Ricerche di Matematica
Vol. LII , fasc. 1°, (2003), 1-

I BOZZE

File: 679.tex
6.luglio.2003

On the symmetry of primes in almost all short intervals

G. COPPOLA (*)

ABSTRACT. – In this paper we study the symmetry of primes in almost all short intervals; by elementary methods (based on the Large Sieve) we give, for $h \gg x^{1/3} \log^c x$ ($c > 0$, suitable), a non-trivial estimate for the mean-square (over $N < x \le 2N$) of an average of "symmetry sums"; these sums control the symmetry of the von-Mangoldt function in short intervals around x. We explicitly remark that our results are out of reach of the classic analytic methods based on explicit formulas and complex integrals

KEY WORDS: **Missing**.

A.M.S. CLASSIFICATION: 11N37, 11N36.

1. Introduction and statement of the results

In this paper we study the "symmetry" of prime numbers in "almost all" short intervals.

As usual, we mean by "short interval" an interval of the type $[x, x+H]$, with $H \to \infty$ and $H = o(x)$, as $x \to \infty$; and we say that something is true for "almost all" short intervals of the kind $[x-h, x+h]$, if, given $N < x \le 2N$, with $N \to \infty$, and given $h = h(N) = o(N)$, with $h(N)$ monotone increasing and $h \to \infty$ as $N \to \infty$, the property holds for all $x \in [N, 2N]$, with a number of (eventual) exceptions which is $o(N)$ (their Lebesgue measure is such). In the sequel, however, our variable x will assume integer values (if not explicitly stated).

(*)**Missing address**

We mean, by studying this symmetry, to study sums of the kind

$$\sum_{|n-x|\le h} \Lambda(n)\,\mathrm{sgn}(n-x)$$

and other sums like this (see our theorems).

(Here $\Lambda(n)$ is the von-Mangoldt function, defined by

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^k, p \text{ a prime and } k \in \mathbf{N} \\ 0 & \text{otherwise} \end{cases}$$

and the sign function is given by $\mathrm{sgn}(t) := t/|t|$, $\forall t \in \mathbf{R} - \{0\}$, $\mathrm{sgn}(0) := 0$.)

In fact, Kaczorowski and Perelli were the first (as far as the Author is aware of) to introduce this type of sums (but with a particular weight function, $G(n, x, T)$, see [6]), in connection with the problem of the estimation of the "Selberg integral", i.e.

$$J(N, H) := \int_N^{2N} \left| \sum_{x<n\le x+H} \Lambda(n) - H \right|^2 \mathrm{d}x$$

(here $[x - H, x + H]$ is always a short interval, but the length H is slightly smaller then our h).

Actually, it can be easily shown that the *Prime Number Theorem* (abbrev. PNT) $\sum_{x<n\le x+H} \Lambda(n) \sim H$ holds for almost all the short intervals $[x, x + H]$ if and only if $J(N, H) = o(NH^2)$, whereas the Brun-Titchmarsh inequality gives (if $H = N^\theta$, $\theta \in]0, 1[$) the "trivial" estimate $\mathcal{O}(NH^2)$.

By classical estimates involving the Riemann-von Mangoldt explicit formula (see [5]), it is possible to prove that $J(N, H) = o(NH^2)$, whenever $H = N^{1/6-\varepsilon}$ (with $\varepsilon = \varepsilon(N) \to 0$, as $N \to \infty$), see [9].

In their work [6] Kaczorowski & Perelli give a link (for H a power of N; the Author & Vitolo [4] did the same for H a power of $\log N$) between the Selberg integral and the mean-square of a variant of our "symmetry sums" (see [6], Theorem 2). This sum arises as a main term in the remainders of their new form of the Riemann-von Mangoldt explicit formula [5].

In particular, if an estimate gaining $L^{1+\varepsilon}$ ($\varepsilon > 0$ enough small) w.r.t. the trivial one (that is, $\mathcal{O}(NH^2)$, see before) for the mean-square of this symmetry sum would be available, it would be possible to get $J(N, H) = o(NH^2)$, whence the "almost-all short intervals" PNT (quoted before).

We remark that the present work is still far from getting such a result; this would have, also, great consequences on the zeros of the Riemann-zeta function, as exploited in [10].

Instead, we will confine our attention in proving a (much weaker) related result; in fact (see our Theorem 1 for the exact statement) we will bound a mean-square of an average of "symmetry sums", calculated in short intervals, going from $[x-h, x+h]$ to the trivial $[\frac{x}{2h}-\frac{1}{2}, \frac{x}{2h}+\frac{1}{2}]$.

