On the ubiquity of Sidon sets

Melvyn B. Nathanson†
Department of Mathematics
Lehman College (CUNY)
Bronx, New York 10468
Email: nathansn@alpha.lehman.cuny.edu

November 13, 2018

Abstract

A Sidon set is a set $A$ of integers such that no integer has two essentially distinct representations as the sum of two elements of $A$. More generally, for every positive integer $g$, a $B_2[g]$-set is a set $A$ of integers such that no integer has more than $g$ essentially distinct representations as the sum of two elements of $A$. It is proved that almost all small sumsets of $\{1, 2, \ldots, n\}$ are $B_2[g]$-sets, in the sense that if $B_2[g](k, n)$ denotes the number of $B_2[g]$-sets of cardinality $k$ contained in the interval $\{1, 2, \ldots, n\}$, then $\lim_{n \to \infty} B_2[g](k, n)/\binom{n}{k} = 1$ if $k = o\left(n^{g/(2g+2)}\right)$.

1 Sidon sets

Let $A$ be a nonempty set of positive integers. The sumset $2A$ is the set of all integers of the form $a_1 + a_2$, where $a_1, a_2 \in A$. The set $A$ is called a Sidon set if every element of $2A$ has a unique representation as the sum of two elements of $A$, that is, if

$$a_1, a_2, a'_1, a'_2 \in A$$

and

$$a_1 + a_2 = a'_1 + a'_2,$$

and if

$$a_1 \leq a_2 \quad \text{and} \quad a'_1 \leq a'_2.$$
then
\[ a_1 = a_1' \quad \text{and} \quad a_2 = a_2'. \]

More generally, for positive integers \( h \) and \( g \), the \( h \)-fold sumset \( hA \) is the set of all sums of \( h \) not necessarily distinct elements of \( A \). The representation function \( r_{A,h}(m) \) counts the number of representations of \( m \) in the form
\[ m = a_1 + a_2 + \cdots + a_h, \]
where
\[ a_i \in A \quad \text{for all} \quad i = 1, 2, \ldots, h, \]
and
\[ a_1 \leq a_2 \leq \cdots \leq a_h. \]

The set \( A \) is called a \( B_{h}[g] \)-set if every element of \( hA \) has at most \( g \) representations as the sum of \( h \) elements of \( A \), that is, if
\[ r_{A,h}(m) \leq g \]
for every integer \( m \). In particular, a \( B_{2}[1] \)-set is a Sidon set, and \( B_{h}[1] \)-sets are usually denoted \( B_{h} \)-sets.

Let \( h \geq 2 \). Let \( A \) be a nonempty set of integers, and \( a \in A \). Then \( r_{h}(m + a) \geq r_{h-1}(m) \). Therefore, if \( r_{A,h-1}(m) > g \) for some \( m \in (h-1)A \), then \( r_{A,h}(m + a) > g \) for every \( a \in A \). It follows that if \( A \) is a \( B_{h}[g] \)-set, then \( A \) is also a \( B_{h-1}[g] \)-set. In particular, every \( B_{h} \)-set is also a \( B_{h-1} \)-set.

Let \( A \) be a subset of \( \{1, 2, \ldots, n\} \), and let \( |A| \) denote the cardinality of \( A \). Then \( hA \subseteq \{h, h+1, \ldots, hn\} \). If \( |A| = k \), then there are exactly \( \binom{k+h-1}{h} \) ordered \( h \)-tuples of the form \((a_1, \ldots, a_h)\), where \( a_i \in A \) for all \( i = 1, \ldots, h \) and \( a_1 \leq \cdots \leq a_h \). If \( A \) is a \( B_{h}[g] \)-set and \( |A| = k \), then
\[ \frac{k^h}{h!} < \binom{k+h-1}{h} = \sum_{m \in hA} r_{A,h}(m) \leq g|hA| < ghn, \]
and so
\[ |A| = k < cn^{1/h} \]
for \( c = (h!gh)^{1/h} \). It follows that if \( A \) is a “large” subset of \( \{1, 2, \ldots, n\} \), then \( A \) cannot be a \( B_{h}[g] \)-set. In this paper we prove that almost all "small" subsets of \( \{1, 2, \ldots, n\} \) are \( B_{2}[g] \)-sets and almost all "small" subsets of \( \{1, 2, \ldots, n\} \) are \( B_{h} \)-sets.

