

## MEAN REPRESENTATION NUMBER OF INTEGERS AS THE SUM OF PRIMES

GAUTAMI BHOWMIK AND  
JAN-CHRISTOPH SCHLAGE-PUCHTA

**Abstract.** Assuming the Riemann hypothesis, we obtain asymptotic estimates for the mean value of the number of representations of an integer as a sum of two primes. By proving a corresponding  $\Omega$ -term, we show that our result is essentially the best possible.

### §1. Introduction and results

When studying the Goldbach conjecture that every even integer larger than 2 is the sum of two primes, it is natural to consider the corresponding problem for the von Mangoldt function  $\Lambda$ . Instead of showing that an even integer  $n$  is the sum of two primes, one aims to show that  $G(n) = \sum_{k_1+k_2=n} \Lambda(k_1)\Lambda(k_2)$  is sufficiently large; more precisely,  $G(n) > C\sqrt{n}$  for a suitable constant  $C$  implies the Goldbach conjecture. It has long been known that this result is true for almost all  $n$ . It is easy to see that if  $f$  is an increasing function such that the Tchebychev function  $\Psi(x) = x + \mathcal{O}(f(x))$ , then the mean value of  $G(n)$  satisfies the relation

$$\sum_{n \leq x} G(n) = x^2/2 + \mathcal{O}(xf(x)).$$

If we consider the contribution of only one zero of the Riemann zeta function  $\zeta$ , an error term of size  $\mathcal{O}(f(x)^2)$  appears, which, based on current knowledge on zero-free regions of  $\zeta$ , would not be significantly better than  $\mathcal{O}(xf(x))$ . Fujii [3] studied the error term of this mean value under the Riemann hypothesis (RH) and obtained

$$\sum_{n \leq x} G(n) = x^2/2 + \mathcal{O}(x^{3/2}),$$

---

Received April 1, 2009. Revised February 5, 2010. Accepted February 14, 2010.  
2000 Mathematics Subject Classification. Primary 11P32; Secondary 11P55.

which he later improved in [4] to

$$(1) \quad \sum_{n \leq x} G(n) = x^2/2 + H(x) + (\mathcal{O}(x \log x)^{4/3}),$$

with  $H(x) = -2 \sum_{\rho} (x^{1+\rho}/\rho(1+\rho))$ , where the summation runs over all nontrivial zeros of  $\zeta$ . In fact, the oscillatory term  $H(x)$  is present even without assuming the RH; however, it is necessary for the error estimate above.

In this article, we prove the following.

**THEOREM 1.1.** *Suppose that the RH is true. Then we have*

$$\sum_{n \leq x} G(n) = \frac{1}{2}x^2 + H(x) + \mathcal{O}(x \log^5 x)$$

and

$$\sum_{n \leq x} G(n) = \frac{1}{2}x^2 + H(x) + \Omega(x \log \log x).$$

This confirms a conjecture of Egami and Matsumoto [2, Conjecture 2.2]. Recently, Granville [6] used (1) to obtain new characterizations of the RH. The innovation of the present work is the idea to use the distribution of primes in short intervals to estimate exponential sums close to the point zero. Note that, using the generalized RH, one could similarly find bounds for the exponential sums in question in certain neighborhoods of Farey fractions. Such a bound, for example, fixes a gap in the proof of [6, Theorem 1C]. This approach can further be used to study the meromorphic continuation of the generating Dirichlet series  $\sum G(n)n^{-s}$ , as introduced by Egami and Matsumoto [2], a topic we deal with elsewhere (see [1]).

The log-power in the error term can be improved, but reaching  $\mathcal{O}(x \log^3 x)$  would probably require some new idea.

## §2. Proofs

To prove the first part of our theorem, we compute the sum using the circle method. We use the following standard notation.

Fix a large real number  $x$ , set  $e(\alpha) = e^{2\pi i\alpha}$ , and let

$$S(\alpha) = \sum_{n \leq x} \Lambda(n)e(\alpha n),$$

$$T_y(\alpha) = \sum_{n \leq y} e(\alpha n),$$

$$T(\alpha) = T_x(\alpha),$$

$$R(\alpha) = S(\alpha) - T(\alpha).$$

The following is due to Selberg [8, (13)].

LEMMA 1. *Assuming the RH, we have*

$$\int_1^x |\Psi(t+h) - \Psi(t) - h|^2 dt \ll xh \log^2 x.$$

The following result is due to Gallagher (see [7, Lemma 1.9]). Put  $T = y^{-1}$ ,  $\delta = y/2$ .

LEMMA 2. *Let  $c_1, \dots, c_N$  be complex numbers, and set  $S(t) = \sum_{n=1}^N c_n \times e(tn)$ . Then*

$$\int_{-1/y}^{1/y} |S(t)|^2 dt \ll y^{-2} \int_{-\infty}^{\infty} |A(x)|^2 dx,$$

where

$$A(x) = \sum_{\substack{n \leq N \\ |n-x| \leq y/4}} c_n.$$

Our main technical result is the following.

