ACTA ARITHMETICA
XCII.4 (2000)

On irregularities of sums of integers

by

Benedek Valkó (Budapest)

1. Introduction. Let $\mathcal{A} \subseteq \{1, \dots, N\}$ and $\eta = |\mathcal{A}|/N$. 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 $\{1, \dots, N\}$ then it is reasonable to think that a well-distributed sequence would have "about" ηt elements in this sequence if k is "big enough" relative to N. Thus if we set

$$V(n,q,k) = \Big| \sum_{\substack{0 \le i < k \\ n+iq \in \mathcal{A}}} 1 - \eta \sum_{\substack{0 \le i < k \\ 1 \le n+iq \le N}} 1 \Big|$$

for a well-distributed sequence, then $V(n, q, \lfloor \lfloor \sqrt{N} \rfloor / 2 \rfloor)$ should be "small" for every integer $n, 1 \le q \le \lfloor \sqrt{N} \rfloor$. (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 $n, 1 \le q \le \lfloor \sqrt{N} \rfloor$ with $V(n, q, \lfloor \sqrt{N}/2 \rfloor) \ge c\sqrt{\eta(1-\eta)}N^{1/4}$ (and the constant c is absolute). This means that unless $\mathcal{A}$ is empty or equal to $\{1, \dots, N\}$ 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 $\mathcal{A} \subseteq \{1, \dots, N\}$ (with a "small" value of $|\eta - 1/2|$) for which $V(n, q, \lfloor \sqrt{N}/2 \rfloor) \le c'\sqrt{\eta(1-\eta)}N^{1/4}$ for every integer $n, 1 \le q \le \lfloor \sqrt{N} \rfloor$ [2].

In the first part of this paper we 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. Define

$$f(n) = \sum_{\substack{a_1 + a_2 = n \\ a_1, a_2 \in \mathcal{A}}} 1 \quad \text{and} \quad g(n) = \sum_{\substack{1 \le k \le N \\ 1 \le n-k \le N}} 1.$$

If $\mathcal{A}$ were constructed using probability methods in a way that the probability of choosing each integer $k \in \{1, \dots, N\}$ in $\mathcal{A}$ were η (independently)

2000 *Mathematics Subject Classification*: Primary 11B25.

then the expected value of $f(n)$ would be "roughly" $\eta^2 g(n)$. Therefore it is reasonable to study

$$A(n,q,k) = \Big|\sum_{i=0}^{k-1}(f(n+iq) - \eta^2 g(n+iq))\Big|.$$

(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 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 $1 \le q \le \lfloor\sqrt{N}\rfloor$ and $A(n, q, \lfloor\lfloor\sqrt{N}\rfloor/2\rfloor) \ge c\eta(1-\eta)N^{3/4}$. Theorem 5 will give an upper estimate on 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: $\mathcal{A} \subseteq \{0, 1, \dots, N-1\}$. The sums $a_1 + a_2$ with $a_1, a_2 \in \mathcal{A}$ are well-distributed if the numbers of sums in each residue class are about the same. Theorem 9 and its corollary states that these sums cannot be well-distributed this way, and inequalities similar to those in the previous part can be shown. Theorems 11 and 12 show that these results are nearly best possible.

2. Irregularities of distribution of sums relative to arithmetic progressions. We use the notations introduced in the introduction.

THEOREM 1. *If $\varepsilon > 0$ and $N > N_\varepsilon$ then for every $\mathcal{A} \subseteq \{1, \dots, N\}$ there exist integers n and q with $1 \le q \le \lfloor\sqrt{N}\rfloor$ and*

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

From now on we denote $\lfloor\sqrt{N}\rfloor$ by Q, $\lfloor\lfloor\sqrt{N}\rfloor/2\rfloor$ by Q_1, and $e^{2\pi i\alpha}$ by $e(\alpha)$. The basis of the proof of Theorem 1 is the following result of [4] which was also the main tool for Roth.

