ON ZERO-SUM FREE SEQUENCES CONTAINED IN RANDOM SUBSETS OF FINITE CYCLIC GROUPS

SANG JUNE LEE AND JUN SEOK OH

Abstract. Let C_n be a cyclic group of order n. A *sequence* S of length ℓ over C_n is a sequence $S = a_1 \cdot a_2 \cdot \ldots \cdot a_\ell$ of ℓ elements in C_n, where a repetition of elements is allowed and their order is disregarded. We say that S is a zero-sum sequence if $\Sigma_{i=1}^{\ell} a_i = 0$ and that S is a zero-sum free sequence if S contains no zero-sum subsequence.

Let R be a random subset of C_n obtained by choosing each element in C_n independently with probability p. Let N_{n-1-k}^R be the number of zero-sum free sequences of length $n-1-k$ in R. Also, let $N_{n-1-k,d}^R$ be the number of zero-sum free sequences of length $n-1-k$ having d distinct elements in R. We obtain the expectation of N_{n-1-k}^R and $N_{n-1-k,d}^R$ for $0 \le k \le \lfloor n/3 \rfloor$. We also show a concentration result on N_{n-1-k}^R and $N_{n-1-k,d}^R$ when k is fixed.

1. Introduction

Let C_n be a cyclic group of order n. A *sequence* S of length ℓ over C_n is a sequence

$$S = a_1 \cdot a_2 \cdot \ldots \cdot a_\ell$$

of ℓ elements in C_n, where a repetition of elements is allowed and their order is disregarded. We say that a sequence S over C_n is contained in $A \subset C_n$ if each element in S is contained in A. For $a \in C_n$, let

$$\mathsf{v}_a(S) = |\{i \in [1, \ell] \mid a_i = a\}|$$

be the *multiplicity* of a in S. A *subsequence* T of S is a sequence over C_n satisfying $\mathsf{v}_a(T) \le \mathsf{v}_a(S)$ for all $a \in C_n$. We say that S is a *zero-sum sequence* if $a_1 + a_2 + \ldots + a_\ell = 0$. A sequence is called *zero-sum free* if it contains no zero-sum subsequence.

Date: Friday 6^{th} March, 2020.
2010 *Mathematics Subject Classification.* 11B50, 11B30, 05D40.
Key words and phrases. zero-sum free sequence, cyclic group, integer partition, Young diagram, hypergraph, Kim–Vu polynomial concentration.
The first author was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1F1A1058860). The second author is the corresponding author. This work was partially done while the second author was visiting Kyung Hee University in South Korea.

An initial study of zero-sum sequences dates back to 1961 when Erdős, Ginzburg, and Ziv [6] proved that $2n-1$ is the smallest positive integer ℓ such that every sequence of length ℓ over C_n has a zero-sum subsequence of length n. Since that time, zero-sum sequences over a finite group have actively studied in additive combinatorics. For more details, see a survey paper by Gao and Geroldinger [8]. Although earlier works often focused on finite abelian groups, an application to factorization theory and invariant theory pushed the object forward to non-abelian groups. The reader can refer to Geroldinger, Grynkiewicz, Zhong, and the second author [9] for recent progress with respect to factorization theory and to Cziszter, Domokos, and Szöllősi [3, 5] for connection with invariant theory.

In this paper, we focus on zero-sum free sequences over a cyclic group. Well-known problems about zero-sum free sequences over a finite group are to determine the maximum length of zero-sum free sequences, which is a combinatorial group invariant known as the *Davenport constant*, and to characterize the structure of zero-sum free sequences. Observe that the maximum length of all zero-sum free sequences over C_n is $n-1$. Also, we have that S is a zero-sum free sequence of length $n-1$ over C_n if and only if

$$S = \underbrace{g \boldsymbol{\cdot} g \boldsymbol{\cdot} \ldots \boldsymbol{\cdot} g}_{n-1}$$

for a generator $g \in C_n$. Gao [7] proved the following result on the structure of long zero-sum free sequences over C_n.

