ON IRREGULARITIES OF SUMS OF INTEGERS

BENEDEK VALKÓ (BUDAPEST)

1. Introduction

Let $\mathcal{A} \subseteq \{1, 2, \dots, N\}$ and $\eta = \frac{|\mathcal{A}|}{N}$. In [4] Roth studied how well can $\mathcal{A}$ be distributed in arithmetic progressions. If the sequence $n, n+q, \dots, n+(k-1)\,q$ has t elements in $\{1, 2, \dots, N\}$ then it is plausible to think that a well-distributed sequence would have "about" ηt elements in this sequence if k is "big enough" relative to N. Thus with $V(n, q, k) = \left|\sum_{\substack{0 \le i < k \\ n+iq \in \mathcal{A}}} 1 - \eta \sum_{\substack{0 \le i < k \\ 1 \le n+iq \le N}} 1\right|$ for a well-distributed sequence $V\left(n, q, \left\lfloor \frac{\lfloor \sqrt{N} \rfloor}{2} \right\rfloor\right)$ "should be small" for every integer $n, 1 \le q \le \left\lfloor \sqrt{N} \right\rfloor$. (It is easy to see that if $\eta = 0$ or 1 then $V(n, q, k) = 0$ for every n, q, k.) Roth proved that there does not exist a non-trivial well-distributed $\mathcal{A}$ in this sense, because for every $\mathcal{A}$ there exist integers $n, 1 \le q \le \left\lfloor \sqrt{N} \right\rfloor$ with $V\left(n, q, \left\lfloor \frac{\sqrt{N}}{2} \right\rfloor\right) \ge c\sqrt{\eta(1-\eta)}N^{\frac{1}{4}}$ (and the constant c is absolute). This means that unless $\mathcal{A}$ is empty or equal to $\{1, 2, \dots, N\}$ it cannot be well-distributed simultaneously among and within all congruence classes. Later it was shown that Roth's estimate is sharp, that is for every N positive integer there exists $\mathcal{A} \subseteq \{1, 2, \dots, N\}$ (with a "small" value of $\left|\eta - \frac{1}{2}\right|$) for which $V\left(n, q, \left\lfloor \frac{\sqrt{N}}{2} \right\rfloor\right) \le c'\sqrt{\eta(1-\eta)}N^{\frac{1}{4}}$ for every integer $n, 1 \le q \le \left\lfloor \sqrt{N} \right\rfloor$ [2].

In the first part of this paper we will prove a similar statement about the distribution of the sums $a_1 + a_2$ where $a_1, a_2 \in \mathcal{A}$. First we have to define what is considered well-distributed in this case. Denote $f(n) = \sum_{\substack{a_1+a_2=n \\ a_1, a_2 \in \mathcal{A}}} 1$ and $g(n) = \sum_{\substack{1 \le k \le N \\ 1 \le n-k \le N}} 1$. If $\mathcal{A}$ would be constructed using probability methods in a way that the probability of choosing each integer $k \in \{1, 2, \dots, N\}$ in $\mathcal{A}$ would be η (independently) then the expected value of $f(n)$ would be "roughly" $\eta^2 g(n)$. Therefore it is plausible to study $A(n, q, k) = \left|\sum_{i=0}^{k-1} \left(f(n+iq) - \eta^2 g(n+iq)\right)\right|$. (If $\eta = 0$ or 1 then $A(n, q, k) = = 0$ for every n, q, k.) In Section 2 two theorems will be proved. Theorem 2.1 will show that unless $\eta = 0$ or 1 the sums $a_1 + a_2$ with $a_1, a_2 \in \mathcal{A}$ cannot be well-distributed in this sense, because there exist integers n and q with $1 \le q \le \left\lfloor \sqrt{N} \right\rfloor$ and $A(n, q, \left\lfloor \frac{\lfloor \sqrt{N} \rfloor}{2} \right\rfloor) \ge c\eta(1-\eta)N^{\frac{3}{4}}$. Theorem 2.5 will give an upper estimate that how good an inequality of this kind can be.

In Section 3 we study the corresponding problem in residue classes. Consider a subset of the residues modulo N: $\mathcal{A} \subseteq \{0, 1, 2, \dots, N-1\}$. The sums $a_1 + a_2$ with $a_1, a_2 \in \mathcal{A}$ are well-distributed if the number of sums in each residue class are about the same. Theorem 3.1 and its corollary states that these sums cannot be well-distributed this way, similar inequalities can be shown as in the previous part. Theorem 3.4 and 3.5 show that these results arenearly best possible.

2. Irregularities of distribution of sums relative to arithmetic progressions

We use the notations introduced in the introduction.

Theorem 2.1. *If $\varepsilon > 0$ and $N > N_\varepsilon$ then for every $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ there exist integers n and q with $1 \le q \le \left\lfloor \sqrt{N} \right\rfloor$ and*

$$A(n, q, \left\lfloor \frac{\left\lfloor \sqrt{N} \right\rfloor}{2} \right\rfloor) \ge \left(\frac{1}{6\sqrt{3}\pi} - \varepsilon \right) \eta(1-\eta) N^{\frac{3}{4}}.$$

From now on we denote $\left\lfloor \sqrt{N} \right\rfloor$ by Q, $\left\lfloor \frac{\lfloor \sqrt{N} \rfloor}{2} \right\rfloor$ by Q_1 and $e^{2\pi i \alpha}$ by $e(\alpha)$. The basis of the proof of Theorem 2.1 is the following result [4] which was the main tool for Roth, too.