THEOREM 2. *Let M be a positive integer, $R \ge 2$ integer, $R_1 = \lfloor R/2\rfloor$, let $s_1, \dots, s_M$ be complex numbers and $s_j = 0$ if $j \le 0$ or $j > M$. 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*

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

Theorem 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. Define

$$S(\alpha)=\sum_{n=1}^{M} s_n e(n\alpha) \quad \text{and} \quad H_M(\alpha)=\sum_{n=1}^{M-1} e(n\alpha).$$

Then $|H_M(\alpha)| = |\sin(\pi M\alpha)/\sin(\pi\alpha)|$ and from that it is easy to prove that $|H_M(\alpha)| \geq (2/\pi)M$ if $|\alpha| \leq 1/(2M)$. It is well known that for every α there exist positive integers p and q with $(p,q)=1$, p and $q \leq R$ and $|\alpha - p/q| \leq 1/(qR)$ (so that $|q\alpha - p| \leq 1/(2R_1)$). From this and the previous inequality we have

$$\sum_{q=1}^{R} |H_{R_1}(q\alpha)|^2 \geq \left(\frac{2}{\pi}R_1\right)^2.$$

Now consider

$$J=\int_0^1 |S(\alpha)|^2 \sum_{q=1}^{R} |H_{R_1}(q\alpha)|^2 \, d\alpha.$$

Clearly,

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

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 follow.

We also need the following lemma:

LEMMA 3. *If* $N \geq 8$ *then*

$$\sum_{i=1}^{2N} (f(i)-\eta^2 g(i))^2 \geq \frac{1}{48}(\eta(1-\eta)N)^2.$$

P r o o f. If $\eta = 0$ or 1 (i.e. $\eta(1-\eta)=0$) then

$$\sum_{i=1}^{2N} (f(i)-\eta^2 g(i))^2 = 0.$$

If $|\mathcal{A}|=1$ (so that $\eta = 1/N$) and $\mathcal{A}=\{a\}$ then

$$\sum_{i=1}^{2N} (f(i)-\eta^2 g(i))^2 \geq (f(2a)-\eta^2 g(2a))^2 \geq \left(1-\frac{1}{N^2}N\right)^2 = (\eta(1-\eta)N)^2.$$

If $|\mathcal{A}| = N-1$ (so that $\eta = 1-1/N$) and $\mathcal{A}=\{1,\dots,N\}\backslash\{a\}$ then $f(a) = g(a) = a-1$ and $f(N+a+1) = g(N+a+1) = \max(N-a-1,0)$.

Clearly,

$$\max(g(a), g(N+a+1)) \ge N/3$$

and thus

$$\sum_{i=1}^{2N}(f(i)-\eta^2 g(i))^2 \ge \left(\frac{N}{3}-\eta^2\frac{N}{3}\right)^2 = \frac{1}{9}((1-\eta^2)N)^2 \ge \frac{1}{9}(\eta(1-\eta)N)^2.$$

Therefore we can suppose $2 \le |\mathcal{A}| \le N-2$. Let

$$F(\alpha)=\sum_{i=1}^{2N} f(i)e(i\alpha) \quad \text{and} \quad G(\alpha)=\sum_{i=1}^{2N}\eta^2 g(i)e(i\alpha).$$

From the definitions of $f(n)$ and $g(n)$ it follows that

$$F(\alpha)=\Big(\sum_{a\in\mathcal{A}} e(a\alpha)\Big)^2$$

and

$$G(\alpha)=\eta^2\Big(\sum_{i=1}^{N} e(i\alpha)\Big)^2=\eta^2\left(\frac{e((N+1)\alpha)-e(\alpha)}{e(\alpha)-1}\right)^2.$$

Clearly,

$$\sum_{i=1}^{2N}(f(i)-\eta^2 g(i))^2=\int_0^1 |F(\alpha)-G(\alpha)|^2\,d\alpha. \tag{1}$$

From the Cauchy–Schwarz inequality we have

$$\Big(\int_0^1 |F(\alpha)-G(\alpha)|^2\,d\alpha\Big)^{1/2} \ge \int_0^1 |F(\alpha)-G(\alpha)|\,d\alpha. \tag{2}$$

As $2 \ge |e(\lfloor N/2\rfloor\alpha)-1|$, we get

$$4\int_0^1 |F(\alpha)-G(\alpha)|\,d\alpha \ge \int_0^1 |F(\alpha)-G(\alpha)||e(\lfloor N/2\rfloor\alpha)-1|^2\,d\alpha. \tag{3}$$

