Probabilistic constructions of $B_2[g]$ sequences

Javier Cilleruelo*

Mathematics Departament
Universidad Autónoma de Madrid, 28049 Madrid, Spain
franciscojavier.cilleruelo@uam.es

June 3, 2008

Abstract

We use the probabilistic method to prove that for any positive integer g there exists an infinite $B_2[g]$ sequence $\mathcal{A} = \{a_k\}$ such that $a_k \ll k^{2+1/g}(\log k)^{1/g+o(1)}$ as $k \to \infty$. The exponent $2 + 1/g$ improves the previous, $2 + 2/g$, obtained by Erdős and Renyi in 1960. We obtain a similar result for $B_2[g]$ sequences of squares.

1 Introduction

Given a sequence $\mathcal{A} = \{a_k\}$ of positive integers the function $r_{\mathcal{A}}(n)$ counts the number of representations of $n = x + y,\ y \le x,\ x, y \in \mathcal{A}$. Those sequences satisfying $r_{\mathcal{A}}(n) \le g$ for all integer $n \ge 1$ are called $B_2[g]$ sequences.

For $B_2[1]$ sequences, which are also called Sidon sets, it is known [4] that the upper bound $a_k \ll k^2$ cannot hold and it is an old open problem to decide whether $a_k \ll k^2$ cannot hold for $B_2[g]$ sequences with $g \ge 2$ either. Erdős has conjectured that for all $\varepsilon > 0$ there exists a Sidon set such that $a_k \ll k^{2+\varepsilon}$.

In 1960, Erdős and Renyi [3] used the probabilistic method to prove that for all positive integers g there exists a $B_2[g]$ sequence $\mathcal{A} = \{a_k\}$ with

$$a_k \le k^{2+2/g+o(1)} \qquad \text{as } k \to \infty. \tag{1}$$

Given a suitable sequence of real numbers $(p_n)_{n\ge 1}$ in the interval $[0, 1]$, they considered the probability space of all sequences $\mathcal{A}$ of positive integers defined by $\mathbb{P}(n \in \mathcal{A}) = p_n$ and

*This work was supported by Grant MTM 2005-04730 of MYCIT (Spain)

proved that almost all sequences in this space have the $B_2[g]$ property by first checking that $\sum_n \mathbb{P}(r_{\mathcal{A}}(n) \geq g+1) < \infty$, and then applying the Borel-Cantelli lemma to conclude that with probability 1, the sequences satisfy the $B_2[g]$ property after removing a finite number of elements. See [6] for an excellent exposition of known results on this problem and other applications of the probabilistic method to additive number theory.

In our approach, we conclude that the same phenomenon happens although in this case the exceptional set is not finite but has only a few elements in each dyadic interval. Then we modify the sequence by removing these bad elements.

Theorem 1 *Let $(p_n)_{n\geq 1}$ be a sequence of numbers in $[0,1]$ with $\lim_{t\to\infty} \frac{1}{\log t} \sum_{n\leq t} p_n = \infty$. Suppose that for a positive integer g we have that*

$$\sum_{k\geq 1} \frac{1 + \sum_{2^k \leq n < 2^{k+2}} \left(\sum_x p_x p_{n-x}\right)^{g+1}}{\sum_{2^{k-1} \leq n < 2^k} p_n} < \infty. \tag{2}$$

Then there exists a $B_2[g]$ sequence $\mathcal{A} \subset \{n,\ p_n > 0\}$ such that $\mathcal{A}(x) \sim \sum_{n\leq x} p_n$.

As a consequence of Theorem 1 we obtain a new upper bound on the growth of infinite $B_2[g]$ sequences, which is an improvement upon (1).

Theorem 2 *For all positive integers g there exists a $B_2[g]$ sequence $\mathcal{A} = \{a_k\}$ such that the inequality $a_k \leq k^{2+1/g} (\log k)^{1/g+o(1)}$ holds as $k \to \infty$.*

Theorem 1 also us also to study the $B_2[g]$ sequences included in some special set S simply by taking $p_n = 0$ for $n \notin S$. We next concentrate on the set S of perfect squares.

The sequence of squares is not a $B_2[g]$ sequence for any given g although the number of representations of a positive integer n as a sum of two squares is bounded by $n^{o(1)}$ as $n \to \infty$. Thus, it is a natural question to ask for dense $B_2[g]$ sequences of squares.

In [1], we adapted the probabilistic method and proved that for any positive integer g there exists a $B_2[g]$ sequence of squares $\mathcal{A} = \{a_k\}$ such that the inequality $a_k \leq k^{2+2/g+o(1)}$ holds as $k \to \infty$. In [2], we removed the term $o(1)$ when $g = 1$.

