On constant-multiple-free sets contained in random sets of integers

Sang June Lee*

Department of Mathematical Sciences,
Korea Advanced Institute of Science and Technology (KAIST)

Received 2/25/2013; Accepted 6/5/2013

Abstract

For a rational number $r > 1$, a set A of positive integers is called an r*-multiple-free set* if A does not contain any solution of the equation $rx = y$. The extremal problem of estimating the maximum possible size of r-multiple-free sets contained in $[n] := \{1, 2, \cdots, n\}$ has been studied in combinatorial number theory for theoretical interest and its application to coding theory. Let a and b be relatively prime positive integers such that $a < b$. Wakeham and Wood showed that the maximum size of (b/a)-multiple-free sets contained in $[n]$ is $\frac{b}{b+1}n + O(\log n)$.

In this note we generalize this result as follows. For a real number $p \in (0, 1)$, let $[n]_p$ be a set of integers obtained by choosing each element $i \in [n]$ randomly and independently with probability p. We show that the maximum possible size of (b/a)-multiple-free sets contained in $[n]_p$ is $\frac{b}{b+p}pn + O(\sqrt{pn} \log n \log \log n)$ with probability that goes to 1 as $n \to \infty$.

1 Introduction

A recent trend in extremal combinatorics transfers extremal problems from *dense* environments to *sparse* environments. It has seen to be a fruitful subject of research. In combinatorial number theory, the following extremal

*Email: sjlee242@gmail.com
This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP)(No.2013042157) and by Korea Institute for Advanced Study (KIAS) grant funded by the Korea government(MEST).

problem in a dense environment has been well-studied and successively extended to sparse settings: Fix an equation and estimate the maximum size of subsets of $[n] := \{1, 2, \cdots, n\}$ containing no non-trivial solutions of a given equation.

As an example of this line of research, Kohayakawa–Łuczak–Rödl [8] transferred Roth's classical theorem [10] on arithmetic progressions of length 3 (i.e. solutions to $x_1 + x_3 = 2x_2$) to show that there are such progressions even in random subsets of the integers. Also, Szemerédi's theorem [12] was transferred to random subsets of integers in Conlon–Gowers [2] and Schacht [11]. The result of Erdős–Turán [4], Chowla [1], and Erdős [3] from the 1940s on the maximum size of Sidon sets in $[n]$ was extended in [6, 7] to sparse random subsets of $[n]$, where a *Sidon set* is a set of positive integers not containing any non-trivial solution of $x_1 + x_2 = y_1 + y_2$.

In this note we transfer the following extremal results to random subsets. For a rational number $r > 1$, a set A of positive integers is called an *r-multiple-free set* if A does not contain any solution of $rx = y$. An interesting problem on r-multiple-free sets is of estimating the maximum possible size $f_r(n)$ of r-multiple-free sets contained in $[n] := \{1, 2, \cdots, n\}$. This extremal problem has been studied in [14, 9, 13] and has applications to coding theory in [5].

Wang [14] showed that $f_2(n) = \frac{2}{3}n + O(\log n)$. Leung and Wei [9] proved that for every integer $r > 1$, $f_r(n) = \frac{r}{r+1}n + O(\log n)$. Wakeham and Wood [13] extended it to rational numbers as follows.

Theorem 1 (Wakeham and Wood [13]). *Let a and b be relatively prime integers with $0 < a < b$. Then*

$$f_{b/a}(n) = \frac{b}{b+1}n + O(\log n).$$

We shall investigate the maximum size of constant-multiple-free sets contained in a random subset of $[n]$. Let $[n]_p$ be a random subset of $[n]$ obtained by choosing each element in $[n]$ independently with probability p. Let $f_r([n]_p)$ denote the maximum size of r-multiple-free sets contained in $[n]_p$. We are interested in the behavior of $f_r([n]_p)$ for every rational number $r > 1$.

Theorem 1 gives the answer of the above question for the case $p = 1$. On the other hand, if $p = o(1)$, then the usual deletion methods give that *with high probability* (that is, with probability that goes to 1 as $n \to \infty$) the maximum size of (b/a)-multiple-free sets contained in $[n]_p$ is $np(1 - o(1))$. Hence, from now on, we consider p as a real number with $0 < p < 1$.

