Functiones et Approximatio
XXIX (2001), 81–88

A NOTE ON THE DISTRIBUTION OF SUMSETS

JÖRG BRÜDERN & ALBERTO PERELLI

1. Introduction

Let $\mathcal{A} \subset \mathbb{N}$ denote a set of natural numbers, and let $\nu(n)$ denote the number of solutions of $a+b=n$ with $a,b \in \mathcal{A}$. In many cases where $\mathcal{A}$ is a specific set, it is conjectured that there is an asymptotic formula for $\nu(n)$. For example, when $\mathcal{A}$ is the sequence of primes, Hardy and Littlewood [1] predict the validity of

$$\nu(n) \sim \frac{n}{(\log n)^2} \prod_{p|n} \frac{p}{p-1} \prod_{p\nmid n} \Big(1 - \frac{1}{(p-1)^2}\Big), \tag{1.1}$$

but this is still not known. Their suggestion is backed by the Siegel-Walfisz-theorem (or any weaker variant thereof) which describes the distribution of primes in arithmetic progressions, so that the contribution of the major arcs in the circle method integral for $\nu(n)$ can be evaluated and yields the right hand side of (1.1).

Returning to the general situation, a similar heuristics applies as soon as a suitable analogue of the Siegel-Walfisz-theorem controls the distribution of $\mathcal{A}$ in arithmetic progressions. One is then lead to expect an asymptotic formula

$$\nu(n) \sim J(n)\mathfrak{S}(n) \tag{1.2}$$

where $J(n)$ and $\mathfrak{S}(n)$ denote the formal singular integral and singular series, respectively, of the problem at hand (for comparison with (1.1), $J(n)$ replaces $n(\log n)^{-2}$, and $\mathfrak{S}(n)$ replaces the Euler product). However, it is well known that the singular series $\mathfrak{S}(n)$ has average value 1 in any plausible concrete case, and we may therefore hope that the sum

$$\sum_{n\in\mathcal{E}} \big(\nu(n) - J(n)\big) \tag{1.3}$$

is small for any sufficiently large "random" set $\mathcal{E}$. The purpose of this note is to show that this is indeed the case for a large class of sets $\mathcal{A}$. It turns out that no information is needed concerning the distribution of $\mathcal{A}$ in arithmetic progressions; a sufficiently "smooth" asymptotic formula for the counting function is enough.

Before we can state the result, we need to introduce the concept of a *regular* arithmetical function. Let $M : \mathbb{N} \to [0,\infty)$ denote an arithmetical function and define $t(n) = M(n) - M(n-1)$ where for convenience we put $M(0) = 0$. The function M is called *regular* when t is monotonically decreasing, non-negative and satisfies the inequalities

$$t(n) \asymp \frac{M(n)}{n}. \tag{1.4}$$

Note that for natural numbers $x \le y \le 2x$ one always has

$$M(x) \asymp M(y) \tag{1.5}$$

when M is a regular function. In fact, (1.4) asserts that $t(n) \le \mathrm{c}M(n)n^{-1}$ holds for all n with an absolute constant $\mathrm{c} > 0$. Hence

$$M(y) - M(x) = \sum_{x<n\le y} t(n) \le \mathrm{c} \sum_{x<n\le y} \frac{M(n)}{n}.$$

From $t(n) \ge 0$ we see that M is increasing, and therefore,

$$M(y) - M(x) \le \mathrm{c}M(y)\frac{y-x}{x}.$$

For $y \le (1+\frac{1}{2\mathrm{c}})x$, this implies $M(x) \le M(y) \le 2M(x)$, and (1.5) follows by repeated application of this.

Typical examples of regular arithmetic functions are

$$n^{\lambda}(\log n)^{\mu}(\log\log n)^{\eta}$$

when $0 < \lambda < 1, \mu \in \mathbb{R}$, or when $\lambda = 1, \mu < 0, \eta \in \mathbb{R}$. If an arithmetic function M is the restriction of a differentiable function $M : [1,\infty) \to [0,\infty)$, then by the mean value theorem, the condition (1.4) may be replaced by $M'(x) \asymp \frac{M(x)}{x}$ for all $x \in (1,\infty)$; this is often useful when checking regularity in concrete cases. We are now ready to state the result.