Here, we apply Theorem 1 and improve upon these upper bounds.

Theorem 3 *For any positive integer g there exists a $B_2[g]$ sequence of* squares $\mathcal{A} = \{a_k\}$ *such that $a_k < k^{2+1/g}(\log k)^{\kappa_g}$ holds for all $k \geq 2$, where κ_g is some positive constant depending on g.*

2 The probabilistic method and applications

As usual, we work with the probability space of all the sequences of positive integers $\mathcal{A}$ defined by

$$\mathbb{P}[n \in \mathcal{A}] = p_n, \tag{3}$$

where $(p_n)_{n\geq 1} \subset [0,1]$ is given.

Definition 1 *Given a sequence of positive integers $\mathcal{A}$, we say that x is $(g+1)$-bad (for $\mathcal{A}$) if $x \in \mathcal{A}$ and there exists $y \in \mathcal{A},\ y \leq x$ such that $r_{\mathcal{A}}(x+y) \geq g+1$.*

In other words, $x \in \mathcal{A}$ is $(g+1)$-bad if x is involved in the representation of a positive integer $n \leq 2x$ which has more than g representations as sum of two elements of $\mathcal{A}$. We observe that $\mathcal{A}$ is a $B_2[g]$ sequence if and only if $\mathcal{A}$ doesn't contain $(g+1)$-bad elements.

Lemma 1 *For any integer $k \geq 0$,*

$$\mathbb{E}(|\mathcal{B}_k|) \leq 2^{g+2} \sum_{2^k \leq n < 2^{k+2}} \left(\sum_y p_y p_{n-y} \right)^{g+1} + 2^{g+2},$$

where $\mathcal{B}_k$ denotes the set of $(g+1)$-bad elements in the interval $[2^k, 2^{k+1})$.

Proof. Write $p_x = \mathbb{P}[x \in \mathcal{A}]$. It is useful to write

$$p(x,y) = \mathbb{P}[x \in \mathcal{A}, y \in \mathcal{A}] = \begin{cases} p_x p_y, & y \neq x \\ p_x, & y = x, \end{cases} \qquad \text{and} \qquad s(n) = \sum_{x+y=n} p(x,y).$$

We have that

$$\begin{aligned} \mathbb{P}(x \text{ is } (g+1)\text{-bad}) &\leq \sum_{y \leq x} p(x,y) \sum_{\substack{x_1,y_1,\dots,x_g,y_g \\ x_i+y_i=x+y}} p(x_1,y_1)\cdots p(x_g,y_g) \\ &\leq \sum_{y\leq x} p(x,y) s^g(x+y). \end{aligned} \tag{4}$$

Then

$$\begin{aligned} \mathbb{E}(|\mathcal{B}_k|) = &\sum_{2^k \leq x < 2^{k+1}} \mathbb{P}(x \text{ is } (g+1)\text{-bad}) \leq \sum_{2^k \leq x < 2^{k+1}} \sum_{y \leq x} p(x,y) s^g(x+y) \\ &\leq \sum_{2^k \leq n < 2^{k+2}} s^g(n) \sum_{\substack{y \leq x \\ x+y=n}} p(x,y) = \sum_{2^k \leq n < 2^{k+2}} s^{g+1}(n) = \sum_{2^k \leq n < 2^{k+2}} \left(p_{n/2} + \sum_y p_y p_{n-y} \right)^{g+1} \\ &\leq 2^{g+1} \sum_{2^{k-1} \leq m < 2^{k+1}} p_m^{g+1} + 2^{g+1} \sum_{2^k \leq n < 2^{k+2}} \left(\sum_y p_y p_{n-y} \right)^{g+1}. \end{aligned} \tag{5}$$

Now note that

$$\left(\sum_{2^{k-1}\le m<2^{k+1}} p_m^{g+1}\right)^2 \le \sum_{2^k\le n<2^{k+2}}\sum_y p_y^{g+1}p_{n-y}^{g+1} \le \sum_{2^k\le n<2^{k+2}}\left(\sum_y p_y p_{n-y}\right)^{g+1},$$

and the lemma follows using the estimate $\sqrt{z}+z\le 2z+2$. ■

Lemma 2 *Given a convergent series of positive numbers $\sum_k c_k<\infty$, there exists a sequence $d_k\to\infty$ such that $\sum_k c_k d_k<\infty$.*

Proof. Let $k_1<\cdots<k_n\cdots$ be an infinite sequence such that $\sum_{k\ge k_n} c_k<2^{-n}$. Define $d_k=1$ for $k\le k_1$ and $d_k=n+1$ for $k_n<k\le k_{n+1}$ and $n\ge 1$. Then $d_k\to\infty$ and

$$\sum_k c_k d_k=\sum_{k\le k_1} c_k+\sum_{1\le n} n\sum_{k_n\le k<k_{n+1}} c_k\le \sum_{k\le k_1} c_k+\sum_n (n+1)2^{-n}<\infty.$$

