View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Digital.CSIC

Generalization of a theorem of Erdős and Rényi on Sidon Sequences

Javier Cilleruelo, Sándor Z. Kiss, Imre Z. Ruzsa, Carlos Vinuesa

October 27, 2009

Abstract

Erdős and Rényi claimed and Vu proved that for all $h \geq 2$ and for all $\varepsilon > 0$, there exists $g = g_h(\varepsilon)$ and a sequence of integers A such that the number of ordered representations of any number as a sum of h elements of A is bounded by g, and such that $|A \cap [1, x]| \gg x^{1/h-\varepsilon}$.

We give two new proofs of this result. The first one consists of an explicit construction of such a sequence. The second one is probabilistic and shows the existence of such a g that satisfies $g_h(\varepsilon) \ll \varepsilon^{-1}$, improving the bound $g_h(\varepsilon) \ll \varepsilon^{-h+1}$ obtained by Vu.

Finally we use the "alteration method" to get a better bound for $g_3(\varepsilon)$, obtaining new bounds on the growth of $B_3[g]$ sequences.

1 Introduction

Given an integer $h \geq 2$, we say that a sequence of integers A is a $B_h[g]$ sequence if every integer n has at most g representations as a sum of h elements of A. We will write

$$r_{h,A}(n) = \left|\{(a_1, a_2, \ldots, a_h) \mid n = a_1 + \cdots + a_h, \quad a_1 \leq \cdots \leq a_h, \quad a_i \in A\}\right|,$$

Thus, A is a $B_h[g]$ sequence if $r_{h,A}(n) \leq g$ for every positive integer n.

As usual, $A(x) = |A \cap [1, x]|$ counts the number of elements of A less than or equal to x. The counting method easily gives $A(x) \ll x^{1/h}$ for any $B_h[g]$ sequence. It is believed that a $B_h[g]$ sequence A cannot satisfy $A(x) \gg x^{1/h}$. However it is only known when $(h, g) = (\text{even}, 1)$.

In a seminal paper, Erdős and Rényi [4] proved, using the probabilistic method (see, for example, [1] for an excellent exposition of the method), that for any $\varepsilon > 0$

there exists $g = g(\varepsilon)$ and a $B_2[g]$ sequence such that $A(x) \gg x^{1/2-\varepsilon}$. In this paper they claimed (but did not prove) that the same method gives the analogous result for $h \geq 3$:

Theorem 1.1. *For any $\varepsilon > 0$ and $h \geq 2$, there exists $g = g_h(\varepsilon)$ and a $B_h[g]$ sequence, A, such that $A(x) \gg x^{1/h-\varepsilon}$.*

Probably they did not notice that when $h \geq 3$ two distinct representations of an integer as a sum of h numbers can share common elements. If $x_1 + \cdots + x_h = y_1 + \cdots + y_h$ and $x_1 = y_1$, then the events $(x_1, \ldots, x_h \in A)$ and $(y_1, \ldots, y_h \in A)$ are not independent (this cannot happen when $h = 2$ since one equal addend implies that the other is also equal). This phenomena makes the cases $h \geq 3$ much more difficult than the case $h = 2$.

In [8] Vu gave the first correct proof of Theorem 1.1. He used ideas from a paper of Erdős and Tetali [5] to solve a similar problem for a related question. The key point is the use of the *"Sunflower lemma"* to prove that if an integer has enough representations then we can select $g+1$ representations which are disjoint. Now the probabilistic method works easily because we deal with independent events. If we follow the details of the proof we can see that Vu obtains $g_h(\varepsilon) \ll \varepsilon^{-h+1}$.

The aim of this paper is to present new proofs of Theorem 1.1 and to obtain better relations between g and ε.

The first one consists of an explicit construction of the sequence claimed in Theorem 1.1. We do it in Section 2.

The second one is a probabilistic but distinct and simpler proof than that presented by Vu. We do not use the *"Sunflower lemma"*, but a simpler one, and we get a better upper bound for $g_h(\varepsilon)$. More precisely, we prove the next theorem in Section 3.

