ACTA ARITHMETICA
XCVI.2 (2000)

A mean value density theorem of additive number theory

by

Friedrich Roesler (München)

Let A be a finite set of integers and

$$A+A=\{a+b:a,b\in A\},\qquad A-A=\{a-b:a,b\in A\}$$

be the sum set and the difference set of A. We denote by

$$S(A)=|A+A|,\qquad D(A)=|A-A|$$

the cardinality of these sets.

There should be intrinsic connections between $A+A$ and $A-A$, for the nontrivial coincidences $a+b=a'+b'$ of sums are equivalent to the nontrivial coincidences $a-a'=b'-b$ of differences.

If A has k elements, then obviously

$$2k-1\le S(A)\le \binom{k+1}{2},\qquad 2k-1\le D(A)\le k^2-k+1.$$

If $A=\{1,\ldots,k\}$ or more generally if A is an arithmetic progression of k integers, then $S(A)=D(A)=2k-1$ and hence

$$\frac{D(A)}{S(A)}=1.$$

If the k elements of A form a sufficiently fast growing sequence, then there are no nontrivial coincidences and thus $S(A)=\binom{k+1}{2}$, $D(A)=k^2-k+1$, and

$$\frac{D(A)}{S(A)}=1+\left(1-\frac{2}{k}\right)\left(1-\frac{2}{k+1}\right)<2.$$

Nevertheless the general conjecture

$$1\le\frac{D(A)}{S(A)}<2 \tag{1}$$

2000 *Mathematics Subject Classification*: Primary 11B05.

is false. G. A. Freĭman and V. P. Pigarev [1] have constructed arbitrarily large sets A and A' such that

$$\frac{D(A)}{S(A)} > D(A)^{0.11} \quad \text{and} \quad \frac{D(A')}{S(A')} < D(A')^{-0.017}.$$

These sets are designed explicitly to violate (1) (comp. also [5]). But even natural born sets like $A_k = \{m^2 : 0 \le m < k\}$, $k = 1, 2, \ldots$, are far from obeying the estimate in (1): E. Landau's theorem [3, p. 643] on the number of integers $n \le x$ which have a representation as a sum of two squares, combined with a theorem of G. Tenenbaum [2, p. 29, Theorem 21(ii)] on the number of integers $n \le x$ having a divisor in the interval $]\sqrt{x}/2, \sqrt{x}\,]$, shows

$$\lim_{k\to\infty} \frac{D(A_k)}{S(A_k)} = \infty$$

for the sequence $A_\infty = (m^2)_{m\ge 0}$ of squares.

Here we will prove a mean value version of (1):

Theorem. *We have*

$$1 \le \frac{D(k,N)}{S(k,N)} < 2 \quad \textit{for } 1 \le k \le N,$$

with

$$S(k,N) := \sum_{A\subset\{0,1,\ldots,N-1\},\,|A|=k} S(A),$$

$$D(k,N) := \sum_{A\subset\{0,1,\ldots,N-1\},\,|A|=k} D(A).$$

Both the lower bound 1 as well as the upper bound 2 in the Theorem are best possible (Remark 3).

The computation of $S(k,N)$ is straightforward (Proposition 1), whereas the treatment of $D(k,N)$ (Propositions 2 and 3) is more delicate. The reason is as follows:

To calculate $S(k,N)$ we have to count the number of subsets A with k elements in $\{0, 1, \ldots, N-1\}$ such that $t \in A + A$ for given values t, i.e.

$$\sigma_t(k,N) := |\{A \subset \{0,1,\ldots,N-1\} : |A| = k,\ t \in A + A\}|.$$

Hence A is counted in $\sigma_t(k,N)$ if and only if A contains one of the sets

$$\{j, t-j\}, \quad 0 \le j \le t/2. \tag{2}$$

Concerning $D(k,N)$ we look at the number of subsets A such that $t \in A-A$, i.e.

$$\delta_t(k,N) := |\{A \subset \{0,1,\ldots,N-1\} : |A| = k,\ t \in A - A\}|.$$

A is counted in $\delta_t(k,N)$ if and only if A contains one of the sets

$$\{j, t+j\}, \quad 0 \le j \le N-1-t. \tag{3}$$

The sets in (2) are pairwise disjoint, and therefore $\sigma_t(k,N)$ is given by a simple combinatorial formula. But the sets in (3) may have nonempty intersections, and this complicates the computation of $\delta_t(k,N)$.

We restrain from developing an exact formula for $D(k,N)$. If k is not too small, then $D_0(k,N)$ in Proposition 3(2) is a fairly good approximation of $D(k,N)$; it is better than indicated by the error term $2\theta\binom{N+1}{k+1}^*$ (comp. Remark 2) and precisely small enough to prove the Theorem.

The technical computations in the proofs suggest introducing the coefficients

$$\binom{N}{k}^* := \binom{N}{k} - \begin{cases} 2^k\dbinom{M}{k} & \text{if } N = 2M, \\ 2^{k-1}\left(\dbinom{M}{k} + \dbinom{M+1}{k}\right) & \text{if } N = 2M+1. \end{cases}$$

A combinatorial interpretation of these numbers is given in Remark 1.

Repeatedly we will have to handle the cases "N even" and "N odd" separately. Then we write $N = 2M + \delta$, $0 \le \delta \le 1$.

The passage from the false estimate (1) to the mean value theorem "kills the arithmetic interest of the question" (J.-M. Deshouillers) which actual value is adopted by the quotient $D(A)/S(A)$ for a given set A. The estimate in the Theorem, and in some more detail the graph of the function $k \mapsto D(k,N)/S(k,N)$, $1 \le k \le N$ (comp. Remark 3), just describes an average density property of finite sets A. But perhaps it might serve as an intuitive clue in the examination of sets as to relative density of their sum and difference set.