■

Proof of Theorem 1. For $k\ge 1$ let $\mathcal{B}_k$ be the set defined in Lemma 1 and write $\mathcal{B}=\cup_k\mathcal{B}_k$. We then have that $\tilde{\mathcal{A}}=\mathcal{A}\setminus\mathcal{B}$ is a $B_2[g]$ sequence. Next we will prove that for almost all sequences $\mathcal{A}$, the sequence $\mathcal{B}$, which depends on $\mathcal{A}$, is a thinner sequence than $\mathcal{A}$. Write

$$c_k=\frac{1+\sum_{2^k\le n<2^{k+2}}\left(\sum_x p_x p_{n-x}\right)^{g+1}}{\sum_{2^{k-1}\le n<2^k} p_n},$$

and let $\{d_k\}$ a sequence as in Lemma 2. We have that

$$\begin{aligned}\sum_k \mathbb{P}\left(|\mathcal{B}_k|>\frac{\sum_{2^{k-1}\le n<2^k} p_n}{d_k}\right)&=\sum_k \mathbb{P}\left(|\mathcal{B}_k|>\frac{1+\sum_{2^k\le n<2^{k+2}}\left(\sum_x p_x p_{n-x}\right)^{g+1}}{c_k d_k}\right)\\&<\sum_k \mathbb{P}\left(|\mathcal{B}_k|>\frac{\mathbb{E}(|\mathcal{B}_k|)}{2^{g+2}c_k d_k}\right)\le 2^{g+2}\sum_k c_k d_k<\infty,\end{aligned}\tag{6}$$

where we have used Lemma 1 and Markov's inequality.

Now the Borell-Cantelli lemma ensures that for almost all sequences $\mathcal{A}$ we have that

$$|\mathcal{B}_k|\le\frac{\sum_{2^{k-1}\le n<2^k} p_n}{d_k}\tag{7}$$

except for a finite number of cases. Since $d_k\to\infty$, it follows that for almost all sequences $\mathcal{A}$ we have that $|\mathcal{B}_k|=o\left(\sum_{2^{k-1}\le n<2^k} p_n\right)$ as $k\to\infty$. Thus, for almost all sequences $\mathcal{A}$ and for all x, letting k be such that $2^k\le x<2^{k+1}$, we have that

$$\mathcal{B}(x)\le\mathcal{B}(2^{k+1})=\sum_{j\le k}|\mathcal{B}_j|=o(\sum_{n\le 2^k} p_n)=o(\sum_{n\le x} p_n)\qquad\text{as } x\to\infty.$$

On the other hand, Chernoff's inequality and the condition $\lim_{x\to\infty}\frac{1}{\log x}\sum_{n\le x}p_n=\infty$ are sufficient to conclude that $\mathcal{A}(x)\sim\sum_{n\le x}p_n$ with probability 1, therefore

$$\tilde{\mathcal{A}}(x)=\mathcal{A}(x)-\mathcal{B}(x)\sim\sum_{n\le x}p_n,$$

which is what we wanted to show. $\blacksquare$

Proof of Theorem 2. Take

$$p_n=n^{-\frac{g+1}{2g+1}}(\log n)^{-\frac{1}{2g+1}}(\log\log n)^{-\frac{1}{2}} \tag{8}$$

for $n>e^e$, and $p_n=1$ otherwise. It is easy to check that

$$\sum_y p_yp_{n-y}\ll cn^{-\frac{1}{2g+1}}(\log n)^{-\frac{2}{2g+1}}(\log\log n)^{-1}\ll 2^{-\frac{k}{2g+1}}k^{-\frac{2}{2g+1}}(\log k)^{-1}$$

for $n\ge 2^{k-1}$. Then, the sum in (2) can be bounded by

$$\sum_k\frac{1+\sum_{2^k\le n<2^{k+2}}2^{-\frac{k(g+1)}{2g+1}}k^{-\frac{2(g+1)}{2g+1}}(\log k)^{-(g+1)}}{\sum_{2^{k-1}\le n<2^k}2^{-\frac{k(g+1)}{2g+1}}k^{-\frac{1}{2g+1}}(\log k)^{-1/2}}\ll\sum_k\frac{1}{k(\log k)^{g+1/2}}<\infty.$$