Theorem. *Let* $1 \le N \le X$ *denote natural numbers. Let* $\mathcal{A} \subset \mathbb{N}$, *write* $A(x) = \#\mathcal{A} \cap [1,x]$, *and let* M *be a regular arithmetic function such that*

$$R(x) = A(x) - M(x)$$

satisfies $R(x) = o\big(M(x)\big)$ *as* $x \to \infty$. *Then*

$$\sum_{\substack{\mathcal{E}\subset\{X+1,\ldots,2X\}\\ \#\mathcal{E}=N}} \Big|\sum_{n\in\mathcal{E}} (\nu(n) - J(n))\Big| \ll N\binom{X}{N}M(X)\left(\frac{1}{\sqrt{N}} + \left(\frac{\max_{y\le 2X}|R(y)|}{X}\right)^{\frac{1}{2}}\right)$$

where

$$J(n) = \sum_{k+l=n} t(k)t(l).$$

For the argument to follow it is useful to have at hand a lower bound for $J(n)$. Since $t(k) \geq 0$ for all k, we have

$$J(n) \geq \sum_{\substack{k+l=n \\ \frac{1}{4}n<k<\frac{3}{4}n}} t(k)t(l).$$

From (1.4) and (1.5), we find

$$J(n) \gg \frac{M(n)^2}{n^2} \sum_{\substack{k+l=n \\ \frac{1}{4}n<k<\frac{3}{4}n}} 1 \gg \frac{M(n)^2}{n}. \tag{1.6}$$

Let $\mathcal{S}(X,N)$ denote the collection of all sets $\mathcal{E} \subset \{X+1,\ldots,2X\}$ with N elements. If we consider the sum (1.3) in the light of the lower bound (1.6), then for a set $\mathcal{E} \in \mathcal{S}(X,N)$ one would aim for

$$\sum_{n\in\mathcal{E}} \big(\nu(n) - J(n)\big) = o\big(NM(X)^2X^{-1}\big) \tag{1.7}$$

as this is then certainly non-trivial.

Corollary. *In addition to the assumptions in the Theorem, suppose that*

$$\max_{y\leq 2x} |R(x)| = o\Big(\frac{M(X)^2}{X}\Big)$$

and that $N = N(X)$ *is an increasing function such that* $\frac{X^2}{N(X)M(X)^2} \to 0$ *as* $X \to \infty$. *Then, for all but* $o\big(\binom{X}{N}\big)$ *of the sets* $\mathcal{E} \subset \mathcal{S}(X,N)$, *the bound* (1.7) *is valid.*

To prove this, it suffices to note that the conditions in the corollary imply that

$$\sum_{\mathcal{E}\in\mathcal{S}(X,N)} \Big|\sum_{n\in\mathcal{E}} \big(\nu(n) - J(n)\big)\Big| = o\Big(N\binom{X}{N}\frac{M(X)^2}{X}\Big)$$

by the Theorem. Note that one cannot expect that (1.3) is small for all sets $\mathcal{E}$ on the sole assumption that N is large. This can be seen, for example, in the case where $\mathcal{A}$ is the set of primes excluding 2. Then $\nu(n) = 0$ whenever $2 \nmid n$, and hence (1.7) certainly fails as soon as a positive proportion of the numbers in $\mathcal{E}$ are odd.

The Theorem and its corollary provide non-trivial results only when $\sqrt{x} = o(M(x))$. This is not surprising since whenever $M(x) = o(\sqrt{x})$, one has $\nu(n) > 0$

for at most $\ll M(x)^2$ of the integers $n \le x$, and hence $\nu(n)$ vanishes for almost all n in this case, forcing the sum $\sum_{n\in\mathcal{E}} \nu(n)$ to vanish also for most sets $\mathcal{E}$ with $\#\mathcal{E} = o(x)$.