Theorem 1.2. *For any $\varepsilon > 0$ and $h \geq 2$, there exists $g = g_h(\varepsilon) \ll \varepsilon^{-1}$ and a $B_h[g]$ sequence, A, such that $A(x) \gg x^{1/h-\varepsilon}$.*

Actually we can check in the proof of the theorem above that we can take any $g_h(\varepsilon) \geq 2^{h-3}h(h-1)!^2\varepsilon^{-1}$. The improvement of this theorem affects to the cases $h \geq 3$, where we have to deal with not independent events. Vu's proof only gives $g_h(\varepsilon) \ll \varepsilon^{-h+1}$ which is worse than our bound when $h \geq 3$. For the case $h = 2$ Erdős and Rényi proved that any $g_2(\varepsilon) > \frac{1}{2\varepsilon} - 1$ satisfies the condition of Theorem 1.2 and the first author [2] used the "alteration method" to improve that bound to $g_2(\varepsilon) > \frac{1}{4\varepsilon} - \frac{1}{2}$.

In the last section we refine Theorem 1.2 when $h = 3$ proving that $g_3(\varepsilon) > \frac{2}{9\varepsilon} - \frac{2}{3}$ works. In other words,

Theorem 1.3. *For every $\varepsilon > 0$ and for every $g \geq 1$ there is a $B_3[g]$ sequence A, such that*

$$A(x) \gg x^{\frac{g}{3g+2}-\varepsilon}.$$

It is also possible to refine Theorem 1.2 for $h \geq 4$ using the "alteration method" but the exponents we would obtain in these cases are not satisfactory enough. For "satisfactory enough" we mean exponents such that when we particularize to $g = 1$, we obtain the same exponent that we get with the greedy algorithm. That is what happens for $h = 2$ ([2]) and $h = 3$ (Theorem 1.3).

2 A constructive proof of Theorem 1.1

Given h and ε, we construct a sequence A with $r_{h,A}(n)$ bounded and we will prove that $A(n) > n^{1/h-\varepsilon}$ for sufficiently large n, which implies Theorem 1.1.

We use the representation of natural numbers in a number system with variable base. It is easy to see that every natural number x can be expressed uniquely in the form

$$x = b_0 + b_1 q_1 + b_2 q_1 q_2 + \ldots + b_s q_1 \ldots q_s + \ldots,$$

where $0 \leq b_i < q_{i+1}$. The b_i's and q_i's are natural numbers, b_i's called the "digits" and q_i's called the "bases".

We consider $l \geq 2$, a large enough number that will be fixed later. We fix the sets $0 \in A_i \subset \left[0, \frac{q_i}{h}\right]$ such that the A_i's are maximal sets with the condition $r_{h,A_i}(n) \leq 1$ for every n. Now we construct the set A in the following way: put that natural numbers in A which digits $b_i \in A_{i+1}$, and for which there is an m such that $b_i = 0$ for $i \notin [m+1, \ldots m+l]$.

First we prove that $r_{h,A}(n) < (h!)^{lh}$. We add up h numbers, $a_1, a_2, \ldots, a_h$. The sum will be $a_1 + a_2 + \cdots + a_h = d_0 + d_1 q_1 + d_2 q_1 q_2 + \ldots + d_m q_1 \ldots q_m$.

Since the j-th digit of each addend is in $\left[0, \frac{q_j}{h}\right]$, each digit of the sum, $d_j \in [0, q_{j+1})$, will be the sum of the j-th digits of the a_i's (in other words, there will be no carries).

And, since the j-th digit of each addend is in a $B_h[1]$ set, the digit d_j of the sum can be obtained in only one way as a sum of h digits. Note that h numbers have $h!$

permutations, so for each digit of the sum we could have the corresponding digits of the h addends distributed in at most $h!$ ways.

Finally, observe that the sum of the number of non zero digits of all the addends is less than or equal to hl, so the number of digits of the sum different from zero will also be less than or equal to hl, and finally we will have $r_{h,A}(n) \leq (h!)^{lh}$ for every n.