If a growing sequence $A_\infty = (a_m)_{m\ge 0}$ of integers is very smooth, then, with $A_k = (a_m)_{0\le m<k}$, one may expect the sequence

$$\frac{D(A_k)}{S(A_k)}, \qquad k = 0, 1, 2, \ldots, \tag{4}$$

to converge. In the case of the squares $a_m = m^2$ it does, even if not to a value between 1 and 2. Similarly, if $a_m = \binom{m}{2}$, then the sequence of quotients in (4) seems to grow in principle, too. On the other hand, if $a_m = [m^{3/2}]$, then the quotients in (4) probably fall to the limit 1. But what kind of arithmetic properties or lack of such properties in A_∞ might cause the sequence $D(A_k)/S(A_k)$ to grow or to fall or to converge at all?

I am grateful to J.-M. Deshouillers for his comments regarding existing results related to this work.

We shall make use of the following combinatorial results:

LEMMA 1. *We have*

(1) $\sum_{j\ge 0}(-1)^j\binom{M}{j}\binom{2M-2j}{2M-k} = 2^k\binom{M}{k}$.

(2) $\sum_{j\ge 0}(-1)^j\binom{m}{j}\binom{N-2j}{k} = \sum_{j\ge 0}2^j\binom{m}{j}\binom{N-2m}{N-k-j}$

for $0\le m\le N/2$.

Proof. (1) Riordan [4, p. 37, line 10]; (2) from part (1) by induction on k and N. ∎

LEMMA 2. *Let* $N=2M+\delta$, $0\le\delta\le 1$.

(1) $\binom{N+2}{k+2}^* = \binom{N}{k+2}^* + 2\binom{N}{k+1}^* + \binom{N}{k}$.

(2) $\binom{N+1}{k+1}^* = \binom{N}{k+1}^* + \binom{N}{k}^* + \delta\cdot 2^{k-1}\binom{M}{k-1}$.

(3) $\binom{N+2}{k+2}^* = \binom{N+1}{k+2}^* + \binom{N}{k+1}^* + \binom{N}{k} - \delta\cdot 2^k\binom{M}{k}$.

(4) $\binom{N}{k}^* = \sum_{j\ge 0}(-1)^j\binom{M}{j+1}\binom{N-2-2j}{N-k}$.

(5) $\binom{2M}{k}^* = \sum_{j\ge 1}2^{k-2j}\binom{k-j}{j}\binom{M}{k-j}$.

(6) $4\binom{N}{k+2}^* + 2\binom{N+1}{k+1}^* \le (2N+3)\binom{N}{k} + 2^{k+2}\binom{M}{k+1}$.

Proof. (1)–(3) immediate; (4) from Lemma 1(1); (5) by induction on k and M; (6) from part (1) by induction on k and N. ∎

First we deal with the mean value $S(k,N)$ for the sum sets.

PROPOSITION 1. (1) *For* $1\le k\le N$ *and* $N=2M+\delta$, $0\le\delta\le 1$,

$$S(k,N) = (2N+1)\binom{N}{k} + 2^k\binom{M}{k} - 2\binom{N+1}{k+2}^* - 2\binom{N+2}{k+2}^*.$$

(2) *For* $k\ge 2, S(k,N)$ *satisfies the recursion*

$$S(k,N) = S(k,N-1) + S(k-1,N-2) + (2N-1)\binom{N-2}{k-2} + 2^{k-1}\binom{M-1}{k-1} - 2\binom{N-2}{k}^*.$$

Proof. (1) By definition of $\sigma_t(k,N)$ we have

$$S(k,N) = \sum_{t=0}^{2N-2} \sigma_t(k,N).$$

If $A \subset \{0,1,\ldots,N-1\}$ and $A' = \{N-1-a : a \in A\}$, then $N-1-i \in A+A$ if and only if $N-1+i \in A'+A'$. Hence

$$\sigma_{N-1-i}(k,N) = \sigma_{N-1+i}(k,N), \quad 0 \le i \le N-1,$$

and therefore

$$S(k,N) = 2\sum_{t=0}^{N-1} \sigma_t(k,N) - \sigma_{N-1}(k,N). \tag{5}$$

Next we compute $\sigma_t(k,N)$. If $t = 2m-1$ is odd, then

$$\sigma_{2m-1}(k,N) = \binom{N}{k} - \sum_{j\ge 0}(-1)^j \binom{m}{j}\binom{N-2j}{N-k}, \tag{6}$$
$$0 \le 2m-1 \le N-1,$$

since $A \subset \{0,1,\ldots,N-1\}$ with $|A| = k$ is counted in $\sigma_{2m-1}(k,N)$ if and only if A contains one of the m pairwise disjoint sets $\{0,2m-1\}$, $\{1,2m-2\},\ldots,\{m-1,m\}$.

If $t = 2m$ is even, then

$$\sigma_{2m}(k,N) = \binom{N}{k} - \sum_{j\ge 0}(-1)^j \binom{m}{j}\binom{N-1-2j}{N-1-k}, \quad 0 \le 2m \le N-1. \tag{7}$$

For $A \subset \{0,1,\ldots,N-1\}$ with $|A| = k$ is counted in $\sigma_{2m}(k,N)$ if and only if A contains one of the pairwise disjoint sets $\{0,2m\}$, $\{1,2m-1\},\ldots$, $\{m-1,m+1\}$, $\{m\}$.

Hence $\sigma_{2m}(k,N)$ counts all $\binom{N-1}{k-1}$ sets A with $m \in A$ and

$$\binom{N-1}{k} - \sum_{j\ge 0}(-1)^j \binom{m}{j}\binom{N-1-2j}{N-1-k}$$

sets A such that $m \notin A$.

Equations (6) and (7) and Lemma 1(1) yield in particular

$$\sigma_{N-1}(k,N) = \binom{N}{k} - 2^k\binom{M}{k}. \tag{8}$$

Finally (5)–(8) show

$$S(k,N) = 2\Big(\sum_{m=0}^{M-1+\delta} \sigma_{2m}(k,N) + \sum_{m=1}^{M} \sigma_{2m-1}(k,N)\Big) - \sigma_{N-1}(k,N)$$

$$= 2\sum_{m=0}^{M-1+\delta}\left(\binom{N}{k} - \sum_{j\ge 0}(-1)^j\binom{m}{j}\binom{N-1-2j}{N-1-k}\right)$$
$$+ 2\sum_{m=1}^{M}\left(\binom{N}{k} - \sum_{j\ge 0}(-1)^j\binom{m}{j}\binom{N-2j}{N-k}\right)$$
$$- \left(\binom{N}{k} - 2^k\binom{M}{k}\right)$$
$$= (2N+1)\binom{N}{k} + 2^k\binom{M}{k}$$
$$- 2\sum_{j\ge 0}(-1)^j\binom{M+\delta}{j+1}\binom{N-1-2j}{N-1-k}$$
$$- 2\sum_{j\ge 0}(-1)^j\binom{M+1}{j+1}\binom{N-2j}{N-k}$$
$$= (2N+1)\binom{N}{k} + 2^k\binom{M}{k} - 2\binom{N+1}{k+2}^{*} - 2\binom{N+2}{k+2}^{*}$$

by Lemma 2(4).