**Notation.** If \( \{u_n\}_{n=1}^{\infty} \) and \( \{v_n\}_{n=1}^{\infty} \) are sequences and \( v_n > 0 \) for all \( n \), we write \( u_n = o(v_n) \) if \( \lim_{n \to \infty} u_n/v_n = 0 \), and \( u_n = O(v_n) \) or \( u_n \ll v_n \) if \( |u_n| \leq cv_n \) for some \( c > 0 \) and all \( n \geq 1 \). The number \( c \) in this inequality is called the implied constant.
2 Random small $B_2[g]$-sets

We require the following elementary lemma.

**Lemma 1** If $n \geq 1$ and $0 \leq j \leq k \leq n$, then

$$\frac{{n-j \choose k-j}}{{n \choose k}} \leq \left(\frac{k}{n}\right)^j.$$

**Proof.** We have

$$\frac{{n-j \choose k-j}}{{n \choose k}} = \frac{(n-j)!k!}{(k-j)!n!} = \prod_{i=0}^{j-1} \frac{k-i}{n-i} \leq \left(\frac{k}{n}\right)^j$$

since

$$\frac{k-i}{n-i} \leq \frac{k}{n}$$

for $i = 0, 1, \ldots, n-1$. □

**Theorem 1** For any positive integers $g, k,$ and $n$, let $B_2[g](k, n)$ denote the number of $B_2[g]$-sets $A$ contained in $\{1, \ldots, n\}$ with $|A| = k$. Then

$$B_2[g](k, n) > {n \choose k} \left(1 - \frac{4k^{2g+2}}{n^g}\right).$$

**Proof.** Let $A$ be a subset of $\{1, 2, \ldots, n\}$ of cardinality $k$. If $A$ is not a $B_2[g]$-set, then there is an integer $m \leq 2n$ such that $r_{A,2}(m) > g$, that is, $m$ has at least $g + 1$ representations as the sum of two elements of $A$. This means that the set $A$ contains $g + 1$ integers $a_1, \ldots, a_{g+1}$ such that

$$1 \leq a_1 < \cdots < a_{g+1} \leq \frac{m}{2},$$

and $A$ also contains the $g + 1$ integers $m - a_i$ for $i = 1, \ldots, g + 1$. If $a_{g+1} < m/2$, then

$$\{|a_i, m - a_i : i = 1, \ldots, g + 1\} = 2g - 2.$$

If $a_{g+1} = m/2$, then

$$\{|a_i, m - a_i : i = 1, \ldots, g + 1\} = 2g - 1.$$

Therefore, for each integer $m$, the number of sets $A \subseteq \{1, \ldots, n\}$ such that $|A| = k$ and $r_{A,2}(m) \geq g + 1$ is at most

$$\left(\begin{bmatrix} m-1 \\n \end{bmatrix} \right) \left(\begin{bmatrix} n-2g-2 \\n \end{bmatrix} \right) + \left(\begin{bmatrix} m-1 \\n \end{bmatrix} \right) \left(\begin{bmatrix} n-2g-1 \\n \end{bmatrix} \right),$$
and so
\[
\binom{n}{k} - B_2[g](k, n) \leq \sum_{m \leq 2n} \binom{\binom{m-1}{2}}{g+1} \binom{n-2g-2}{k-2g-2} + \sum_{m \leq 2n} \binom{\binom{m-1}{2}}{g} \binom{n-2g-1}{k-2g-1}.
\]

Observing that
\[
\sum_{m \leq 2n} \left( \binom{\binom{m-1}{2}}{g+1} \binom{n}{k} \right) \leq 2n^{g+2}
\]
and
\[
\sum_{m \leq 2n} \left( \binom{\binom{m-1}{2}}{g} \binom{n}{k} \right) < 2n^{g+1}
\]
and applying Lemma 1 we obtain
\[
1 - \frac{B_2[g](k, n)}{\binom{n}{k}} \leq \sum_{m \leq 2n} \left( \binom{\binom{m-1}{2}}{g+1} \binom{n-2g-2}{k-2g-2} \binom{n}{k} + \binom{\binom{m-1}{2}}{g} \binom{n-2g-1}{k-2g-1} \binom{n}{k} \right)
\]
\[
\leq \sum_{m \leq 2n} \left( \binom{\binom{m-1}{2}}{g+1} \binom{k}{n}^{2g+2} + \sum_{m \leq 2n} \binom{\binom{m-1}{2}}{g} \binom{k}{n}^{2g+1} \right)
\]
\[
< 2n^{g+2} \binom{k}{n}^{2g+2} + 2n^{g+1} \binom{k}{n}^{2g+1}
\]
\[
\leq \frac{4k^{2g+2}}{n^g}.
\]