Now, we give an estimation of the value of $A(n)$. Given n, we know that there exists j such that

$$q_1 q_2 \dots q_j \leq n < q_1 q_2 \dots q_{j+1}. \tag{1}$$

It is clear that those integers which digits

$$b_0 = b_1 = \dots = b_{j-l-1} = 0 \text{ and } b_i \in A_{i+1},\ i = j-l, \dots, j-1,$$

are in A. Let N denote the number of such integers. We define $r = \frac{\log_2 l}{l}$. Let $q_1 = \lfloor e \rfloor = 2$ and

$$q_i = \lfloor e^{(1+r)^{i-1}} \rfloor. \tag{2}$$

We know (Bose-Chowla[1]; see, for example, [6]) that

$$|A_i| > \frac{1}{2} \left(\frac{q_i}{h} \right)^{1/h}. \tag{3}$$

Since $\frac{e^{(1+r)^{i-1}}}{2} \leq q_i \leq e^{(1+r)^{i-1}}$ and $2(2h)^{1/h} \leq 2e^{2/e} < e^2$ we have

$$|A_i| > \frac{1}{2} \left(\frac{e^{(1+r)^{i-1}}}{2h} \right)^{1/h} > e^{\frac{(1+r)^{i-1}}{h} - 2} \tag{4}$$

First we give an upper bound for $\log n$. It follows from (1) and (2) that

$$\log n < \log(q_1 \dots q_{j+1}) \leq 1 + (1+r) + \dots + (1+r)^j < \frac{(1+r)^{j+1}}{r}. \tag{5}$$

In the next step we will give a lower estimation for $\log N$. Applying (4) we have

$$\begin{aligned} \log N = \sum_{i=j-l+1}^{j} \log |A_i| &> \frac{(1+r)^{j-l} + \dots + (1+r)^{j-1}}{h} - 2l \\ &= \frac{(1+r)^j}{hr}(1 - (1+r)^{-l}) - 2l. \end{aligned} \tag{6}$$

[1]The result of Bose and Chowla implies that for every q which is a power of a prime, and for every $h \geq 2$ there is a set $C = \{c_1, \dots, c_q\} \subseteq [0, q^h - 1)$ with $r_{h,C}(n) \leq 1$. Using the well known Bertrand's postulate (for every n there is a prime $p \in (n, 2n)$) we deduce that for every M there is a set in $[0, M]$ with more than $\frac{M^{1/h}}{2}$ elements with the condition $r_h(n) \leq 1$.

In view of (5) and (6) we have

$$\begin{aligned}\frac{h\log N}{\log n} &> \frac{(1+r)^j(1-(1+r)^{-l})}{(1+r)^{j+1}}-\frac{2lrh}{(1+r)^{j+1}}\\ &= \frac{1-(1+r)^{-l}}{1+r}-\frac{2lrh}{(1+r)^{j+1}}.\end{aligned}\tag{7}$$

Using that $\frac{1}{1+r}>1-r$ and that $\left(1+\frac{\log_2 l}{l}\right)^{\frac{l}{\log_2 l}}\geq 2$ for any $l\geq 2$, we have

$$\begin{aligned}\frac{1-(1+r)^{-l}}{1+r} &> 1-r-(1+r)^{-l}=1-\frac{\log_2 l}{l}-\left(1+\frac{\log_2 l}{l}\right)^{-l}\\ &> 1-\frac{\log_2 l}{l}-\frac{1}{l}>1-\frac{2\log_2 l}{l}.\end{aligned}\tag{8}$$

On the other hand, since $\lim_{j\to\infty}\frac{2lrh}{(1+r)^{j+1}}=0$ we have that, for sufficiently large j,

$$\frac{2lrh}{(1+r)^{j+1}}<\frac{\log_2 l}{l}.\tag{9}$$

Finally, from (7), (8) and (9) we have

$$\frac{h\log N}{\log n}>1-\frac{3\log_2 l}{l},$$

for sufficiently large n.

We finish the proof of Theorem 1.1 taking, for a given $\varepsilon>0$, a large enough integer l such that $\frac{3\log_2 l}{l}<h\varepsilon$, because then $\log N>\left(\frac{1}{h}-\varepsilon\right)\log n$, i. e. $N>n^{1/h-\varepsilon}$.

Just a little comment about the dependence of g on ε. Observe that our g is $(h!)^{lh}$ and that, given ε, we need to choose a large value of l, say $l\gg\varepsilon^{-1}\log\varepsilon^{-1}$. This makes the dependence of g on ε very bad. The value of g we get with this construction depends more than exponentially on ε^{-1}. We will try to improve this in the next section and, for the case $h=3$, even more in the last one.

