A PROBABILISTIC THRESHOLD FOR MONOCHROMATIC ARITHMETIC PROGRESSIONS

Aaron Robertson
Department of Mathematics, Colgate University, Hamilton, New York
arobertson@colgate.edu

Abstract

Let $f_r(k) = \sqrt{k} \cdot r^{k/2}$ (where $r \geq 2$ is fixed) and consider r-colorings of $[1, n_k] = \{1, 2, \ldots, n_k\}$. We show that $f_r(k)$ is a threshold function for k-term arithmetic progressions in the following sense: if $n_k = \omega(f_r(k))$, then $\lim_{k\to\infty} P([1, n_k]$ contains a monochromatic k-term arithmetic progression$) = 1$; while, if $n_k = o(f_r(k))$, then $\lim_{k\to\infty} P([1, n_k]$ contains a k-term monochromatic arithmetic progression$) = 0$.

1. Introduction

For $k, r \in \mathbb{Z}^+$, let $w(k; r)$ be the minimum integer such that *every* r-coloring of $[1, w(k; r)]$ admits a monochromatic k-term arithmetic progression. The existence of such an integer was shown by van der Waerden [10], and these integers are referred to as van der Waerden numbers. Current knowledge (for r fixed) places $w(k; r)$ somewhere between $\frac{r^{k-1}}{ek}(1 + o(1))$ and

$$2^{2^{r^{2^{2^{k+9}}}}},$$

with the upper bound being from one of Gowers' seminal works [5] (with slightly better lower bounds when $r = 2$). A matching of upper and lower bounds appears unlikely in the near (or distant?) future. However, by loosening the restriction that *every* r-coloring must have a certain property to *almost every* (in the probabilistic sense), we are able to home in on the rate of growth of the associated numbers.

Let $\mathcal{C}_n$ be the collection of all r-colorings of $[1, n]$ and define $\mathcal{C} = \bigcup_{i\geq 1} C_i$. A *threshold function* on $\mathcal{C}$ for a property $\mathcal{P}$ is a function $t(n)$ such that the following both hold:

- given $f(n) = \omega(t(n))$ and $S \in \mathcal{C}$ with $|S| = f(n)$ we have $\lim_{n\to\infty} P(S$ has property $\mathcal{P}) = 1$;
- given $g(n) = o(t(n))$ and $S \in \mathcal{C}$ with $|S| = g(n)$ we have $\lim_{n\to\infty} P(S$ has property $\mathcal{P}) = 0$.

Although the majority of threshold functions studied have concerned graphs (which would require a change in the definition of $\mathcal{C}$ above), work on integer Ramsey structures has been investigated. Conlon and Gowers [3] proved a probability threshold function in the area of random subsets containing a k-term arithmetic progression. Schacht [8] has a similar result, derived independently using different methods. Balogh, Morris, and Samotij [1] provide yet another method addressing the same problem, as do Saxton and Thomason [9]. Friedgut, Rödl, and Schacht [4] provide probability thresholds for Rado's Theorem.

From a different persepective, a threshold function for the size of a random subset needed to admit a monochromatic arithmetic progressions is given by Rödl and Ruciński [6, 7]. The present article is more in-line with this direction insomuch as we are coloring integers at random rather than searching for the probability threshold with which we include (i.e., "color") a particular integer.

In this article, we assume that every r-coloring of a given interval is equally likely, i.e., for each integer in the interval, the probability that it is a given color is $\frac{1}{r}$. We refer to a k-term arithmetic progression as a k-ap and will use the notation $\langle a, d\rangle_k$ to represent $a, a+d, a+2d, \ldots, a+(k-1)d$, where we refer to d as the *gap* of the k-ap. We use the standard notation $[1, n] = \{1, 2, \ldots, n\}$.

For ease of exposition, we separate the definition of a threshold function into parts via the following two definitions.