(2) Direct computation with Lemma 2(2, 3). ∎

Now we start to estimate the mean value $D(k,N)$ for the difference sets.

PROPOSITION 2. (1) *For* $2 \le k \le N$ *and* $1 \le t \le N-1$,

$$\delta_t(k,N) = \delta_t(k,N-1) + \delta_t(k-1,N-2) + \binom{N-2}{k-2} + E_t(k,N)$$

with the error term

$$E_t(k,N) = |\mathcal{B}_t| - |\mathcal{B}'_t|,$$

where

$$\mathcal{B}_t = \{B : \{1,t+1\} \subset B \subset \{1,\ldots,t-1,t+1,\ldots,N-1\},\ |B| = k-1,$$
$$t = (t+1)-1 \text{ is the only representation of } t \text{ in } B-B\},$$
$$\mathcal{B}'_t = \{B' : \{t,2t\} \subset B' \subset \{2,3,\ldots,N-1\},\ |B'| = k-1,$$
$$t = 2t-t \text{ is the only representation of } t \text{ in } B'-B'\}.$$

(2) *For* $1 \le t < N/2$,

$$E_t(k,N) = |\mathcal{C}_t| - |\mathcal{C}'_t|$$

with

$$\mathcal{C}_t = \{B \in \mathcal{B}_t : 3t \in B\}, \quad \mathcal{C}'_t = \{B' \in \mathcal{B}'_t : 2t+1 \in B'\}.$$

Proof. (1) We divide the sets $B \subset \{0,1,\ldots,N-1\}$ with $|B| = k$ into three classes:

(i) the sets B such that $0 \notin B$,

(ii) the sets B such that $0 \in B$ and $t \in B$,

(iii) the sets B such that $0 \in B$ and $t \notin B$.

The number of sets in (i) which are counted in $\delta_t(k,N)$ is $\delta_t(k,N-1)$. The $\binom{N-2}{k-2}$ sets in (ii) are all counted in $\delta_t(k,N)$. The fact that the sets in (iii) contain 0 is irrelevant because t is not in B. So we can cancel 0, and hence the number of sets in (iii) which are counted in $\delta_t(k,N)$ is equal to

$$|\{B \subset \{1,\ldots,t-1,t+1,\ldots,N-1\} : |B| = k-1,\ t \in B-B\}|. \tag{9}$$

Now the sets B in (9) are divided into two classes:

(i′) the sets B which have a representation $t = b-a$ with $a,b \in B-\{1\}$,

(ii′) the sets B which have no such representation, i.e. for which $t = (t+1)-1$ is the only representation of t in $B-B$.

The number of sets in (ii′) is $|\mathcal{B}_t|$ by definition. For the description of the number of sets in (i′) we use the map

$$\phi : \{1,\ldots,t-1,t+1,\ldots,N-1\} \to \{2,3,\ldots,N-1\},$$
$$\phi(1) := t, \quad \phi(x) := x \quad \text{otherwise.}$$

The bijectivity of ϕ carries over to the map

$$\begin{aligned}\Phi : \{B \subset \{1,\ldots,t-1,t+1,\ldots,N-1\} :&\\ |B| = k-1, \exists a,b \in B-\{1\} : b-a = t\}&\\ \to \{B' \subset \{2,3,\ldots,N-1\} : |B'| = k-1, \exists a',b' \in B'-\{t\} : b'-a' = t\},&\\ \Phi(B) := \{\phi(b) : b \in B\}.&\end{aligned}$$

Hence the number of sets in (i′) is

$$\begin{aligned}|\{B' \subset \{2,3,\ldots,N-1\} : |B'| = k-1,\ \exists a',b' \in B'-\{t\} : b'-a' = t\}|&\\ = \delta_t(k-1,N-2) - |\mathcal{B}'_t| \quad \text{by definition of } \mathcal{B}'_t.&\end{aligned}$$

Together we get the recursion formula for $\delta_t(k,N)$ with the error term $E_t(k,N) = |\mathcal{B}_t| - |\mathcal{B}'_t|$.

(2) For $1 \le t < N/2$ we use the bijective map

$$\psi : \{1,\ldots,t-1,t+1,\ldots,N-1\} \to \{2,3,\ldots,N-1\},$$
$$\psi(1) := t, \quad \psi(t+1) := 2t, \quad \psi(2t) := t+1, \quad \psi(x) := x \quad \text{otherwise}$$

and show:

$$\Psi : \mathcal{B}_t - \mathcal{C}_t \to \mathcal{B}'_t - \mathcal{C}'_t, \quad \Psi(B) := \{\psi(b) : b \in B\}, \text{ is bijective.} \tag{10}$$

Then part (1) and assertion (10) give at once

$$\begin{aligned}E_t(k,N) = |\mathcal{B}_t| - |\mathcal{B}'_t| &= |\mathcal{B}_t - \mathcal{C}_t| + |\mathcal{C}_t| - (|\mathcal{B}'_t - \mathcal{C}'_t| + |\mathcal{C}'_t|)\\ &= |\mathcal{C}_t| - |\mathcal{C}'_t|.\end{aligned}$$

For the proof of (10) we have to show:

(I) $B \in \mathcal{B}_t - \mathcal{C}_t$ implies $\Psi(B) \in \mathcal{B}'_t - \mathcal{C}'_t$,

(II) $B' \in \mathcal{B}'_t - \mathcal{C}'_t$ implies $\Psi^{-1}(B') \in \mathcal{B}_t - \mathcal{C}_t$.