Note that in [3] we can find an explicit Sidon sequence with $A(x)\gg x^{1/3-o(1)}$. But this construction can not be generalized.

3 A new probabilistic proof of Theorem 1.1

Definition 3.1. *Given $0<\alpha<1$ we define $S(\alpha,m)$ as the probability space of the sequences of positive integers defined by*

$$P(x\in A)=\begin{cases}0 & \text{if } x<m\\ x^{-\alpha} & \text{if } x\geq m\end{cases}.$$

Theorem 3.2. *For any m, a random sequence A in $S(\alpha, m)$ satisfies $A(x) \gg x^{1-\alpha}$ with probability* 1.

Proof. First of all, we calculate

$$\mathbb{E}(A(x)) = \sum_{n \le x} \mathbb{P}(n \in A) = \sum_{m \le n \le x} n^{-\alpha} = \frac{x^{1-\alpha}}{1-\alpha} + O_{\alpha,m}(1),$$

when $x \to \infty$. Now, we use Chernoff's Lemma[2] to get

$$\mathbb{P}\left(A(x) \le \tfrac{1}{2}\mathbb{E}\left(A(x)\right)\right) \le \mathbb{P}\left(|A(x) - \mathbb{E}(A(x))| \ge \tfrac{1}{2}\mathbb{E}(A(x))\right) \le 2e^{-\frac{1}{16}\left(\frac{x^{1-\alpha}}{1-\alpha}+O(1)\right)}.$$

Since $\sum_{x=1}^{\infty} 2e^{-\frac{1}{16}\left(\frac{x^{1-\alpha}}{1-\alpha}+O(1)\right)} < \infty$, Borel-Cantelli Lemma[3] says that $A(x) \gg x^{1-\alpha}$ with probability 1. □

Notation 3.3. *We denote the set which elements are the coordinates of the vector $\bar{x}$ as $Set(\bar{x})$. Of course, if two or more coordinates of $\bar{x}$ are equal, this value appears only once in $Set(\bar{x})$.*

Notation 3.4. *We define*

$$R_h(n) = \{(n_1, n_2, \ldots, n_h) \mid n = n_1 + \cdots + n_h, \quad n_1 \le \cdots \le n_h, \quad n_i \in \mathbb{N}\}.$$

Lemma 3.5. *For a sequence A in $S(\alpha, m)$, for every h and n*

$$\mathbb{E}(r_{h,A}(n)) \le C_{h,\alpha}\ n^{h(1-\alpha)-1}$$

where $C_{h,\alpha}$ depends only on h and α.

Proof.

$$\begin{aligned}\mathbb{E}(r_{h,A}(n)) &= \sum_{\bar{x} \in R_h(n)} \prod_{x \in Set(\bar{x})} \mathbb{P}(x \in A) \\ &= \sum_{j=1}^{h} \sum_{\substack{\bar{x} \in R_h(n) \\ |Set(\bar{x})| = j}} \prod_{x \in Set(\bar{x})} \mathbb{P}(x \in A).\end{aligned}$$

[2] Let $X = t_1 + \cdots + t_n$ where the t_i are independent Boolean random variables. **Chernoff's Lemma** says that for every $0 < \varepsilon < 2$, $\mathbb{P}(|X - \mathbb{E}(X)| \ge \varepsilon \mathbb{E}(X)) \le 2e^{-\varepsilon^2 \mathbb{E}(X)/4}$.

[3] **Borel-Cantelli Lemma.** Let $(E_n)_{n=1}^{\infty}$ be a sequence of events in a probability space. If $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ then the probability that infinitely many of them occur is 0.

