#A03 INTEGERS 9 (2009), 41-45

ON TWO-POINT CONFIGURATIONS IN A RANDOM SET

Hoi H. Nguyen[1]
Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA
hoi@math.rutgers.edu

Received: 8/2/08, Accepted: 1/3/09

Abstract

We show that with high probability, a random subset of $\{1, \ldots, n\}$ of size $\Theta(n^{1-1/k})$ contains two elements a and $a + d^k$, where d is a positive integer. As a consequence, we prove an analogue of the Sárközy-Fürstenberg theorem for a random subset of $\{1, \ldots, n\}$.

1. Introduction

Let $\wp$ be a general additive configuration, $\wp = (a, a+P_1(d), \ldots, a+P_{k-1}(d))$, where $P_i \in \mathbb{Z}[d]$ and $P_i(0) = 0$. Let $[n]$ denote the set of positive integers up to n. A natural question is:

Question 1. *How is $\wp$ distributed in $[n]$?*

Roth's theorem [6] says that for $\delta > 0$ and sufficiently large n, any subset of $[n]$ of size δn contains a nontrivial instance of $\wp = (a, a + d, a + 2d)$ (here nontrivial means $d \neq 0$). In 1975, Szemerédi [8] extended Roth's theorem for general linear configurations $\wp = (a, a + d, \ldots, a + (k - 1)d)$. For a configuration of type $\wp = (a, a + P(d))$, Sárközy [7] and Fürstenberg [2] independently discovered a similar phenomenon.

Theorem 2 (Sárközy-Fürstenberg theorem, quantitative version; [9, Theorem 3.2], [4, Theorem 3.1])**.** *Let δ be a fixed positive real number, and let P be a polynomial of integer coefficients satisfying $P(0) = 0$. Then there exists an integer $n = n(\delta, P)$ and a positive constant $c(\delta, P)$ with the following property. If $n \geq n(\delta, P)$ and $A \subset [n]$ is any subset of cardinality at least δn, then*

- *A contains a nontrivial instance of $\wp$.*
- *A contains at least $c(\delta, P)|A|^2 n^{1/\deg(P)-1}$ instances of $\wp = (a, a + P(d))$.*

In 1996, Bergelson and Leibman [1] extended this result for all configurations $\wp = (a, a + P_1(d), \ldots, P_{k-1}(d))$, where $P_i \in \mathbb{Z}[d]$ and $P_i(0) = 0$ for all i.

Following Question 1, one may consider the distribution of $\wp$ in a "pseudo-random" set.

[1]This work was written while the author was supported by a DIMACS summer research fellowship, 2008.

Question 3. *Does the set of primes contain a nontrivial instance of $\wp$? How is $\wp$ distributed in this set?*

The famous Green–Tao theorem [3] says that any subset of positive upper density of the set of primes contains a nontrivial instance of $\wp = (a, a+d, \dots, a+(k-1)d)$ for any k. This phenomenon also holds for more general configurations $(a, a+P_1(d), \dots, a+P_{k-1}(d))$, where $P_i \in \mathbb{Z}[d]$ and $P_i(0) = 0$ for all i (cf. [9]).

The main goal of this note is to consider a similar question.

Question 4. *How is $\wp$ distributed in a typical random subset of $[n]$?*

Let $\wp$ be an additive configuration and let δ be a fixed positive real number. We say that a set A is $(\delta, \wp)$-dense if any subset of cardinality at least $\delta|A|$ of A contains a nontrivial instance of $\wp$. In 1991, Kohayakawa–Łuczak–Rödl [5] showed the following result.

Theorem 5. *Almost every subset R of $[n]$ of cardinality $|R| = r \gg_\delta n^{1/2}$ is $(\delta, (a, a+d, a+2d))$-dense.*

The assumption $r \gg_\delta n^{1/2}$ is tight, up to a constant factor. Indeed, a typical random subset R of $[n]$ of cardinality r contains about $\Theta(r^3/n)$ three-term arithmetic progressions. Hence, if $(1-\delta)r \gg r^3/n$, then there is a subset of R of cardinality δr which does not contain any nontrivial 3-term arithmetic progression.

Motivated by Theorem 5, Łaba and Hamel [4] studied the distribution of $\wp = (a, a+d^k)$ in a typical random subset of $[n]$, as follows.

