Csaba SÁNDOR

Non-degenerate Hilbert cubes in random sets

<http://jtnb.cedram.org/item?id=JTNB_2007__19_1_249_0>

© Université Bordeaux 1, 2007, tous droits réservés.

L’accès aux articles de la revue « Journal de Théorie des Nombres de Bordeaux » (http://jtnb.cedram.org/), implique l’accord avec les conditions générales d’utilisation (http://jtnb.cedram.org/legal/). Toute reproduction en tout ou partie cet article sous quelque forme que ce soit pour tout usage autre que l’utilisation à fin strictement personnelle du copiste est constitutive d’une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.
Non-degenerate Hilbert cubes in random sets

par Csaba SÁNDOR

Abstract. A slight modification of the proof of Szemerédi’s cube lemma gives that if a set $S \subset [1, n]$ satisfies $|S| \geq \frac{n}{2}$, then $S$ must contain a non-degenerate Hilbert cube of dimension $\lfloor \log_2 \log_2 n - 3 \rfloor$. In this paper we prove that in a random set $S$ determined by $\Pr\{s \in S\} = \frac{1}{2}$ for $1 \leq s \leq n$, the maximal dimension of non-degenerate Hilbert cubes is a.e. nearly $\log_2 \log_2 n + \log_2 \log_2 \log_2 n$ and determine the threshold function for a non-degenerate $k$-cube.

1. Introduction

Throughout this paper we use the following notations: let $[1, n]$ denote the first $n$ positive integers. The coordinates of the vector $A_{(k,n)} = (a_0, a_1, \ldots, a_k)$ are selected from the positive integers such that $\sum_{i=0}^{k} a_i \leq n$. The vectors $B_{(k,n)}$, $A_{(k,n)}$ are interpreted similarly. The set $S_n$ is a subset of $[1, n]$. The notations $f(n) = o(g(n))$ means $\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$. An arithmetic progression of length $k$ is denoted by $AP_k$. The rank of a matrix $A$ over the field $\mathbb{F}$ is denoted by $r_{\mathbb{F}}(A)$. Let $\mathbb{R}$ denote the set of real numbers, and let $\mathbb{F}_2$ be the finite field of order 2.

Let $n$ be a positive integer, $0 \leq p_n \leq 1$. The random set $S(n, p_n)$ is the random variable taking its values in the set of subsets of $[1, n]$ with the law determined by the independence of the events $\{k \in S(n, p_n)\}$, $1 \leq k \leq n$ with the probability $\Pr\{k \in S(n, p_n)\} = p_n$. This model is often used for

Manuscrit reçu le 30 novembre 2005.
Supported by Hungarian National Foundation for Scientific Research, Grant No. T 49693 and 61908.
proving the existence of certain sequences. Given any combinatorial number theoretic property $P$, there is a probability that $S(n, p_n)$ satisfies $P$, which we write $\Pr\{S(n, p_n) = P\}$. The function $r(n)$ is called a threshold function for a combinatorial number theoretic property $P$ if

(i) When $p_n = o(r(n))$, $\lim_{n \to \infty} \Pr\{S(n, p_n) = P\} = 0$,

(ii) When $r(n) = o(p(n))$, $\lim_{n \to \infty} \Pr\{S(n, p_n) = P\} = 1$,

or visa versa. It is clear that threshold functions are not unique. However, threshold functions are unique within factors $m(n)$, $0 < \lim \inf_{n \to \infty} m(n) \leq \lim \sup_{n \to \infty} m(n) < \infty$, that is if $p_n$ is a threshold function for $P$ then $p'_n$ is also a threshold function iff $p_n = O(p'_n)$ and $p'_n = O(p_n)$. In this sense we can speak of the threshold function of a property.

We call $H \subset [1, n]$ a Hilbert cube of dimension $k$ or, simply, a $k$-cube if there is a vector $A^{(k, n)}$ such that

$$H = H_A^{(k, n)} = \{a_0 + \sum_{i=1}^k \epsilon_i a_i : \epsilon_i \in \{0, 1\}\}.$$ 

The positive integers $a_1, \ldots, a_k$ are called the generating elements of the Hilbert cube. The $k$-cube is non-degenerate if $|H| = 2^k$ i.e. the vertices of the cube are distinct, otherwise it is called degenerate. The maximal dimension of a non-degenerate Hilbert cube in $S_n$ is denoted by $H_{max}(S_n)$, i.e. $H_{max}(S_n)$ is the largest integer $l$ such that there exists a vector $A^{(l, n)}$ for which the non-degenerate Hilbert cube $H_A^{(l, n)} \subset S_n$.

Hilbert originally proved that if the positive integers are colored with finitely many colors then one color class contains a $k$-cube. The density version of theorem was proved by Szemerédi and has since become known as "Szemerédi’s cube lemma". The best known result is due to Gunderson and Rödl (see [3]):

**Theorem 1.1** (Gunderson and Rödl). For every $d \geq 3$ there exists $n_0 \leq (2^d - 2/\ln 2)^2$ so that, for every $n \geq n_0$, if $A \subset [1, n]$ satisfies $|A| \geq 2n^{1 - \frac{1}{2d-1}}$, then $A$ contains a $d$-cube.