(I) Let $B \in \mathcal{B}_t - \mathcal{C}_t$. Then $1, t+1 \in B$ and $2t+1, 3t \notin B$. Application of ψ for $B' := \Psi(B)$ shows that

$$t, 2t \in B' \quad \text{and} \quad 2t+1, 3t \notin B'.$$

Hence $t = 2t - t$ is a representation of t in $B' - B'$, and $B' \notin \mathcal{C}'_t$ because $2t+1 \notin B'$. It remains to show that $t = 2t - t$ is the only representation of t in $B' - B'$. So let $t = b' - a'$ be any representation of t with $a', b' \in B'$. Then

$$\{a', b'\} \cap \{t, t+1, 2t\} \neq \emptyset. \tag{11}$$

For otherwise a' and b' would be invariant under ψ^{-1}, and $t = \psi^{-1}(b') - \psi^{-1}(a')$ would be a representation of t in $B - B$ which is different from $t = (t+1) - 1$. But $a' \in \{t+1, 2t\}$ would imply $a' + t = b' \in B'$, and if $b' \in \{t, t+1\}$, then $b' - t = a' \in B'$, which both are impossible. Hence by (11), $a' = t$ or $b' = 2t$, which means $a' = t$ and $b' = 2t$ because $t = b' - a'$.

The proof of (II) is exactly the same. Just exchange

$$t+1 \leftrightarrow 2t, \quad 2t+1 \leftrightarrow 3t, \quad B, \mathcal{B}_t, \mathcal{C}_t, \psi, \Psi \leftrightarrow B', \mathcal{B}'_t, \mathcal{C}'_t, \psi^{-1}, \Psi^{-1}$$

everywhere and $1 \leftrightarrow t$ at the "right" places, i.e. where ψ is involved. ∎

Remark 1. The sets $\mathcal{B}_1$ and $\mathcal{B}'_1$ in Proposition 2(1) are identical, hence $E_1(k, N) = 0$, and then the recursion formula yields via induction

$$\delta_1(k, N) = \binom{N}{k} - \binom{N+1-k}{k}.$$

Similarly one can show

$$\delta_2(k, N) = \binom{N}{k} - \binom{N+1-k}{k} - \binom{N-1-k}{k-2} - \binom{N-3-k}{k-4},$$

with $\binom{m}{n} := 0$ if $m < 0$. Further,

$$\delta_M(k, 2M) = \binom{2M}{k}^* \quad \text{and} \quad \delta_{M+1}(k, 2M+1) = \binom{2M+1}{k}^*,$$

which furnishes a combinatorial interpretation of the coefficients $\binom{N}{k}^*$.

Presumably the sequences $(\delta_t(k, N))_{0 \le t < N}$ are almost decreasing. This is easy to show within the interval $N/2 \le t < N$, whereas in $0 \le t \le N/2$ there is at least the exception $\delta_{M-1}(M+1, 2M) < \delta_M(M+1, 2M)$, $M \ge 3$.

Now we develop a concept to estimate the error terms $E_t(k, N)$ of Proposition 2. Let A be a set with N elements. We arrange these elements in a *scheme*

$$\mathrm{Sch}(A;N) = (a_{ij})_{1\le j\le \ell(i),\,1\le i\le t} = \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & a_{1,\ell(1)} \\ a_{21} & a_{22} & \dots & \dots & a_{2,\ell(2)} \\ \vdots & & & & \\ a_{t1} & a_{t2} & \dots & \dots & a_{t,\ell(t)} \end{pmatrix}$$

consisting of t rows of possibly different lengths $\ell(i)$.

In particular, $\mathrm{Sch}(N/t)$ will denote the standard scheme

$$\mathrm{Sch}(N/t) = \begin{pmatrix} 0 & t & 2t & 3t & \dots \\ 1 & t+1 & 2t+1 & 3t+1 & \dots \\ 2 & t+2 & 2t+2 & 3t+2 & \dots \\ \vdots & & & & \\ t-1 & 2t-1 & 3t-1 & 4t-1 & \dots \end{pmatrix}$$

on the set $A = \{0, 1, \dots, N-1\}$.

Schemes $\mathrm{Sch}(A;N)$ with r rows of length 1 and all rows of length at most 2 will be denoted by $\mathrm{Sch}^*(A;N,r)$. For instance the standard scheme $\mathrm{Sch}(N/t)$ is of type $\mathrm{Sch}^*(\{0,1,\dots,N-1\};N,2t-N)$ if $(N-1)/2 < t \le N$.

Two schemes

$$\mathrm{Sch}_1(A;N) = (a_{ij})_{1\le j\le \ell(i),\,1\le i\le t} \quad \text{and} \quad \mathrm{Sch}_2(A';N) = (a'_{ij})_{1\le j\le \ell'(i),\,1\le i\le t}$$

with the same number N of elements and the same number t of rows are called *similar* if their rows have the same lengths, i.e. if there exists a permutation π on $\{1,\dots,t\}$ such that $\ell'(i) = \ell(\pi(i))$ for $1 \le i \le t$.

We call $\mathrm{Sch}_2(A;N)$ *finer* than $\mathrm{Sch}_1(A;N)$ if $\mathrm{Sch}_2(A;N)$ results from $\mathrm{Sch}_1(A;N)$ by dissection of a row $a_{i1}\cdots a_{i,\ell(i)}$ of $\mathrm{Sch}_1(A;N)$ into two rows $a_{i1}\cdots a_{i,m}$ and $a_{i,m+1}\cdots a_{i,\ell(i)}$. Further we require the relation "finer than" to be transitive.

The sets $B \in \mathcal{B}_t$ and $B' \in \mathcal{B}'_t$ in Proposition 2 have the following property:

Except 1 and $t+1$ (resp. t and $2t$), B and B' do not contain two numbers which are neighbours in any row of the standard scheme $\mathrm{Sch}(N/t)$. Hence generally a subset $B \subset A$ will be called *admissible* for a given scheme $\mathrm{Sch}(A;N)$ if and only if B does not contain two elements which are neighbours in any of the rows of $\mathrm{Sch}(A;N)$.