Now

$$|F(\alpha)-G(\alpha)||e(\lfloor N/2\rfloor\alpha)-1|^2 \ge |F(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2|-|G(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2|.$$

But

$$\begin{aligned}
&\int_0^1 |G(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2|\,d\alpha \\
&\qquad = \int_0^1 \eta^2 \left| \left(\frac{e((N+1)\alpha)-e(\alpha)}{e(\alpha)-1} \right)^2 (e(\lfloor N/2\rfloor\alpha)-1)^2 \right| d\alpha \\
&\qquad = \eta^2 \int_0^1 \left| \frac{e(\lfloor N/2\rfloor\alpha)-1}{e(\alpha)-1} (e((N+1)\alpha)-e(\alpha)) \right|^2 d\alpha \\
&\qquad = \eta^2 \int_0^1 \Big| \sum_{i=0}^{\lfloor N/2\rfloor-1} e(i\alpha) \Big|^2 |e((N+1)\alpha)-e(\alpha)|^2\,d\alpha \\
&\qquad = \eta^2 \int_0^1 \Big| -\sum_{i=1}^{\lfloor N/2\rfloor} e(i\alpha) + \sum_{i=N+1}^{N+\lfloor N/2\rfloor} e(i\alpha) \Big|^2 d\alpha = 2\eta^2\lfloor N/2\rfloor \le \eta^2 N.
\end{aligned}$$

Therefore

$$\int_0^1 |F(\alpha)-G(\alpha)||e(\lfloor N/2\rfloor\alpha)-1|^2\,d\alpha \ge \int_0^1 |F(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2|\,d\alpha - \eta^2 N.$$

Clearly,

$$\begin{aligned}
&F(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2 \\
&\qquad = \Big(\Big(\sum_{a\in\mathcal{A}} e(a\alpha) \Big) (e(\lfloor N/2\rfloor\alpha)-1) \Big)^2 = \Big(\sum_{i=1}^{N+\lfloor N/2\rfloor} a_i e(i\alpha) \Big)^2
\end{aligned}$$

where

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

We have

$$\int_0^1 |F(\alpha)(e(\lfloor N/2\rfloor\alpha)-1)^2|\,d\alpha = \sum_{i=1}^{N+\lfloor N/2\rfloor} |a_i|^2 = \sum_{i=1}^{N+\lfloor N/2\rfloor} |a_i|,$$

therefore the value of the integral equals the number of integers i for which $a_i \neq 0$. If $1 \le i \le \lfloor N/2\rfloor$ and $i\in\mathcal{A}$ then $a_i = -1$ because $i-\lfloor N/2\rfloor \le 0$ and hence $i-\lfloor N/2\rfloor \notin \mathcal{A}$. Similarly, if $N-\lfloor N/2\rfloor < i \le N$ and $i\in\mathcal{A}$ then $a_{i+\lfloor N/2\rfloor} = 1$. If $i\in\mathcal{A}$ then $1\le i\le \lfloor N/2\rfloor$ or $N-\lfloor N/2\rfloor < i \le N$ and if N is odd i may also be equal to $\lfloor N/2\rfloor + 1$. From the previous arguments it follows that the value of the integral is at least $|\mathcal{A}| - 1 = \eta N - 1$ and

therefore

$$\int_0^1 |F(\alpha)-G(\alpha)||e(\lfloor N/2\rfloor\alpha)-1|^2\,d\alpha \ge \eta N-\eta^2 N-1=\eta(1-\eta)N-1.$$

Since $2\le|\mathcal{A}|\le N-2$ and $N\ge 8$, we have $\eta(1-\eta)N\ge 2(N-2)/N\ge 3/2$ and $\eta(1-\eta)N-1\ge\frac{1}{3}\eta(1-\eta)N$. Thus, according to (1)–(3), the statement of Lemma 3 follows.

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

$$|A(n,q,Q_1)|\ge\frac{2}{\pi}Q_1\Big(\sum_{m=1}^{2N}|s_m|^2\Big)^{1/2}\Big(2N+\frac{Q^2}{4}\Big)^{-1/2}Q^{-1/2}.$$

From Lemma 3,

$$\Big(\sum_{m=1}^{2N}|s_m|^2\Big)^{1/2}\ge\frac{1}{4\sqrt{3}}\eta(1-\eta)N\quad\text{ if } N\ge 8.$$

Thus

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

if $N\ge N_\varepsilon$ and Theorem 1 follows. (Indeed, the proof also shows that the inequality in Theorem 1 is valid in a mean square sense.)