Definition 1. Let $t(k)$ be a function defined on $\mathbb{Z}^+$ with some property $\mathcal{P}$. We say that $t(k)$ is a *minimal function* (with respect to $\mathcal{P}$) if for every function $s(k)$ defined on $\mathbb{Z}^+$ with property $\mathcal{P}$ we have

$$\liminf_{k\to\infty} \frac{t(k)}{s(k)} \leq 1.$$

We say that $t(k)$ is a *maximal function* (with respect to $\mathcal{P}$) if for every function $s(k)$ defined on $\mathbb{Z}^+$ with property $\mathcal{P}$ we have

$$\limsup_{k\to\infty} \frac{t(k)}{s(k)} \geq 1.$$

Definition 2. Let $k, r \in \mathbb{Z}^+$ with r fixed. Denote by $N^+(k;r)$ a minimal function such that the probability that a randomly chosen r-coloring of $[1, N^+(k;r)]$ admits a monochromatic k-ap tends to 1 as $k \to \infty$. Denote by $N^-(k;r)$ a maximal function such that the probability that a randomly chosen r-coloring of $[1, N^-(k;r)]$ admits a monochromatic k-ap tends to 0 as $k \to \infty$.

Brown [2] showed that $N^+(k;2) \leq (\log k)2^k g(k)$, while Vijay [11] made a significant improvement by showing that $N^+(k;2) \leq k^{3/2}2^{k/2}g(k)$, where, in each bound, $g(k)$ is any function tending to ∞. Vijay [11], using the linearity of expectation, also provided a lower bound for $N^-(k;2)$ that is not much smaller than his given upper bound. The generalization to $N^-(k;r)$ is straightforward, but included for completeness.

Remark 3. For the remainder of the article, r will be fixed and all limits will be taken as $k \to \infty$.

Theorem 4. (Vijay) *Let $f(k) \to 0$ arbitrarily slowly. Then $N^-(k;r) \geq \sqrt{k} \cdot r^{k/2} f(k)$.*

Proof. Let $n = \sqrt{k} \cdot r^{k/2} f(k)$. We will show that the probability that a random r-coloring of $[1, n]$ admits a monochromatic k-ap tends to 0 (as $k \to \infty$) by calculating the expected number of monochromatic k-aps in two ways. Consider a random r-coloring χ of $[1, n]$. Let X_i equal 1 if the i^{th} k-ap is monochromatic under χ; otherwise, let X_i equal 0. Then

$$X = \sum_{i=1}^{\frac{n^2}{2(k-1)}} X_i$$

is the number of monochromatic k-aps under χ (where we are suppressing the asymptotic $(1 + o(1))$ in the upper limit). By the linearity of expectation, we have

$$E(X) = \sum_{i=1}^{\frac{n^2}{2(k-1)}} E(X_i) = \sum_{i=1}^{\frac{n^2}{2(k-1)}} P(X_i = 1) = \frac{n^2}{2(k-1)} \cdot \frac{1}{r^{k-1}} = \frac{rk}{2(k-1)} \cdot f^2(k).$$

Using the definition of $E(X)$, we have

$$E(X) = \sum_{i=0}^{\frac{n^2}{2(k-1)}} i \cdot P(X = i) \geq \sum_{i=0}^{\frac{n^2}{2(k-1)}} P(X = i) - P(X = 0) = 1 - P(X = 0).$$

Hence, we have $\frac{rk}{2(k-1)} \cdot f^2(k) \geq 1 - P(X = 0)$, so that

$$P(X = 0) \geq 1 - \frac{rk}{2(k-1)} \cdot f^2(k) \underset{k\to\infty}{\longrightarrow} 1,$$

which completes the proof. □

We note here that if we consider the set of k-aps with gaps that are primes larger than k, we can follow Vijay's argument for an upper bound on $N^+(k;2)$ very closely to match his bound: $N^+(k) \leq k^{3/2}\, 2^{k/2} g(k)$ for any function $g(k) \to \infty$ (we will use the notation $g(k) \to \infty$ as opposed to $g(k) \to 0$ as found in [11]). In the next section, we construct a larger family of k-aps (than Vijay's and than those k-aps with prime gap larger than k) with the aim of lowering this upper bound on $N^+(k;2)$ by a factor of k, while also generalizing to an arbitrary number of colors.