Since the largest element of every $\bar{x} \in R_h(n)$ is $\geq n/h$ we have

$$\begin{aligned}
\mathbb{E}(r_{h,A}(n)) &\leq \left(\frac{n}{h}\right)^{-\alpha} \sum_{j=1}^{h} \left(\sum_{x<n} x^{-\alpha} \right)^{j-1} \\
&\leq \left(\frac{n}{h}\right)^{-\alpha} \sum_{j=1}^{h} \left(\int_0^n x^{-\alpha} \, dx \right)^{j-1} \\
&\leq \left(\frac{n}{h}\right)^{-\alpha} \sum_{j=1}^{h} \left(\frac{n^{1-\alpha}}{1-\alpha} \right)^{j-1} \\
&\leq C_{h,\alpha} \, n^{h(1-\alpha)-1}.
\end{aligned}$$

□

Definition 3.6. *We say that two vectors $\bar{x}$ and $\bar{y}$ are disjoint if $Set(\bar{x})$ and $Set(\bar{y})$ are disjoint sets. We define $r^*_{l,A}(n)$ as the maximum number of pairwise disjoint representations of n as sum of l elements of A, i. e. the maximum number of pairwise disjoint vectors of $R_l(n)$ with their coordinates in A. We say that A is a $B^*_l[g]$ sequence if $r^*_{l,A}(n) \leq g$ for every n.*

Lemma 3.7. *For a sequence A in $S(\alpha, m)$, for every h and n*

$$\mathbb{P}(r^*_{h,A}(n) \geq s) \leq C_{h,\alpha,s} \, n^{(h(1-\alpha)-1)s}$$

where $C_{h,\alpha,s}$ depends only on h, α and s.

Proof. Using the independence given by the pairwise disjoint condition

$$\begin{aligned}
\mathbb{P}(r^*_{h,A}(n) \geq s) &= \sum_{\substack{\{\bar{x}_1,\dots,\bar{x}_s\} \\ \bar{x}_i \in R_h(n) \\ \bar{x}_1,\dots,\bar{x}_s \\ \text{pairwise disjoint}}} \prod_{i=1}^{s} \prod_{x \in Set(\bar{x}_i)} \mathbb{P}(x \in \mathcal{A}) \\
&\leq \left(\sum_{\bar{x} \in R_h(n)} \prod_{x \in Set(\bar{x})} \mathbb{P}(x \in \mathcal{A}) \right)^s \\
&= E(r_{h,A}(n))^s
\end{aligned}$$

and using Lemma 3.5 we conclude the proof. □

Proposition 3.8. *Given $h \geq 2$ and $0 < \varepsilon < 1/h$, a random sequence in $S(1 - \frac{1}{h} + \varepsilon, m)$ is a $B^*_h[g]$ sequence for every $g \geq \frac{2}{h\varepsilon}$ with probability $1 - O(\frac{1}{m})$.*

Proof. ¿From Lemma 3.7, and taking into account the value of $\alpha = 1 - \frac{1}{h} + \varepsilon$, we have

$$\mathbb{P}(r^*_{h,A}(n) \geq g+1) \leq C_{h,\varepsilon,g} n^{-h\varepsilon(g+1)}.$$

Since $r^*_{h,A}(n) = 0$ for $n < m$ we have

$$\begin{aligned} \mathbb{P}\left(r^*_{h,A}(n) \geq g+1 \text{ for some } n\right) &\leq \sum_{n\geq m} \mathbb{P}\left(r^*_{h,A}(n) \geq g+1\right) \\ &\leq \sum_{n\geq m} C_{h,\varepsilon} n^{-h\varepsilon(g+1)}. \end{aligned}$$

If $g \geq \frac{2}{h\epsilon}$, the last sum is $O(1/m)$. Thus, if it is the case,

$$\begin{aligned} \mathbb{P}\left(r^*_{h,A}(n) \leq g \text{ for every } n\right) &= 1 - \mathbb{P}\left(r^*_{h,A}(n) \geq g+1 \text{ for some } n\right) \\ &\geq 1 - O\left(\frac{1}{m}\right). \end{aligned}$$

□

The next "simple" lemma is the key idea of the proof of Theorem 1.2.

Lemma 3.9.

$$B^*_h[g] \cap B_{h-1}[k] \subseteq B_h[hkg].$$

Remark 3.10. *In fact, the "true" lemma, which is a little more ugly, is*

$$B^*_h[g] \cap B_{h-1}[k] \subseteq B_h[g(h(k-1)+1)]$$

and this is what we will prove. As an example, we can have in mind that a sequence A with $r^*_{3,A}(n) \leq g$ which is a Sidon sequence is also a $B_3[g]$ sequence, i. e. $B^*_3[g] \cap B_2[1] \subseteq B_3[g]$, since in this case two representations that share one element share the three of them.