This completes the proof. □

**Theorem 2** Let \( \{k_n\}_{n=1}^\infty \) be a sequence of positive integers such that \( k_n \leq n \) for all \( n \) and
\[
k_n = o \left( n^{g/(2g+2)} \right).
\]

Then
\[
\lim_{n \to \infty} \frac{B_2[g](k_n, n)}{\binom{n}{k_n}} = 1.
\]

**Proof.** This follows immediately from Theorem 1 □

**Theorem 3** Let \( B_2(k, n) \) denote the number of Sidon sets of cardinality \( k \) contained in \( \{1, \ldots, n\} \). If \( k_n = o \left( n^{1/4} \right) \), then
\[
\lim_{n \to \infty} \frac{B_2(k_n, n)}{\binom{n}{k_n}} = 1.
\]
Proof. This follows immediately from Theorem 2 with $g = 1$. □

Theorem 4 Let $\{k_n\}_{n=1}^\infty$ be a sequence of positive integers such that $k_n \leq n$ for all $n$ and

$$k_n = o \left( n^{g/(2g+3)} \right).$$

Then

$$\lim_{n \to \infty} \frac{\sum_{k \leq k_n} B_2[g](k, n)}{\sum_{k \leq k_n} \binom{n}{k}} = 1.$$ 

Proof. It suffices to show that

$$\lim_{n \to \infty} \frac{\sum_{k \leq k_n} \left( \binom{n}{k} - B_2[g](k, n) \right)}{\sum_{k \leq k_n} \binom{n}{k}} = 0,$$

where $f(k)$ is defined in the proof of Theorem 1. If $a_1, \ldots, a_\ell, b_1, \ldots, b_\ell$ are positive real numbers and $B = \max(b_1, \ldots, b_\ell)$, then

$$\frac{a_1 + \cdots + a_\ell}{b_1 + \cdots + b_\ell} \leq \frac{a_1 + \cdots + a_\ell}{B} \leq \frac{a_1}{b_1} + \cdots + \frac{a_\ell}{b_\ell}.$$

This implies that

$$\frac{\sum_{k \leq k_n} \left( \binom{n}{k} - B_2[g](k, n) \right)}{\sum_{k \leq k_n} \binom{n}{k}} \leq \sum_{k \leq k_n} \frac{4k^{2g+2}}{n^g}$$

$$\leq \frac{4k_n^{2g+3}}{n^g}$$

$$= 4 \left( \frac{k_n}{n^{g/(2g+3)}} \right)^{2g+3}$$

$$= o(n),$$

and the proof is complete. □

We can restate our results in the language of probability. Let $\Omega$ be the probability space consisting of the $\binom{n}{k}$ subsets of $\{1, \ldots, n\}$ of cardinality $k$, where the probability of choosing $A \in \Omega$ is $1/\binom{n}{k}$. If $P_{h,g}(k, n)$ denotes the probability that a random set $A \in \Omega$ is a $B_{h}[g]$-set, then Theorem 4 states that

$$\lim_{n \to \infty} P_{2,g}(k_n, n) = 1$$

if $k_n = o \left( n^{g/(2g+2)} \right)$. 

Similarly, Theorem 4 states that if $k_n = o \left( n^{g/(2g+3)} \right)$ and if $P_{h,g}(k_n, n)$ denotes the probability that a random set $A \subseteq \{1, \ldots, n\}$ of cardinality $|A| \leq k_n$ is a $B_{h}[g]$-set, then

$$\lim_{n \to \infty} P_{2,g}(k_n, n) = 1.$$
3 Random small $B_h$-sets

A set $A$ is called a $B_h$-set if $r_{A,h}(m) = 1$ for all $m \in hA$. Let $B_h(k,n)$ denote the number of $B_h$-sets of cardinality $k$ contained in $\{1, \ldots, n\}$. Since every set of integers is a $B_1$-set, and every $B_h$ set is a $B_{h-1}$-set, it follows that $\binom{n}{k} = B_1(k,n) \geq \cdots \geq B_{h-1}(k,n) \geq B_h(k,n) \geq \cdots$