Using Chernoff bounds (for example, see Lemma 11), Theorem 1 easily implies the following:

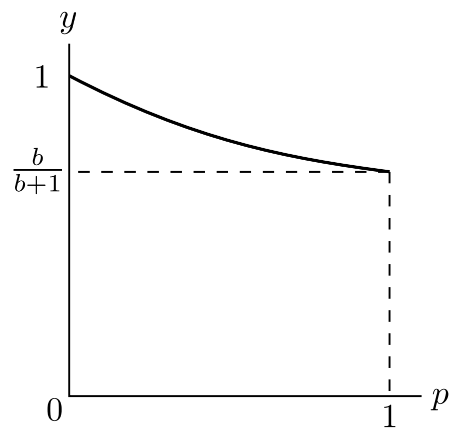

Figure 1: The graph of $y = b/(b+p)$ for $0 \le p \le 1$

Fact 2. *Let $p \in (0,1)$ and let a and b be relatively prime integers such that $0 < a < b$. Let ω be a function of n that goes to ∞ arbitrarily slowly as $n \to \infty$. With high probability, there is a (b/a)-multiple-free set in $[n]_p$ of size*

$$\frac{b}{b+1}pn + \omega\sqrt{pn}.$$

The lower bound on $f_{b/a}([n]_p)$ given by Fact 2 is not tight. The main result of this note improves it:

Theorem 3. *Let $p \in (0,1)$ and let a and b be relatively prime integers such that $0 < a < b$. Then, with high probability,*

$$f_{b/a}\left([n]_p\right) = \frac{b}{b+p}pn + O\left(\sqrt{pn}\log n \log\log n\right).$$

The ratio $\frac{f_{b/a}([n]_p)}{np}$ goes from 1 to $\frac{b}{b+1}$ as p varies from 0 to 1 (See Figure 1). The proof of Theorem 3 is given in Sections 2 and 3. It is graph theoretic.

2 Proof of Theorem 3

In order to show Theorem 3, we use a graph theoretic approach which was used in Wakeham and Wood [13]. Let $r = b/a > 1$ be a rational number. Let $D = (V, E)$ be the directed graph with the vertex set $V = [n]$ in which the set E of arcs (or directed edges) is $\{(x, y) : \ rx = y\}$. Let $D[[n]_p]$ be the subgraph of D induced on $[n]_p$. Observe that $f_r([n]_p)$ is the same as the independence number $\alpha(D[[n]_p])$ of $D[[n]_p]$.

We consider the structure of $D[[n]_p]$. The in-degree and out-degree of each vertex in D are both at most 1. Also, there is no directed cycle in D because $(x, y) \in E$ implies $x < y$. Therefore, each component of D or $D[[n]_p]$ is a directed path.

In order to obtain an independent set of $D[[n]_p]$ of maximum size, we find independent sets in each component. Let C be a component of $D[[n]_p]$. As we mentioned above, C is a directed path. Let $V(C) = \{u_0, u_1, \cdots, u_i, \cdots, u_l\}$ be the vertex set of C such that $u_j < u_{j+1}$ and $(u_j, u_{j+1}) \in E$ for $0 \le j \le l-1$. Observe that $V^*(C) := \{u_0, u_2, u_4, \cdots\} \subset V(C)$ forms an independent set of C of maximum size. Therefore, the set

$$T^* := \bigcup_C V^*(C),$$

where C runs over all components of $D[[n]_p]$, forms an independent set of $D[[n]_p]$ of maximum size. Hence, we have the following.

Lemma 4. $f_r([n]_p) = |T^*|$.

Thus, in order to show Theorem 3, it suffices to show the following.

Lemma 5. *Let $p \in (0, 1)$ and let a and b be relatively prime integers such that $0 < a < b$. Then, with high probability,*

$$|T^*| = \frac{b}{b+p} pn + O\left(\sqrt{pn} \log n \log \log n\right).$$

The proof of Lemma 5 is given in Section 3.

3 Proof of Lemma 5

In the remainder of this note, we prove Lemma 5. For positive integers b and k, let k be an i-th subpower of b if $k = b^i l$ for some $l \not\equiv 0 \pmod b$. Let T_i be the set of i-th subpowers of b in $[n]$. Let $T_i^* \subset T_i$ denote the set of i-th subpowers v of b in $[n]_p$ such that v is at an even distance from the smallest vertex of the component of $D[[n]_p]$ containing v. Observe that $T^* = \bigsqcup_i T_i^*$, and hence,

$$|T^*| = \sum_i |T_i^*|. \tag{1}$$

In Section 3.1, we estimate the expected value $\mathbb{E}(|T^*|)$. Section 3.2 deals with a concentration result about $|T^*|$ with high probability.