Theorem 1 (Theorem 4.3 in [8], Lemma 2.5 in [7]). *Let $n \geq 2$ and $0 \leq k \leq \lfloor \frac{n}{3} \rfloor$. Then S is a zero-sum free sequence of length $n-1-k$ over C_n if and only if*

$$S = \underbrace{g \boldsymbol{\cdot}\ g \boldsymbol{\cdot}\ \ldots \boldsymbol{\cdot}\ g}_{n-1-2k} \boldsymbol{\cdot}\ (x_1 g) \boldsymbol{\cdot}\ (x_2 g) \boldsymbol{\cdot}\ \ldots \boldsymbol{\cdot}\ (x_k g),$$

where g is a generator of C_n and $x_1, x_2, \ldots, x_k$ are positive integers such that

$$1 \leq x_1 \leq x_2 \leq \ldots \leq x_k \quad \textit{and} \quad x_1 + x_2 + \ldots + x_k \leq 2k. \tag{1}$$

Theorem 1 was generalized by Savchev and Chen [17] on the zero-sum free sequences of length at least $(n+1)/2$ over C_n. Theorem 1 and the result by Savchev and Chen were applied to the number of minimal zero-sum sequences of long length by Ponomarenko [15] and Cziszter, Domokos, and Geroldinger [4], respectively.

Remark that Theorem 4.3 in [8] only gives the statement in Theorem 1 from the left-hand side to the right-hand side. The proof from the right-hand side to the left-hand side is obvious since g is a generator of C_n and all subsequences T of S satisfy $\sigma(T) = \ell g \neq 0$ for some integer $0 < \ell < n$, where $\sigma(T)$ is the sum of all elements in T.

In this paper, we are interested in zero-sum free sequences of a given length contained in a random subset of C_n. Investigating how classical extremal results in *dense* environments transfer to *sparse* settings has become a deep line of research. For example, Roth's theorem on 3-term arithmetic progressions [16] was generalized for random subsets of integers [13], and there are recent generalizations about various classical extremal results by Schacht [18] and Conlon and Gowers [2].

Let R be a random subset of C_n obtained by choosing each element in C_n independently with probability p. Let N_{n-1-k} be the number of zero-sum free sequences of length $n-1-k$ over C_n. Also, let N^R_{n-1-k} be the number of zero-sum free sequences of length $n-1-k$ in R. The result on the expectation of N^R_{n-1-k} is as follows.

Theorem 2. *Let $n \geq 2$ and $0 \leq k \leq \left\lfloor \frac{n}{3} \right\rfloor$. The expected number of zero-sum free sequences of length $n-1-k$ in a random subset R of C_n is*

$$\mathbb{E}\left(N^R_{n-1-k}\right) = \varphi(n)\left[p + \sum_{d=2}^{D} p^d \left(\sum_{j=\frac{(d-1)d}{2}}^{k} q(j, d-1)\right)\right],$$

where

- $D = \left\lfloor \frac{1+\sqrt{1+8k}}{2} \right\rfloor$,
- *$\varphi(n)$ denotes the number of generators in C_n, and*
- *$q(j, d-1)$ is the number of partitions of j having $d-1$ distinct parts.*

The number $q(j, d-1)$ can be computed in two ways: The first way is based on its generating function (see Section 2.1 for details). Second, we provide a recursive formula for computing $X_{k,d-1} = \sum_{j=(d-1)d/2}^{k} q(j, d-1)$ (see Section 2.2).

If k is fixed, then we can obtain a simpler statement as follows.

Corollary 3. *If k is fixed and $p = o(1)$ as $n \to \infty$, then*

$$\mathbb{E}\left(N^R_{n-1-k}\right) = p\varphi(n)\left(1 + O_k(p)\right),$$

where the constant in O_k depends only on k.

Next, we have a concentration result on N^R_{n-1-k} when k is fixed.