Before we move on to establish the theorem, it perhaps worth to stress again that the estimates in the Theorem do not depend on the distribution of $\mathcal{A}$ in arithmetic progressions. If, on the contrary, one has a result of Siegel-Walfisz type available for $\mathcal{A}$, then it also possible to study the sums

$$\sum_{n\in\mathcal{E}} \big(\nu(n) - \mathfrak{S}(n)J(n)\big). \tag{1.8}$$

The correction by the singular series should make the individual terms smaller. Indeed, if the asymptotic formula (1.2) holds for almost all n, then it is easy to count the sets $\mathcal{E} \in \mathcal{S}(X,N)$ where (1.8) exceeds $\varepsilon NM(X)^2X^{-1}$ in size: let $\mathcal{B}$ be the set of all $n \le X$ for which (1.2) fails whence $\#\mathcal{B} = o(X)$; then for any $\mathcal{E} \in \mathcal{S}(X,N)$ where (1.8) is large, one must have $\#(\mathcal{E}\cap\mathcal{B}) \ge \varepsilon N$. A simple combinatorial counting argument gives an estimate for the number of all such $\mathcal{E} \in \mathcal{S}(X,N)$ in terms of ε, N and $\#\mathcal{B}$, which is non-trivial throughout the range $1 \le N \le X$, and is much superior to the Theorem in the ranges where the Theorem is applicable.

We illustrate this last point with an example and consider the set $\mathcal{A}$ of all natural numbers that are the sum of two cubes of natural numbers. In this case, $\nu(n)$ is intrinsically related to Waring's problem for four cubes. Therefore, we also introduce the functions $r_s(n)$ to denote the number of solutions of $n = x_1^3 + x_2^3 + \dots + x_s^3$ in natural numbers x_i. In particular, we have $\mathcal{A} = \{n : r_2(n) > 0\}$. A recent result of Heath-Brown [2] (improving earlier work of Hooley [3, 4]) shows that $r_2(n) = 2$ holds for all but $O(X^{4/9+\varepsilon})$ of the numbers $n \le X$ with $n \in \mathcal{A}$. Since $r_2(n) \ll n^\varepsilon$ holds for any $\varepsilon > 0$, one finds that

$$A(X) = \frac{1}{2}\sum_{n\le X} r_2(n) + O(X^{4/9+\varepsilon}) = \frac{3\Gamma(\frac{4}{3})^2}{4\Gamma(\frac{2}{3})}X^{\frac{2}{3}} + O\big(X^{\frac{4}{9}+\varepsilon}\big)$$

with the aid of Gauss lattice point argument to evaluate the sum of $r_2(n)$. Returning now to the function $\nu(n)$ in the special case under consideration, we have

$$\nu(n) = \frac{1}{4}r_4(n) + E(n)$$

where

$$E(n) \ll n^\varepsilon \#\{(a,b)\in\mathcal{A}^2 : a+b=n, r_2(b)\ne 2\}.$$

The aforementioned result of Heath-Brown then shows that

$$\sum_{n\le X} |E(n)| \ll A(X)X^{4/9+\varepsilon} \ll X^{10/9+\varepsilon}. \tag{1.9}$$

Moreover, as a consequence of Theorem 2 of Vaughan [5], the asymptotic formula

$$r_4(n) = \Gamma\Big(\frac{4}{3}\Big)^3 \mathfrak{S}(n) n^{1/3} + O\big(n^{1/3}(\log n)^{-1/4}\big),$$

where

$$\mathfrak{S}(n) = \sum_{q=1}^{\infty} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} q^{-4} \left(\sum_{x=1}^{q} e\Big(\frac{ax^3}{q}\Big) \right)^4 e\Big(-\frac{an}{q}\Big)$$

is the singular series for four cubes, holds for all but $O\big(X(\log X)^{-\frac{1}{4}}\big)$ of the natural numbers $n \le X$. Combining this with (1.9), it follows that

$$\nu(n) - \frac{1}{4}\Gamma\Big(\frac{4}{3}\Big)^3 \mathfrak{S}(n) n^{1/3} \ll n^{1/3}(\log n)^{-1/4} \tag{1.10}$$

holds for all but $O\big(X(\log X)^{-\frac{1}{4}}\big)$ of the natural numbers $n \le X$. We now carry out the counting argument alluded to in the previous paragraph. Let E denote the exact number of n in the interval $X < n \le 2X$ for which (1.10) fails. Then, for any $\varepsilon > 0$, the inequality

$$\Big| \sum_{n \in \mathcal{E}} \big(\nu(n) - \mathfrak{S}(n) n^{1/3}\big) \Big| > \varepsilon N X^{1/3}$$

can hold for sets $\mathcal{E} \in \mathcal{S}(X,N)$ only if at least εN elements of $\mathcal{E}$ are counted by E. Thus, the number of such sets $\mathcal{E} \in \mathcal{S}(X,N)$ does not exceed

$$\sum_{j > \varepsilon N} \binom{E}{j} \binom{X-E}{N-j} \ll \binom{X}{N} 2^N (E/X)^{\varepsilon N}.$$

2. A simple lemma

In this section, we consider the mean square of the exponential sums

$$K_{\mathcal{E}}(\alpha) = \sum_{n \in \mathcal{E}} e(\alpha n)$$

when $\mathcal{E}$ varies over $\mathcal{S}(X,N)$.