LEMMA 3. *Suppose that the RH is true. Then we have for  $y \leq x$  the estimate*

$$\int_{-y^{-1}}^{y^{-1}} |R(\alpha)|^2 d\alpha \ll \frac{x}{y} \log^4 x.$$

*Proof.* We put  $N = x$  and  $c_n = \Lambda(n) - 1$  into Lemma 2. Putting

$$B(t) = \sum_{\substack{n \leq x \\ t < n \leq t+y/2}} c_n,$$

we obtain

$$\int_{-y^{-1}}^{y^{-1}} |R(\alpha)|^2 d\alpha \ll y^{-2} \int_{-\infty}^{\infty} |B(t)|^2 dt = y^{-2} \int_{-y/2}^N |B(t)|^2 dt.$$

In the range  $-y/2 < t < 0$ , we have

$$\int_{-y/2}^0 |B(t)|^2 dt = \int_0^{y/2} |\Psi(t) - [t]|^2 dt \ll y^2 \log^4 y.$$

For  $0 \leq t \leq x - y/2$ , we have  $B(t) = \Psi(t + y/2) - \Psi(t) - y/2 + \mathcal{O}(1)$ ; thus, we can apply Lemma 1 to obtain

$$\int_0^{x-y/2} |B(t)|^2 dt \ll x + \int_0^{x-y/2} |\Psi(t + y/2) - \Psi(t) - y/2|^2 dt \ll xy \log^2 x.$$

Finally, for  $x - y/2 \leq x \leq N$  we have  $B(x) = \Psi(x) - \Psi(t) - (x - t) + \mathcal{O}(1)$ . Since the RH is equivalent to  $\Psi(x) = x + \mathcal{O}(x^{1/2} \log^2 x)$ , this implies that  $B(x) \ll x^{1/2} \log^2 x$ , and therefore that

$$\int_{x-y/2}^x |B(t)|^2 dt \ll xy \log^4 x.$$

Collecting our estimates, our claim follows.  $\square$

Note that no nontrivial unconditional version of Lemma 3 can be proven without better understanding of the zeros of the Riemann  $\zeta$ -function, since the existence of a single zero close to 1 would already blow up the left-hand side.

Writing  $S^2(\alpha)$  as  $(T(\alpha) + R(\alpha))^2$ , we have

$$\begin{aligned} \sum_{n \leq x} G(n) &= \int_0^1 T(-\alpha) S^2(\alpha) d\alpha \\ &= \frac{1}{2} x^2 + 2 \int_0^1 |T(\alpha)|^2 R(\alpha) d\alpha + \int_0^1 T(-\alpha) R^2(\alpha) d\alpha + \mathcal{O}(x). \end{aligned}$$

We claim that the second term yields  $H(x)$  and that the last one yields an error of admissible size. In fact, the second term can be written as

$$\begin{aligned} 2 \int_0^1 |T(\alpha)|^2 S(\alpha) d\alpha - 2 \int_0^1 |T(\alpha)|^2 T(\alpha) d\alpha &= 2 \sum_{n \leq x} (\Lambda(n) - 1)([x] - n) \\ &= 2 \sum_{n \leq x-1} (\Psi(n) - n). \end{aligned}$$

We now insert the explicit formula for  $\Psi(n)$  and replace the sum over  $n$  by an integral to find that the second term is indeed  $H(x) + \mathcal{O}(x)$ .

We now consider the third term. We split the integral into an integral over  $[-x^{-1}, x^{-1}]$  and integrals, of the form  $[2^k x^{-1}, 2^{k+1} x^{-1}]$ . On each interval, we bound  $T(\alpha)$  by  $\min\{x, 1/\|\alpha\|\}$ , where  $\|\alpha\|$  is the distance of  $\alpha$  to the

nearest integer, and we bound  $R(\alpha)$  using Lemma 3. For the first interval this yields

$$\int_{-x^{-1}}^{x^{-1}} T(-\alpha)R^2(\alpha) d\alpha \ll x \int_{-x^{-1}}^{x^{-1}} R^2(\alpha) d\alpha \ll x \log^4 x,$$

while for the other intervals we obtain

$$\begin{aligned} \int_{2^k x^{-1}}^{2^{k+1} x^{-1}} T(-\alpha)R^2(\alpha) d\alpha &\ll 2^{-k} x \int_{2^k x^{-1}}^{2^{k+1} x^{-1}} R^2(\alpha) d\alpha \\ &\ll 2^{-k} x \frac{x}{2^{-k} x} \log^4 x \ll x \log^4 x. \end{aligned}$$

There are  $\mathcal{O}(\log x)$  summands; hence, the contribution of  $R^2$  to the whole integral is  $\mathcal{O}(x \log^5 x)$ , and the first part of our theorem is proven.

We now turn to the proof of the  $\Omega$ -result. To do so we show that  $G(n) = \Omega(n \log \log n)$ ; hence, the left-hand side of (1) has jumps of order  $\Omega(n \times \log \log n)$ . Since  $x^2/2$  and  $H(x)$  are continuous, the error term cannot be  $o(x \log \log x)$ . By considering the average behavior of  $H(n) - H(n-1)$ , one can even show that the error term is of order  $\Omega(x \log \log x)$  for integral  $x$ ; however, we will do only the easier case of real  $x$  here.