Theorem 2.2. *Let M be a positive integer, $R \ge 2$ integer, $R_1 = \left\lfloor \frac{R}{2} \right\rfloor$, $s_1, s_2, ..., s_M$ are complex numbers and $s_j = 0$ if $j \le 0$ or $j > M$. We denote $\sum_{i=0}^{k-1} s_{n+iq}$ by $D(n, q, k)$. Then there exist integers n and q with $1 \le q < R$ and*

$$|D(n, q, R_1)| \ge \frac{2}{\pi} R_1 \sqrt{\sum_{m=1}^{M} |s_m|^2} \left(M + \frac{R^2}{4} \right)^{-\frac{1}{2}} R^{-\frac{1}{2}}.$$

Theorem 2.2 is a corollary of the following inequality:

$$\sum_{q=1}^{R} \sum_{n=1-(R_1-1)q}^{M} |D(n, q, R_1)|^2 \ge \left(\frac{2}{\pi} R_1 \right)^2 \sum_{m=1}^{M} |s_m|^2 .$$

We only give a short outline of the proof, details are left to the reader. Denote $S(\alpha) = \sum_{n=1}^{M} s_n e(n\alpha)$ and $H_M(\alpha) = \sum_{n=1}^{M-1} e(n\alpha)$. $|H_M(\alpha)| = \left| \frac{\sin(\pi M \alpha)}{\sin(\pi \alpha)} \right|$ and from that it is easy to prove that $|H_M(\alpha)| \ge \frac{2}{\pi} M$ if $|\alpha| \le \frac{1}{2M}$. It is well-known that for every α there exist positive integers p and q with $(p, q) = 1$, p and $q \le R$ and $\left| \alpha - \frac{p}{q} \right| \le \frac{1}{qR}$ (so that $|q\alpha - p| \le \frac{1}{2R_1}$). From this and the previous inequality we have $\sum_{q=1}^{R} |H_{R_1}(q\alpha)|^2 \ge \left(\frac{2}{\pi} R_1 \right)^2$. Now consider $J = \int_0^1 |S(\alpha)|^2 \sum_{q=1}^{R} |H_{R_1}(q\alpha)|^2 \, d\alpha$. Clearly

$$J \ge \min_{\alpha} \sum_{q=1}^{R} |H_{R_1}(q\alpha)|^2 \int_0^1 |S(\alpha)|^2 \ge \left(\frac{2}{\pi} R_1 \right)^2 \sum_{m=1}^{M} |s_m|^2 .$$

On the other hand by Parseval's formula, $J = \sum_{q=1}^{R} \sum_{n=1-(R_1-1)q}^{M} |D(n, q, R_1)|^2$. From this the desired inequality and the theorem follows.

We also need the following lemma:

Lemma 2.3. *If $N \geq 8$ then*
$$\sum_{i=1}^{2N} \left(f(i) - \eta^2 g(i)\right)^2 \geq \frac{1}{48} \left(\eta \left(1-\eta\right) N\right)^2 .$$

Proof of Lemma 2.3. If $\eta = 0$ or 1 (i.e. $\eta\,(1-\eta) = 0$) then $\sum\limits_{i=1}^{2N} \left(f(i) - \eta^2 g(i)\right)^2 = 0$. If $|\mathcal{A}| = 1$ (so that $\eta = \frac{1}{N}$) and $\mathcal{A} = \{a\}$ then
$$\sum_{i=1}^{2N} \left(f(i) - \eta^2 g(i)\right)^2 \geq \left(f(2a) - \eta^2 g(2a)\right)^2 \geq \left(1 - \frac{1}{N^2} N\right)^2 = \left(\eta \left(1-\eta\right) N\right)^2 .$$
If $|\mathcal{A}| = N-1$ (so that $\eta = 1 - \frac{1}{N}$) and $\mathcal{A} = \{1, 2, \dots, N\} \setminus \{a\}$ then $f(a) = g(a) = a - 1$ and $f(N+a+1) = g(N+a+1) = \max\left(N-a-1, 0\right)$. Clearly $\max(g(a), g(N+a+1)) \geq \frac{N}{3}$ thus
$$\sum_{i=1}^{2N} \left(f(i) - \eta^2 g(i)\right)^2 \geq \left(\frac{N}{3} - \eta^2 \frac{N}{3}\right)^2 = \frac{1}{9}\left(\left(1-\eta^2\right) N\right)^2 \geq \frac{1}{9} \left(\eta \left(1-\eta\right) N\right)^2 .$$
Therefore we can suppose $2 \leq |\mathcal{A}| \leq N-2$.
Let $F(\alpha) = \sum\limits_{i=1}^{2N} f(i) e(i\alpha)$ and $G(\alpha) = \sum\limits_{i=1}^{2N} \eta^2 g(i) e(i\alpha)$. From the definitions of $f(n)$ and $g(n)$ it follows that $F(\alpha) = \left(\sum\limits_{a \in \mathcal{A}} e(a\alpha)\right)^2$ and $G(\alpha) = \eta^2 \left(\sum\limits_{i=1}^{N} e(i\alpha)\right)^2 = \eta^2 \left(\frac{e((N+1)\alpha) - e(\alpha)}{e(\alpha)-1}\right)^2$.
Clearly,
$$\sum_{i=1}^{2N} \left(f(i) - \eta^2 g(i)\right)^2 = \int_0^1 \left|F(\alpha) - G(\alpha)\right|^2 d\alpha. \tag{2.1}$$
From the Cauchy-Schwarz inequality we have
$$\sqrt{\int_0^1 \left|F(\alpha) - G(\alpha)\right|^2 d\alpha} \geq \int_0^1 \left|F(\alpha) - G(\alpha)\right| d\alpha. \tag{2.2}$$
$2 \geq \left|e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1\right|$ therefore
$$4 \int_0^1 \left|F(\alpha) - G(\alpha)\right| d\alpha \geq \int_0^1 \left|F(\alpha) - G(\alpha)\right| \left|e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1\right|^2 d\alpha. \tag{2.3}$$
Now
$$\left|F(\alpha) - G(\alpha)\right| \left|e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1\right|^2 \geq \left|F(\alpha)(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1)^2\right| - \left|G(\alpha)(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1)^2\right|.$$
But
$$\int_0^1 \left|G(\alpha)(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1)^2\right| d\alpha = \int_0^1 \eta^2 \left|\left(\frac{e((N+1)\alpha) - e(\alpha)}{e(\alpha)-1}\right)^2 \left(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1\right)^2\right| d\alpha =$$