Theorem 1 gave a lower estimate for $\max_{n,\,1\le q\le Q}A(n,q,Q_1)$. 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 $\mathcal{A}\subseteq\{1,\dots,N\}$ (with $\eta\approx 1/2$) for which $A(n,q,Q_1)\le cN$ if n and q are integers with $1\le q\le Q$ and for "most of the pairs" n and q, $A(n,q,Q_1)\le c_1N^{5/6}\log N$ (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 4. *Let N be a positive integer, $Q=\lfloor\sqrt{N}\rfloor$, $Q_1=\lfloor Q/2\rfloor$. Let p be a prime number with $N^{2/3}<p\le 2N^{2/3}$. Define $\mathcal{A}\subset\{1,\dots,N\}$ by letting $a\in\mathcal{A}$ exactly if $a\in\{1,\dots,N\}$ 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 $y\equiv 0 \bmod p$). For this $\mathcal{A}$, $|\eta-1/2|\le c_1N^{-2/3}\log N$ and $A(n,q,Q_1)\le c_2N$ for all integers n and q with $1\le q\le Q$ (the constants are absolute). Also if $N>N_\varepsilon$ then for at least $(1-\varepsilon)\cdot 100\%$ of the pairs (n,q) with $1\le q\le Q$ and $1-(Q-1)q\le n\le N$ we have $A(n,q,Q_1)\le c_3N^{5/6}\log N$ (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 4:

THEOREM 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 $0 < Y \le p$ then*

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

LEMMA 6. *For the set $\mathcal{A}$ in Theorem 4 we have*

$$|\eta - 1/2| \le c_1 N^{-2/3}\log N.$$

P r o o f. It is well known that if $x \ge 1$ then there exists a prime p with $x < p \le 2x$. This proves the existence of a proper prime for our construction.

Consider the following sets:

$$B_j = \left\{ i \;\middle|\; \left(\frac{i}{p}\right) = j,\ 1 \le i \le N \right\} \quad \text{with } j = -1, 0, 1.$$

Clearly,

$$|B_0| = \left\lfloor \frac{N}{p} \right\rfloor, \quad B_1 = \mathcal{A}, \quad \sum_{i=1}^{N} \left(\frac{i}{p}\right) = |B_1| - |B_{-1}| = 2\eta N - N + \left\lfloor \frac{N}{p} \right\rfloor.$$

With $\sum_{i=a}^{a+p} \left(\frac{i}{p}\right) = 0$ and using Theorem 5 (or the Pólya–Vinogradov inequality) we have

$$\left| \sum_{i=1}^{N} \left(\frac{i}{p}\right) \right| = \left| \sum_{i=\lfloor N/p\rfloor p}^{N} \left(\frac{i}{p}\right) \right| \le 9\sqrt{p}\log p.$$

Thus $|2\eta N - N| \le 9\sqrt{p}\log p + \lfloor N/p \rfloor$ and

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

With $N^{2/3} < p \le 2N^{2/3}$ Lemma 6 follows.

LEMMA 7. *If $2 \le n \le 2N$ then*

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

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.)

P r o o f. First let n be fixed with $2 \le n \le N$. Then

$$f(n) = \sum_{\substack{a_1+a_2=n \\ a_1,a_2 \in \mathcal{A}}} 1 = \sum_{k=1}^{n-1} \alpha_k$$

where $\alpha_k = 1$ if $k, n-k \in \mathcal{A}$ and $\alpha_k = 0$ otherwise.

Define $C_1 = \{k \mid k \text{ or } n-k \equiv 0 \bmod p,\ 1 \le k \le n-1\}$ and $C = \{1, \dots, n-1\} \setminus C_1$. Then $f(n) = \sum_{k \in C} \alpha_k + \sum_{k \in C_1} \alpha_k$. 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

$$\begin{aligned}\left| f(n) - \sum_{k=1}^{n-1} \left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right| &= \left| \sum_{k \in C_1} \left(\alpha_k - \left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right)\right)\right| \\ &= |C_1| \left|\frac{\left(\frac{n}{p}\right)+1}{4}\right| \le 2\left\lfloor \frac{n-1}{p} \right\rfloor \left|\frac{\left(\frac{n}{p}\right)+1}{4}\right|.\end{aligned}$$

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 since $N^{2/3} < p \le 2N^{2/3}$ Lemma 7 follows with $c = 1$.