3.1 Expectation

We first estimate $\mathbb{E}(|T_i^*|)$ for each i, and their sum $\mathbb{E}(|T^*|)$. Recall that T_i denotes the set of i-th subpowers of b in $[n]$. Note that since $1 \le b^i \le n$, the range of i is $0 \le i \le \log_b n$. It is clear that

$$T_i = \left\{ b^i x \;\middle|\; 1 \le x \le \frac{n}{b^i}, \quad x \not\equiv 0 \pmod{b} \right\}.$$

Hence we have the following:

Fact 6.

$$|T_i| = \frac{b-1}{b}\frac{n}{b^i} \pm 1. \tag{2}$$

We consider two cases separately, based on the parity of i.

Lemma 7. *For $0 \le j \le (\log_b n)/2$, we have*

$$\mathbb{E}\left(|T_{2j}^*|\right) = \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^{2j}} + \left(\frac{p}{b}\right)^{2j} p \right) \pm 1.$$

Proof. First we consider $\Pr\left[v \in T_{2j} \text{ is in } T_{2j}^*\right]$. Let $\{v_0, v_1, v_2, \cdots\}$, where $v_i < v_{i+1}$, be the vertex set of the component of D containing v. Observe that $v_i \in T_i$, and hence, $v = v_{2j}$. The event that $v \in T_{2j}$ is in T_{2j}^* happens only when one of the following holds:

- There is some r with $0 \le r \le j-1$ such that $v_{2j-1-2r} \notin [n]_p$ and $v_i \in [n]_p$ for all $2j - 2r \le i \le 2j$.
- The vertices $v_0, v_1, \cdots, v_{2j}$ are in $[n]_p$.

Hence, we have

$$\Pr\left[v \in T_{2j} \text{ is in } T_{2j}^*\right] = p\left((1-p) + p^2(1-p) + \cdots + p^{2j-2}(1-p) + p^{2j}\right). \tag{3}$$

Thus we infer

$$\begin{aligned} \mathbb{E}\left(|T_{2j}^*|\right) &= |T_{2j}| \cdot \Pr\left[v \in T_{2j} \text{ is in } T_{2j}^*\right] \\ &\overset{(2),(3)}{=} \left(\frac{b-1}{b}\frac{n}{b^{2j}} \pm 1\right) p \left((1-p)\frac{1-p^{2j}}{1-p^2} + p^{2j}\right) \\ &= \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^{2j}} + \frac{p^{2j}}{b^{2j}} p\right) \pm 1, \end{aligned}$$

which completes the proof of Lemma 7. □

Lemma 8. *For $1 \le j \le (\log_b n)/2$, we have*

$$\mathbb{E}\left(|T_{2j-1}^*|\right) = \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^{2j-1}} - \left(\frac{p}{b}\right)^{2j-1} p \right) \pm 1.$$

Proof. Using an argument similar to the proof of (3), one may obtain that

$$\Pr\left[v \in T_{2j-1} \text{ is in } T^*_{2j-1}\right] = p\left((1-p) + p^2(1-p) + \cdots + p^{2j-2}(1-p)\right). \tag{4}$$

Thus we infer

$$\begin{aligned}
\mathbb{E}\left(|T^*_{2j-1}|\right) &= |T_{2j-1}| \cdot \Pr\left[v \in T_{2j-1} \text{ is in } T^*_{2j-1}\right] \\
&\overset{(2),(4)}{=} \left((b-1)\frac{n}{b^{2j}} \pm 1\right) p(1-p)\frac{1-p^{2j}}{1-p^2} \\
&= \frac{b-1}{1+p} pn \left(\frac{1}{b^{2j}} - \left(\frac{p}{b}\right)^{2j}\right) \pm 1,
\end{aligned}$$

which completes the proof of Lemma 8. □

Lemmas 7 and 8 immediately imply the following.

Corollary 9. *For* $0 \le i \le \log_b n$, *we have*

$$\mathbb{E}\left(|T^*_i|\right) = \frac{b-1}{b(1+p)} pn \left(\frac{1}{b^i} + \left(-\frac{p}{b}\right)^i p\right) \pm 1. \tag{5}$$

Summing over all i with $0 \le i \le \log_b n$, we have the following.