Lemma. *For $\alpha \in \mathbb{R}$ we have*

$$\sum_{\mathcal{E} \in \mathcal{S}(X,N)} |K_{\mathcal{E}}(\alpha)|^2 \ll \binom{X}{N} \big(N + N^2(1 + X\|\alpha\|)^{-2}\big)$$

where $\|\alpha\|$ denotes the distance of α to the nearest integer.

Proof. For brevity, all sums over $\mathcal{E}$ are over all $\mathcal{E} \in \mathcal{S}(X,N)$. We open the square and start from

$$\sum_{\mathcal{E}} |K_{\mathcal{E}}(\alpha)|^2 = \binom{X}{N} N + \sum_{\mathcal{E}} \sum_{\substack{n,m\in\mathcal{E} \\ n\neq m}} \mathrm{e}(\alpha(n-m)). \tag{2.1}$$

The first term on the right is acceptable. In the remaining sum, we exchange summation and note that for any pair $n \neq m$ with $X < n, m \leq 2X$ there are exactly $\binom{X-2}{N-2}$ sets $\mathcal{E} \in \mathcal{S}(X,N)$ with $n \in \mathcal{E}, m \in \mathcal{E}$. It follows that

$$\sum_{\mathcal{E}} \sum_{\substack{n,m\in\mathcal{E} \\ n\neq m}} \mathrm{e}(\alpha(n-m)) = \sum_{\substack{X<n,m\leq 2X \\ n\neq m}} \mathrm{e}(\alpha(n-m)) \binom{X-2}{N-2}.$$

We add terms with $n = m$ to the right hand side. Then, by a standard estimate,

$$\begin{aligned}\sum_{\mathcal{E}} \sum_{\substack{n,m\in\mathcal{E} \\ n\neq m}} \mathrm{e}(\alpha(n-m)) &= \binom{X-2}{N-2}\Big(\Big|\sum_{X<n\leq 2X} \mathrm{e}(\alpha n)\Big|^2 - X\Big) \\ &\ll \binom{X-2}{N-2}\Big(X^2(1+X\|\alpha\|)^{-2}\Big).\end{aligned}$$

The Lemma now follows from (2.1) on noting that

$$\binom{X-2}{N-2} X^2 = \frac{XN(N-1)}{X-1}\binom{X}{N} \ll N^2 \binom{X}{N}.$$

3. Proof of the theorem

We shall compare the exponential sums

$$S(\alpha) = \sum_{\substack{n\in\mathcal{A} \\ n\leq 2X}} \mathrm{e}(\alpha n), \qquad T(\alpha) = \sum_{n\leq 2X} t(n)\mathrm{e}(\alpha n)$$

in various ways. From $S(0) = A(2X)$ and $T(0) = M(2X)$ we see that $S(0)$ and $T(0)$ are close to each other. Partial summation shows that

$$S(\alpha) - T(\alpha) = \mathrm{e}(2\alpha X)R(2X) - 2\pi i\alpha \int_1^{2X} \mathrm{e}(\alpha\tau)R([\tau])d\tau$$

where $[\tau]$ is the integer part of τ. On writing

$$R^*(X) = \max_{m\leq 2X} |R(m)|$$

we infer that

$$S(\alpha) - T(\alpha) \ll (1 + X|\alpha|)R^*(X). \tag{3.1}$$

It will also be convenient to have at hand the mean square of $S(\alpha)$ and $T(\alpha)$. By Parseval's identity and (1.5), we have

$$\int_{-1/2}^{1/2} |S(\alpha)|^2 d\alpha = A(2X) \ll M(X). \tag{3.2}$$

We may argue similarly for $T(\alpha)$, recalling that $t(n)$ is decreasing and non-negative. This leads to the bound

$$\int_{-1/2}^{1/2} |T(\alpha)|^2 d\alpha = \sum_{n \le 2X} t(n)^2 \le t(1) \sum_{n \le 2X} t(n) \ll M(X). \tag{3.3}$$