2. A Structured Family of Arithmetic Progressions

For $k, n \in \mathbb{Z}^+$, let $AP_k(n) = \{\langle a, d\rangle_k \subseteq [1, n] : a, d \in \mathbb{Z}^+\}$, i.e., the set of k-aps in $[1, n]$ and let $A_j(n) = \{\langle a, d\rangle_k \in AP_k(n) : d = j\}$ so that $AP_k(n)$ is the disjoint union of the $A_j(n)$:

$$AP_k(n) = \bigsqcup_{d=1}^{\lfloor \frac{n-1}{k-1} \rfloor} A_d(n).$$

We will now sieve out elements from each $A_d(n)$. Order the elements of $A_d(n)$ by their initial term. Remove every k-ap in $A_d(n)$ with initial term in

$$\bigcup_{j=1}^{\lceil \frac{n}{3d} - \frac{k-1}{3} \rceil} \big[(3j-2)d + 1, 3jd\big].$$

Let $\overline{A}_d(n)$ be the set of elements in $A_d(n)$ which were not removed and define

$$\overline{AP}_k(n) = \bigsqcup_{d=1}^{\left\lfloor \frac{n-1}{k-1} \right\rfloor} \overline{A}_d(n).$$

Notice that, for each d, every two *intersecting* k-aps in $\overline{A}_d(n)$ have initial terms that are more than $2d$ apart, by construction.

We next provide a useful lemma that gives bounds on the size of $\overline{AP}_k(n)$.

Lemma 5. *Let $k \geq 5$. Then*

$$\frac{|AP_k(n)|}{3} \leq |\overline{AP}_k(n)| \leq \frac{|AP_k(n)|}{2}. \tag{1}$$

Hence,

$$\frac{n^2}{6(k-1)}(1+o(1)) \leq \left|\overline{AP}_k(n)\right| \leq \frac{n^2}{4(k-1)}(1+o(1)). \tag{2}$$

Proof. It is a standard exercise to show that $|AP_k(n)| = \frac{n^2}{2(k-1)}(1+o(1))$, so that the inequalities in (2) follow immediately once we prove the inequalities in (1). The lower bound in (1) follows easily from the sieving construction, hence we focus on the upper bound in (1).

For $d \in \left[1, \frac{n}{k+6}\right)$ we have $\overline{A}_d(n) \leq \frac{3}{7}|A_d(n)|$. To see this, note that since $d < \frac{n}{k+6}$ we have $7d + (k-1)d \leq n$, so that $A_d(n)$ has at least $7d$ elements. Of the k-aps in $A_d(n)$ with initial term at most $7d$, our sieve admits exactly $\frac{3}{7}^{\text{ths}}$ of them to $\overline{A}_d(n)$. By continuing the sieving process, starting with the k-ap with initial term $7d+1$ (if it exists), we will remove at least twice as many k-aps from $A_d(n)$ as we add to $\overline{A}_d(n)$; hence, $\overline{A}_d(n) \leq \frac{3}{7}|A_d(n)|$ holds.

For $d \in \left[\frac{n}{k+6}, \frac{n-1}{k-1}\right]$ we use the trivial bound $\overline{A}_d(n) \leq A_d(n)$. Lastly, we note that $|A_d(n)| = n-(k-1)d$. Hence,

$$\begin{aligned}
\left|\overline{AP}_k(n)\right| = \sum_{d=1}^{\left\lfloor \frac{n-1}{k-1} \right\rfloor} |\overline{A}_d(n)| &= \sum_{d=1}^{\left\lfloor \frac{n}{k+6} \right\rfloor - 1} |\overline{A}_d(n)| + \sum_{d=\left\lfloor \frac{n}{k+6} \right\rfloor}^{\left\lfloor \frac{n-1}{k-1} \right\rfloor} |\overline{A}_d(n)| \\
&\leq \sum_{d=1}^{\left\lfloor \frac{n}{k+6} \right\rfloor - 1} \frac{3|A_d(n)|}{7} + \sum_{d=\left\lfloor \frac{n}{k+6} \right\rfloor}^{\left\lfloor \frac{n-1}{k-1} \right\rfloor} |A_d(n)| \\
&= \frac{3k^3 + 33k^2 + 23k + 284}{14(k-1)^2(k+6)^2} \cdot n^2(1+o(1)) \\
&\leq \frac{1}{4(k-1)} n^2(1+o(1)) \qquad \text{(for } k \geq 5\text{)} \\
&= \frac{|AP_k(n)|}{2},
\end{aligned}$$

which proves the upper bound in (2). ∎

We next give a result on how pairs of k-aps from $\overline{AP}_k(n)$ intersect.