Theorem 6. *Let $k \geq 2$ be an integer. Then there exists a positive real number $\varepsilon(k)$ with the following property. Let δ be a fixed positive real number, then almost every subset R of $[n]$ of cardinality $|R| = r \gg_\delta n^{1-\varepsilon(k)}$ is $(\delta, (a, a+d^k))$-dense.*

It was shown that $\varepsilon(2) = 1/110$, and $\varepsilon(3) \gg \varepsilon(2)$, etc. Although the method used in [4] is strong, it seems to fall short of obtaining relatively good estimates for $\varepsilon(k)$. On the other hand, one can show that $\varepsilon(k) \leq 1/k$. Indeed, a typical random subset of $[n]$ of size r contains $\Theta(n^{1+1/k}r^2/n^2)$ instances of $(a, a+d^k)$. Thus if $(1-\delta)r \gg n^{1+1/k}r^2/n^2$ (which implies $r \ll_\delta n^{1-1/k}$) then there is a subset of size δr of R which does not contain any nontrivial instance of $(a, a+d^k)$.

In this note we shall sharpen Theorem 6 by showing that $\varepsilon(k) = 1/k$.

Theorem 7 (Main theorem). *Almost every subset R of $[n]$ of size $|R| = r \gg_\delta n^{1-1/k}$ is $(\delta, (a, a+d^k))$-dense.*

Our method to prove Theorem 7 is elementary. We will invoke a combinatorial lemma and the quantitative Sárközy-Fürstenberg theorem (Theorem 2). As the reader will see later on, the method also works for more general configurations $(a, a+P(d))$, where $P \in \mathbb{Z}[d]$ and $P(0) = 0$.

2. A Combinatorial Lemma

Let $G(X,Y)$ be a bipartite graph. We denote the number of edges going through X and Y by $e(X,Y)$. The average degree $\bar{d}(G)$ of G is defined to be $e(X,Y)/(|X||Y|)$.

Lemma 8. *Let $\{G = G([n],[n])\}_{n=1}^{\infty}$ be a sequence of bipartite graphs. Assume that for any $\varepsilon > 0$ there exist an integer $n(\varepsilon)$ and a number $c(\varepsilon) > 0$ such that $e(A,A) \geq c(\varepsilon)|A|^2\bar{d}(G)/n$ for all $n \geq n(\varepsilon)$ and all $A \subset [n]$ satisfying $|A| \geq \varepsilon n$. Then for any $\alpha > 0$ there exist an integer $n(\alpha)$ and a number $C(\alpha) > 0$ with the following property. If one chooses a random subset S of $[n]$ of cardinality s, then the probability of $G(S,S)$ being empty is at most α^s, providing that $|S| = s \geq C(\alpha)n/\bar{d}(G)$ and $n \geq n(\alpha)$.*

Proof. For short we denote the ground set $[n]$ by V. We shall view S as an ordered random subset, whose elements will be chosen in order, v_1 first and v_s last. We shall verify the lemma within this probabilistic model. Deduction of the original model follows easily.

For $1 \leq k \leq s-1$, let N_k be the set of neighbors of the first k chosen vertices, i.e., $N_k = \{v \in V, (v_i, v) \in E(G) \text{ for some } i \leq k\}$. Since $G(S,S)$ is empty, we have $v_{k+1} \notin N_k$. Next, let B_{k+1} be the set of possible choices for v_{k+1} (from $V \backslash \{v_1, \ldots, v_k\}$) such that $N_{k+1} \backslash N_k \leq c(\varepsilon)\varepsilon\bar{d}(G)$, where ε will be chosen to be small enough ($\varepsilon = \alpha^2/6$ is fine) and $c(\varepsilon)$ is the constant from Lemma 8. We observe the following.

Claim 9. $|B_{k+1}| \leq \varepsilon|V|$.

To prove this claim, we assume for contradiction that $|B_{k+1}| \geq \varepsilon|V| = \varepsilon n$. Since $B_{k+1} \cap N_k = \emptyset$, we have $e(B_{k+1}, B_{k+1}) \leq e(B_{k+1}, V\backslash N_k) \leq c(\varepsilon)\varepsilon\bar{d}(G)|B_{k+1}| < c(\varepsilon)|B_{k+1}|^2\bar{d}(G)/n$. This contradicts the property of G assumed in Lemma 8, provided that n is large enough.