$$= \eta^2 \int_0^1 \left| \frac{e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1}{e(\alpha) - 1} \left(e((N+1)\alpha) - e(\alpha) \right) \right|^2 d\alpha = \eta^2 \int_0^1 \left| \sum_{i=0}^{\left\lfloor \frac{N}{2} \right\rfloor - 1} e(i\alpha) \right|^2 \left| e((N+1)\alpha) - e(\alpha) \right|^2 d\alpha =$$

$$= \eta^2 \int_0^1 \left| - \sum_{i=1}^{\left\lfloor \frac{N}{2} \right\rfloor} e(i\alpha) + \sum_{i=N+1}^{N+\left\lfloor \frac{N}{2} \right\rfloor} e(i\alpha) \right|^2 d\alpha = 2\eta^2 \left\lfloor \frac{N}{2} \right\rfloor \leq \eta^2 N.$$

Therefore

$$\int_0^1 |F(\alpha) - G(\alpha)| \left| e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right|^2 d\alpha \geq \int_0^1 \left| F(\alpha) \left(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right)^2 \right| d\alpha - \eta^2 N.$$

Clearly,

$$F(\alpha) \left(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right)^2 = \left(\left(\sum_{a \in \mathcal{A}} e(a\alpha) \right) \left(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right) \right)^2 = \left(\sum_{i=1}^{i=N+\left\lfloor \frac{N}{2} \right\rfloor} a_i e(i\alpha) \right)^2$$

where

$$a_i = \begin{cases} -1 & \text{if} \quad i \in \mathcal{A} \text{ and } i - \left\lfloor \frac{N}{2} \right\rfloor \notin \mathcal{A} \\ \ \ 1 & \text{if} \quad i \notin \mathcal{A} \text{ and } i - \left\lfloor \frac{N}{2} \right\rfloor \in \mathcal{A} \\ \ \ 0 & \quad\ \text{otherwise} \end{cases}.$$

$\int_0^1 \left| F(\alpha) \left(e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right)^2 \right| d\alpha = \sum_{i=1}^{N+\left\lfloor \frac{N}{2} \right\rfloor} |a_i|^2 = \sum_{i=1}^{N+\left\lfloor \frac{N}{2} \right\rfloor} |a_i|$, therefore the value of the integral equals the number of integers i for which $a_i \neq 0$. If $1 \leq i \leq \left\lfloor \frac{N}{2} \right\rfloor$ and $i \in \mathcal{A}$ then $a_i = -1$ because $i - \left\lfloor \frac{N}{2} \right\rfloor \leq 0$ and hence $i - \left\lfloor \frac{N}{2} \right\rfloor \notin \mathcal{A}$. Similarly, if $N - \left\lfloor \frac{N}{2} \right\rfloor < i \leq N$ and $i \in \mathcal{A}$ then $a_{i+\left\lfloor \frac{N}{2} \right\rfloor} = 1$. If $i \in \mathcal{A}$ then $1 \leq i \leq \left\lfloor \frac{N}{2} \right\rfloor$ or $N - \left\lfloor \frac{N}{2} \right\rfloor < i \leq N$ and if N is odd i may be equal to $\left\lfloor \frac{N}{2} \right\rfloor + 1$, too. From the previous arguments it follows that the value of the integral is at least $|\mathcal{A}| - 1 = \eta N - 1$ and therefore

$$\int_0^1 |F(\alpha) - G(\alpha)| \left| e(\left\lfloor \frac{N}{2} \right\rfloor \alpha) - 1 \right|^2 d\alpha \geq \eta N - \eta^2 N - 1 = \eta \left(1 - \eta\right) N - 1.$$

$2 \leq |\mathcal{A}| \leq N - 2$ and $N \geq 8$ hence $\eta \left(1 - \eta\right) N \geq \frac{2(N-2)}{N} \geq \frac{3}{2}$ and $\eta \left(1 - \eta\right) N - 1 \geq$ $\geq \frac{1}{3} \eta \left(1 - \eta\right) N$. Thus, according to (2.1), (2.2) and (2.3), the statement of Lemma 2.3 follows.

Proof of Theorem 2.1. We will use Theorem 2.2. Let $M = 2N$, $s_j = f(j) - \eta^2 g(j)$ if $1 \leq j \leq M$, and $R = Q = \left\lfloor \sqrt{N} \right\rfloor$. Then there exist integers n and q with $1 \leq q < Q$ and

$$|A(n, q, Q_1)| \geq \frac{2}{\pi} Q_1 \sqrt{\sum_{m=1}^{2N} |s_m|^2} \left(2N + \frac{Q^2}{4} \right)^{-\frac{1}{2}} Q^{-\frac{1}{2}}.$$

From Lemma 2.3 $\sqrt{\sum\limits_{m=1}^{2N} |s_m|^2} \geq \frac{1}{4\sqrt{3}}\eta\,(1-\eta)\,N$ if $N \geq 8$. Thus

$$\frac{2}{\pi}Q_1\sqrt{\sum_{m=1}^{2N} |s_m|^2}\left(2N+\frac{Q^2}{4}\right)^{-\frac{1}{2}} Q^{-\frac{1}{2}} \geq \left(\frac{1}{6\sqrt{3}\pi}-\varepsilon\right)\eta(1-\eta)N^{\frac{3}{4}}$$

if $N \geq N_\varepsilon$ and Theorem 2.1 follows. (Indeed, the proof also gives that the inequality in Theorem 2.1 follows in a mean square sense.)