Corollary 10.

$$\mathbb{E}\left(|T^*|\right) = \sum_{i=0}^{\log_b n} \mathbb{E}(|T^*_i|) = \frac{b}{b+p} pn + O(\log n).$$

Proof. One may easily see that for $|x| \ge b \ge 2$,

$$\sum_{i=0}^{\log_b n} \frac{1}{x^i} = \frac{x}{x-1} + O\left(\frac{1}{n}\right). \tag{6}$$

Corollary 9 yields that for $b \ge 2$

$$\begin{aligned}
\sum_{i=0}^{\log_b n} \mathbb{E}(|T^*_j|) &\overset{(5)}{=} \sum_{i=0}^{\log_b n} \left[\frac{b-1}{b(1+p)} pn \left(\frac{1}{b^i} + \left(-\frac{p}{b}\right)^i p\right) \pm 1\right] \\
&\overset{(6)}{=} \frac{b-1}{b(1+p)} pn \left[\frac{b}{b-1} + O\left(\frac{1}{n}\right) + \frac{-b/p}{-b/p-1} p + O\left(\frac{1}{n}\right)\right] \\
&\quad + O(\log n) \\
&= \frac{b}{b+p} pn + O(\log n),
\end{aligned}$$

which completes the proof of Corollary 10. □

3.2 Concentration

Next we consider a concentration result about $|T_i^*|$. In other words, we show that $|T_i^*|$ is close to its expectation with high probability. We will apply the following version of Chernoff bounds.

Lemma 11 (Chernoff bound)**.** *Let X_i be independent random variables such that $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$, and let $X = \sum_{i=1}^n X_i$. Then for any $\lambda \geq 0$,*

$$\Pr\left[X \geq (1+\lambda)\mathbb{E}(X)\right] \leq e^{-\frac{\lambda^2}{2+\lambda}\mathbb{E}(X)}, \tag{7}$$

$$\Pr\left[X \leq (1-\lambda)\mathbb{E}(X)\right] \leq e^{-\frac{\lambda^2}{2}\mathbb{E}(X)}. \tag{8}$$

In particular, for $0 \leq \lambda \leq 1$,

$$\Pr\left[|X - \mathbb{E}(X)| \geq \lambda\mathbb{E}(X)\right] \leq 2e^{-\frac{\lambda^2}{3}\mathbb{E}(X)}. \tag{9}$$

We first consider the case when $0 \leq i \leq 0.9 \log_b n$.

Lemma 12. *For $0 \leq i \leq 0.9 \log_b n$, we have*

$$|T_i^*| = \mathbb{E}\left(|T_i^*|\right) + O\left(\sqrt{pn}\log\log n\right) \tag{10}$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log\log n)^2}$.

Proof. Fix i. If $k \in T_i \subset [n]$, then let

$$X_k = \begin{cases} 1 & \text{with probability } p^* \\ 0 & \text{with probability } 1 - p^*. \end{cases}$$

where $p^* = \Pr\left[v \in T_i \text{ is in } T_i^*\right]$. Otherwise, let $X_k = 0$ with probability 1. Let $X = \sum_{k=1}^n X_k$. Observe that

$$X = |T_i^*| \tag{11}$$

as random variables.

Note that for each $k \in T_i$, the event that $k \in T_i^*$ depends only on the events that $v \in [n]_p$, where the vertices v are in the component of D containing k and $v \leq k$. Hence, X_k are independent for all $k \in T_i$. Therefore we are able to use Chernoff bounds (Lemma 11) for a concentration result on X.

Set $\lambda = \dfrac{\log\log n}{\sqrt{\mathbb{E}(X)}}$. Note that $0 \leq \lambda \leq 1$ for $0 \leq i \leq 0.9 \log_b n$ since

$$\mathbb{E}(X) \geq \Omega\left(pn\frac{\varepsilon_p}{b^i}\right) \geq \Omega\left(\frac{\varepsilon_p}{n^{0.9}}\right) = \Omega\left(\varepsilon_p pn^{0.1}\right),$$

where ε_p is a positive constant such that $\varepsilon_p \to 0$ as $p \to 1$. The inequality (9) yields that

$$\Pr\left[|X - \mathbb{E}(X)| \geq \sqrt{\mathbb{E}\left(X\right)} \log\log n\right] \leq 2e^{-\frac{1}{3}(\log\log n)^2}. \tag{12}$$