Theorem 4. *Let k be fixed, and let p be such that*

$$\frac{(\log n)^{2d} \log\log n}{n} \ll p \ll 1.$$

Then, asymptotically almost surely (a.a.s.)

$$N^R_{n-1-k} = p\varphi(n) + O_k\left(p^2\varphi(n) + \sqrt{p\varphi(n)}(\log n)^d\right),$$

where the constant in O_k depends only on k.

Moreover, we have a refined result. Let $N_{n-1-k,d}$ be the number of zero-sum free sequences of length $n-1-k$ having d distinct elements over C_n. Also, let $N^R_{n-1-k,d}$ be the number of zero-sum free sequences of length $n-1-k$ having d distinct elements contained in a random subset R of C_n. We show a concentration result on $N^R_{n-1-k,d}$.

Theorem 5. *If* $0 \le k \le \left\lfloor \frac{n}{3} \right\rfloor$ *and*

$$p \gg \frac{\log\log n}{n},$$

then we have that a.a.s.

$$p\varphi(n) - \omega\sqrt{p\varphi(n)} \le N^R_{n-1-k,1} \le p\varphi(n) + \omega\sqrt{p\varphi(n)},$$

where ω *tends to* ∞ *arbitrarily slowly as* $n \to \infty$.

Let $d \ge 2$. *If* k *is fixed and*

$$p \gg \frac{(\log n)^2(\log\log n)^{1/d}}{n^{1/d}},$$

then we have that a.a.s.

$$N^R_{n-1-k,d} = p^d\varphi(n)\left(\sum_{j=\frac{(d-1)d}{2}}^{k} q(j,d-1)\right) + O_k\left(\sqrt{p^d\varphi(n)}(\log n)^d\right).$$

The organization of this paper is as follows. In Section 2, we consider expectations and prove Theorem 2 and Corollary 3. Then, we deal with our concentration results and prove Theorems 4 and 5 in Section 3.

2. Expectation

In this section, we prove Theorem 2 and Corollary 3. Also, we provide a recursive formula to compute the important value $X_{k,d-1} = \sum_{j=(d-1)d/2}^{k} q(j,d-1)$ given in Theorem 2.

2.1. Proofs of Theorem 2 and Corollary 3.

It turns out that the number of distinct elements in a zero-sum free sequence plays an important role since each element in C_n is contained in a random set R with probability p. Recall that $N_{n-1-k,d}$ is the number of zero-sum free sequences of length $n-1-k$ having d distinct elements over C_n, and $N^R_{n-1-k,d}$ is the number of zero-sum free sequences over C_n of length $n-1-k$ having d distinct elements contained in a random set R.

Clearly, the expectation of $N^R_{n-1-k,d}$ is

$$\mathbb{E}\left(N^R_{n-1-k,d}\right) = p^d N_{n-1-k,d}.$$

Based on Theorem 1, the numbers $N_{n-1-k,d}$ and $N^R_{n-1-k,d}$ are related to the number of

$$(x_1, x_2, \dots, x_k)$$

satisfying that $x_1, x_2, \dots, x_k$ are positive integers such that (1) holds and the number of distinct $x_i \neq 1$ is $d-1$. With $x'_i := x_i - 1$, the number can be simplified as follows.

Definition 6. Let $X_{k,d}$ be the number of $(x'_1, x'_2, \dots x'_k)$ such that

$$0 \leq x'_1 \leq x'_2 \leq \dots \leq x'_k, \quad x'_1 + x'_2 + \dots + x'_k \leq k,$$

and the number of distinct positive x'_i is d.