Theorem 2.1 gave a lower estimate for $\max\limits_{n,1\leq q\leq Q} A(n,q,Q_1)$. The aim of the following theorem is to show how far we can go with the lower estimate. It states that for every sufficiently large positive integer N there exists $\mathcal{A} \subseteq \{1,2,\ldots,N\}$ (with $\eta \approx \frac{1}{2}$) for which $A(n,q,Q_1) \leq cN$ if n and q are integers with $1 \leq q \leq Q$ and for "most of the pairs" n and q $A(n,q,Q_1) \leq c_1 N^{\frac{5}{6}} \log N$ (and the constants are absolute). For the proof we adapt Sárközy's construction used for an upper estimate related to Roth's problem [1] .

Theorem 2.4. *Let N be a positive integer, $Q = \left\lfloor\sqrt{N}\right\rfloor$, $Q_1 = \left\lfloor\frac{Q}{2}\right\rfloor$. Let p be a prime number with $N^{\frac{2}{3}} < p \leq 2N^{\frac{2}{3}}$. We construct an $\mathcal{A}$ subset of $\{1,2,\ldots,N\}$ the following way: $a \in \mathcal{A}$ exactly if $a \in \{1,2,\ldots,N\}$ and $\left(\frac{a}{p}\right) = 1$ where $\left(\frac{y}{p}\right)$ is the Legendre symbol (with $\left(\frac{y}{p}\right) = 0$ if $y \equiv 0 \mod p$).*

For this subset $\left|\eta - \frac{1}{2}\right| \leq c_1 N^{-\frac{2}{3}} \log N$ and $A(n,q,Q_1) \leq c_2 N$ for each integer n and q with $1 \leq q \leq Q$ (the constants are absolute).

Also if $N > N_\varepsilon$ then for at least $(1-\varepsilon)$ percentage of the pairs (n,q) with $1 \leq q \leq$ $\leq Q$ and $1-(Q-1)\,q \leq n \leq N$ we have $A(n,q,Q_1) \leq c_3 N^{\frac{5}{6}} \log N$ (with an absolute constant c_3).

The following lemma of Mauduit and Sárközy [3] (a corollary of A. Weil's theorem [6]) will be used several times during the proof of Theorem 2.4:

Theorem 2.5. *If p is a prime number, $h(x) \in F_p[x]$ is a polynomial of degree k such that it is not of the form $b\,(h_1(x))^2$ with $b \in F_p$ and $h_1(x) \in F_p[x]$, and X, Y are integers with $0 < Y \leq p$ then*

$$\left|\sum_{X<n\leq X+Y}\left(\frac{h(n)}{p}\right)\right| < 9k\sqrt{p}\log p.$$

Lemma 2.6. *For the set $\mathcal{A}$ in Theorem 2.4: $\left|\eta - \frac{1}{2}\right| \leq c_1 N^{-\frac{2}{3}} \log N$.*

Proof of Lemma 2.6. It is well-known that if $x \geq 1$ then there exists a prime p with $x < p \leq 2x$. This proves the existence of a proper prime for our construction.

Consider the following sets: $B_j = \left\{i \,\middle|\, \left(\frac{i}{p}\right) = j,\ 1 \leq i \leq N\right\}$ with $j = -1, 0, 1$. Clearly $|B_0| = \left\lfloor\frac{N}{p}\right\rfloor$, $B_1 = \mathcal{A}$ and $\sum\limits_{i=1}^{N}\left(\frac{i}{p}\right) = |B_1| - |B_{-1}| = 2\eta N - N + \left\lfloor\frac{N}{p}\right\rfloor$. With $\sum\limits_{i=a}^{a+p}\left(\frac{i}{p}\right) = 0$ and using Theorem 2.5 (or the Pólya-Vinogradov inequality) we have

$\left|\sum\limits_{i=1}^{N}\left(\frac{i}{p}\right)\right| = \left|\sum\limits_{i=\left\lfloor\frac{N}{p}\right\rfloor p}^{N}\left(\frac{i}{p}\right)\right| \le 9\sqrt{p}\log p$. Thus $|2\eta N - N| \le 9\sqrt{p}\log p + \left\lfloor\frac{N}{p}\right\rfloor$ and

$$\left|\eta - \frac{1}{2}\right| \le \frac{1}{2N}\left(9\sqrt{p}\log p + \left\lfloor\frac{N}{p}\right\rfloor\right).$$

With $N^{\frac{2}{3}} < p \le 2N^{\frac{2}{3}}$ Lemma 2.6 follows.