The idea of the proof is that if an integer  $n$  is divisible by many small primes, then  $G(n)$  should be large. Let  $q_1$  be the exceptional modulus for which a Siegel zero for moduli up to  $Q$  might exist, and let  $p_1$  be some prime divisor of  $q_1$ . For the sake of determinacy, we put  $p_1 = 2$  if no Siegel zero exists. We now use the following result due to Gallagher [5, Theorem 7].

LEMMA 4. *There exists a positive constant  $c$ , such that for  $x/Q \leq h \leq x$ , and  $\exp(\log^{1/2} x) \leq Q \leq x^c$  we have*

$$\left| x - \sum_{x \leq n \leq x+h} \Lambda(n) \right| + \sum_{q \leq Q} \sum_{\chi}^* \left| \sum_{x \leq n \leq x+h} \Lambda(n) \chi(n) \right| \ll h \exp\left(-c \frac{\log x}{\log Q}\right),$$

Here  $\sum^*$  denotes summation over primitive characters modulo  $q$ , and if there exists an exceptional character, for which a Siegel zero exists, this character has to be left out of the summation.

We put  $Q = q = \prod_{p < h, p \neq p_1} p$ . Then all characters  $\chi$  modulo  $q$  are induced by some primitive character  $\chi'$  modulo  $q' \leq q$ , and

$$\left| \sum_{x \leq n \leq x+h} \Lambda(n) \chi(n) - \sum_{x \leq n \leq x+h} \Lambda(n) \chi'(n) \right| \leq \sum_{d|q} \Lambda(d) \leq \log q,$$

which is negligible. Hence, it follows from Lemma 2 that

$$\left| x - \sum_{x \leq n \leq 2x} \Lambda(n) \chi_0(n) \right| + \sum_{\substack{\chi \pmod{q} \\ \chi \neq \chi_0}} \left| \sum_{x \leq n \leq 2x} \Lambda(n) \chi(n) \right| \leq \frac{x}{2},$$

where  $\chi_0$  is the principal character, provided that  $q < x^{c'}$  for some absolute constant  $c'$ . It follows that, for  $(a, q) = 1$ , we have

$$S(x, q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \Lambda(n) \geq \frac{x}{\varphi(q)}.$$

Now

$$\sum_{\substack{n \leq 4x \\ q|n}} G(n) \geq \sum_{(a, q)=1} S(x, q, a) S(x, q, q-a) \geq \frac{x^2}{4\varphi(q)}.$$

On the left we take the average over  $\ll x/q$  integers; hence, we obtain

$$\max_{n \leq 4x} G(n) \gg \frac{x}{2\varphi(q)} = (1 - p_1^{-1}) \prod_{p \leq h} (1 - p^{-1})^{-1} x \gg x \log \log x,$$

and our claim follows.

**Acknowledgment.** We would like to thank the referee for suggesting the use of Lemma 2, which led to a substantial improvement.

#### REFERENCES

- [1] G. Bhowmik and J.-C. Schlage-Puchta, *Meromorphic continuation of the Goldbach generating function*, to appear in *Funct. Approx. Comment. Math.*, arXiv:1002.4806v1 [math.NT]
- [2] S. Egami and K. Matsumoto, “Convolutions of the von Mangoldt function and related Dirichlet series” in *Proceedings of the 4th China-Japan Seminar on Number Theory*, World Sci. Publ., Hackensack, N.J., 2007, 1–23.
- [3] A. Fujii, *An additive problem of prime numbers*, *Acta Arith.* **58** (1991), 173–179.
- [4] ———, *An additive problem of prime numbers, II*, *Proc. Japan Acad.* **67A** (1991), 248–252.
- [5] P. X. Gallagher, *A large sieve density estimate near  $\sigma = 1$* , *Invent. Math.* **11** (1970), 329–339.
- [6] A. Granville, *Refinements of Goldbach’s conjecture, and the generalized Riemann hypothesis*, *Funct. Approx. Comment. Math.* **37** (2007), 159–173; *Corrigendum*, *Funct. Approx. Comment. Math.* **38** (2008), 125–127.
- [7] H. Montgomery, *Topics in Multiplicative Number Theory*, *Lecture Notes in Math.* **227**, Springer, New York, 1971.

- [8] A. Selberg, *On the normal density of primes in short intervals, and the difference between consecutive primes*, Arch. Math. Naturvidenskab **47** (1943), 87–105.

Gautami Bhowmik  
*Université de Lille 1*  
*Laboratoire Paul Painlevé UMR CNRS 8524*  
*59655 Villeneuve d'Ascq Cedex*  
*France*  
[bhowmik@math.univ-lille1.fr](mailto:bhowmik@math.univ-lille1.fr)

Jan-Christoph Schlage-Puchta  
*University of Gent*  
*Department of Pure Mathematics and Computer Algebra*  
*9000 Gent*  
*Belgium*  
[jcsp@cage.ugent.be](mailto:jcsp@cage.ugent.be)