Theorem 1 and Definition 6 give that

$$N_{n-1-k,d} = \varphi(n) X_{k,d-1}, \tag{2}$$

where $\varphi(n)$ is the number of generators in C_n. Therefore, the expectation of $N^R_{n-1-k,d}$ is

$$\mathbb{E}(N^R_{n-1-k,d}) = p^d \varphi(n) X_{k,d-1}. \tag{3}$$

From now on, we focus on estimating $X_{k,d-1}$. To this end, we use the definition of a partition of an integer. A *partition* of a positive integer k is a non-decreasing sequence whose sum equals k. A partition λ of k can be shortly expressed by

$$1^{r_1}\ 2^{r_2}\ \cdots\ t^{r_t}$$

meaning that

$$k = (\underbrace{1 + 1 + \ldots + 1}_{r_1}) + (\underbrace{2 + 2 + \ldots + 2}_{r_2}) + \ldots + (\underbrace{t + t + \ldots + t}_{r_t}).$$

If λ is a partition of k, then we denote $\lambda \vdash k$. Let $|\lambda| = k$ if $\lambda \vdash k$.

Let $q(k,d)$ be the number of partitions of k having d distinct parts. For example, all partitions of 7 are as follows:

- $1^7,\ 7^1,$
- $1^5 2^1,\ 1^3 2^2,\ 1^1 2^3,\ 1^4 3^1,\ 1^1 3^2,\ 2^2 3^1,\ 1^3 4^1,\ 3^1 4^1,\ 1^2 5^1,\ 2^1 5^1,\ 1^1 6^1,$
- $1^2 2^1 3^1,\ 1^1 2^1 4^1.$

We have that $q(7,1) = 2$, $q(7,2) = 11$, $q(7,3) = 2$, and $q(7,d) = 0$ for $d \geq 4$.

Recalling Definition 6, we have that $X_{k,d}$ is the same as the number of partitions λ of at most k having d distinct parts. Observe that if λ is counted for $X_{k,d}$, then

$$\frac{d(d+1)}{2} \leq |\lambda| \leq k \tag{4}$$

because λ contains parts with at least $1, 2, \ldots, d$. Thus, we have

$$X_{k,d} = \sum_{j=\frac{d(d+1)}{2}}^{k} q(j, d). \tag{5}$$

Remark that the number $q(j, d)$ can be found in A116608 of the on-line encyclopedia of integer sequences (OEIS), and it can be computed from its generating function

$$Q(x, t) = -1 + \prod_{i=1}^{\infty} \left(1 + \frac{tx^i}{1 - x^i}\right),$$

where

$$Q(x, t) = \sum_{j,d \geq 1} q(j, d) x^j t^d.$$

There are related results on $q(j, d)$. Kim [11] constructed a generating function with one variable for $q(j, d)$ when d is fixed. Also, Goh and Schmutz [10] obtained the asymptotic distribution of the number of distinct part sizes in a random integer partition. On the other hand, $X_{k,d}$ is not found in OEIS.

We are ready to prove Theorem 2.

Proof of Theorem 2. Trivially, the expected number of zero-sum free sequences of length $n - 1 - k$ with same elements in R is

$$\mathbb{E}\left(N^R_{n-1-k,1}\right) = \varphi(n)p.$$

Next, for $d \geq 2$, we infer that

$$\mathbb{E}\left(N^R_{n-1-k,d}\right) = p^d N_{n-1-k,d} \overset{(2)}{=} p^d \varphi(n) X_{k,d-1} \overset{(5)}{=} p^d \varphi(n) \left(\sum_{j=\frac{(d-1)d}{2}}^{k} q(j, d-1) \right).$$

Next we consider the range of d. If $X_{k,d-1}$ is positive, then (4) gives that

$$k \geq \frac{(d-1)d}{2}.$$

Hence, let D be the lagest integer d satisfying $(d-1)d/2 \leq k$, and then,

$$d \leq D = \left\lfloor \frac{1 + \sqrt{1 + 8k}}{2} \right\rfloor.$$

This completes our proof of Theorem 2. □

Now we are ready to prove Corollary 3 using Theorem 2.