Proof. We proceed by contradiction. Suppose that $A \in B^*_h[g] \cap B_{h-1}[k]$ and suppose that there is an n with $g(h(k-1)+1)+1$ distinct representations as a sum of h elements of A.

Fix one of these representations, say $n = x_1 + x_2 + \ldots + x_h$. How many representations of n can intersect with it? Well, the number of representations of n that involve x_1 is at most k, since $A \in B_{h-1}[k]$. So we have at most $k-1$ more representations with x_1. The same thing happens for $x_2, \ldots, x_h$. So, finally, the maximun number of representations that can intersect with the one we fixed is $h(k-1)$.

Now we fix a second representation of n that does not intersect with the first one that we chose. Again, there are at most $h(k-1)$ representations that intersect with our new choice.

After g disjoint choices, counting them and all the representations that intersect with each one of them, we have at most $gh(k-1)+g$ representations. By hipothesis, there is at least one representation of n left that does not intersect with any of our g choices. But this means that we have $g+1$ disjoint representations of n, which contradicts the fact that $A \in B_h^*[g]$. $\square$

Proposition 3.11. *For every $h \geq 2$ and $0 < \varepsilon < 1/h$ a random sequence in $S(1-\frac{1}{h}+\varepsilon, m)$ is a $B_h[g]$ sequence for every $g \geq c_h/\varepsilon$ with probability $1-O(\frac{1}{m})$, where $c_h = 2^{h-3}h(h-1)!^2$.*

Proof. We proceed by induction on h.

For $h=2$, and using Proposition 3.8, the result is true since a $B_2^*[g]$ sequence is the same that a $B_2[g]$ sequence.

Now suppose that the result is true for $h-1$. Let $\alpha = 1-\frac{1}{h}+\varepsilon$. From Proposition 3.8 we know that a random sequence in $S(\alpha, m)$ is $B_h^*[g_1]$ for every $g_1 \geq \frac{2}{h\varepsilon}$ with probability $1-O(\frac{1}{m})$. But, since $\alpha > 1-\frac{1}{h-1}+\frac{1}{h(h-1)}$, by the induction hypothesis we know that this random sequence is also $B_{h-1}[g_2]$ for every $g_2 \geq h(h-1)c_{h-1} = c_h/2$ with probability $1-O(\frac{1}{m})$. So, with probability $1-O(\frac{1}{m})$ the two things happen at the same time, i. e. the random sequence is in $B_h^*[g_1] \cap B_{h-1}[g_2]$ for every $g_1 \geq \frac{2}{h\varepsilon}$ and $g_2 \geq c_h/2$.

Lemma 3.9 concludes the proof. $\square$

Lemma 3.2 and Proposition 3.11 imply Theorem 1.2.

4 Sequences with $r_{3,A}(n)$ bounded

Now[4], we will try to find a more precise relation between g and ε. In fact, the result of Erdős and Rényi in [4] is more precise than what we said in the Introduction. They proved that for every $g > \frac{1}{2\varepsilon}-1$ there is a $B_2[g]$ sequence, A, with $A(x) \gg x^{1/2-\varepsilon}$. Stated perhaps in a more convenient way, what they proved is that for every positive integer g there is a sequence A such that $r_{2,A}(n) \leq g$ with $A(x) \geq x^{\frac{1}{2+2/g}-o(1)}$, as $x \to \infty$.

[4]Of course, our Proposition 3.11 gives a relation between g and ε, but observe that for $g=1$ it gives values of $\varepsilon \geq 1$, so it does not give any useful information. In our terminology, this is not a "satisfactory enough exponent".

In [2] the first author used the "alteration method" (perhaps our random sequences do not satisfy what we want but they do if we remove "a few" elements) to prove that for every $g > \frac{1}{4\varepsilon} - \frac{1}{2}$ there is a $B_2[g]$ sequence, A, with $A(x) \gg x^{1/2-\varepsilon}$. In other words, for every positive integer g there is a sequence A such that $r_{2,A}(n) \leq g$ with $A(x) \gg x^{\frac{1}{2+1/g}-o(1)}$ as $x \to \infty$.