A direct consequence is the following:

**Corollary 1.2.** Every subset $S_n$ such that $|S_n| \geq \frac{n}{2}$ contains a $\lfloor \log_2 \log_2 n \rfloor$-cube.

A slight modification of the proof gives that the above set $S_n$ must contain a non-degenerate $\lfloor \log_2 \log_2 n - 3 \rfloor$-cube.

Obviously, a sequence $S$ has the Sidon property (that is the sums $s_i + s_j$, $s_i \leq s_j$, $s_i, s_j \in S$ are distinct) iff $S$ contains no 2-cube. Godbole, Janson, Locantore and Rapoport studied the threshold function for the Sidon property and gave the exact probability distribution in 1999 (see [2]):
Theorem 1.3 (Godbole, Janson, Locantore and Rapoport). Let $c > 0$ be arbitrary. Let $P$ be the Sidon property. Then with $p_n = cn^{-3/4}$,

$$\lim_{n \to \infty} Pr\{S(n, p_n) \models P\} = e^{-c/12}.$$ 

Clearly, a subset $H \subset [1, n]$ is a degenerate 2-cube iff it is an $AP_3$. Moreover, an easy argument gives that the threshold function for the event "$AP_3$-free" is $p_n = n^{-2/3}$. Hence

Corollary 1.4. Let $c > 0$ be arbitrary. Then with $p_n = cn^{-3/4}$,

$$\lim_{n \to \infty} Pr\{S(n, p_n) \text{ contains no non-degenerate 2-cube}\} = e^{-c/12}.$$ 

In Theorem 1.5 we extend the previous Corollary.

Theorem 1.5. For any real number $c > 0$ and any integer $k \geq 2$, if $p_n = cn^{k+1/2^k}$,

$$\lim_{n \to \infty} Pr\{S(n, p_n) \text{ contains no non-degenerate } k\text{-cube}\} = e^{-c^k/(k+1)!k!}.$$ 

In the following we shall find bounds on the maximal dimension of non-degenerate Hilbert cubes in the random set $S(n, 1/2)$. Let

$$D_n(\epsilon) = \lfloor \log_2 \log_2 n + \log_2 \log_2 \log_2 n + \frac{(1 - \epsilon) \log_2 \log_2 \log_2 n}{\log_2 \log_2 \log_2 n} \rfloor$$

and

$$E_n(\epsilon) = \lfloor \log_2 \log_2 n + \log_2 \log_2 \log_2 n + \frac{(1 + \epsilon) \log_2 \log_2 \log_2 n}{\log_2 \log_2 \log_2 n} \rfloor.$$ 

The next theorem implies that for almost all $n$, $H_{\max}(S(n, 1/2))$ concentrates on a single value because for every $\epsilon > 0$, $D_n(\epsilon) = E_n(\epsilon)$ except for a sequence of zero density.

Theorem 1.6. For every $\epsilon > 0$

$$\lim_{n \to \infty} Pr\{D_n(\epsilon) \leq H_{\max}(S(n, 1/2)) \leq E_n(\epsilon)\} = 1.$$ 

2. Proofs

In order to prove the theorems we need some lemmas.

Lemma 2.1. For $k_n = o\left(\frac{\log n}{\log \log n}\right)$ the number of non-degenerate $k_n$-cubes in $[1, n]$ is $(1 + o(1))(\frac{n}{k_n+1})^{1/k_n!}$, as $n \to \infty$. 

Proof. All vectors $A^{(k_n,n)}$ are in 1-1 correspondence with all vectors $(v_0, v_1, \ldots, v_{k_n})$ with $1 \leq v_1 < v_2 < \cdots < v_{k_n} \leq n$ in $\mathbb{R}^{k_n+1}$ according to the formulas $(a_0, a_1, \ldots, a_{k_n}) \mapsto (v_0, v_1, \ldots, v_{k_n}) = (a_0, a_0 + a_1, \ldots, a_0 + a_1 + \cdots + a_{k_n})$; and $(v_0, v_1, \ldots, v_{k_n}) \mapsto (a_0, a_1, \ldots, a_{k_n}) = (v_0, v_1 - v_0, \ldots, v_{k_n} - v_{k_n-1})$. Consequently,

$$\binom{n}{k_n + 1} = |\{A^{(k_n,n)} : H_A^{(k_n,n)} \text{ is non-degenerate}\}|$$

$$+ |\{A^{(k_n,n)} : H_A^{(k_n,n)} \text{ is degenerate}\}|.$$

By the definition of a non-degenerate cube the cardinality of the set $\{A^{(k_n,n)} : H_A^{(k_n,n)} \text{ is non-degenerate}\}$ is equal to

