the area of the triangle in two different ways. By direct strategy, we have that the area of the triangle, denoted $\mathcal{A}$, is given by

$$\mathcal{A}=1/2\bigg(\sum_{i=1}^{n}r_i\bigg)\bigg(\sum_{i=1}^{n}h_i\bigg).$$

On the other hand, we compute the area of the triangle by computing the area of each trapezium and the one remaining triangle and sum them together. That is,

$$\mathcal{A}=h_n/2\bigg(\sum_{i=1}^{n}r_i+\sum_{i=1}^{n-1}r_i\bigg)+h_{n-1}/2\bigg(\sum_{i=1}^{n-1}r_i+\sum_{i=1}^{n-2}r_i\bigg)+\cdots+1/2r_1h_1.$$

By comparing the area of the second argument, and linking this to the first argument, the result follows immediately. □

Corollary 2.1. Let $f:\mathbb{N}\longrightarrow\mathbb{C}$, then we have the decomposition

$$\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j)=\sum_{2\leq n\leq x}f(n)\sum_{m\leq n-1}f(m).$$

Proof. Let us take $f(j)=r_j=h_j$ in Theorem 2.1, then we denote by $\mathcal{G}$ the partial sums

$$\mathcal{G}=\sum_{j=1}^{n}f(j)$$

and we notice that

$$\begin{aligned}\sum_{j=1}^{n}\sqrt{(h_j^2+r_j^2)}&=\sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}\\&=\sum_{j=1}^{n}\sqrt{(f(j)^2+f(j)^2}\\&=\sqrt{2}\sum_{j=1}^{n}f(j).\end{aligned}$$

Since $\sqrt{(\mathcal{G}^2+\mathcal{G}^2)}=\mathcal{G}\sqrt{2}=\sqrt{2}\sum\limits_{j=1}^{n}f(j)$ our choice of sequence is valid and, therefore the decomposition is valid for any arithmetic function. □

3. Main results

In this section we state the main Lemma and establish our main result.

Theorem 3.1. *Let* $f:\mathbb{N}\longrightarrow\mathbb{C}$. *Suppose there exists some constant* $1\leq\mathcal{N}<x$ *such that*

$$\sum_{n\leq x}f(n)f(n+l_o)=\frac{\mathcal{N}}{x}\sum_{n\leq x-1}\sum_{j\leq x-n}f(n)f(n+j)$$

for arbitrary l_o with $1 \leq l_o < x$ then

$$\sum_{n\leq x} f(n)f(n+l_o) = \frac{\mathcal{N}}{x} \sum_{2\leq n\leq x} f(n) \sum_{m\leq n-1} f(m).$$

Proof. This is an easy consequence of Corollary 2.1. □

Remark 3.2. The function $\frac{\mathcal{N}}{x}$ in the statement of Theorem 3.1 can more be thought of as the local density function of the correlation

$$\sum_{n\leq x} f(n)f(n+l_o)$$

for arbitrary l_o in the interval $[1,x]$. Indeed this function will always exists for any arithmetic function so long as it depends on the size of the arbitrary shift $l_o \in \mathbb{N}$ and consequently on the range of summation $[1,x]$.

Theorem 3.3. *Let $\mathbb{P}$ denotes the set of all prime numbers, then we have the estimate*

$$\#\left\{p\leq x \mid 2p+1, p\in\mathbb{P}\right\} \geq (1+o(1))\frac{\mathcal{D}}{(2+2\log 2)}\frac{x}{\log^2 x}$$

where $\mathcal{D} \geq 1$.

Proof. Let us consider the function $\vartheta : \mathbb{N} \longrightarrow \mathbb{C}$ defined as

$$\vartheta(n) := \begin{cases} \log p & \textbf{if} \quad n = p \in \mathbb{P} \\ 0 & \textbf{otherwise} \end{cases}$$

so that by virtue of Corollary 2.1 we obtain the decomposition

$$\sum_{n\leq x} \vartheta(n)\vartheta(n+(n+1)) = \frac{\mathcal{D}}{x} \sum_{2\leq n\leq x} \vartheta(n) \sum_{m\leq n-1} \vartheta(m) \tag{3.1}$$

for $\mathcal{D} \geq 1$. Now using the weaker estimate found in the literature

$$\sum_{n\leq x} \vartheta(n) = (1+o(1))x$$

we obtain the following estimates by an appeal to summation by parts

$$\begin{aligned}
\sum_{2\leq n\leq x} \vartheta(n) \sum_{m\leq n-1} \vartheta(m) &= (1+o(1)) \sum_{2\leq n\leq x} \vartheta(n)n \\
&= (1+o(1))x \sum_{2\leq n\leq x} \theta(n) - (1+o(1)) \int_2^x \left(\sum_{2\leq n\leq t} \vartheta(n)\right)dt \\
&= (1+o(1))x^2 - (1+o(1)) \int_2^x (1+o(1))t dt \\
&= (1+o(1))x^2 - (1+o(1))\frac{x^2}{2} + O(1) \\
&= (1+o(1))\frac{x^2}{2}. \qquad (3.2)
\end{aligned}$$

By plugging (3.2) into (3.1) we obtain the estimate

$$\sum_{n\leq x}\vartheta(n)\vartheta(n+(n+1)) = \frac{\mathcal{D}}{x}(1+o(1))\frac{x^2}{2} = (1+o(1))\frac{\mathcal{D}}{2}x.$$

On the other hand, we can write

$$\sum_{n\leq x}\vartheta(n)\vartheta(n+(n+1)) = \sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}}\log p\log(2p+1) \approx \sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}}\log^2 p + (\log 2)\sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}}\log p \leq (1+\log 2)\sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}}\log^2 p \tag{3.3}$$

so that by an application of partial summation we have

$$\sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}}\log^2 p \leq \log^2 x\sum_{\substack{p\leq x\\ 2p+1\in\mathbb{P}}} 1. \tag{3.4}$$

By combining (3.2), (3.1) and (3.4) the lower bound follows as a consequence. □

Corollary 3.1. There are infinitely many primes $p\in\mathbb{P}$ such that $2p+1\in\mathbb{P}$.

Proof. This is a consequence of Theorem 3.3. □

References

1. Agama, Theophilus, *The area method and applications*, arXiv preprint arXiv:1903.09257, 2019.
2. Rivest, RL and Silverman, RD, *Are StrongPrimes Needed for RSA?*, Available from World Wide Web: http://theory. lcs. mit. edu, 1999.
3. Agrawal, Manindra and Kayal, Neeraj and Saxena, Nitin, *PRIMES is in P*, Annals of Mathematics, vol. 160, 2004, 781–798.

Department of Mathematics, African institute for mathematical sciences, Ghana, Cape-coast

E-mail address: Theophilus@ims.edu.gh/emperordagama@yahoo.com