If $N+1 \le n \le 2N$ then

$$f(n) = \sum_{\substack{a_1+a_2=n \\ a_1,a_2 \in \mathcal{A}}} 1 = \sum_{k=N-n}^{N} \alpha_k$$

where $\alpha_k = 1$ if $k, n-k \in \mathcal{A}$ and $\alpha_k = 0$ otherwise. The proof can be finished as in the previous case.

Lemma 8. *If $2 \le n \le 2N$ and $p \nmid n$ then*

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

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*

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

P r o o f. 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

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

It is well known that if $(a,p)=1$ then $\sum_{k=1}^{p}\left(\frac{k(a-k)}{p}\right) = -\left(\frac{-1}{p}\right)$ (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 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 $\sum_{k=a}^{a+p}\left(\frac{k}{p}\right) = 0$ and Theorem 5,

$$\left|\sum_{k=1}^{n-1}\left(\frac{k}{p}\right)\right| \le 9\sqrt{p}\log p.$$

From these inequalities and $N^{2/3} < p \le 2N^{2/3}$ 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 as in the previous case.

If n is an integer with $1 \le n \le 2N$ and $p \mid n$ then

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

therefore we cannot use the previous proof. But from

$$0 \le \sum_{k=x_0}^{x_1} \left(\frac{\left(\frac{k}{p}\right)+1}{2}\right)\left(\frac{\left(\frac{n-k}{p}\right)+1}{2}\right) \le N$$

and $g(n) \le N$ we have

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

and this concludes the proof of Lemma 8.

Proof of Theorem 4. Using Lemma 6 and $g(n) \le N$ we have

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

Now from Lemmas 7 and 8 it follows that if $2 \le n \le 2N$ and $p \nmid n$ then $|f(n) - \eta^2 g(n)| \le cN^{1/3}\log N$ (the constant c is absolute) and if $p \mid n$ then $|f(n) - \eta^2 g(n)| \le N$. 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 $Q_1 \le \sqrt{N}/2$ and p is a prime with $N^{2/3} < p$, at most one of the integers $n, n+q, \dots, n+(Q_1-1)q$ is divisible by p. Therefore

$$\begin{aligned}A(n,q,Q_1) &= \Big|\sum_{k=0}^{Q_1-1}(f(n+kq) - \eta^2 g(n+kq))\Big| \\ &\le cN^{1/3}Q\log N + N \le cN^{5/6}\log N + N.\end{aligned}$$

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 $n, n+q, \dots, n+(Q_1-1)q$ is divisible by p then $A(n,q,Q_1) \le cN^{5/6}\log N$.

If we consider all sequences of the form $n, n+q, \dots, n+(Q_1-1)q$ with $1 \le q \le Q$ which intersect $\{1,\dots,N\}$ (these are the ones with $1-(Q-1)q \le n \le N$) it can be seen that at least $(1-2Q/p)\cdot 100\%$ of them do not contain an element divisible by p. Since $2Q/p \le 4N^{-1/3} \le \varepsilon$ if $N \ge N_\varepsilon$ this proves the last statement of Theorem 4.

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

3. Irregularities of distribution of sums of residues. Let $\mathcal{A} \subseteq \{0,1,\dots,q-1\}$ and $\eta = |\mathcal{A}|/q$. Set

$$h(i) = \sum_{\substack{a_1+a_2 \equiv i \bmod q \\ a_1,a_2 \in \mathcal{A}}} 1.$$

Clearly, $(1/q)\sum_{i=0}^{q-1} h(i) = \eta^2 q$ and thus if the sums $a_1 + a_2$ are well-distributed modulo q then the value of $\max_{0\le i\le q-1} |h(i) - \eta^2 q|$ is small. The following theorem and its corollary give a lower estimation for this maximum.