However, there is an arithmetical interest for such a result, due to the intrinsic elementary nature of the arguments used; our proof is only based, in fact, on the Large Sieve.

Also, due to the low cancelation of the remainders in the (quoted) explicit formulas for Λ, we see that the classical methods (using complex integrals and Dirichlet series) are unuseful to prove our results.

Our present paper continues, from our previous papers [2] and [3], the study of the symmetry in almost all short intervals; in particular, it takes inspiration much from Lemma 2 of [3].

In [3] we study the symmetry of the divisor function $d(n)$, while in [2] we have studied the prime-divisors $\omega(n)$.

In all our papers we use the properties of the function $\chi_q(x)$ (see the following) in the classic environment of the Large-Sieve.

We will explore the set of arithmetical functions f for which the method works in future papers.

Here we confine to $f = \Lambda$, the von-Mangoldt function (now on $L := \log N$).

Theorem 1. – *Let $A > 2$ and let N, h be natural numbers, such that $h = h(N) \to \infty$ and $h \ll NL^{-A}$ as $N \to \infty$. Then*

$$\sum_{x\sim N} \left| \sum_{j\le 2h} \sum_{|d-\frac{x}{j}|\le \frac{h}{j}} \Lambda(d)\,\mathrm{sgn}\left(d-\frac{x}{j}\right) \right|^2 \ll Nh^2L^{-A},$$

provided $h \ge N^{1/3}L^B$, where $B = B(A) > 0$ can be explicitly computed.

(Actually, a good value is $B = A + 6$.)

We explicitly remark that it's non-trivial w.r.t. the same without the "sgn".

Actually, the same methods (see the proof of our Theorem 1) allow us to prove that the same is true for such "symmetry sums" of, also, P_K numbers, not only primes (here P_K stands for an integer product of exactly K prime factors, counted with their multiplicities).

As Λ serves as a substitute for the characteristic function of primes, Λ_K serves as a substitute for the characteristic function of P_K numbers:

$$\Lambda_K(n) := \sum_{d|n} \mu(d) \log^K \left(\frac{n}{d}\right) .$$

As regards these generalized von-Mangoldt functions Λ_K, we have the following

THEOREM 2. – *Let $A > 2$ and $N, h \in \mathbf{N}$, with $h = h(N) \to \infty$ as $N \to \infty$. Assume $K > 1$ is a (fixed) natural number and $h \ll NL^{1-K-A}$. Then*

$$\sum_{x \sim N} \left| \sum_{j \le 2h} \sum_{|d - \frac{x}{j}| \le \frac{h}{j}} \Lambda_K(d) \,\mathrm{sgn}\left(d - \frac{x}{j}\right) \right|^2 \ll N h^2 L^{-A} ,$$

provided $h \ge N^{1/3} L^B$, where $B = B(A,K) > 0$ can be explicitly computed.

Also this bound is non-trivial w.r.t. the one without the "sgn".

Let's define the function $\chi_d(x)$ (not a Dirichlet character!):

$$\chi_d(x) := \left\{\frac{x+h}{d}\right\} + \left\{\frac{x-h}{d}\right\} - 2\left\{\frac{x}{d}\right\} = - \sum_{\substack{|n-x| \le h \\ n \equiv 0 (\mathrm{mod}\, d)}} \mathrm{sgn}(n-x);$$

as a function of $x (\in \mathbf{N})$ with period $d \in \mathbf{N}$ has a finite Fourier expansion.

We explicitly remark that in [3] we use the "flipping" property of the divisors d to apply the Large Sieve to the divisor function (through $\chi_d(x)$).

(In the sequel we'll talk about "moduli" and "divisors" meaning the same.)

The moduli d we will use in Theorem 1 proof are "low" and "high"; by *low* we mean that we can treat them *by the Large Sieve* (here "we can go beyond $\sqrt{N}$", due to the low mean-square of $\chi_d(x)$ Fourier coefficients); by it high we mean the Large Sieve's out of reach, but *by* a *weak flipping* (instead of the flipping, not applicable) we make them *become sporadic* terms, which can be *estimated by* the *"Fragmented Large Sieve"*.