$$k_n!|\{\text{non-degenerate } k_n\text{-cubes in } [1, n]\}|,$$

because permutations of $a_1, \ldots, a_k$ give the same $k_n$-cube. It remains to verify that the number of vectors $A^{(k_n,n)}$ which generate degenerate $k_n$-cubes is $o\left(\binom{n}{k_n+1}\right)$. Let $A^{(k_n,n)}$ be a vector for which $H_A^{(k_n,n)}$ is a degenerate $k_n$-cube. Then there exist integers $1 \leq u_1 < u_2 < \cdots < u_s \leq k_n$, $1 \leq v_1 < v_2 < \cdots < v_t \leq k_n$ such that

$$a_0 + a_{u_1} + \cdots + a_{u_s} = a_0 + a_{v_1} + \cdots + a_{v_t},$$

where we may assume that the indices are distinct, therefore $s + t \leq k_n$. Then the equation

$$x_1 + x_2 + \cdots + x_s - x_{s+1} - \cdots - x_{s+t} = 0$$

can be solved over the set $\{a_1, a_2, \ldots, a_{k_n}\}$. The above equation has at most $n^{s+t-1} \leq n^{k_n-1}$ solutions over $[1, n]$. Since we have at most $k_n^2$ possibilities for $(s, t)$ and at most $n$ possibilities for $a_0$, therefore the number of vectors $A^{(k_n,n)}$ for which $H_A^{(k_n,n)}$ is degenerate is at most $k_n^2 n^{k_n} = o\left(\binom{n}{k_n+1}\right)$. \hfill $\square$

In the remaining part of this section the Hilbert cubes are non-degenerate.

The proofs of Theorem 1.5 and 1.6 will be based on the following definition. For two intersecting $k$-cubes $H_A^{(k,n)}, H_B^{(k,n)}$ let $H_A^{(k,n)} \cap H_B^{(k,n)} = \{c_1, \ldots, c_m\}$ with $c_1 < \cdots < c_m$, where

$$c_d = a_0 + \sum_{l=1}^{k} \alpha_{d,l}a_l = b_0 + \sum_{l=1}^{k} \beta_{d,l}b_l, \quad \alpha_{d,l}, \beta_{d,l} \in \{0, 1\}$$

for $1 \leq d \leq m$ and $1 \leq l \leq k$. The rank of the intersection of two $k$-cubes $H_A^{(k,n)}, H_B^{(k,n)}$ is defined as follows: we say that $r(H_A^{(k,n)}, H_B^{(k,n)})=(s,t)$ if for the matrices $A = (\alpha_{d,l})_{m \times k}, B = (\beta_{d,l})_{m \times k}$ we have $r_R(A) = s$ and $r_R(B) = t$. The matrices $A$ and $B$ are called matrices of the common vertices of $H_A^{(k,n)}, H_B^{(k,n)}$. 

Lemma 2.2. The condition \( r(H_{A^{(k,n)}}, H_{B^{(k,n)}}) = (s, t) \) implies that
\[
|H_{A^{(k,n)}} \cap H_{B^{(k,n)}}| \leq 2^{\min\{s,t\}}.
\]

Proof. We may assume that \( s \leq t \). The inequality \( |H_{A^{(k,n)}} \cap H_{B^{(k,n)}}| \leq 2^s \) is obviously true for \( s = k \). Let us suppose that \( s < k \) and the number of common vertices is greater than \( 2^s \). Then the corresponding \((0-1)\)-matrices \( A \) and \( B \) have more than \( 2^s \) different rows, therefore \( r_{\mathbb{R}_2}(A) > s \), but we know from elementary linear algebra that for an arbitrary \((0-1)\)-matrix \( M \) we have \( r_{\mathbb{R}_2}(M) \geq r_{\mathbb{R}}(M) \), which is a contradiction. \( \square \)

Lemma 2.3. Suppose that the sequences \( A^{(k,n)} \) and \( B^{(k,n)} \) generate non-degenerate \( k \)-cubes. Then
\[
\begin{align*}
(1) \quad & |\{(A^{(k,n)}, B^{(k,n)}) : r(H_{A^{(k,n)}}, H_{B^{(k,n)}}) = (s, t)\}| \\
& \leq 2^{2k^2(\frac{n}{k+1})n^{k+1-max\{s,t\}}}
\end{align*}
\]
for all \( 0 \leq s, t \leq k \);
\[
(2) \quad |\{(A^{(k,n)}, B^{(k,n)}) : r(H_{A^{(k,n)}}, H_{B^{(k,n)}}) = (r, r), |H_{A^{(k,n)}} \cap H_{B^{(k,n)}|} = 2^r\}| \leq 2^{2k^2(\frac{n}{k+1})n^{k-r}}
\]
for all \( 0 \leq r < k \);
\[
(3) \quad |\{(A^{(k,n)}, B^{(k,n)}) : r(H_{A^{(k,n)}}, H_{B^{(k,n)}}) = (k, k), |H_{A^{(k,n)}} \cap H_{B^{(t,n)}|} > 2^{k-1}\}| \leq 2^{2k^2+2k(\frac{n}{k+1})}.
\]