Lemma 6. *Let $A = \langle a, b\rangle_k$ and $C = \langle c, d\rangle_k$ belong to $\overline{AP}_k(n)$. Then $|A \cap C| \le k - 3$. Furthermore,*

(i) $|A \cap C| > \lceil \frac{k}{2} \rceil$ *only if* $b = d$*;*

(ii) $\lceil \frac{k}{3} \rceil \le |A \cap C| \le \lceil \frac{k}{2} \rceil$ *only if* $\frac{b}{d} \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 2, 3\}$.

Proof. We first argue that in order to have $|A \cap C| \ge \lceil \frac{k}{3} \rceil$ we must have $\frac{b}{d} \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 2, 3\}$. Consider $b \le d$. We have $a + ib = c + j_1 d$ and $a + (i + x)b = c + j_2 d$ for some $i \in [0, k - 2]$ and $x \in \{1, 2, 3\}$ (else we cannot have enough intersections) and $j_1 < j_2$. Thus, $xb = (j_2 - j_1)d$. Since $d \ge b$ we must have $j_2 - j_1 \le x$. This leaves $\frac{b}{d} = \frac{j_2 - j_1}{x} \in \{\frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1\}$. For $d > b$, we take reciprocals and achieve the stated goal.

Now, in order for A and C to intersect in more than $\lceil \frac{k}{2} \rceil$ places, there must be two consecutive elements of, say, A in the intersection. Let $a + ib$ and $a + (i + 1)b$ be two such elements. We must have $d \le b$ in order for C to intersect both of these. So, let $a + ib = c + jd$ and $a + (i + 1)b = c + \ell d$. These imply that $b = (\ell - j)d$. If $\ell - j > 1$ then $d \le \frac{b}{2}$. In this situation, C intersects A in at most $\lceil \frac{k}{2} \rceil$ places since for every two consecutive terms of A, there exists a term of C between them. Thus, $\ell - j = 1$ and $b = d$ as stated.

To show that $|A \cap C| \le k - 3$, note that we have proved part (i) so we need only consider k-aps with the same gap, i.e., those in the same $\overline{A}_g(n)$ for some gap g. In order for two such k-aps to intersect in more than $k - 3$ places, their starting elements must be within $2g$ of each other. As noted before, by construction of $\overline{A}_g(n)$, this is not possible. □

Lemma 7. *For a given $A = \langle a, b\rangle_k \in \overline{AP}_k(n)$, the number of $\langle c, d\rangle_k \in \overline{AP}_k(n)$ with $c \ge a$ that intersect A in p places is*

(i) 0 *for* $p > k - 3$*;*

(ii) 1 *for each* $p \in \left[\lceil \frac{k}{2} \rceil + 1, k - 3\right]$*;*

(iii) at most 7 *for each* $p \in \left[\lceil \frac{k}{3} \rceil, \lceil \frac{k}{2} \rceil\right]$.

Proof. Part (i) is just a restatement of part of Lemma 6. For part (ii), by Lemma 6(i), we must have $b = d$. For a given p, we have $c = a + (k - p)d$ and the result follows. For part (iii), since b is fixed, d must be one of the 7 gaps that adhere to Lemma 6(ii). In order to intersect in exactly p places, c is determined. □

With these lemmas under our belt, we are now ready to move onto the main result.

3. Main Result

We incorporate Theorem 4 into the main result, which we now state.

Theorem 8. *Let $f(k) \to 0$ and $g(k) \to \infty$ arbitrarily slowly. Then,*

$$\sqrt{k} \cdot r^{k/2} f(k) \leq N^-(k;r) < N^+(k;r) \leq \sqrt{k} \cdot r^{k/2} g(k).$$

Proof. We need only to prove the upper bound on $N^+(k;r)$ and do so by using the family defined in the previous section, along with techniques from [2] and [11]. To this end, let $n = \sqrt{k} \cdot r^{k/2} g(k)$ and partition $[1,n]$ into intervals of length $s = \left\lceil \frac{n}{g^{4/3}(k)} \right\rceil$, where the last interval may be shorter (and won't be used). Let I_j, $1 \leq j \leq \lfloor \frac{n}{s} \rfloor$, be these intervals.