2. Proof of Theorem 1

First of all,

$$\sum_{|n-x|\le h} \mathrm{sgn}(n-x) \sum_{d|n} \Lambda(d) = \sum_{|n-x|\le h} \mathrm{sgn}(n-x) \log n$$
$$= \sum_{|n-x|\le h} \mathrm{sgn}(n-x) \log\left(1+\frac{n-x}{x}\right) \ll \frac{h^2}{x}\,;$$

writing the first sum in terms of $\chi_d(x)$ this is equivalent in saying

$$\sum_{d\le x+h} \Lambda(d)\chi_d(x) \ll \frac{h^2}{x}\,;$$

and this will be our start point to prove Theorem 1; in fact, we will be able to treat all the relevant ranges of d, except for the one that will be in our final statement (see the sum on d in Theorem 1), but we know that the whole range of summation on d gives a small contribution (even individually, rather than on average over x !); so we will get the Theorem.

In order to distinguish between "low" and "high" modules d, we then introduce a (free) parameter $M = M(N,h) \to \infty$ as $N \to \infty$

$$\Sigma_1(x) := \sum_{d\le M\sqrt{N}} \chi_d(x)\Lambda(d) \qquad \Sigma_2(x) := \sum_{M\sqrt{N}<d\le x+h} \chi_d(x)\Lambda(d)\,.$$

Before to apply the Large-Sieve to $\Sigma_1(x)$ we first need to rearrange the inner exponential sum, as follows.

Then, $\chi_d(x)$ definition (apply the additive characters orthogonality [8] to get the finite Fourier expansion: see [2] and [3]) gives its Fourier coefficients property $c_{at,bt} = \frac{1}{t}c_{a,b}$

$$\chi_d(x) = \sum_{j<d} c_{j,d}e_d(jx) = \sum_{\substack{j<d\\(j,d)=t}} c_{j,d}e_d(jx)$$
$$= \sum_{t|d} \frac{1}{t} \sum_{\substack{j'<(d/t)\\(j',(d/t))=1}} c_{j',d/t}e_{d/t}(j'x) = \sum_{t|d} \frac{t}{d} \sum_{\substack{r\le t\\(r,t)=1}} c_{r,t}e_t(rx)\,,$$

whence

$$\Sigma_1(x) = \sum_{t\le M\sqrt{N}} \left(\sum_{n\le \frac{M\sqrt{N}}{t}} \frac{\Lambda(tn)}{n} \right) \sideset{}{^*}\sum_{j\le t} c_{j,t}e_t(jx)\,,$$

($\sum^*$ indicates, as usual, the summation over reduced classes, i.e. $(j,t)=1$).

Now we can apply the Large-Sieve inequality (see [1])

$$\sum_{x\sim N}\left|\sum_{q\le M\sqrt{N}}\alpha_q{\sum_{j\le t}}^{*}c_{j,t}e_t(jx)\right|^2 \ll NM^2\sum_{q\le M\sqrt{N}}|\alpha_q|^2\sum_{j<t}|c_{j,t}|^2\ ,$$

where, by definition of $\chi_q(x)$

$$\alpha_q:=\sum_{n\le\frac{M\sqrt{N}}{q}}\frac{\Lambda(qn)}{n}\qquad \sum_{j<q}|c_{j,q}|^2\ll\frac{h}{q}\,;$$

then we obtain (by the upper bound $\sum_{n\le X}\Lambda(qn)/n\ll\log X$, see [7])

$$\sum_{x\sim N}\left|\sum_{d\le M\sqrt{N}}\Lambda(d)\chi_d(x)\right|^2\ll NM^2hL^3\,.$$

Hence we know $\Sigma_1(x)$-mean-square.

Recalling the definition of $\chi_d(x)$, we get the series

$$\Sigma_2(x)=\sum_{j\in\mathbf{N}}\left(\sum_{\substack{M\sqrt{N}<d\le x+h\\ \frac{x}{j}<d<\frac{x+h}{j}}}\Lambda(d)-\sum_{\substack{M\sqrt{N}<d\le x+h\\ \frac{x-h}{j}<d<\frac{x}{j}}}\Lambda(d)\right)+ \\ +\mathcal{O}\left(\left(d(x-h)+d(x)+d(x+h)\right)L\right)$$

($d=\frac{x-h}{j}$, $d=\frac{x}{j}$ and $d=\frac{x+h}{j}$ give the divisor-functions-remainders; for the moment they will be neglected as this mean-square contributes $\mathcal{O}(NL^5)$).