Proof. (1) We may assume that \( s \leq t \). In this case we have to prove that the number of corresponding pairs \((A^{(k,n)}, B^{(k,n)})\) is at most \((\frac{n}{k+1})\) \(2^{2k^2n^{k+1-t}}\).
We have already seen in the proof of Lemma 2.1 that the number of vectors \( A^{(k,n)} \) is at most \((\frac{n}{k+1})\). Fix a vector \( A^{(k,n)} \) and count the suitable vectors \( B^{(k,n)} \). Then the matrix \( B \) has \( t \) linearly independent rows, namely \( r_{\mathbb{R}}((\beta_{d_i,l})_{t \times k}) = t \), for some \( 1 \leq d_1 < \cdots < d_t \leq m \), where

\[
a_0 + \sum_{l=1}^{k} \alpha_{d_i,l}a_l = b_0 + \sum_{l=1}^{k} \beta_{d_i,l}b_l, \quad \alpha_{d_i,l}, \beta_{d_i,l} \in \{0, 1\} \quad \text{for} \quad 1 \leq i \leq t.
\]

The number of possible \( b_l \)s is at most \( n \). For fixed \( b_0, \alpha_{d_i,l}, \beta_{d_i,l} \) let us study the system of equations

\[
a_0 + \sum_{l=1}^{k} \alpha_{d_i,l}a_l = b_0 + \sum_{l=1}^{k} \beta_{d_i,l}x_l, \quad \alpha_{d_i,l}, \beta_{d_i,l} \in \{0, 1\} \quad \text{for} \quad 1 \leq i \leq t.
\]

The assumption \( r_{\mathbb{R}}((\beta_{d_i,l})_{t \times k}) = t \) implies that the number of solutions over \([1, n]\) is at most \( n^{k-t} \). Finally, we have at most \( 2^{kt} \) possibilities on the left-hand side for \( \alpha_{d_i,l} \)s and, similarly, we have at most \( 2^{kt} \) possibilities on the right-hand side for \( \beta_{d_i,l} \)s, therefore the number of possible systems of equations is at most \( 2^{2k^2} \).

(2) The number of vectors \( A^{(k,n)} \) is \((\frac{n}{k+1})\) as in (1). Fix a vector \( A^{(k,n)} \) and count the suitable vectors \( B^{(k,n)} \). It follows from the assumptions
\( r(\mathbf{H}_{\mathbf{A}(k,n)}, \mathbf{H}_{\mathbf{B}(k,n)}) = (r, r), \) \(|\mathbf{H}_{\mathbf{A}(k)} \cap \mathbf{H}_{\mathbf{B}(k)}| = 2^r \) that the vectors \((\alpha_{d,1}, \ldots, \alpha_{d,k})\), \(d = 1, \ldots, 2^r\) and the vectors \((\beta_{d,1}, \ldots, \beta_{d,k})\), \(d = 1, \ldots, 2^r\), respectively form \(r\)-dimensional subspaces of \(\mathbb{F}_2^{k}\). Considering the zero vectors of these subspaces we get \(a_0 = b_0\). The integers \(b_1, \ldots, b_k\) are solutions of the system of equations

\[
a_0 + \sum_{l=1}^{k} \alpha_{d,l} a_l = b_0 + \sum_{l=1}^{k} \beta_{d,l} x_l, \quad \alpha_{d,l}, \beta_{d,l} \in \{0, 1\} \quad \text{for } 1 \leq d \leq 2^r.
\]

Similarly to the previous part this system of equation has at most \(n^{k-r}\) solutions over \([1, n]\) and the number of choices for the \(r\) linearly independent rows is at most \(2^{k^2}\).

(3) Fix a vector \(\mathbf{A}(k,n)\). Let us suppose that for a vector \(\mathbf{B}(k,n)\) we have \(r(\mathbf{H}_{\mathbf{A}(k,n)}, \mathbf{H}_{\mathbf{B}(k,n)}) = (k, k)\) and \(|\mathbf{H}_{\mathbf{A}(k,n)} \cap \mathbf{H}_{\mathbf{B}(k,n)}| > 2^{k-1}\). Let the common vertices be

\[
a_0 + \sum_{l=1}^{k} \alpha_{d,l} a_l = b_0 + \sum_{l=1}^{k} \beta_{d,l} b_l, \quad \alpha_{d,l}, \beta_{d,l} \in \{0, 1\} \quad \text{for } 1 \leq d \leq m,
\]

where we may assume that the rows \(d_1, \ldots, d_k\) are linearly independent, i.e. the matrix \(B_k = (\beta_{d,i})_{k \times k}\) is regular. Write the rows \(d_1, \ldots, d_k\) in matrix form as

\[
(1) \quad \mathbf{a} = b_0 \mathbf{1} + B_k \mathbf{b},
\]

with vectors \(\mathbf{a} = (a_0 + \sum_{l=1}^{k} \alpha_{d,l} a_l)_{k \times 1}\), \(\mathbf{b} = (b_i)_{k \times 1}\). It follows from (1) that

\[
b = B_k^{-1}(\mathbf{a} - b_0 \mathbf{1}) = B_k^{-1} \mathbf{a} - b_0 B_k^{-1} \mathbf{1}.
\]