We are now ready for the main argument. Let $\mathcal{E} \in \mathcal{S}(X, N)$. Then, by orthogonality,

$$\sum_{n \in \mathcal{E}} (\nu(n) - J(n)) = \int_{-1/2}^{1/2} (S(\alpha)^2 - T(\alpha)^2) K_{\mathcal{E}}(-\alpha) d\alpha.$$

However, by Cauchy's inequality and the Lemma, we have

$$\sum_{\mathcal{E} \in \mathcal{S}(X,N)} |K_{\mathcal{E}}(-\alpha)| \ll \binom{X}{N} (\sqrt{N} + N(1 + X|\alpha|)^{-1})$$

whenever $|\alpha| \le \frac{1}{2}$. Since (3.2) and (3.3) imply that

$$\int_{-1/2}^{1/2} |S(\alpha)^2 - T(\alpha)^2| d\alpha \ll M(X),$$

it follows that

$$\sum_{\mathcal{E} \in \mathcal{S}(X,N)} \Big| \sum_{n \in \mathcal{E}} (\nu(n) - J(n)) \Big|$$
$$\ll \binom{X}{N} M(X) \sqrt{N} + \binom{X}{N} N \int_{-1/2}^{1/2} \frac{|S(\alpha)^2 - T(\alpha)^2|}{1 + X|\alpha|} d\alpha. \tag{3.4}$$

We are now reduced to estimate the integral on the right hand side. Let $\delta \ge 1$ be a parameter to be chosen later. We split the integral into the ranges $|\alpha| \le \delta/X$ and $\delta/X \le |\alpha| \le \frac{1}{2}$. In the first case, (3.1) yields

$$\frac{|S(\alpha)^2 - T(\alpha)^2|}{1 + X|\alpha|} \ll R^*(X)(|S(\alpha)| + |T(\alpha)|) \ll R^*(X) M(X);$$

here we used the trivial bounds $|S(\alpha)| \le S(0), |T(\alpha)| \le T(0)$. This shows that

$$\int_{-\delta/X}^{\delta/X} \frac{|S(\alpha)^2 - T(\alpha)^2|}{1+X|\alpha|} d\alpha \ll \delta X^{-1} R^*(X) M(X).$$

On the complementary part, we have

$$\int_{\delta/X \le |\alpha| \le \frac{1}{2}} \frac{|S(\alpha)^2 - T(\alpha)^2|}{1+X|\alpha|} d\alpha \le \delta^{-1} \int_{-1/2}^{1/2} |S(\alpha)^2 - T(\alpha)^2| d\alpha \ll \frac{M(X)}{\delta}.$$

Hence we choose δ by $\delta^2 = XR^*(X)^{-1}$ to deduce that

$$\int_{-1/2}^{1/2} \frac{|S(\alpha)^2 - T(\alpha)^2|}{1+X|\alpha|} d\alpha \ll M(X) R^*(X)^{\frac{1}{2}} X^{-\frac{1}{2}} \tag{3.5}$$

(here it is essential to note that $M(X) \ll X$, and so $R^*(X) = o\big(M(X)\big)$ gives $R^*(X) = o(X)$ whence $\delta = \delta(X) \to \infty$ as $X \to \infty$). The Theorem is now available from (3.4) and (3.5).

References

[1] G.H. Hardy and J.E. Littlewood, *Some problems of 'Partitio Numerorum': III. The expression of a number as a sum of primes*, Acta Math. **44** (1922), 1–70.

[2] D.R. Heath-Brown, *The density of rational points on cubic surfaces*, Acta Arith. **79** (1997), 17–30.

[3] C. Hooley, *On the representation of a number as the sum of two cubes*, Math. Z. **82** (1963), 259–266.

[4] C. Hooley, *On the numbers that are representable as the sum of two cubes*, J. Reine Angew. Math. **314** (1980), 146–173.

[5] R.C. Vaughan, *On Waring's problem for cubes*, J. Reine Angew. Math. **365** (1986), 122–170.

Address: Jörg Brüdern, Mathematisches Institut A, Pfaffenwaldring 57, D-70567 Stuttgart
Alberto Perelli, Dipartimento di Matematica, Via Dodecaneso 35, I-16146 Genova
E-mail: bruedern@mathematik.uni-stuttgart.de; perelli@dima.unige.it
Received: 1 October 2001