We shall prove that almost all “small” subsets of $\{1, \ldots, n\}$ are $B_h$-sets. The method is similar to that used to prove Theorem $\mathbb{2}$ but, for $h \geq 3$, we have to consider the possible dependence between different representations of an integer $n$ as the sum of $h$ elements of $A$. This means the following: Let $(a_1, \ldots, a_h)$ and $(a'_1, \ldots, a'_h)$ be $h$-tuples of elements of $A$ such that

$$a_1 + \cdots + a_h = a'_1 + \cdots + a'_h,$$

$$a_1 \leq \cdots \leq a_h,$$

$$a'_1 \leq \cdots \leq a'_h,$$

and

$$\{a_1, \ldots, a_h\} \cap \{a'_1, \ldots, a'_h\} \neq \emptyset.$$

If $h = 2$, then $a_i = a'_i$ for $i = 1,2$, but if $h \geq 3$, then it is not necessarily true that $a_i = a'_i$ for all $i = 1, \ldots, h$. For example, in the case $h = 3$ we have $1 + 3 + 4 = 1 + 2 + 5$ but $(1,3,4) \neq (1,2,5)$. In the case $h = 5$ we have $1+1+2+3+3 = 1+2+2+2+3$ but $(1,1,2,3,3) \neq (1,2,2,2,3)$, even though $\{1,1,2,3,3\} = \{1,2,2,2,3\}$.

Because of the lack of independence, we need a careful description of a representation of $m$ as the sum of $h$ not necessarily distinct integers. We introduce the following notation. Let $A$ be a set of positive integers. Corresponding to each representation of $m$ in the form

$$m = a_1 + \cdots + a_h,$$

(1)

where $a_1, \ldots, a_h \in A$ and $a_1 \leq \cdots \leq a_h$, there is a unique triple

$$(r, (h_j), (a'_j)),$$

(2)

where

(i) $r$ is the number of distinct summands in this representation,

(ii) $(h_j) = (h_1, \ldots, h_r)$ is an ordered partition of $h$ into $r$ positive parts, that is, an $r$-tuple of positive integers such that

$$h = h_1 + \cdots + h_r,$$
(iii) \((a'_j) = (a'_1, \ldots, a'_r)\) is an \(r\)-tuple of pairwise distinct elements of \(A\) such that
\[
1 \leq a'_1 < \cdots < a'_r \leq m
\]
and
\[
\{a_1, \ldots, a_h\} = \{a'_1, \ldots, a'_r\},
\]
(iv)
\[
m = h_1 a'_1 + \cdots + h_r a'_r,
\]
where each integer \(a'_j\) occurs exactly \(h_j\) times in the representation \((i)\).

There is a one-to-one correspondence between distinct representations of an integer \(m\) in the form \((i)\) and triples of the form \((ii)\). Moreover, for each \(r\) and \(m\), the integer \(a'_r\) is completely determined by the ordered partition \((h_j)\) of \(h\) and the \((r-1)\)-tuple \((a'_1, \ldots, a'_{r-1})\). Therefore, for positive integers \(m\) and \(r\), the number of triples of the form \((ii)\) does not exceed \(\pi_r(h)m^{r-1}\), where \(\pi_r(h)\) is number of ordered partitions of \(h\) into exactly \(r\) positive parts.

**Theorem 5** Let \(h \geq 2\). For all positive integers \(k \leq n\),
\[
B_{h-1}(k, n) - B_h(k, n) \ll \left( \frac{n}{k} \right)^{\frac{k^2 h^2}{n}}
\]
and
\[
B_h(k, n) \gg \left( \frac{n}{k} \right) \left( 1 - \frac{k^{2h}}{n} \right),
\]
where the implied constants depend only on \(h\).

**Proof.** Let \(A\) be a \(B_{h-1}\)-set contained in \(\{1, \ldots, n\}\). Then \(hA \subseteq \{h, h + 1, \ldots, hn\}\). If \(m \in hA\) and \(m\) has two distinct representations as the sum of \(h\) elements of \(A\), then there exist positive integers \(r_1\) and \(r_2\) and triples
\[
(r_1, (h_1,j), (a'_1,j)) \quad \text{and} \quad (r_2, (h_2,j), (a'_2,j))
\]
such that, for \(i = 1\) and \(2\), we have
\[
m = \sum_{j=1}^{r_i} h_{i,j} a'_{i,j},
\]
\[
h = \sum_{j=1}^{r_i} h_{i,j},
\]
and
\[
1 \leq a'_{i,1} < \cdots < a'_{i,r_i} \leq m.
\]
The number of pairs of triples of the form \((ii)\) for fixed positive integers \(m\), \(r_1 \leq h\), and \(r_2 \leq h\) is at most
\[
\pi_{r_1}(h)m^{r_1-1} \pi_{r_2}(h)m^{r_2-1} \ll m^{r_1+r_2-2},
\]
where the implied constant depends only on \( h \). Moreover, since \( A \) is a \( B_{h-1} \)-set, no number can have two representations as the sum of \( h-1 \) elements of \( A \). This implies that