Let \(B_k^{-1} \mathbf{1} = (d_i)_{k \times 1}\) and \(B_k^{-1} \mathbf{a} = (c_i)_{k \times 1}\). Obviously, the number of subsets \(\{i_1, \ldots, i_t\} \subset \{1, \ldots, k\}\) for which \(d_{i_1} + \ldots + d_{i_t} \neq 1\) is at least \(2^{k-1}\), therefore there exist \(1 \leq u_1 < \ldots < u_s \leq k\) and \(1 \leq v_1 < \ldots < v_t \leq k\) such that \(a_0 + a_{u_1} + \ldots + a_{u_s} = b_0 + b_{v_1} + \ldots + b_{v_t}\), and \(d_{v_1} + \ldots + d_{v_t} \neq 1\). Hence

\[
a_0 + a_{u_1} + \ldots + a_{u_s} = b_0 + b_{v_1} + \ldots + b_{v_t} = b_0 + c_{v_1} + \ldots + c_{v_t} - b_0 (d_{v_1} + \ldots + d_{v_t})
\]

\[
b_0 = \frac{a_0 + a_{u_1} + \ldots + a_{u_s} - c_{v_1} - \ldots - c_{v_t}}{1 - (d_{v_1} + \ldots + d_{v_t})}.
\]

To conclude the proof we note that the number of sets \(\{u_1, \ldots, u_s\}\) and \(\{v_1, \ldots, v_t\}\) is at most \(2^{2k}\) and there are at most \(2^{k^2}\) choices for \(B_k\) and \(\mathbf{a}\), respectively. Finally, for given \(B_k, \mathbf{a}, b_0\), \(1 \leq u_1 < \ldots < u_s \leq k\) and \(1 \leq v_1 < \ldots < v_t \leq k\), the vector \(\mathbf{B}(k,n)\) is determined uniquely. \(\Box\)

In order to prove the theorems we need two lemmas from probability theory (see e.g. [1] p. 41, 95-98.). Let \(X_i\) be the indicator function of the event \(A_i\) and \(S_N = X_1 + \ldots + X_N\). For indices \(i, j\) write \(i \sim j\) if \(i \neq j\) and
the events $A_i, A_j$ are dependant. We set $\Gamma = \sum_{i \sim j} \Pr\{A_i \cap A_j\}$ (the sum over ordered pairs).

**Lemma 2.4.** If $E(S_n) \to \infty$ and $\Gamma = o(E(S_n)^2)$, then $X > 0$ a.e.

In many instances, we would like to bound the probability that none of the bad events $B_i, i \in I$, occur. If the events are mutually independent, then $\Pr\{\cap_{i \in I} B_i\} = \prod_{i \in I} \Pr\{B_i\}$. When the $B_i$ are "mostly" independent, the Janson’s inequality allows us, sometimes, to say that these two quantities are "nearly" equal. Let $\Omega$ be a finite set and $R$ a random subset of $\Omega$ given by $\Pr\{r \in R\} = p_r$, these events being mutually independent over $r \in \Omega$. Let $E_i, i \in I$ be subsets of $\Omega$, where $I$ a finite index set. Let $B_i$ be the event $E_i \subset R$. Let $X_i$ be the indicator random variable for $B_i$ and $X = \sum_{i \in I} X_i$ be the number of $E_i$s contained in $R$. The event $\cap_{i \in I} B_i$ and $X = 0$ are then identical. For $i, j \in I$, we write $i \sim j$ if $i \neq j$ and $E_i \cap E_j \neq \emptyset$. We define $\Delta = \sum_{i \sim j} \Pr\{B_i \cap B_j\}$, here the sum is over ordered pairs. We set $M = \prod_{i \in I} \Pr\{B_i\}$.

**Lemma 2.5** (Janson’s inequality). Let $\varepsilon \in ]0, 1[$ and let $B_i, i \in I, \Delta, M$ be as above and assume that $\Pr\{B_i\} \leq \varepsilon$ for all $i$. Then

$$M \leq \Pr\{\cap_{i \in I} B_i\} \leq M e^{\frac{1-\varepsilon}{2}\frac{\Delta}{\varepsilon}}.$$