Thus we conclude that if $G(S,S)$ is empty then $|B_{k+1}| \leq \varepsilon|V|$ for $1 \leq k \leq s-1$.

Now let s be sufficiently large, say $s \geq 2(c(\varepsilon)\varepsilon)^{-1}n/\bar{d}(G)$, and assume that the vertices $v_1, \ldots, v_s$ have been chosen. Let s' be the number of vertices v_{k+1} that do not belong to B_{k+1}. Then we have

$$n \geq |N_s| \geq \sum_{v_{k+1} \notin B_{k+1}} |N_{k+1}\backslash N_k| \geq s' c(\varepsilon)\varepsilon\bar{d}(G).$$

Hence, $s' \leq (c(\varepsilon)\varepsilon)^{-1}n/\bar{d}(G) \leq s/2$.

As a result, there are $s - s'$ vertices v_{k+1} that belong to B_{k+1}. But since $|B_{k+1}| \leq \varepsilon n$, we see that the number of subsets S of V such that $G(S,S)$ is empty is bounded by

$$\sum_{s' \le s/2} \binom{s}{s'} n^{s'} (\varepsilon n)^{s-s'} \le (6\varepsilon)^{s/2} n(n-1)\dots(n-s+1) \le \alpha^s n(n-1)\dots(n-s+1),$$

thereby completing the proof. □

3. Proof of Theorem 7

First, we define a bipartite graph G on $[n] \times [n] = V_1 \times V_2$ by connecting $u \in V_1$ to $v \in V_2$ if $v - u = d^k$ for some integer $d \in [1, n^{1/k}]$. Notice that $\bar{d}(G) \approx Cn^{1/k}$ for some absolute constant C.

Let us restate the Sárközy-Fürstenberg theorem (Theorem 2, for $P(d) = d^k$) in terms of the graph G.

Theorem 10. *Let $\varepsilon > 0$ be a positive constant. Then there exists a positive integer $n(\varepsilon, k)$ and a positive constant $c(\varepsilon, k)$ such that $e(A, A) \ge c(\varepsilon, k)|A|^2 n^{1/k-1}$ for all $n \ge n(\varepsilon, k)$ and all $A \subset [n]$ satisfying $|A| \ge \varepsilon n$.*

Now let S be a subset of $[n]$ of size s. We call S *bad* if it does not contain any nontrivial instance of $(a, a + d^k)$. In other words, S is bad if $G(S, S)$ contains no edges. By Lemma 8 and Theorem 10, the number of bad subsets of $[n]$ is at most $\alpha^s \binom{n}{s}$, provided that $s \ge C(\alpha)n/\bar{d}(G)$. This condition is satisfied if we assume that

$$s \ge 2C(\alpha)C^{-1}n^{1-1/k}.$$

Next, let $r = s/\delta$ and consider a random subset R of $[n]$ of size r. The probability that R contains a bad subset of size s is at most

$$\alpha^s \binom{n}{s} \binom{n-s}{r-s} / \binom{n}{r} = o(1),$$

provided that $\alpha = \alpha(\delta)$ is small enough.

To finish the proof, we note that if R does not contain any bad subset of size δr, then R is $(\delta, (a, a + d^k))$-dense.

References

[1] V. Bergelson and A. Leibman, *Polynomial extensions of Van Der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725-753.

[2] H. Fürstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, Princeton, 1981.

[3] B. J. Green and T. Tao, *Primes contain arbitrarily long arithmetic progression*, to appear in Ann. Math.

[4] M. Hamel and I. Łaba, *Arithmetic structures in random sets*, Integers: Electronic J. Comb. Number Theory **8** (2008).

[5] Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets on a random set*, Acta Arith. **75** (1996), 133-163.

[6] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245-252.

[7] A. Sárközy, *On difference sets of sequences of integers III*, Acta Math. Sci. Hungar. **31** (1978), 125-149.

[8] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progressions*, Acta Arith. **27** (1975), 299-345.

[9] T. Tao and T. Ziegler *The primes contain arbitrarily long polynomial progressions*, to appear in Acta Math.