Lemma 2.7. *If* $2 \le n \le 2N$ *then* $\left|f(n) - \sum\limits_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| \le cN^{\frac{1}{3}}$ *where* $x_0 = 1$ *and* $x_1 = n-1$ *if* $2 \le n \le N$ *and* $x_0 = n-N$ *and* $x_1 = N$ *if* $N+1 \le n \le 2N$. *(The constant c is absolute.)*

Proof of Lemma 2.7. First let n be fixed with $2 \le n \le N$. Then $f(n) = \sum_{\substack{a_1+a_2=n\\ a_1,a_2\in\mathcal{A}}} 1 =$
$= \sum\limits_{k=1}^{n-1}\alpha_k$ where $\alpha_k = 1$ if $k, n-k \in \mathcal{A}$ and $\alpha_k = 0$ otherwise.
Let $C_1 = \{k\,|k$ or $n-k \equiv 0 \mod p$, $1 \le k \le n-1\}$ and $C = \{1, 2, \dots, n-1\} \setminus C_1$. Then $f(n) = \sum\limits_{k\in C}\alpha_k + \sum\limits_{k\in C_1}\alpha_k$. If $k \in C$ then $\left(\frac{k}{p}\right) = \pm 1$ and $\alpha_k = 1$ if and only if $\left(\frac{k}{p}\right) = = \left(\frac{n-k}{p}\right) = 1$. Hence it is easy to see that if $k \in C$ then $\alpha_k = \left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)$and if $k \in C_1$ then $\alpha_k = 0$ and $\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right) = \frac{\left(\frac{n}{p}\right)+1}{4}$.
Therefore

$$\left|f(n) - \sum_{k=1}^{n-1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| = \left|\sum_{k\in C_1}\left(\alpha_k - \left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right)\right| =$$

$$= |C_1|\left|\frac{\left(\frac{n}{p}\right)+1}{4}\right| \le 2\left\lfloor\frac{n-1}{p}\right\rfloor\left|\frac{\left(\frac{n}{p}\right)+1}{4}\right|.$$

Thus

$$\left|f(n) - \sum_{k=1}^{n-1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| \le \left\lfloor\frac{n-1}{p}\right\rfloor \le \frac{N}{p}$$

and by $N^{\frac{2}{3}} < p \le 2N^{\frac{2}{3}}$ Lemma 2.7 follows with $c = 1$.

If $N+1 \le n \le 2N$ then $f(n) = \sum_{\substack{a_1+a_2=n\\ a_1,a_2\in\mathcal{A}}} 1 = \sum\limits_{k=N-n}^{N}\alpha_k$ where $\alpha_k = 1$ if $k, n-k \in \mathcal{A}$ and $\alpha_k = 0$ otherwise. The proof can be finished similarly as in the previous case.

Lemma 2.8. *If* $2 \le n \le 2N$ *and* $p \nmid n$ *then*

$$\left|\frac{1}{4}g(n) - \sum_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| \le cN^{\frac{1}{3}}\log N$$

where $x_0 = 1$ *and* $x_1 = n-1$ *if* $2 \le n \le N$ *and* $x_0 = n-N$ *and* $x_1 = N$ *if* $N+1 \le n \le 2N$. *(The constant c is absolute.) If* $p \mid n$ *then*

$$\left|\frac{1}{4}g(n) - \sum_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| \le N.$$

Proof of Lemma 2.8. First let n be fixed with $2 \le n \le N$ and $p \nmid n$.
Then $g(n) = \sum_{\substack{1\le k\le N\\ 1\le n-k\le N}} 1 = n-1$, and

$$\frac{1}{4}g(n) - \sum_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right) = \frac{1}{4}(n-1) - \frac{1}{4}\sum_{k=1}^{n-1}\left(\frac{k(n-k)}{p}\right) - \frac{1}{2}\sum_{k=1}^{n-1}\left(\frac{k}{p}\right) - \frac{1}{4}(n-1) =$$

$$= -\frac{1}{4}\sum_{k=1}^{n-1}\left(\frac{k(n-k)}{p}\right) - \frac{1}{2}\sum_{k=1}^{n-1}\left(\frac{k}{p}\right).$$

It is well-known that if $(a,p)=1$ then $\sum\limits_{k=1}^{p}\left(\frac{k(a-k)}{p}\right) = -\left(\frac{-1}{p}\right)$ (cf. [5, Ch. 5.]). Also, if $p \nmid n$ then $x(n-x)$ is not of the form $b\,(h_1(x))^2$ in $F_p[x]$ and hence we can use Theorem 2.5. It follows that

$$\left|\sum_{k=1}^{n-1}\left(\frac{k(n-k)}{p}\right)\right| \le \left\lfloor\frac{n-1}{p}\right\rfloor + 18\sqrt{p}\log p \le \frac{N}{p} + 18\sqrt{p}\log p$$

From $\sum\limits_{k=a}^{a+p}\left(\frac{k}{p}\right) = 0$ and Theorem 2.5 $\left|\sum\limits_{k=1}^{n-1}\left(\frac{k}{p}\right)\right| \le 9\sqrt{p}\log p$. Using these inequalities and $N^{\frac{2}{3}} < p \le 2N^{\frac{2}{3}}$ the desired inequality follows in this case.

If n is an integer with $N+1 \le n \le 2N$ and $p \nmid n$ then $g(n) = \sum_{\substack{1\le k\le N\\ 1\le n-k\le N}} 1 = 2N-n+1$. The proof can be finished similarly as in the previous case.

If n is an integer with $1 \le n \le 2N$ and $p \mid n$ then $\sum\limits_{k=1}^{n-1}\left(\frac{k(n-k)}{p}\right) = \sum\limits_{k=1}^{n-1}\left(\frac{-k^2}{p}\right) = (n-1)\left(\frac{-1}{p}\right)$ therefore we cannot use the previous proof. But from $0 \le \sum\limits_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right) \le N$ and $g(n) \le N$ we have $\left|\frac{1}{4}g(n) - \sum\limits_{k=x_0}^{x_1}\left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| \le N$, and this concludes the proof of Lemma 2.8.