**Proof of Theorem 1.5.** Let $H^{(k,n)}_1, \ldots, H^{(k,n)}_N$ be the distinct non-degenerate $k$-cubes in $[1, n]$. Let $B_i$ be the event $H^{(k,n)}_i \subset S(n, cn^{k+1}2^k)$. Then $\Pr\{B_i\} = c^{2^k} n^{-k-1} = o(1)$ and $N = (1 + o(1))\left(\frac{n}{k+1}\right)^{\frac{1}{k!}}$. It is enough to prove

$$\Delta = \sum_{i \sim j} \Pr\{B_i \cap B_j\} = o(1)$$

since then Janson’s inequality implies

$$\Pr\{S(n, cn^{k+1}2^k)\text{ does not contain any } k\text{-cubes}\}$$

$$= \Pr\{\cap_{i=1}^N B_i\}$$

$$= (1 + o(1))(1 - (cn^{k+1}2^k)^{k+1} + o(1))\left(\frac{n}{k+1}\right)^{\frac{1}{k!}}$$

$$= (1 + o(1))e^{-\frac{c^k}{(k+1)!}}.$$
It remains to verify that \( \sum_{i \sim j} \Pr\{B_i \cap B_j\} = o(1) \). We split this sum according to the ranks in the following way:

\[
\sum_{i \sim j} \Pr\{B_i \cap B_j\} = \sum_{s=0}^{k-1} \sum_{t=0}^{k-s} \Pr\{B_i \cap B_j\}
\]

\[
= 2 \sum_{s=1}^{k} \sum_{t=0}^{s-1} \Pr\{B_i \cap B_j\}
\]

\[
+ \sum_{r=0}^{k-1} \sum_{i \sim j} \Pr\{B_i \cap B_j\}
\]

\[
= \sum_{s=1}^{k} \sum_{t=0}^{s-1} \Pr\{B_i \cap B_j\}
\]

\[
+ \sum_{r=1}^{k-1} \sum_{i \sim j} \Pr\{B_i \cap B_j\} + \sum_{i \sim j} \Pr\{B_i \cap B_j\}.
\]

The first sum can be estimated by Lemmas 2 and 2.3(1):

\[
\sum_{s=1}^{k} \sum_{t=0}^{s-1} \Pr\{B_i \cap B_j\}
\]

\[
\leq \sum_{s=1}^{k} \sum_{t=0}^{s-1} 2^{2k^2} \left( \frac{n}{k+1} \right)^{n^{k+1-s} \left( cn \frac{k+1}{2k} \right)^{2k^2 - 2t}}
\]

\[
= n^{o(1)} \sum_{s=1}^{k} n^{s^{k+1-s} \frac{k+1}{2k} - s} = n^{o(1)} \left( n^{\frac{k+1}{2k} - 1} + n^{\frac{k+1}{2k} - k} \right) = o(1),
\]

since the sequence \( a_s = 2^{s-1} \frac{k+1}{2k} - s \) is decreasing for \( 1 \leq s \leq k + 1 - \log_2(k+1) \) and increasing for \( k + 1 - \log_2(k+1) < s \leq k \).
To estimate the second sum we apply Lemma 2.3(2)

$$\sum_{r=0}^{k-1} \sum_{i \sim j} \Pr\{B_i \cap B_j\}$$

$$r(H_{A_i}^{(k,n)} H_{A_j}^{(k,n)}) = (r,r)$$

$$|H_{A_i}^{(k,n)} \cap H_{A_j}^{(k,n)}| = 2^r$$

$$\leq \sum_{r=0}^{k-1} 2^{2k^2} \left( \frac{n}{k+1} \right) n^{k-r} (cn^{-\frac{k+1}{2k}})^{2 \cdot 2^k - 2^r}$$

$$= n^{-1+o(1)} \sum_{r=0}^{k-1} n^{r \frac{k+1}{2k}-r} = n^{-1+o(1)} \left( n^{\frac{k+1}{2k}} + n^{\frac{k+1}{2}- (k-1)} \right) = o(1).$$

The third sum can be bounded using Lemma 2.3(1):

$$\sum_{r=1}^{k-1} \sum_{i \sim j} \Pr\{B_i \cap B_j\}$$

$$r(H_{A_i}^{(k,n)} H_{A_j}^{(k,n)}) = (r,r)$$

$$|H_{A_i}^{(k,n)} \cap H_{A_j}^{(k,n)}| \leq 2^r$$

$$\leq \sum_{r=1}^{k-1} 2^{2k^2} \left( \frac{n}{k+1} \right) n^{k+1-r} (cn^{-\frac{k+1}{2k}})^{2 \cdot 2^k - 2^r + 1}$$

$$\leq n^{-1+o(1)} \sum_{r=1}^{k-1} n^{r \frac{k+1}{2k}-r} = n^{-1+o(1)} \left( n^{\frac{k+1}{2k}} + n^{\frac{k+1}{2}- (k-1)} \right) = o(1).$$

Similarly, for the fourth sum we apply Lemma 2.3(1)

$$\sum_{r=1}^{k-1} \sum_{i \sim j} \Pr\{B_i \cap B_j\} \leq n^{-1+o(1)} n^{k+2} (cn^{-\frac{k+1}{2k}})^{1.5 \cdot 2^k} = o(1).$$