Our concern will be the cardinality of the sets

$$\mathcal{P}_k(\mathrm{Sch}(A;N)) := \{B \subset A : |B| = k, B \text{ admissible for } \mathrm{Sch}(A;N)\}.$$

LEMMA 3. (1) *If two schemes* Sch_1 *and* Sch_2 *are similar*, *then*

$$|\mathcal{P}_k(\mathrm{Sch}_1)| = |\mathcal{P}_k(\mathrm{Sch}_2)|.$$

(2) *If the scheme* Sch_2 *is finer than* Sch_1, *then*

$$\mathcal{P}_k(\mathrm{Sch}_2) \supset \mathcal{P}_k(\mathrm{Sch}_1).$$

(3) *In particular* $r_2 \geq r_1$ *implies*

$$|\mathcal{P}_k(\mathrm{Sch}^*(A; N, r_2))| \geq |\mathcal{P}_k(\mathrm{Sch}^*(A'; N, r_1))|.$$

Proof. Immediate consequences of the definitions. ∎

LEMMA 4. $E_t(k, N) = 0$ *if* $t \mid N - 1$.

Proof. Proposition 2 shows at once $\mathcal{B}_t = \mathcal{B}'_t$ for $t = 1$ and $t = N-1$, and $\mathcal{C}_t = \mathcal{C}'_t = \emptyset$ if N is odd and $t = (N-1)/2$. Thus let $2 \leq t = (N-1)/r \in \mathbb{N}$ and $r \geq 3$. We consider the standard scheme

$$\mathrm{Sch}(N/t) = \begin{pmatrix} 0 & t & 2t & 3t & 4t & \ldots & (r-1)t & rt \\ 1 & t+1 & 2t+1 & 3t+1 & 4t+1 & \ldots & (r-1)t+1 & \\ 2 & t+2 & 2t+2 & 3t+2 & 4t+2 & \ldots & (r-1)t+2 & \\ \vdots & & & & & & & \\ t-1 & 2t-1 & 3t-1 & 4t-1 & 5t-1 & \ldots & rt-1 & \end{pmatrix}$$

and apply Proposition 2(2): $\mathcal{C}_t$ contains all subsets B of $\{0, 1, \ldots, N-1\} - \{0, t\}$ with $|B| = k-1$ and $\{1, t+1, 3t\} \subset B$, which—except 1 and $t+1$—do not contain two neighbouring elements in any of the rows of $\mathrm{Sch}(N/t)$. In particular these sets B do not contain any of the numbers $2t$, $4t$, $2t+1$. Hence we cancel $0, t, 2t, 3t, 1, t+1, 2t+1$ and if possible $4t$ in $\mathrm{Sch}(N/t)$ and see: $|\mathcal{C}_t|$ counts the sets $B_0 = B - \{1, t+1, 3t\}$ with $|B_0| = k - 4$ which are admissible for the scheme

$$\mathrm{Sch}_1 = \begin{pmatrix} & & & & & 5t & \ldots & (r-1)t & rt \\ & & & 3t+1 & 4t+1 & 5t+1 & \ldots & (r-1)t+1 & \\ 2 & t+2 & 2t+2 & 3t+2 & 4t+2 & 5t+2 & \ldots & (r-1)t+2 & \\ \vdots & & & & & & & & \\ t-1 & 2t-1 & 3t-1 & 4t-1 & 5t-1 & 6t-1 & \ldots & rt-1 & \end{pmatrix}.$$

Similarly $\mathcal{C}'_t$ contains all subsets B' of $\{0, 1, \ldots, N-1\} - \{0, 1\}$ with $|B'| = k-1$ and $\{t, 2t, 2t+1\} \subset B'$, which—except t and $2t$—do not contain two neighbouring numbers in any of the rows of $\mathrm{Sch}(N/t)$. In particular, these sets B' do not contain $3t, t+1$, and $3t+1$. Hence $|\mathcal{C}'_t|$ counts the sets $B'_0 = B' - \{t, 2t, 2t+1\}$ with $|B'_0| = k - 4$ which are admissible for the scheme

$$\mathrm{Sch}_2=\begin{pmatrix} &&&&4t&\dots&(r-1)t&rt\\ &&&&4t+1&\dots&(r-1)t+1&\\ 2&t+2&2t+2&3t+2&4t+2&\dots&(r-1)t+2&\\ \vdots&&&&&&&\\ t-1&2t-1&3t-1&4t-1&5t-1&\dots&rt-1& \end{pmatrix}.$$

The first resp. second row of Sch_1 has the same length as the second resp. first row of Sch_2. All other rows of Sch_1 and Sch_2 coincide. Thus Sch_1 and Sch_2 are similar, and Lemma 3(1) asserts

$$E_t(k,N)=|\mathcal{C}_t|-|\mathcal{C}_t'|=|\mathcal{P}_{k-4}(\mathrm{Sch}_1)|-|\mathcal{P}_{k-4}(\mathrm{Sch}_2)|=0. \ \blacksquare$$

REMARK 2. A refinement of the argument in the proof of Lemma 4 shows

$$(-1)^r E_t(k,N)\ge 0 \quad \text{for} \quad \frac{N-1}{r+1}<t<\frac{N-1}{r},\ r=1,2,\dots$$

This change of signs in the error terms $E_t(k,N)$ makes it difficult to derive an upper bound for $\sum_{t=1}^{N-1} E_t(k,N)$ which would be essentially better than the one given in Lemma 7.

LEMMA 5. *We have*

(1) $E_t(k,N)=|\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,2t-N))|$ *for* $(N-1)/2<t<N-1$.

(2) $$\sum_{(N-1)/2<t<N} E_t(k,N)=\binom{N-2}{k-1}^{*}.$$

Proof. (1) For $(N-1)/2<t<N-1$ Proposition 2(1) shows $\mathcal{B}_t'=\emptyset$ and hence

$$E_t(k,N)=|\mathcal{B}_t|.$$

Again we consider the standard scheme $\mathrm{Sch}(N/t)$, which is now of type $\mathrm{Sch}^*(\{0,1,\dots,N-1\};N,2t-N)$.