THEOREM 9. *For every $\mathcal{A} \subseteq \{0,1,\dots,q-1\}$ (with $q \ge 2$),*

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

COROLLARY 10. *For every $\mathcal{A} \subseteq \{0,1,\dots,q-1\}$ (with $q \ge 2$) there exists an integer i with $0 \le i \le q-1$ and $|h(i) - \eta^2 q| \ge \sqrt{q}\,\eta(1-\eta)$.*

Proof (of Theorem 9). Define $S(\alpha) = \sum_{a\in\mathcal{A}} e(a\alpha)$. It is clear that if k is an integer then

$$(S(k/q))^2 = \sum_{i=0}^{q-1} h(i) e(ik/q).$$

It is well known that if $B(\alpha) = \sum_{j=0}^{q-1} b_j e(j\alpha)$ then

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

With $b_j = h(j) - \eta^2 q$ $(0 \le j \le q-1)$ we have

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

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

$$\begin{aligned} B\left(\frac{k}{q}\right) &= \sum_{j=0}^{q-1} (h(j) - \eta^2 q) e\left(j\frac{k}{q}\right) = \sum_{j=0}^{q-1} h(j) e\left(j\frac{k}{q}\right) - \eta^2 q \sum_{j=0}^{q-1} e\left(j\frac{k}{q}\right) \\ &= \sum_{j=0}^{q-1} h(j) e\left(j\frac{k}{q}\right) = \left(S\left(\frac{k}{q}\right)\right)^2. \end{aligned}$$

Thus

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

As $S(0) = |\mathcal{A}| = \eta q$ we have

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

Using Cauchy's inequality we obtain

$$\begin{aligned}\sum_{i=0}^{q-1}(h(i)-\eta^2 q)^2 &= \frac{1}{q}\sum_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^4 \ge \frac{1}{q(q-1)}\left(\sum_{k=1}^{q-1}\left|S\left(\frac{k}{q}\right)\right|^2\right)^2 \\ &= \frac{1}{q(q-1)}q^4\eta^2(1-\eta)^2 = \frac{q^3}{q-1}\eta^2(1-\eta)^2.\end{aligned}$$

This completes the proof of Theorem 9.

Remark 2. As a generalization of Theorem 9 we may study the distribution of sums of the form of $\sum_{i=1}^{k} a_i$ in residues where $a_1, \dots, a_k \in \mathcal{A}$ and $k \ge 2$ is a fixed integer. Define

$$h_k(n) = \sum_{\substack{a_1+\dots+a_k \equiv n \bmod q \\ a_1,\dots,a_k \in \mathcal{A}}} 1.$$

Clearly,

$$\frac{1}{q}\sum_{i=0}^{q-1} h_k(i) = \eta^k q^{k-1}$$

and using a similar argument to that for Theorem 9 it can be proved that

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

The following theorems show how sharp Theorem 9 and its corollary are. Theorem 11 shows that Theorem 9 is sharp for an infinite number of integers and the next theorem proves that Corollary 10 is sharp up to a log power. (From this it also follows that Theorem 9 is sharp up to a log power for every integer.)

Theorem 11. *Let p be a prime. Define $\mathcal{A} \subset \{0,1,\dots,p-1\}$ by setting $a \in \mathcal{A}$ exactly if $a \in \{0,1,\dots,p-1\}$ and $\left(\frac{a}{p}\right) = 1$. For this subset we have*

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

Proof. Clearly, we have $|\mathcal{A}| = (p-1)/2$, $\eta = 1/2 - 1/(2p)$ and $\eta^2 p = (p-1)^2/(4p) = (p-2)/4 + 1/(4p)$. 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 $h(0) = (p-1)/2$ if $\left(\frac{-1}{p}\right) = 1$. Thus

$$(h(0) - \eta^2 p)^2 \le \left(\frac{p-1}{2} - \frac{(p-1)^2}{4p}\right)^2 = \frac{p^2}{16} - \frac{1}{8} + \frac{1}{16p^2}.$$

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

$$\begin{aligned} h(a) &= \sum_{i=0}^{p-1} \frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4} - \frac{\left(\frac{a}{p}\right)+1}{2} \\ &= \sum_{\substack{0\le i\le p-1 \\ i\ne 0,a}} \frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4}. \end{aligned}$$