To estimate the fifth sum we note that $|H_{A_i}^{(k,n)} \cup H_{A_j}^{(k,n)}| \geq 2^k + 1$. It follows from Lemma 2.3(3) that

$$\sum_{r=1}^{k-1} \Pr\{B_i \cap B_j\} \leq 2^{2k^2 + 2k} n^{k+1} (cn^{-\frac{k+1}{2k}})^{2^k + 1} = o(1),$$

which completes the proof. \qed
Proof of Theorem 1.6. Let \( \epsilon > 0 \) and for simplicity let \( D_n = D_n(\epsilon) \) and \( E_n = E_n(\epsilon) \). In the proof we use the estimations

\[
2^{2D_n} \leq 2^{2^{\log_2 \log_2 n + \log_2 \log_2 n + \frac{(1-\epsilon) \log_2 \log_2 n}{\log_2 \log_2 n}}}
= n^{\log_2 \log_2 n + (1-\epsilon + o(1)) \log_2 \log_2 n}
\]

and

\[
2^{2E_{n+1}} \geq 2^{2^{\log_2 \log_2 n + \log_2 \log_2 n + \frac{(1+\epsilon) \log_2 \log_2 n}{\log_2 \log_2 n}}}
= n^{\log_2 \log_2 n + (1+\epsilon + o(1)) \log_2 \log_2 n}
\]

In order to verify Theorem 1.6 we have to show that

(2) \( \lim_{n \to \infty} \Pr\{S(n, \frac{1}{2}) \text{ contains a } D_n\text{-cube}\} = 1 \)

and

(3) \( \lim_{n \to \infty} \Pr\{S(n, \frac{1}{2}) \text{ contains an } (E_n + 1)\text{-cube}\} = 0 \).

To prove the limit in (4) let \( H_{A_i^{(D_n, n)}}(D_n, n), \ldots, H_{A_N^{(D_n, n)}}(D_n, n) \) be the different non-degenerate \( D_n \)-cubes in \([1, n]\), \( B_i \) be the event \( H_{A_i^{(D_n, n)}}(D_n, n) \subset S(n, \frac{1}{2}) \), \( X_i \) be the indicator random variable for \( B_i \) and \( S_N = X_1 + \ldots + X_N \) be the number of \( H_{A_i^{(D_n, n)}}(D_n, n) \subset S(n, \frac{1}{2}) \). The linearity of expectation gives by Lemma 2.1 and inequality (2)

\[
E(S_N) = NE(X_i) = (1 + o(1)) \left( \frac{n}{D_n + 1} \right) \frac{1}{D_n!} 2^{-2D_n}
\geq n^{\log_2 \log_2 n + (1+o(1)) \log_2 \log_2 n - \log_2 \log_2 n - (1-\epsilon + o(1)) \log_2 \log_2 n}
= n^{(\epsilon + o(1)) \log_2 \log_2 n}.
\]

Therefore \( E(S_N) \to \infty \), as \( n \to \infty \). By Lemma 2.4 it remains to prove that

\[
\sum_{i \sim j} \Pr\{B_i \cap B_j\} = o(E(S_N)^2)
\]

where \( i \sim j \) means that the events \( B_i, B_j \) are not independent i.e. the cubes \( H_{A_i^{(D_n, n)}}, H_{A_j^{(D_n, n)}} \) have common vertices. We split this sum according to
the ranks

\[
\sum_{i \sim j} \Pr\{B_i \cap B_j\} = \sum_{s=0}^{D_n} \sum_{t=0}^{D_n} \sum_{i \sim j} \Pr\{B_i \cap B_j\} \quad r(\mathcal{H}_{A_i}^{(D_n,n)}, \mathcal{H}_{A_j}^{(D_n,n)}) = (s,t)
\]

\[
\leq \sum_{i \sim j} \Pr\{B_i \cap B_j\} \quad r(\mathcal{H}_{A_i}^{(D_n,n)}, \mathcal{H}_{A_j}^{(D_n,n)}) = (0,0)
\]

\[
+ 2 \sum_{s=1}^{D_n} \sum_{t=0}^{s} \sum_{i \sim j} \Pr\{B_i \cap B_j\} \quad r(\mathcal{H}_{A_i}^{(D_n,n)}, \mathcal{H}_{A_j}^{(D_n,n)}) = (s,t)
\]

The condition \(r(\mathcal{H}_{A_i}^{(D_n,n)}, \mathcal{H}_{A_j}^{(D_n,n)}) = (0,0)\) implies that

\[
|\mathcal{H}_{A_i}^{(D_n,n)} \cup \mathcal{H}_{A_j}^{(D_n,n)}| = 2^{D_n+1} - 1,
\]

thus by Lemma 2.3(2)

\[
\sum_{i \sim j} \Pr\{B_i \cap B_j\} \leq 2^{2D_n^2} \left( \frac{n}{D_n + 1} \right) n^{D_n} 2^{-2^{D_n+1}+1}
\]

\[
= o\left( \left( \frac{n}{D_n + 1} \right) \frac{1}{D_n!} 2^{-2^{D_n}} \right)^2 \quad = o(ES_N^2).
\]