Fix an interval I_ℓ and randomly color each element of I_ℓ with one of r colors, where each color is equally likely. Let X_i be the event that the i^{th} k-ap in I_ℓ is monochromatic, where $1 \leq i \leq \frac{s^2}{6(k-1)}$ holds by the lower bound in (2) (and we suppress lower order terms). Denote by p the probability that a monochromatic k-ap exists. Via one of the Bonferroni inequalities, we have

$$p = \left| P\left(\bigcup_{i=1}^{\frac{s^2}{6(k-1)}} X_i \right) \right| \geq \sum_{i=1}^{\frac{s^2}{6(k-1)}} P(X_i) \quad - \sum_{1 \leq i < j \leq \frac{s^2}{6(k-1)}} P(X_i \cap X_j).$$

Hence,

$$p \geq \frac{s^2}{6(k-1)} \cdot \frac{1}{r^{k-1}} \quad - \sum_{1 \leq i < j \leq \frac{s^2}{6(k-1)}} P(X_i \cap X_j).$$

We now focus on the double summation. With a slight abuse of notation, we rewrite this as

$$\sum_{b \in \overline{AP}_k(s)} \sum_{a \in \overline{AP}_k(s)} P(X_a \cap X_b),$$

where the initial term of b is at most as large as the initial term of a. For a given $b \in \overline{AP}_k(s)$ with gap g, we define:

$$S_b = \{a \in \overline{A}_g(s) : a \cap b \neq \emptyset\};$$

$$T_b = \left\{a \in \overline{A}_h(s) : \left\lceil \tfrac{k}{3} \right\rceil \leq |a \cap b| \leq \left\lceil \tfrac{k}{2} \right\rceil \text{ and } \tfrac{g}{h} \in \{\tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, \tfrac{3}{2}, 2, 3\}\right\}$$

$$Q_b = \left\{a \in \overline{AP}_k(s) : a \cap b \neq \emptyset \text{ and } a \notin S_b \sqcup T_b\right\};$$

$$R_b = \{a \in \overline{AP}_k(s) : a \cap b = \emptyset\}.$$

Note that $\overline{AP}_k(s)$ is the disjoint union $S_b \sqcup T_b \sqcup Q_b \sqcup R_b$. With these definitions, the double summation becomes

$$\sum_{b \in \overline{AP}_k(s)} \left(\sum_{a \in S_b} P(X_a \cap X_b) + \sum_{a \in T_b} P(X_a \cap X_b) + \sum_{a \in Q_b} P(X_a \cap X_b) + \sum_{a \in R_b} P(X_a) P(X_b) \right).$$

Appealing to Lemmas 5 and 6, we find the following (for k sufficiently large):

$$\sum_{a \in S_b} P(X_a \cap X_b) \leq \sum_{i=2}^{k-1} \frac{1}{r^{k+i}} = \frac{1}{r^{k+1}} - \frac{1}{r^{2k-1}} \leq \frac{1}{r^{k+1}}; \tag{3}$$

$$\sum_{a \in T_b} P(X_a \cap X_b) \leq \frac{7\left(\left\lceil \frac{k}{2} \right\rceil - \left\lceil \frac{k}{3} \right\rceil + 1\right)}{r^{k+k/2}} \leq \frac{1}{3 \cdot r^{k+1}}; \tag{4}$$

$$\sum_{a \in Q_b} P(X_a \cap X_b) \leq \frac{sk^2}{r^{k+2k/3}} \leq \frac{1}{3 \cdot r^{k+1}}; \tag{5}$$

$$\sum_{a \in R_b} P(X_a \cap X_b) \leq \frac{s^2}{4(k-1)r^{2k-2}} \leq \frac{1}{3 \cdot r^{k+1}}, \tag{6}$$

where: (3) holds since there is exactly one such k-ap that intersects b in $k-1-i$ places and Lemma 7(i) gives $i \geq 2$; (4) follows from Lemma 6(ii) and Lemma 7(iii); (5) holds from the lower bound in Lemma 6(ii) and since sk^2 or fewer k-aps intersect a given k-ap (there are k choices for which element of the given k-ap is intersected, k choices for which element of an intersecting k-ap is used, and less that s choices for the gap of an intersecting k-ap); and (6) holds by using the upper bound in (2) along with independence since the two k-aps do not share any element.