Thus, Theorem 1 implies that there exists a $B_2[g]$ sequence $\mathcal{A}=\{a_k\}$ such that $\mathcal{A}(x)\sim\sum_{n\le x}p_n$. It can be checked easily that this last sum is

$$\sim\frac{2g+1}{g}\frac{x^{\frac{g}{2g+1}}}{(\log x)^{1/(2g+1)}(\log\log x)^{1/2}}.$$

Inverting the above estimate we get that $a_k\sim\left(\frac{g}{2g+1}k\right)^{2+1/g}(\log k)^{1/g}(\log\log k)^{1+1/(2g)}$. $\blacksquare$

Proof of Theorem 3. Write $p_n=\begin{cases}q_m, & n=m^2,\\ 0, & n \text{ is not an square}\end{cases}$ and denote by $r(n)$ the number of representations of n as a sum of two squares. For n with $r(n)\ne 0$ we define $s_n=\min\{s,\ s^2+t^2=n\}$ and t_n by $n=s_n^2+t_n^2$. Then, if the sequence q_t is decreasing, we can apply Hölder's inequality to get

$$\begin{aligned}\sum_{2^{k-1}\le n<2^k}\left(\sum_x p_xp_{n-x}\right)^{g+1} &= \sum_{2^{k-1}\le n<2^k}\left(\sum_{\substack{l,m,\\ l^2+m^2=n}} q_lq_m\right)^{g+1} \ll \sum_{2^{k-1}\le n<2^k}(q_{s_n}q_{t_n}r(n))^{g+1}\\ &\ll \left(\sum_{2^{k-1}\le n<2^k} q_{s_n}^{2g+1}q_{t_n}^{2g+1}\right)^{\frac{g+1}{2g+1}}\left(\sum_{2^{k-1}\le n<2^k} r(n)^{\frac{(g+1)(2g+1)}{g}}\right)^{\frac{g}{2g+1}}\\ &\ll \left(\sum_{2^{(k-1)/2}\le t<2^{k/2}} q_t^{2g+1}\sum_{s<2^{(k-1)/2}} q_s^{2g+1}\right)^{\frac{g+1}{2g+1}}\left(\sum_{2^{k-1}\le n<2^k} r(n)^{2g+4}\right)^{\frac{g}{2g+1}}.\end{aligned} \tag{9}$$

Taking $q_t = t^{-\frac{1}{2g+1}}(\log t)^{-\beta_g}$ with $\beta_g = 2^{2g+3}$, we get $\sum_s q_s^{2g+1} < \infty$, and the first factor of the last formula above is $\ll k^{-\beta_g(g+1)}$. To estimate the second factor, note that $\sum_{n\le x} r^l(n) \ll x(\log x)^{2^{l-1}-1}$, so this factor is $\ll 2^{k\frac{g}{2g+1}}k^{(2^{2g+3}-1)\frac{g}{2g+1}}$. Thus,

$$1+\sum_{2^{k-1}\le n<2^{k+2}}\left(\sum_x p_xp_{n-x}\right)^{g+1} \ll 2^{k\frac{g}{2g+1}}k^{(2^{2g+3}-1)\frac{g}{2g+1}-\beta_g(g+1)}.$$

On the other hand it is easy to check that $\sum_{2^{k-1}\le n<2^k} p_n \gg 2^{k\frac{g}{2g+1}}k^{-\beta_g}$. Then we have that

$$(2) \ll \sum_k k^{(2^{2g+3}-1)\frac{g}{2g+1}-g\beta_g} < \infty.$$

To finish the proof, note that with probability 1 we have that

$$\mathcal{A}(x) \sim \sum_{n\le x} p_n \sim \sum_{m\le x^{1/2}} q_m \gg x^{g/(2g+1)}(\log x)^{-\beta_g},$$

which implies that $a_k \ll x^{2+1/g}(\log k)^{\kappa_g}$ for the terms of the sequence $\mathcal{A} = \{a_k\}$ with a positive constant κ_q depending on g.

$\blacksquare$

References

[1] J. Cilleruelo, B_2 sequences whose terms are squares, *Acta Arithmetica* **55** (1990),

[2] J.Cilleruelo, $B_2[g]$ sequences whose terms are squares, *Acta Math. Hungar.* **67** (1995),

[3] P. Erdős and A. Renyi, Additive properties of random sequences of positive integers, *Acta Arith.* **6** (1960), 83–110.

[4] P. Erdős and P. Turan, On a problem of Sidon in additive number theory, and on some related problems, *J. London Math. Soc.* **16** (1941), 212-215.

[5] H. Halberstam and K.F. Roth, Sequences, Second edition, *Springer Verlag* (1983).

[6] M. Kolountzakis, Some Applications of Probability to Additive Number Theory, New York Number Theory Seminar, CUNY, (1994).