In this section we will use the ideas from [2] to prove Theorem 1.3, which is a refinement on the dependence between g and ε, for sequences with $r_{3,A}(n) \leq g$.

Definition 4.1. *Given a sequence of positive integers, A, we say that x is $(g+1)_h-$bad (for A) if $x \in A$ and there exist $x_1, \dots, x_{h-1} \in A$, $x_1 \leq \cdots \leq x_{h-1} \leq x$, such that $r_{h,A}(x_1 + \cdots + x_{h-1} + x) \geq g+1$.*

In other words, $x \in A$ is $(g+1)_h-$bad if it is the largest element in a representation of an element that has more than g representations as a sum of h elements of A. Observe that A is a $B_h[g]$ sequence if and only if it does not contain $(g+1)_h-$bad elements.

Definition 4.2. *A sequence of positive integers, A, is in $\tilde{B}_h[g]$ if the number of $(g+1)_h-$bad elements less than or equal to x for A, say $\mathcal{B}(x)$, is*

$$\mathcal{B}(x) = o(A(x)) \text{ when } x \to \infty.$$

So, $A \in \tilde{B}_h[g]$ if removing a few elements from it ("a little o"), it is a $B_h[g]$ sequence.

Notation 4.3. *We denote by $\mathcal{B}_{k,h}(g+1)$ the set of $(g+1)_h-$bad elements for A in the interval $[h^k, h^{k+1})$.*

Substituting $r_{h,A}$ by $r^*_{h,A}$, we define the $(g+1)^*_h-$bad elements for A. Analogously, we define $A \in \tilde{B}^*_h[g]$ and $\mathcal{B}^*_{k,h}(g+1)$.

Obviously, the "tilde" version of Lemma 3.9 is also true. In particular, from Remark 3.10:

Lemma 4.4. *$\tilde{B}^*_3[g] \cap \tilde{B}_2[1] \subseteq \tilde{B}_3[g]$.*

Now we can write the next theorem, which we will use only in the cases $h = 2$ and $h = 3$.

Theorem 4.5. *Given $0 < \delta < \frac{1}{2h-3}$ and $h \geq 2$, a random sequence A in $S\left(\frac{2h-4}{2h-3} + \delta, m\right)$ is $\tilde{B}^*_h[g]$ for every $g > \frac{\frac{h-1}{2h-3} - (h-1)\delta}{\frac{h-3}{2h-3} + h\delta}$ with probability $1 - O\left(\frac{1}{\log m}\right)$.*

Proof. We consider a random sequence A in $S(\alpha, m)$.

$$\begin{aligned}
\mathbb{E}(|\mathcal{B}^*_{k,h}(g+1)|) &= \sum_{h^k \le x < h^{k+1}} \mathbb{P}(x \text{ is } (g+1)^*_h - \text{bad}) \\
&\le \sum_{h^k \le x < h^{k+1}} \sum_{\substack{\bar{x}_{g+1}=(y_1,\dots,y_{h-1},x) \\ y_1 \le \dots \le y_{h-1} \le x}} \prod_{z \in Set(\bar{x}_{g+1})} \mathbb{P}(z \in A) \cdot \\
&\quad \cdot \sum_{\substack{\{\bar{x}_1,\dots,\bar{x}_g\} \\ \bar{x}_i \in R_h(y_1+\dots+y_{h-1}+x) \\ \bar{x}_1,\dots,\bar{x}_g,\bar{x}_{g+1} \\ \text{pairwise disjoint}}} \prod_{i=1}^{g} \prod_{z \in Set(\bar{x}_i)} \mathbb{P}(z \in A) \\
&\le \sum_{h^k \le n < h^{k+2}} \left(\sum_{\bar{x} \in R_h(n)} \prod_{x \in Set(\bar{x})} \mathbb{P}(x \in A) \right)^{g+1} \\
&\le \sum_{h^k \le n < h^{k+2}} \left(C n^{h(1-\alpha)-1} \right)^{g+1} \\
&\ll \sum_{h^k \le n < h^{k+2}} n^{(h-1-h\alpha)(g+1)} \\
&\ll h^{k((h-1-h\alpha)(g+1)+1)}
\end{aligned}$$

when $k \to \infty$, where we have used Lemma 3.5.