Proof of Theorem 2.4. Using Lemma 2.6 and $g(n) \le N$ we have

$$\left|\frac{1}{4}g(n) - \eta^2 g(n)\right| = g(n)\left|\frac{1}{2}+\eta\right|\left|\frac{1}{2}-\eta\right| \le N\frac{3}{2}c_1 N^{-\frac{2}{3}}\log N \le c_1' N^{\frac{1}{3}}\log N$$

Now from Lemma 2.7 and 2.8 it follows that if $2 \le n \le 2N$ and $p \nmid n$ then $\left|f(n) - \eta^2 g(n)\right| \le\ \le cN^{\frac{1}{3}}\log N$ (the constant c is absolute) and if $p \mid n$ then $\left|f(n) - \eta^2 g(n)\right| \le N$. It is also clear, that if $n < 2$ or $2N < n$ then $f(n) = g(n) = 0$. First let n and q be fixed with $1 \le q \le Q$. Since $Q_1 \le \frac{\sqrt{N}}{2}$ and p is a prime with $N^{\frac{2}{3}} < p$, from the integers $n, n+q, \dots, n+(Q_1-1)q$ at most one is divisible by p. Therefore

$$A(n,q,Q_1) = \left|\sum_{k=0}^{Q_1-1}\left(f(n+kq) - \eta^2 g(n+kq)\right)\right| \le cN^{\frac{1}{3}}Q\log N + N \le cN^{\frac{5}{6}}\log N + N.$$

This shows that $A(n,q,Q_1) \le c'N$ where the constant is absolute. But it also shows that if none of the integers $n, n+q, \dots, n+(Q_1-1)q$ are divisible by p then $A(n,q,Q_1) \le cN^{\frac{5}{6}}\log N$.

If we consider all sequences of the form $n, n+q, \ldots, n+(Q_1-1)q$ with $1 \le q \le Q$ which intersect with $\{1,2,\ldots,N\}$ (these are the ones with $1-(Q-1)\,q \le n \le N$) it can be seen that at least $1-\frac{2Q}{p}$ percentage of them does not contain an element divisible by p. Since $\frac{2Q}{p} \le 4N^{-\frac{1}{3}} \le \varepsilon$ if $N \ge N_\varepsilon$ this proves the last statement of Theorem 2.5.

Remark 2.9. The author conjectures that Theorem 2.1 is sharp (or at least it is sharp apart from log power), i.e. for every sufficiently large positive integer N there exists an $\mathcal{A} \subseteq \{1,2,\ldots,N\}$ for which $\varepsilon < \frac{|\mathcal{A}|}{n} < 1-\varepsilon$ and $A(n,q,Q_1) \le cN^{\frac{3}{4}}$ (or $cN^{\frac{3}{4}}(\log N)^{c_1}$) if n and q are integers with $1 \le q \le Q$.

3. Irregularities of distribution of sums of residues

Let $\mathcal{A} \subseteq \{0,1,\ldots,q-1\}$ and $\eta = \frac{|\mathcal{A}|}{q}$. Denote $h(i) = \sum_{\substack{a_1+a_2\equiv i \mod q\\ a_1,a_2\in\mathcal{A}}} 1$. Clearly, $\frac{1}{q}\sum\limits_{i=0}^{q-1} h(i) = \eta^2 q$ thus if the sums a_1+a_2 are well-distributed modulo q then the value of $\max\limits_{0\le i\le q-1} \left|h(i)-\eta^2 q\right|$ is small. The following theorem and its corollary will give a lower estimation for this maximum.

Theorem 3.1. *For every $\mathcal{A} \subseteq \{0,1,\ldots,q-1\}$ (with $q \ge 2$)*

$$\sum_{i=0}^{q-1}\left(h(i)-\eta^2 q\right)^2 \ge \frac{q^3}{q-1}\eta^2(1-\eta)^2.$$

Corollary 3.2. *For every $\mathcal{A} \subseteq \{0,1,\ldots,q-1\}$ (with $q \ge 2$) there exists an integer i with $0 \le i \le q-1$ and $\left|h(i)-\eta^2 q\right| \ge \sqrt{q}\eta(1-\eta)$.*

Proof of Theorem 3.1. Denote $S(\alpha) = \sum\limits_{a\in\mathcal{A}} e(a\alpha)$. It is clear that if k is an integer then $\left(S(\frac{k}{q})\right)^2 = \sum\limits_{i=0}^{q-1} h(i)\,e(i\frac{k}{q})$. It is well-known that if $B(\alpha) = \sum\limits_{j=0}^{q-1} b_i e(j\alpha)$ then $\frac{1}{q}\sum\limits_{k=0}^{q-1}\left|B\left(\frac{k}{q}\right)\right|^2 = \sum\limits_{j=0}^{q-1}|b_j|^2$. With $b_j = h(j)-\eta^2 q$ $(0 \le j \le q-1)$ we have

$$\sum_{i=0}^{q-1}\left(h(i)-\eta^2 q\right)^2 = \frac{1}{q}\sum_{k=0}^{q-1}\left|B\left(\frac{k}{q}\right)\right|^2.$$

$B(0) = \sum\limits_{i=0}^{q-1} h(i)-\eta^2 q^2 = 0$ and if $1 \le k \le q-1$ then

$$B\left(\frac{k}{q}\right) = \sum_{j=0}^{q-1}\left(h(j)-\eta^2 q\right)e\left(j\frac{k}{q}\right) = \sum_{j=0}^{q-1} h(j)\,e\left(j\frac{k}{q}\right) - \eta^2 q\sum_{j=0}^{q-1} e\left(j\frac{k}{q}\right) =$$

$$= \sum_{j=0}^{q-1} h(j)\,e\left(j\frac{k}{q}\right) = \left(S(\frac{k}{q})\right)^2.$$

Thus $\sum\limits_{i=0}^{q-1}\left(h(i)-\eta^2 q\right)^2 = \frac{1}{q}\sum\limits_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^4$.