\[
\{a'_{1,1}, a'_{1,2}, \ldots, a'_{1,r_1}\} \cap \{a'_{2,1}, a'_{2,2}, \ldots, a'_{2,r_2}\} = \emptyset,
\]

and so the set

\[
\{a'_{i,j} : i = 1, 2 \text{ and } j = 1, \ldots, r_i\}
\]

contains exactly \( r_1 + r_2 \) elements of \( A \). Therefore, given positive integers \( r_1 \leq h \), \( r_2 \leq h \) and \( m \leq hn \), there are

\[
\ll m r_1 r_2 - 2 \left( \frac{n - r_1 - r_2}{k - r_1 - r_2} \right)
\]

sets \( A \) for which \( m \) has two representations as the sum of \( h \) elements of \( A \), and in which one representation uses \( r_1 \) distinct integers and the other representation uses \( r_2 \) distinct integers. Summing over \( m \leq hn \), we obtain

\[
\ll n^{r_1 + r_2 - 1} \left( \frac{n - r_1 - r_2}{k - r_1 - r_2} \right).
\]

Applying Lemma 1, we obtain

\[
B_{h-1}(k, n) - B_h(k, n) \ll \binom{n}{k} \sum_{r_1, r_2 \leq h} \frac{n^{r_1 + r_2 - 1} \left( \frac{n - r_1 - r_2}{k - r_1 - r_2} \right)}{\binom{n}{k}^{r_1 + r_2}}
\]

\[
\ll \binom{n}{k} \sum_{r_1, r_2 \leq h} n^{r_1 + r_2 - 1} \left( \frac{k}{n} \right)^{r_1 + r_2}
\]

\[
= \binom{n}{k} \sum_{r_1, r_2 \leq h} k^{r_1 + r_2} \frac{1}{n}
\]

\[
\ll \binom{n}{k} \frac{k^{2h}}{n}.
\]

It follows that

\[
\binom{n}{k} - B_h(k, n) = B_1(k, n) - B_h(k, n)
\]

\[
= \sum_{j=2}^{h} \left( B_{j-1}(k, n) - B_j(k, n) \right)
\]

\[
\ll \sum_{j=2}^{h} \binom{n}{k} \frac{k^{2j}}{n}
\]

\[
\ll \binom{n}{k} \frac{k^{2h}}{n},
\]

8
and so

\[ B_h(k, n) \gg \binom{n}{k} \left(1 - \frac{k^{2h}}{n}\right). \]

This completes the proof. □

**Theorem 6** Let \( B_h(k, n) \) denote the number of \( B_h \)-sets \( A \) contained in \( \{1, \ldots, n\} \) with \( |A| = k \). Let \( \{k_n\}_{n=1}^\infty \) be a sequence of positive integers such that

\[ k_n = o\left(n^{1/2h}\right). \]

Then

\[ \lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = 1. \]

**Proof.** By Theorem 5

\[ 1 \geq \frac{B_h(k_n, n)}{\binom{n}{k_n}} \gg 1 - \left(\frac{k_n}{n^{1/2h}}\right)^{2h}, \]

and so

\[ \lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = 1. \]

This completes the proof. □

## 4 Remarks added in proof

A variant of Theorem 3 appears in Nathanson [2, p. 37, Exercise 14]. Godbole, Janson, Locantore, and Rapoport [1] used probabilistic methods to obtain a converse of Theorem 6. They proved that if

\[ \lim_{n \to \infty} \frac{k_n}{n^{1/2h}} = \infty, \]

then

\[ \lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = 0. \]

They also analyzed the threshold behavior of \( B_h(k, n) \), and proved that if

\[ \lim_{n \to \infty} \frac{k_n}{n^{1/2h}} = \Lambda > 0, \]

then

\[ \lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = e^{-\lambda}. \]
where $\lambda = \kappa h^2$.

It is natural to conjecture that analogous results hold for the function $B_h[g](k, n)$, namely, if

$$\lim_{n \to \infty} \frac{k_n}{n^{g/(gh+h)}} = 0,$$

then

$$\lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = 1,$$

and if

$$\lim_{n \to \infty} \frac{k_n}{n^{g/(gh+h)}} = \infty,$$

then

$$\lim_{n \to \infty} \frac{B_h(k_n, n)}{\binom{n}{k_n}} = 0.$$