Using these bounds, we have

$$\sum_{a \in S_b} P(X_a \cap X_b) + \sum_{a \in T_b} P(X_a \cap X_b) + \sum_{a \in Q_b} P(X_a \cap X_b) + \sum_{a \in R_b} P(X_a \cap X_b) \leq \frac{2}{r^{k+1}} \leq \frac{1}{r^k}.$$

Hence, using the upper bound in (2), we have

$$\sum_{b \in \overline{AP}_k(s)} \sum_{a \in \overline{AP}_k(s)} P(X_a \cap X_b) \leq \frac{s^2}{4(k-1)} \cdot \frac{1}{r^k}$$

so that

$$p \geq \frac{s^2}{6(k-1)} \cdot \frac{1}{r^{k-1}} - \frac{s^2}{4(k-1)} \cdot \frac{1}{r^k} = \frac{(2r-3)s^2}{12(k-1)} \cdot \frac{1}{r^k}$$

This gives us that the probability that I_ℓ has no monochromatic k-ap from $\overline{AP}_k(s)$ is at most

$$1 - \frac{(2r-3)s^2}{12(k-1)r^k} \leq 1 - \frac{(2r-3)n^2}{12g^{8/3}(k) \cdot (k-1)r^k} = 1 - \frac{1}{g^{2/3}(k)} \cdot \frac{k}{k-1} \cdot \frac{2r-3}{12} \leq 1 - \frac{2r-3}{12g^{2/3}(k)}.$$

Thus, the probability that none of the intervals I_j contains a monochromatic k-ap from $\overline{AP}_k(n)$ is, for k sufficiently large, at most

$$\left(1 - \tfrac{(2r-3)}{12g^{2/3}(k)}\right)^{g^{4/3}(k)-1} \lessapprox \exp\left(-\tfrac{(g^{4/3}(k)-1)(2r-3)}{12g^{2/3}(k)}\right) \leq \exp\left(-\tfrac{2r-3}{12} g^{1/3}(k)\right) \underset{k\to\infty}{\longrightarrow} 0.$$

It follows that the probability that one of the intervals I_j contains a monochromatic k-ap from $\overline{AP}_k(n)$ tends to 1. Hence, the probability that a random r-coloring of $[1,n]$ admits a monochromatic k-ap also tends to 1, thereby completing the proof. □

Acknowledgement The author would like to thank an anonymous referee for a detailed reading, stylistic suggestions, and for calling attention to many of the references in the Introduction.

References

[1] József Balogh, Robert Morris, and Wojciech Smotij, Independent Sets in Hypergraphs, *J. Amer. Math. Soc.* **28** (2015), 669-709.

[2] Tom Brown, A pseudo upper bound for the van der Waerden function, *J. Combin. Theory Ser. A* **87** (1999), 233-238.

[3] David Conlon and W.T. Gowers, Combinatorial theorems in sparse random sets, submitted.

[4] Ehud Friedgut, Vojtěch Rödl, Mathias Schacht, Ramsey properties of random discrete structures, *Random Structures and Algorithms* **37** (2010), 407-436.

[5] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), no. 3, 465-588.

[6] Vojtěch Rödl and Andrzej Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917-942.

[7] Vojtěch Rödl and Andrzej Ruciński, Rado partition theorem for random subsets of integers, *Proc. London Math. Soc. (3)* **74** (1997), 481-502.

[8] Mathias Schacht, Extremal results for random discrete structures, preprint.

[9] David Saxton and Andrew Thomason, Hypergraph containers, to appear in *Invent. Math.*

[10] B. L. Van der Waerden, Beweis einer baudetschen Vermutung, *Nieuw Archief voor Wiskunde* **15** (1927), 212-216.

[11] Sujith Vijay, Monochromatic progressions in random colorings, *J. Combin. Theory Ser. A* **119** (2012), 1078-1080.