$\mathcal{B}_t$ contains all subsets B of $\{0,1,\dots,N-1\}-\{0,t\}$ with $|B|=k-1$ and $\{1,t+1\}\subset B$, which—except 1 and $t+1$—do not contain two numbers in any of the rows of $\mathrm{Sch}(N/t)$. Hence $|\mathcal{B}_t|$ is the number of subsets $B_0=B-\{1,t+1\}$ of $A=\{0,1,\dots,N-1\}-\{0,t,1,t+1\}$ with $|B_0|=k-3$, admissible for the scheme $\mathrm{Sch}^*(A;N-4,2t-N)$, which results from $\mathrm{Sch}(N/t)$ by cancellation of the first two rows.

(2) Part (1) shows

$$E_t(k,N)=\sum_{j=0}^{k-3}2^j\binom{N-t-2}{j}\binom{2t-N}{k-3-j},\qquad (N-1)/2<t<N-1,$$

for if $B\in\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,2t-N))$ contains j elements out of the $N-t-2$ rows of length 2, for which there are $2^j\binom{N-t-2}{j}$ possibilities, then there are $\binom{2t-N}{k-3-j}$ possibilities left for the remaining $k-3-j$ elements of B in the $2t-N$ rows of length 1.

Hence with $N = 2M+\delta$, $0 \le \delta \le 1$, and in view of Lemma 4,

$$\begin{aligned}\sum_{(N-1)/2<t<N} E_t(k,N) &= \sum_{M+\delta\le t\le N-2}\ \sum_{j\ge 0} 2^j \binom{N-t-2}{j}\binom{2t-N}{k-3-j}\\ &= \sum_{0\le m\le M-2}\ \sum_{j\ge 0} 2^j \binom{m}{j}\binom{(N-4)-2m}{(N-4)-(N-k-1)-j}\\ &= \sum_{0\le m\le M-2}\ \sum_{j\ge 0}(-1)^j \binom{m}{j}\binom{N-4-2j}{N-k-1}\\ &\qquad\qquad\qquad\qquad\qquad\qquad\text{(by Lemma 1(2))}\\ &= \sum_{j\ge 0}(-1)^j \binom{M-1}{j+1}\binom{N-2-2(j+1)}{N-k-1}\\ &= \binom{N-2}{k-1} - \sum_{j\ge 0} 2^j \binom{M-1}{j}\binom{\delta}{k-1-j}\\ &\qquad\qquad\qquad\qquad\qquad\qquad\text{(by Lemma 1(2))}\\ &= \binom{N-2}{k-1}^{*}. \qquad ■\end{aligned}$$

Lemma 6. *We have*

$$\sum_{t=1}^{N-1} E_t(k,N) \ge 0.$$

P r o o f. Let $N = 2M+\delta$, $0 \le \delta \le 1$, and $2 \le t \le M-1$. By Proposition 2(1) we have

$$-E_t(k,N) \le |\mathcal{B}'_t|. \tag{12}$$

To estimate $|\mathcal{B}'_t|$ we start again by regarding $\mathrm{Sch}(N/t)$. All sets $B \in \mathcal{B}'_t$ contain $k-1$ numbers, in particular t and $2t$, and certainly not 0 and 1. Hence if we cancel $0, 1, t, 2t$ in $\mathrm{Sch}(N/t)$ we obtain a scheme $\mathrm{Sch}(A; N-4)$ with at most t rows and such that

$$|\mathcal{B}'_t| \le |\mathcal{P}_{k-3}(\mathrm{Sch}(A; N-4))|. \tag{13}$$

Now we refine this scheme by cutting every row of length $l \ge 3$ into rows of length 2 and possibly one row of length 1. The resulting scheme is of type $\mathrm{Sch}^*(A; N-4, \tau(t))$ with some $\tau(t) \le t$, and Lemma 3(2) asserts

$$|\mathcal{P}_{k-3}(\mathrm{Sch}(A; N-4))| \le |\mathcal{P}_{k-3}(\mathrm{Sch}^*(A; N-4, \tau(t)))|. \tag{14}$$

Therefore, by Lemma 4,

$$\begin{aligned}\sum_{t=1}^{N-1} E_t(k,N) &= \sum_{2\le t\le M-1} E_t(k,N) + \sum_{M+\delta\le t\le N-2} E_t(k,N)\\ &\ge -\sum_{2\le t\le M-1} |\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,\tau(t)))|\\ &\quad + \sum_{M+1+\delta\le t\le N-2} |\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,2t-N))|\\ &\qquad\qquad \text{(by (12)–(14), and Lemma 5(1))}\\ &= \sum_{2\le t\le M-1} (|\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,2t-2+\delta))|\\ &\qquad\qquad -|\mathcal{P}_{k-3}(\mathrm{Sch}^*(A;N-4,\tau(t)))|),\end{aligned}$$

and here all summands are nonnegative by Lemma 3(3), since

$$2t-2+\delta \ge t \ge \tau(t) \quad \text{for } 2\le t\le M-1. \ \blacksquare$$

LEMMA 7. *We have*

$$\sum_{t=1}^{N-1} E_t(k,N) \le \binom{N-1}{k-1}^*.$$

Proof. We already know that

$$\begin{aligned} E_t(k,N) &= 0 && \text{for } t\mid N-1 \text{ (by Lemma 4)},\\ E_t(k,N) &\le |\mathcal{C}_t| && \text{for } 2\le t<(N-1)/3 \text{ (by Proposition 2(2))},\\ E_t(k,N) &\le 0 && \text{for } N/3\le t<N/2 \text{ (by Proposition 2(2))},\end{aligned}$$

and

$$\sum_{(N-1)/2<t<N} E_t(k,N) = \binom{N-2}{k-1}^* \quad \text{(by Lemma 5(2))}.$$