$S(0) = |\mathcal{A}| = \eta q$ therefore

$$\frac{1}{q}\sum_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^2 = \frac{1}{q}\sum_{k=0}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^2 - \frac{1}{q}\eta^2 q^2 = \eta q - \eta^2 q = q\eta(1-\eta).$$

Using Cauchy's inequality:

$$\sum_{i=0}^{q-1}\left(h(i)-\eta^2 q\right)^2 = \frac{1}{q}\sum_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^4 \geq \frac{1}{q(q-1)}\left(\sum_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^2\right)^2 =$$

$$= \frac{1}{q(q-1)}q^4\eta^2(1-\eta)^2 = \frac{q^3}{q-1}\eta^2(1-\eta)^2.$$

This completes the proof of Theorem 3.1.

Remark 3.3. As a generalization of Theorem 3.1 we may study the distribution of sums of the form of $\sum_{i=1}^{k} a_i$ in residues where $a_1, a_2, \ldots, a_k \in \mathcal{A}$ and $k \geq 2$ is a fixed integer. Denote $h_k(n) = \sum_{\substack{a_1+a_2+\ldots+a_k \equiv n \mod q \\ a_1,a_2,\ldots,a_k\in\mathcal{A}}} 1$. Clearly $\frac{1}{q}\sum_{i=0}^{q-1} h_k(i) = \eta^k q^{k-1}$ and using a similar argument as in Theorem 3.1 it can be proved that

$$\sum_{i=0}^{q-1}\left(h_k(i)-\eta^k q^{k-1}\right)^2 \geq \frac{q^{2k-1}}{(q-1)^{k-1}}\eta^k(1-\eta)^k.$$

The following theorems show how sharp Theorem 3.1 and its corollary are. Theorem 3.4 shows that Theorem 3.1 is sharp for an infinite number of integers and the next theorem proves that Corollary 3.2 is sharp apart from log power. (From this it also follows that Theorem 3.1 is sharp apart from log power for every integer.)

Theorem 3.4. *Let p be a prime. Consider the following subset $\mathcal{A}$ of $\{0,1,2,\ldots,p-1\}$: $a\in\mathcal{A}$ exactly if $a\in\{0,1,2,\ldots,p-1\}$ and $\left(\frac{a}{p}\right)=1$. Then for this subset we have*

$$\sum_{i=0}^{p-1}\left(h(i)-\eta^2 p\right)^2 \leq \frac{p^3}{p-1}\eta^2(1-\eta)^2 + p.$$

Proof of Theorem 3.4. Clearly $|\mathcal{A}| = \frac{p-1}{2}$, $\eta = \frac{1}{2}-\frac{1}{2p}$ and $\eta^2 p = \frac{(p-1)^2}{4p} = \frac{p-2}{4}+\frac{1}{4p}$. Since $\left(\frac{a}{p}\right)=\left(\frac{-1}{p}\right)\left(\frac{p-a}{p}\right)$ we have $h(0)=0$ if $\left(\frac{-1}{p}\right)=-1$ and $h(0)=\frac{p-1}{2}$ if $\left(\frac{-1}{p}\right)=1$. Thus $\left(h(0)-\eta^2 p\right)^2 \leq \left(\frac{p-1}{2}-\frac{(p-1)^2}{4p}\right)^2 = \frac{p^2}{16}-\frac{1}{8}+\frac{1}{16p^2}$.

Suppose $1\leq a\leq p-1$. We will show that

$$h(a) = \sum_{i=0}^{p-1}\frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4} - \frac{\left(\frac{a}{p}\right)+1}{2} = \sum_{\substack{0\leq i\leq p-1\\ i\neq 0,a}}\frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4}.$$

With $f(n) = \sum_{\substack{a_1+a_2=n\\ a_1,a_2\in\mathcal{A}}} 1$ it is clear that $h(a)=f(a)+f(a+p)$. i and $a-i$ are elements of $\mathcal{A}$ simultaneously if and only if $\frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4}=1$, and otherwise $\frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4} = = 0$ if $i\neq 0,a$. Similarly i and $a+p-i$ are elements of $\mathcal{A}$

simultaneously if and only if $\frac{((\frac{i}{p})+1)((\frac{a-i}{p})+1)}{4}=1$, and otherwise $\frac{((\frac{i}{p})+1)((\frac{a-i}{p})+1)}{4}=0$ if $i\neq 0,a$. Thus it follows that $\sum_{\substack{0\le i\le p-1\\ i\neq 0,a}}\frac{((\frac{i}{p})+1)((\frac{a-i}{p})+1)}{4}=f(a)+f(a+p)=h(a)$.

$\sum\limits_{i=0}^{p-1}\frac{((\frac{i}{p})+1)((\frac{a-i}{p})+1)}{4}-\frac{(\frac{a}{p})+1}{2}=\frac{1}{4}\sum\limits_{i=0}^{p-1}\left(\frac{i(a-i)}{p}\right)+\frac{1}{2}\sum\limits_{i=0}^{p-1}\left(\frac{i}{p}\right)+\frac{p-2}{4}-\frac{1}{2}\left(\frac{a}{p}\right)$. Since $\sum\limits_{i=0}^{p-1}\left(\frac{i}{p}\right)=0$ and $\sum\limits_{i=0}^{p-1}\left(\frac{i(a-i)}{p}\right)=-\left(\frac{-1}{p}\right)$ thus $h(a)=\frac{p-2}{4}-\frac{1}{4}\left(\frac{-1}{p}\right)-\frac{1}{2}\left(\frac{a}{p}\right)$ if $1\le a\le p-1$. Then $\left(h(a)-\eta^2p\right)^2=\left(\frac{p-2}{4}-\frac{1}{4}\left(\frac{-1}{p}\right)-\frac{1}{2}\left(\frac{a}{p}\right)-\frac{p-2}{4}-\frac{1}{4p}\right)^2\le 1$. Therefore

$$\sum_{i=0}^{p-1}\left(h(i)-\eta^2p\right)^2\le\frac{p^2}{16}-\frac{1}{8}+\frac{1}{16p^2}+p-1.$$

But $\frac{p^3}{p-1}\eta^2\left(1-\eta\right)^2+p=\frac{1}{16}p^2+\frac{17}{16}p-\frac{1}{16}-\frac{1}{16p}$ and from this the statement of the theorem follows.