In the light of Lemmas 2 and 2.3(1) the second term in (6) can be estimated as

\[
\sum_{s=1}^{D_n} \sum_{t=0}^{s} \sum_{i \sim j} \Pr\{B_i \cap B_j\} \quad r(\mathcal{H}_{A_i}^{(D_n,n)}, \mathcal{H}_{A_j}^{(D_n,n)}) = (s,t)
\]

\[
\leq \sum_{s=1}^{D_n} \sum_{t=0}^{s} \left( \frac{n}{D_n + 1} \right) 2^{2D_n^2} n^{D_n+1-s} 2^{-2^{D_n}+2^t}
\]

\[
= \left( \frac{n}{D_n + 1} \right) \frac{1}{D_n!} 2^{-2^{D_n}} \sum_{s=1}^{D_n} \sum_{t=0}^{s} \frac{2^t}{n^s}
\]

\[
= \left( \frac{n}{D_n + 1} \right) \frac{1}{D_n!} 2^{-2^{D_n}} \sum_{s=1}^{D_n} \frac{2^s}{n^s}.
\]
Finally, the function $f(x) = \frac{2^x}{n^x}$ decreases on $(-\infty, \log_2 \log n - 2 \log_2 \log 2]$ and increases on $[\log_2 \log n - 2 \log_2 \log 2, \infty)$, therefore by (2)

$$\sum_{s=1}^{D_n} \frac{2^s}{n^s} = n^{o(1)} \left( \frac{4}{n} + \frac{2^{2D_n}}{n^{D_n}} \right) = n^{-1+o(1)},$$

which proves the limit in (4).

In order to prove the limit in (5) let $H_{C_1^{(E_{n+1}, n)}} \ldots H_{C_k^{(E_{n+1}, n)}}$ be the distinct $(E_{n+1})$-cubes in $[1, n]$ and let $F_i$ be the event $H_{C_i^{(E_{n+1}, n)}} \subset S(n, \frac{1}{2})$.

By (3) we have

$$\Pr \{ S_n \text{ contains an } (E_{n+1}) \text{-cube} \} = \Pr \{ \bigcup_{i=1}^{K} F_i \} \leq \sum_{i=1}^{K} \Pr \{ F_i \} \leq$$

$$\left( \frac{n}{E_n + 2} \right)^{2^{-2E_{n+1}}} \leq \frac{n^{\log_2 \log n + (1+o(1)) \log_2 \log_2 \log_2 n}}{n^{\log_2 \log_n + (1+\epsilon+o(1)) \log_2 \log_2 \log_2 n}} = o(1),$$

which completes the proof. □

3. Concluding remarks

The aim of this paper is to study non-degenerate Hilbert cubes in a random sequence. A natural problem would be to give analogous theorems for Hilbert cubes, where degenerate cubes are allowed. In this situation the dominant terms may come from arithmetic progressions. An $AP_{k+1}$ forms a $k$-cube. One can prove by the Janson inequality (see Lemma 2.5) that for a fixed $k \geq 2$

$$\lim_{n \to \infty} \Pr \{ S(n, cn^{-\frac{2}{k+1}}) \text{ contains no } AP_{k+1} \} = e^{-\frac{k+1}{2k}}.$$ 

An easy argument shows (using Janson’s inequality again) that for all $c > 0$, with $p_n = cn^{-2/5}$

$$\lim_{n \to \infty} \Pr \{ S(n, p_n) \text{ contains no } 4 \text{-cubes} \} = e^{-\frac{5}{8}}.$$

**Conjecture 3.1.** For $k \geq 4$

$$\lim_{n \to \infty} \Pr \{ S(n, cn^{-\frac{2}{k+1}}) \text{ contains no } k \text{-cubes} \} = e^{-\frac{k+1}{2k}}.$$ 

A simple calculation implies that in the random sequence $S(n, \frac{1}{2})$ the length of the longest arithmetic progression is a.e. nearly $2 \log_2 n$, therefore it contains a Hilbert cube of dimension $(2 - \epsilon) \log_2 n$.

**Conjecture 3.2.** For every $\epsilon > 0$

$$\lim_{n \to \infty} \Pr \left\{ \begin{array}{l} \text{the maximal dimension of Hilbert cubes} \\ \text{in } S(n, \frac{1}{2}) \text{ is } < (2 + \epsilon) \log_2 n \end{array} \right\} = 1.$$
N. Hegyvári (see [5]) studied the special case where the generating elements of Hilbert cubes are distinct. He proved that in this situation the maximal dimension of Hilbert cubes is a.e. between $c_1 \log n$ and $c_2 \log n \log \log n$. In this problem the lower bound seems to be the correct magnitude.

References


Csaba Sándor
Institute of Mathematics
Budapest University of Technology and Economics
Egry J. u. 1., H-1111 Budapest, Hungary
E-mail: csandor@math.bme.hu