Now, we can use Markov's Inequality[5] to have:

$$\mathbb{P}(|\mathcal{B}^*_{k,h}(g+1)| \ge k^2 \mathbb{E}(|\mathcal{B}^*_{k,h}(g+1)|)) \le \frac{1}{k^2}.$$

Since $|\mathcal{B}^*_{k,h}(g+1)| = 0$ for $h^{k+1} < m$ we have

$$\mathbb{P}\left(|\mathcal{B}^*_{k,h}(g+1)| \ge k^2 \mathbb{E}(|\mathcal{B}^*_{k,h}(g+1)|) \text{ for some } k\right) \le \sum_{k \ge \log_h m - 1} \frac{1}{k^2} = O\left(\frac{1}{\log m}\right),$$

so with probability $1 - O\left(\frac{1}{\log m}\right)$ we have that $|\mathcal{B}^*_{k,h}(g+1)| \ll k^2 h^{k((h-1-h\alpha)(g+1)+1)}$ for every k.

On the other hand, by Theorem 3.2 we know that $A(x) \gg x^{1-\alpha}$ with probability 1.

So, given x, we will have $h^l \le x < h^{l+1}$ for some l and with probability $1 - O\left(\frac{1}{\log m}\right)$ the number of bad elements less than or equal to x will be

$$\mathcal{B}(x) \le \sum_{k=0}^{l} |\mathcal{B}^*_{k,h}(g+1)| \ll l^2 h^{l((h-1-h\alpha)(g+1)+1)}$$

[5]**Markov's Inequality.** For a random variable X and $a > 0$, $\mathbb{P}(|X| \ge a) \le \frac{\mathbb{E}(|X|)}{a}$.

while the number of elements in A less than or equal to x will be

$$A(x) \gg h^{l(1-\alpha)}.$$

Since, in order for A to be in $\tilde{B}_h^*[g]$, we want $\mathcal{B}(x) = o(A(x))$ we need

$$(h-1-h\alpha)(g+1)+1 < 1-\alpha$$

and so, with $\alpha = \frac{2h-4}{2h-3} + \delta$ we have

$$g > \frac{\frac{h-1}{2h-3} - (h-1)\delta}{\frac{h-3}{2h-3} + h\delta}.$$

□

In particular, for $h = 2$, since a $\tilde{B}_2^*[g]$ sequence is also a $\tilde{B}_2[g]$ sequence, we deduce that given $0 < \varepsilon < \frac{1}{3}$, a random sequence A in $S\left(\frac{2}{3} + \varepsilon, m\right)$ is $\tilde{B}_2[1]$ with probability $1 - O\left(\frac{1}{\log m}\right)$.

Also, for $h = 3$, we deduce that given $0 < \delta < \frac{1}{3}$, a random sequence A in $S\left(\frac{2}{3} + \delta, m\right)$ is $\tilde{B}_3^*[g]$ for every $g > \frac{2}{9\delta} - \frac{2}{3}$ with probability $1 - O\left(\frac{1}{\log m}\right)$.

Lemma 4.4 gives the proof of Theorem 1.3.

References

[1] N. Alon, J. H. Spencer, The Probabilistic Method, Wiley-Interscience (2000).

[2] J. Cilleruelo, Probabilistic constructions of $B_2[g]$ sequences, To appear in Acta Mathematica Sinica (2009).

[3] J. Cilleruelo, Sidon sets in $\mathbb{N}^d$, Preprint (2009).

[4] P. Erdős and A. Rényi, Additive properties of random sequences of positive integers, Acta Arithmetica, 6, 85-110 (1960).

[5] P. Erdős and P. Tetali, Representations of Integers as the Sum of k terms, Random Structures and Algorithms, Vol. 1, N. 3 (1990).

[6] H. Halberstam and K. F. Roth, Sequences, Springer - Verlag, New York (1983).

[7] T. Tao, V. H. Vu, Additive Combinatorics, Cambridge University Press (2006).

[8] V. H. Vu, On a refinement of Waring's problem, Duke Math. J. Volume 105, Number 1 (2000), 107-134.