Thus all we need is an appropriate upper bound for $|\mathcal{C}_t|$, $2\le t<(N-1)/3$. So let us look once more at the standard scheme $\mathrm{Sch}(N/t)$. $\mathcal{C}_t$ contains the subsets $B\subset\{0,1,\ldots,N-1\}-\{0,t\}$ with $|B|=k-1$ and $\{1,t+1,3t\}\subset B$, which—except 1 and $t+1$—do not contain two neighbouring numbers in any of the rows of $\mathrm{Sch}(N/t)$. In particular these sets B do not contain the numbers $2t$ and $2t+1$. Hence $|\mathcal{C}_t|$ counts certain subsets $B_0=B-\{1,t+1,3t\}$ of $A=\{0,1,\ldots,N-1\}-\{0,t,1,t+1,3t,2t,2t+1\}$ with $|B_0|=k-4$ which are admissible for the scheme $\mathrm{Sch}(A;N-7)$, resulting from $\mathrm{Sch}(N/t)$ by cancellation of 0, t, 1, $t+1$, $3t$, $2t$, and $2t+1$:

$$|\mathcal{C}_t| \le |\mathcal{P}_{k-4}(\mathrm{Sch}(A;N-7))|.$$

We refine this scheme by cutting every row of length $l\ge 3$ into rows of length 2 and possibly one row of length 1. Then the resulting scheme is of type $\mathrm{Sch}^*(A;N-7,\tau(t))$ with some $\tau(t)\le t$, for $\mathrm{Sch}(A;N-7)$ has at most t rows. Then Lemma 3(2) asserts

$$|\mathcal{P}_{k-4}(\mathrm{Sch}(A;N-7))| \le |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,\tau(t)))|.$$

Hence for $2\le t<(N-1)/3$,

$$|\mathcal{C}_t| \le |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,\tau(t)))|, \qquad \tau(t)\le t. \tag{15}$$

On the other hand, Lemma 5 with $N-3$ and $k-1$ instead of N and k yields

$$\begin{aligned}\binom{N-5}{k-2}^* &= \sum_{(N-4)/2<t<N-4} |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,2t-N+3))| \\ &= \sum_{2-\delta\le t\le M-2} |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,2t+\delta-3))| \\ &\qquad \text{(by substitution } t\mapsto t+M+\delta-3 \text{ with } N=2M+\delta,\ 0\le\delta\le 1) \\ &\ge \sum_{2\le t<(N-1)/3} |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,2t+\delta-3))|.\end{aligned} \tag{16}$$

Therefore by (15) and (16),

$$\begin{aligned}\binom{N-5}{k-2}^* - \sum_{2\le t<(N-1)/3} |\mathcal{C}_t| &\ge \sum_{2\le t<(N-1)/3} (|\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,2t+\delta-3))| \\ &\qquad - |\mathcal{P}_{k-4}(\mathrm{Sch}^*(A;N-7,\tau(t)))|),\end{aligned}$$

and by Lemma 3(3), all summands here are nonnegative, since $t\ge\tau(t)$ and thus $2t+\delta-3\ge\tau(t)$. This is obvious for $t\ge 3$ and also for $t=2$ and $\delta=1$. But if $t=2$ and $\delta=0$, then $N-7$ is odd and hence $\tau(2)=1$.

This shows

$$\sum_{2\le t<(N-1)/3} |\mathcal{C}_t| \le \binom{N-5}{k-2}^*,$$

and combined with the estimates at the beginning of the proof and with Lemma 2(2) we finally get

$$\sum_{t=1}^{N-1} E_t(k,N) \le \binom{N-2}{k-1}^* + \binom{N-5}{k-2}^* \le \binom{N-1}{k-1}^*. \ \blacksquare$$

Proposition 3. (1) *For $2\le k\le N$ and suitable $\theta\in[0,1]$,*

$$D(k,N) = D(k,N-1) + D(k-1,N-2) + (2N-1)\binom{N-2}{k-2} + 2\theta\binom{N-1}{k-1}^*.$$

(2) *Explicitly for* $1 \le k \le N$ *and* $\theta \in [0,1]$,

$$D(k,N) = D_0(k,N) + 2\theta \binom{N+1}{k+1}^*$$

with

$$D_0(k,N) = (2N+1)\binom{N}{k} - 2\binom{N}{k+1} - 2\binom{N+2}{k+2} + 2\binom{N+2-k}{k+2}.$$

Proof. (1) Clearly

$$D(k,N) = \binom{N}{k} + 2\sum_{k=1}^{N-1} \delta_t(k,N), \tag{17}$$

since $D(k,N) = \sum_{-N+1\le t\le N-1} \delta_t(k,N)$ by definition and

$$\delta_0(k,N) = \binom{N}{k}, \qquad \delta_{-t}(k,N) = \delta_t(k,N).$$

Therefore the recursion formula in Proposition 2(1) gives

$$\begin{aligned} D(k,N) &= \binom{N}{k} + 2\sum_{t=1}^{N-1} (\delta_t(k,N-1) + \delta_t(k-1,N-2)) \\ &\quad + 2(N-1)\binom{N-2}{k-2} + 2\sum_{t=1}^{N-1} E_t(k,N) \\ &= 2\sum_{t\ge 1} \delta_t(k,N-1) + 2\sum_{t\ge 1} \delta_t(k-1,N-2) \\ &\quad + (2N-2)\binom{N-2}{k-2} + \binom{N-1}{k} \\ &\quad + \binom{N-2}{k-1} + \binom{N-2}{k-2} + 2\theta\binom{N-1}{k-1}^* \\ &\qquad\qquad \text{(with } \theta \in [0,1]\text{, by Lemmata 6 and 7)} \\ &= D(k,N-1) + D(k-1,N-2) \\ &\quad + (2N-1)\binom{N-2}{k-2} + 2\theta\binom{N-1}{k-1}^* \qquad \text{(by (17))}. \end{aligned}$$

(2) The initial values are

$$D(1,N) = N = D_0(1,N), \qquad N \ge 1,$$

and

$$D(k,k) = 2k-1 = D_0(k,k), \qquad k \ge 1.$$

The rest is straightforward induction on k and N with the use of the recursion formula of part (1) and Lemma 2(2) for the θ-terms. ∎

Now we use the recursion formulae of $S(k,N)$ and $D(k,N)$ to prove:

THEOREM. *We have*

$$1 \le \frac{D(k,N)}{S(k,N)} < 2 \quad \textit{for } 1 \le k \le N.$$

Proof. First we show

$$\Delta_1(k,N) := D(k,N) - S(k,N) \ge 0 \quad \text{for } 1 \le k \le N$$

by induction on k and N. The initial values are

$$\Delta_1(1,N) = D(1,N) - S(1,N) = N - N = 0,$$
$$\Delta_1(k,k) = D(k,k) - S(k,k) = (2k-1) - (2k-1) = 0,$$

and the induction step $N-1 \mapsto N$ with $N = 2M + \delta > k \ge 2$ is

$$\begin{aligned}
\Delta_1(k,N) &= D(k,N) - S(k,N) \\
&\ge D(k,N-1) + D(k-1,N-2) + (2N-1)\binom{N-2}{k-2} \\
&\quad - S(k,N-1) - S(k-1,N-2) \\
&\quad - (2N-1)\binom{N-2}{k-2} - 2^{k-1}\binom{M-1}{k-1} + 2\binom{N-2}{k}^{*} \\
&\qquad\qquad \text{(by Proposition 3(1) and Proposition 1(2))} \\
&= \Delta_1(k,N-1) + \Delta_1(k-1,N-2) - 2^{k-1}\binom{M-1}{k-1} + 2\binom{N-2}{k}^{*} \\
&\ge 2\binom{2M-2}{k}^{*} - 2^{k-1}\binom{M-1}{k-1} \\
&\qquad\qquad \text{(by induction hypothesis and Lemma 2(2))} \\
&\ge 2^{k-1}(k-2)\binom{M-1}{k-1} \quad \text{(by Lemma 2(5))} \\
&\ge 0.
\end{aligned}$$

Finally we prove

$$\Delta_2(k,N) := 2S(k,N) - D(k,N) > 0 \quad \text{for } 1 \le k \le N,$$

again by induction on k and N, and by Propositions 1(2) and 3(1):

$$\Delta_2(1,N) = 2N - N = N > 0,$$
$$\Delta_2(k,k) = 2(2k-1) - (2k-1) = 2k-1 > 0,$$

and the induction step $N-1 \mapsto N$ with $N = 2M + \delta > k \ge 2$ is

$$\begin{aligned}
\Delta_2(k,N) &= 2S(k,N) - D(k,N) \\
&\ge 2S(k,N-1) + 2S(k-1,N-2) + 2(2N-1)\binom{N-2}{k-2}
\end{aligned}$$

$$
\begin{aligned}
&\quad + 2^k\binom{M-1}{k-1} - 4\binom{N-2}{k}^* \\
&\quad - D(k, N-1) - D(k-1, N-2) \\
&\quad - (2N-1)\binom{N-2}{k-2} - 2\binom{N-1}{k-1}^* \\
&= \Delta_2(k, N-1) + \Delta_2(k-1, N-2) + (2N-1)\binom{N-2}{k-2} \\
&\quad + 2^k\binom{M-1}{k-1} - 4\binom{N-2}{k}^* - 2\binom{N-1}{k-1}^* \\
&> 0
\end{aligned}
$$

by induction hypothesis and Lemma 2(6). ■

REMARK 3. The development of $S(k,N)$ in Proposition 1(1) and of $D_0(k,N)$ in Proposition 3(2) in powers of N yields, as $N \to \infty$,

$$
S(k,N) = \binom{k+1}{2}\binom{N}{k} + O(N^{k-1}), \tag{18}
$$

$$
D_0(k,N) = (k^2 - k + 1)\binom{N}{k} + O(N^{k-1}). \tag{19}
$$

The appearance of the coefficients $\binom{k+1}{2}$ and $k^2 - k + 1$ is not surprising: If N is large compared to k, then within most of the $\binom{N}{k}$ subsets A of $\{0, 1, \ldots, N-1\}$ with $|A| = k$ there are only very few nontrivial coincidences $a + b = a' + b'$. In particular sets with $|A| = k$ and without such coincidences have $S(A) = \binom{k+1}{2}$ and $D(A) = k^2 - k + 1$ (comp. introduction). Therefore the upper bound $D(k,N) \le (k^2 - k + 1)\binom{N}{k}$ is obvious. On the other hand, $D(k,N) \ge D_0(k,N)$ by Proposition 3(2), which together with (19) yields

$$
D(k,N) = (k^2 - k + 1)\binom{N}{k} + O(N^{k-1}).
$$

This and (18) imply at once

$$
\lim_{N\to\infty} \frac{D(k,N)}{S(k,N)} = 1 + \left(1 - \frac{2}{k}\right)\left(1 - \frac{2}{k+1}\right)
$$

and

$$
\lim_{k\to\infty} \lim_{N\to\infty} \frac{D(k,N)}{S(k,N)} = 2.
$$

On the other hand, the explicit formulae for $S(k,N)$ and $D(k,N)$ show immediately that there exists a positive constant c_0 such that for all $N \ge 1$,

$$
1 \le \frac{D(k,N)}{S(k,N)} < 1 + \frac{c_0}{N} \quad \text{for } N/2 < k \le N.
$$

Hence the lower bound 1 as well as the upper bound 2 of the Theorem are best possible, and their values are not caused by accidental irregularities of the quotient $D(k,N)/S(k,N)$ for small values of k and N.

References

[1] G. A. Freĭman and V. P. Pigarev, *The relation between the invariants R and T*, in: Number-Theoretic Studies in the Markov Spectrum and in the Structural Theory of Set Addition, Kalinin. Gos. Univ., Moscow, 1973, 172–174 (in Russian).
[2] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Univ. Press, 1988.
[3] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, 2. Band, Teubner, Leipzig, 1909.
[4] J. Riordan, *Combinatorial Identities*, Krieger, Huntington, 1979.
[5] I. Z. Ruzsa, *Sets of sums and differences*, in: Séminaire de Théorie des Nombres de Paris, 1982–83, Birkhäuser, Boston, 1984, 267–273.

Zentrum Mathematik
Technische Universität München
D-80290 München, Germany
E-mail: roeslerf@mathematik.tu-muenchen

Received on 17.8.1998 (3448)