This series is genuinely finite, due to the inner constraints on d; in fact

$$\begin{aligned}\Sigma_2(x)&=\sum_{j<\frac{x+h}{M\sqrt{N}}}\left(\sum_{\substack{M\sqrt{N}<d\le x+h\\ \frac{x}{j}<d<\frac{x+h}{j}}}\Lambda(d)-\sum_{\substack{M\sqrt{N}<d\le x+h\\ \frac{x-h}{j}<d<\frac{x}{j}}}\Lambda(d)\right)\\ &=\sum_{j\le\frac{x-h}{M\sqrt{N}}}\ \sum_{\frac{x-h}{j}<d\le\frac{x+h}{j}}\mathrm{sgn}\left(d-\frac{x}{j}\right)\Lambda(d)+\mathcal{O}\left(L\sum_{\frac{x-h}{M\sqrt{N}}<j\le\frac{x+h}{M\sqrt{N}}}\left(\frac{h}{j}+1\right)\right)\\ &=\sum_{j\le\frac{x-h}{M\sqrt{N}}}\ \sum_{\frac{x-h}{j}<d\le\frac{x+h}{j}}\mathrm{sgn}\left(d-\frac{x}{j}\right)\Lambda(d)+\mathcal{O}\left(\left(\frac{h^2}{N}+\frac{h}{M\sqrt{N}}+1\right)L\right)\,.\end{aligned}$$

We isolate the first $2h$ terms of $\Sigma_2(x)$ (appearing in Theorem 1), say $\Sigma(x)$

$$\Sigma_2(x) := \Sigma(x) + S(x) + \mathcal{O}\left(\left(\frac{h^2}{N} + \frac{h}{M\sqrt{N}} + 1\right) L\right),$$

where

$$S(x) := \sum_{2h<j\le \frac{x-h}{M\sqrt{N}}} \sum_{\frac{x-h}{j}<d\le \frac{x+h}{j}} \mathrm{sgn}\left(d - \frac{x}{j}\right) \Lambda(d) .$$

Thus (the terms coming from $d(x)$ and $d(x \pm h)$ contribute in mean-square $\mathcal{O}(NL^5)$, see [7]) we will prove that (see at the end for the choice of M)

$$(*) \qquad \sum_{x\sim N} |S(x)|^2 \ll \frac{N^{3/2}}{M} hL^7 .$$

We explicitly remark that in $S(x)$ (since $j > 2h$) the inner sum on d becomes *sporadic*, i.e. present with at most *one* term (for which the estimate $\mathcal{O}(1)$ is non-optimal, since it's "rather always" 0).

Henceforth, the following argument holds: this sporadic term is present if and only if the interval $](x-h)/j, (x+h)/j]$ contains a prime power (and then it's unique); if so, this prime power is exactly $[\frac{x+h}{j}]$; the function $\chi_j(x)$ detects the presence of integers in this interval, also giving (by the sign) the side of x/j (right or left of).

Then

$$S(x) = \sum_{2h<j\le \frac{x-h}{M\sqrt{N}}} \chi_j(x) \Lambda\left(\left[\frac{x+h}{j}\right]\right) ;$$

we use, now, a dyadic dissection of the j-range to get

$$S(x) \ll L \max_{2h<J\le \frac{N-h}{2M\sqrt{N}}} S_J(x) + S^*(x) ,$$

where, say,

$$S_J(x) := \sum_{q\sim J} \chi_q(x) \Lambda\left(\left[\frac{x+h}{q}\right]\right) ,$$

and

$$S^*(x) := \sum_{\frac{N-h}{2M\sqrt{N}}<q\le \frac{x-h}{2M\sqrt{N}}} \chi_q(x) \Lambda\left(\left[\frac{x+h}{q}\right]\right) ;$$

we remark, here, that everything valid for $S_J(x)$ will be also valid for $S^*(x)$, in particular the following estimate over the x-exponential sum (see [3]).