Theorem 3.5. *Let $q>1$ be an integer and p a prime with $2q<p\le 4q$. Consider the following subset $\mathcal{A}$ of $\{1,2,\dots,q\}$: $a\in\mathcal{A}$ exactly if $a\in\{1,2,\dots,q\}$ and $\left(\frac{a}{p}\right)=1$ (this will give us a subset of the residue classes modulo q). Then for this subset we have $\left|\eta-\frac{1}{2}\right|\le 9\frac{\log(4q)}{\sqrt{q}}$ and*

$$\left|h(i)-\eta^2q\right|\le 50\sqrt{q}\log(4q)$$

for every $0\le i\le q-1$.

Proof of Theorem 3.5. It is known that there exists a prime p with $2q<p\le 4q$. From the definition of the set $\mathcal{A}$ it is easy to see that $\sum\limits_{i=1}^{q}\left(\frac{i}{p}\right)=q\eta-q(1-\eta)=q(2\eta-1)$ and by Theorem 2.5 (or the Pólya-Vinogradov inequality) we have

$$q\left|2\eta-1\right|=\left|\sum_{i=1}^{q}\left(\frac{i}{p}\right)\right|\le 9\sqrt{p}\log p\le 18\sqrt{q}\log(4q).$$

and thus

$$\left|\eta-\frac{1}{2}\right|\le 9\frac{\log(4q)}{\sqrt{q}}.$$

Let a be an integer with $1\le a\le q$. With the already used technique we can prove that $f(a)=\sum\limits_{i=1}^{a-1}\frac{((\frac{i}{p})+1)((\frac{a-i}{p})+1)}{4}=\frac{1}{4}\sum\limits_{i=1}^{a-1}\left(\frac{i(a-i)}{p}\right)+\frac{1}{2}\sum\limits_{i=1}^{a-1}\left(\frac{i}{p}\right)+\frac{a-1}{4}$. $1\le a<p$ thus $x(a-x)$ is not of the form $b\left(h_1(x)\right)^2$ in $F_p\left[x\right]$ therefore we can use Theorem 2.5 for the first sum, and clearly we can use it for the second one, too:

$$\left|\frac{1}{4}\sum_{i=1}^{a-1}\left(\frac{i(a-i)}{p}\right)+\frac{1}{2}\sum_{i=1}^{a-1}\left(\frac{i}{p}\right)\right|\le\left(\frac{18}{4}+\frac{9}{2}\right)\sqrt{p}\log p\le 18\sqrt{q}\log(4q).$$

Hence $\left|f(a)-\frac{a-1}{2}\right|\le 18\sqrt{q}\log(4q)$.
Similarly

$$f(a+q)=\sum_{i=a}^{q}\frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{q+a-i}{p}\right)+1\right)}{4}=\frac{1}{4}\sum_{i=a}^{q}\left(\frac{i(q+a-i)}{p}\right)+\frac{1}{2}\sum_{i=a}^{q}\left(\frac{i}{p}\right)+\frac{q-a+1}{4}.$$

$1 \le a+q < p$ thus it can be shown similarly as before that $\left|f(a+q) - \frac{q-a+1}{4}\right| \le$ $\le 18\sqrt{q}\log(4q)$. If $1 \le i \le q$ then $h(i) = f(i) + f(q+i)$ (and $h(q) = h(0)$) and with the just proved inequalities $\left|h(i) - \frac{q}{4}\right| \le 36\sqrt{q}\log(4q)$ for every $1 \le i \le q$.
Clearly

$$\left|\frac{q}{4} - \eta^2 q\right| = \frac{q}{4}\,|2\eta-1|\,|2\eta+1| \le \frac{3}{4}q\,|2\eta-1| \le \frac{27}{2}\sqrt{q}\log(4q).$$

Therefore

$$\left|h(i) - \eta^2 q\right| \le \left|h(i) - \frac{q}{4}\right| + \left|\frac{q}{4} - \eta^2 q\right| \le \frac{99}{2}\sqrt{q}\log(4q)$$

for every $0 \le i \le q-1$ and this concludes the proof of Theorem 3.5 .

Acknowledgments. The author wishes to thank Prof. A. Sárközy for his helpful comments and for simplifying the proof of Theorem 3.1.

References

[1] P. Erdős and A. Sárközy, *Some solved and unsolved problems in combinatorial number theory,* Math. Slovaca 28 (1978), 407-421.
[2] J. Matoušek and J. Spencer, *Discrepancy in arithmetic progressions,* J. Amer. Math. Soc. 9 (1996), 195-204.
[3] C. Mauduit and A. Sárközy, *On finite binary sequences I: Measure of pseudorandomness, the Legendre symbol,* Acta Arith. 82 (1997), 365-377.
[4] K.F. Roth, *Remark concerning integer sequences,* Acta Arith. 9 (1964), 257-260.
[5] I. M. Vinogradov, *Elements of Number Theory*, Dover (1954).
[6] A. Weil, *Sur les courbes algébriques et les variétés qui en déduisent,* Actualités Sci. et Ind., no. 1041, Hermann, Paris (1948).

Valkó Benedek
1211 Budapest, Pezt F. u. 28.
valko@ludens.elte.hu