With $f(n) = \sum_{a_1+a_2=n,\, a_1,a_2\in\mathcal{A}} 1$ it is clear that $h(a) = f(a) + f(a+p)$. Now, i and $a-i$ are both in $\mathcal{A}$ if and only if $\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)/4 = 1$, and otherwise $\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)/4 = 0$ if $i \ne 0, a$. Similarly i and $a+p-i$ are both in $\mathcal{A}$ if and only if $\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)/4 = 1$, and otherwise $\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)/4 = 0$ if $i \ne 0, a$. Thus it follows that

$$\sum_{\substack{0\le i\le p-1 \\ i\ne 0,a}} \frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4} = f(a) + f(a+p) = h(a).$$

We have

$$\begin{aligned} &\sum_{i=0}^{p-1} \frac{\left(\left(\frac{i}{p}\right)+1\right)\left(\left(\frac{a-i}{p}\right)+1\right)}{4} - \frac{\left(\frac{a}{p}\right)+1}{2} \\ &\qquad = \frac{1}{4}\sum_{i=0}^{p-1}\left(\frac{i(a-i)}{p}\right) + \frac{1}{2}\sum_{i=0}^{p-1}\left(\frac{i}{p}\right) + \frac{p-2}{4} - \frac{1}{2}\left(\frac{a}{p}\right). \end{aligned}$$

Since $\sum_{i=0}^{p-1}\left(\frac{i}{p}\right) = 0$ and $\sum_{i=0}^{p-1}\left(\frac{i(a-i)}{p}\right) = -\left(\frac{-1}{p}\right)$, we get

$$h(a) = \frac{p-2}{4} - \frac{1}{4}\left(\frac{-1}{p}\right) - \frac{1}{2}\left(\frac{a}{p}\right) \quad \text{if } 1 \le a \le p-1.$$

Then

$$(h(a) - \eta^2 p)^2 = \left(\frac{p-2}{4} - \frac{1}{4}\left(\frac{-1}{p}\right) - \frac{1}{2}\left(\frac{a}{p}\right) - \frac{p-2}{4} - \frac{1}{4p}\right)^2 \le 1.$$

Therefore

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

But

$$\frac{p^3}{p-1}\eta^2(1-\eta)^2+p=\frac{1}{16}p^2+\frac{17}{16}p-\frac{1}{16}-\frac{1}{16p}$$

and from this the statement of the theorem follows.

THEOREM 12. *Let $q>1$ be an integer and p a prime with $2q<p\le 4q$. Define $\mathcal{A}\subset\{1,\dots,q\}$ by setting $a\in\mathcal{A}$ exactly if $a\in\{1,\dots,q\}$ and $\left(\frac{a}{p}\right)=1$ (this gives a subset of the residue classes modulo q). For this subset we have $|\eta-1/2|\le 9\log(4q)/\sqrt{q}$ and*

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

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

P r o o f. 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

$$\sum_{i=1}^{q}\left(\frac{i}{p}\right)=q\eta-q(1-\eta)=q(2\eta-1)$$

and by Theorem 5 (or the Pólya–Vinogradov inequality) we have

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

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 technique already used we can prove that

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

Since $1\le a<p$, $x(a-x)$ is not of the form $b(h_1(x))^2$ in $F_p[x]$ therefore we can use Theorem 5 for the first sum, and clearly also for the second:

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

Hence

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

Similarly

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

As $1 \le a+q < p$, it can be shown as before that

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

If $1 \le i \le q$ then $h(i) = f(i) + f(q+i)$ (and $h(q) = h(0)$) and from the inequalities just proved, $|h(i) - q/4| \le 36\sqrt{q}\log(4q)$ 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

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

for every $0 \le i \le q-1$ and this concludes the proof of Theorem 12.

Acknowledgments. The author wishes to thank Prof. A. Sárközy for his helpful comments and for simplifying the proof of Theorem 9.

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 pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol*, Acta Arith. 82 (1997), 365–377.
[4] K. F. Roth, *Remark concerning integer sequences*, ibid. 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 s'en déduisent*, Actualités Sci. Indust. 1041, Hermann, Paris, 1948.

Department of Algebra and Number Theory
Eötvös University
Rákóczi u. 5
H-1088 Budapest, Hungary
E-mail: valko@ludens.elte.hu

Received on 10.12.1998 (3528)