We apply here the "fragmented Large-Sieve", i.e. we prove a Large-Sieve type inequality in which the x-summation range is fragmented. In fact, as already seen before, by $\chi_q(x)$ properties we can write

$$S_J(x) = \sum_{d\le 2J} \sum_{n\sim \frac{J}{d}} \frac{1}{n}\Lambda\left(\left[\frac{x+h}{nd}\right]\right) \sum_{j\le d}{}^{*} c_{j,d} e_d(jx)\,.$$

Expanding the square we get

$$\sum_{x\sim N} |S_J(x)|^2 = \sum_{d_1,d_2\le 2J} \sum_{\substack{n_1\sim J/d_1\\ n_2\sim J/d_2}} \frac{1}{n_1}\frac{1}{n_2} \sum_{j_1\le d_1}{}^{*} \sum_{j_2\le d_2}{}^{*} c_{j_1,d_1}\overline{c_{j_2,d_2}}\times$$
$$\times \sum_{x\sim N} \Lambda\left(\left[\frac{x+h}{n_1 d_1}\right]\right)\Lambda\left(\left[\frac{x+h}{n_2 d_2}\right]\right) e\left(\left(\frac{j_1}{d_1}-\frac{j_2}{d_2}\right)x\right)\,.$$

This last sum over x is estimated trivially, when $j_1/d_1 = j_2/d_2$, as $\mathcal{O}(NL^2)$; while can be bounded by means of Lemma 2 of [3] when j_1/d_1 and j_2/d_2 are distinct Farey fractions. In fact, we can break the range of x into $\mathcal{O}(N/J)$ "fragmented" ranges (each of length at most $\mathcal{O}(J)$) in which the Λ-functions are constant (being constant their arguments); then, we argue as in Lemma 2 of [3], with $k := j/d$ and $C_k := c_{j,d}$ (and we use the classical estimate for exponential sums, see [8]):

$$\sum_{x\sim N} |S_J(x)|^2 \ll L^2 \sum_{n_1,n_2\le 2N} \frac{1}{n_1}\frac{1}{n_2} \sum_{k_r,k_s} |C_{k_r}||C_{k_s}| \min\left(N, \frac{N}{J\|k_r-k_s\|}\right)$$
$$\ll L^2 \sum_{n_1,n_2\le 2N} \frac{1}{n_1 n_2}\left(N\sum_k |C_k|^2 + \frac{N}{J}\sum_{k_r\ne k_s} |C_{k_r}||C_{k_s}|\frac{1}{\|k_r-k_s\|}\right)$$

which is, by Lemma 2 of [3], with $\delta \ge 1/J^2$, at most

$$\ll L^2 \sum_{n_1,n_2\le 2N} \frac{1}{n_1 n_2} L(N+NJ)\sum_k |C_k|^2 \ll NJhL^5\,,$$

since $\sum_k |C_k|^2 \ll h$. Hence we get $(*)$:

$$\sum_{x\sim N} |S(x)|^2 \ll \max_{2h<J\le \frac{\sqrt{N}}{M}} NJhL^7 \ll \frac{N^{3/2}}{M} hL^7\,.$$

Collecting the estimates for $\Sigma_1(x)$ and $S(x)$, together with the optimal choice $M = N^{1/6}L^{4/3}$, we get (with $B = A+6$) the theorem.

REFERENCES

[1] E. Bombieri: "Le Grande Crible dans la théorie analytique des nombres", *Asterisque 18, Société Mathématique de France*, Paris, 1974. MR 51 # 8057.

[2] G. Coppola: "On the symmetry of distribution of the prime-divisors function in almost all short intervals", to appear.

[3] G. Coppola, S. Salerno: "On the symmetry of the divisor function in almost all short intervals", to appear.

[4] G. Coppola, A. Vitolo: "On the distribution of primes in intervals of length $\log^{\Theta} N$", *Acta Math. Hungar.* **70** (2) (1996), 151-166. MR 96i:11103.

[5] J. Kaczorowski, A. Perelli: "A new form of the Riemann-von Mangoldt explicit formula", *Boll. Un. Mat. Ital. B* **10** (1) (1996), 51-66. MR 97g:11098.

[6] J. Kaczorowski, A. Perelli: "On the distribution of primes in short intervals", *J. Math. Soc. Japan* **45** (3) (1993), 447-458. MR 94e:11100.

[7] G. Tenenbaum: "Introduction to analytic and probabilistic number theory", *Cambridge University Press*, Cambridge, 1995. MR 97e:11005b.

[8] I. M. Vinogradov: "The Method of Trigonometrical Sums in the Theory of Numbers", *Interscience Publishers LTD.*, London, 1954. MR 15,941b.

[9] A. Zaccagnini: "Primes in almost all short intervals", *Acta Arith.* **84** (3) (1998), 225-244. MR 2001k:11105.

[10] A. Zaccagnini: "A conditional density theorem for the zeros of the Riemann zeta-function", *Acta Arith.* **93** (3) (2000), 293-301. MR 2001k:11166.

Pervenuto alla Redazione il 4 gennaio 2002