Corollary 9 yields that $\mathbb{E}\left(|X|\right) = O(pn)$, and hence, we infer that

$$X = \mathbb{E}\left(X\right) + O\left(\sqrt{pn}\log\log n\right)$$

with probability at least $1 - 2e^{-\frac{1}{3}(\log\log n)^2}$. This together with (11) completes the proof of Lemma 12. □

Next we consider the remaining case when $0.9\log_b n \leq i \leq \log_b n$.

Lemma 13.

$$\sum_{i=\lfloor 0.9\log_b n\rfloor+1}^{\log_b n} |T_i^*| = O(\sqrt{pn})$$

with probability at least $1 - o(1)$.

Proof. Corollary 9 implies that

$$\mathbb{E}(|T_i^*|) = O\left(pn\frac{1}{b^i}\right) = O\left(pn^{0.1}\right) = O\left((pn)^{0.1}\right), \tag{13}$$

where the second inequality holds for $i \geq 0.9\log_b n$. Markov's inequality completes the proof of Lemma 13. □

Now we are ready to show Lemma 5.

Proof of Lemma 5. We have that

$$|T^*| \quad = \quad \sum_{i=1}^{\log_b n} |T_i^*| = \sum_{i=1}^{\lfloor 0.9\log_b n\rfloor} |T_i^*| + \sum_{i=\lfloor 0.9\log_b n\rfloor+1}^{\log_b n} |T_i^*|.$$

Lemmas 12 and 13 give that

$$|T^*| \quad = \quad \sum_{i=1}^{\log_b n} \mathbb{E}\left(|T_i^*|\right) + O\left(\sqrt{pn}\log n\log\log n\right),$$

with probability at least

$$\begin{aligned} & 1 - (\log_b n)\cdot 2e^{-\frac{1}{3}(\log\log n)^2} - o(1) \\ = \quad & 1 - 2e^{\log\log_b n - \frac{1}{3}(\log\log n)^2} - o(1) = 1 - o(1). \end{aligned}$$

This together with Corollary 10 implies that with high probability

$$|T^*| \quad = \quad \frac{b}{b+p} pn + O\left(\sqrt{pn}\log n \log\log n\right),$$

which completes the proof of Lemma 5. □

Acknowledgement. *The author thanks Yoshiharu Kohayakawa and Mark Siggers for helpful comments and corrections, and thanks Jaigyoung Choe for his support while the author was at KIAS (Korea Institute for Advanced Study).*

References

[1] S. Chowla. Solution of a problem of Erdös and Turan in additive-number theory. *Proc. Nat. Acad. Sci. India. Sect. A.*, 14:1–2, 1944.

[2] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. submitted, 70pp, 2010.

[3] P. Erdős. On a problem of Sidon in additive number theory and on some related problems. Addendum. *J. London Math. Soc.*, 19:208, 1944.

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

[5] M. Jimbo, M. Mishima, S. Janiszewski, A. Y. Teymorian, and V. D. Tonchev. On conflict-avoiding codes of length $n = 4m$ for three active users. *IEEE Trans. Inform. Theory*, 53(8):2732–2742, 2007.

[6] Y. Kohayakawa, S. Lee, and V. Rödl. The maximum size of a Sidon set contained in a sparse random set of integers. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 159–171, Philadelphia, PA, 2011. SIAM.

[7] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers. *Random Structures & Algorithms*, accepted, 25pp, 2013. DOI 10.1002/rsa

[8] Y. Kohayakawa, T. Łuczak, and V. Rödl. Arithmetic progressions of length three in subsets of a random set. *Acta Arith.*, 75(2):133–163, 1996.

[9] J. Y.-T. Leung and W.-D. Wei. Maximal k-multiple-free sets of integers. *Ars Combin.*, 38:113–117, 1994.

[10] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[11] M. Schacht. Extremal results for random discrete structures. submitted, 27pp, 2009.

[12] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[13] D. Wakeham and D. Wood. On multiplicative Sidon sets. *INTEGERS*, 13:#A26, 2013.

[14] E. T. H. Wang. On double-free sets of integers. *Ars Combin.*, 28:97–100, 1989.