Proof of Corollary 3. For a fixed k, Theorem 2 gives that

$$\begin{aligned}\mathbb{E}\left(N^R_{n-1-k}\right) &= \varphi(n)\left[p+\sum_{d=2}^{D}p^d\left(\sum_{j=\frac{(d-1)d}{2}}^{k}q(j,d-1)\right)\right]\\ &= p\varphi(n)\left(1+\sum_{d=2}^{D}O_k(p^{d-1})\right)\\ &= p\varphi(n)\left(1+O_k(p)\right),\end{aligned}$$

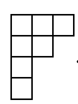
where the constant in O_k depends only on k, which completes the proof of Corollary 3. □

2.2. **Recursive formula for $X_{k,d}$.** Here, we give another way to compute the important value

$$X_{k,d-1}=\sum_{j=\frac{(d-1)d}{2}}^{k}q(j,d-1)$$

given in Theorem 2 using a recursive formula.

A partition of an integer can be illustrated by a *Young diagram* (also called a *Ferrers diagram*), which is a useful way to understand a partition in combinatorics. A Young diagram corresponding to a partition $\lambda \vdash k$ is a collection of left-justified rows of k boxes piled up in non-decreasing order of row lengths from parts. For example, the partition $1^2 2^1 3^1 \vdash 7$ corresponds to the Young diagram


.

Let $Y_{b,c,d}$ be the number of partitions of at most b with at most c parts having d distinct parts. Equivalently, $Y_{b,c,d}$ is the number of Young diagrams with at most b boxes, at most c rows, and d distinct rows. See Figure 1 (a).

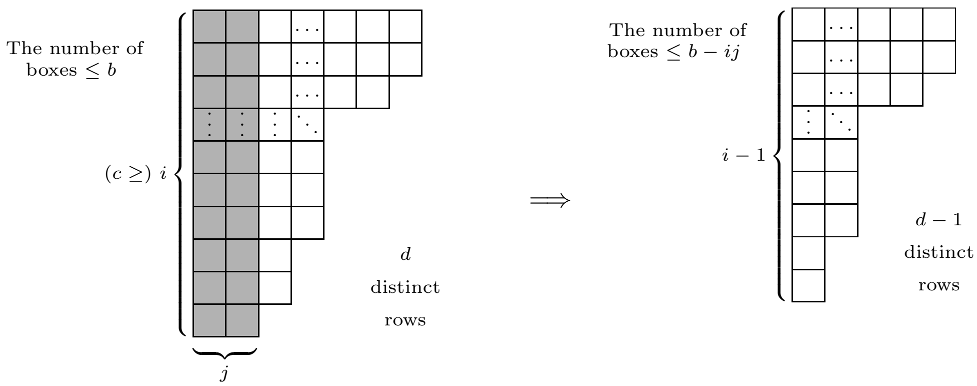
Note that $Y_{b,c,d}>0$ if and only if $b \geq \frac{d(d+1)}{2}$ and $c \geq d$, where the first inequaity follows from (4). Observe that

$$X_{k,d}=Y_{k,k,d}.$$

A recursive formula for $Y_{b,c,d}$ is as follows. (Hence we have a recursive formula for $X_{k,d}$.)

Lemma 7. *We have that, for $b \geq \frac{d(d+1)}{2}$ and $c \geq d \geq 2$,*

$$Y_{b,c,d}=\sum_{i=1}^{c}\sum_{j=1}^{\left\lfloor\frac{b-(d-1)d/2}{i}\right\rfloor}Y_{b-ij,i-1,d-1} \tag{6}$$



(a) $Y_{b,c,d}$ (b) $Y_{b-ij,i-1,d-1}$

FIGURE 1. Definition of $Y_{b,c,d}$ and the deletion process

and, for $b \geq 1$ *and* $c \geq 1$*,*

$$Y_{b,c,1} = \sum_{i=1}^{c} \left\lfloor \frac{b}{i} \right\rfloor. \tag{7}$$

Proof. We first show (6). We delete the gray retangle in Figure 1 from a Young diagram counted for $Y_{b,c,d}$, and then we have a Young diagram with at most $b - ij$ boxes, at most $i-1$ rows, and $d-1$ distinct rows.

We consider the ranges of i and j. Clearly, the range of i is $1 \leq i \leq c$. Then the remaining Young diagram after the deletion has $d-1$ distinct rows, and hence, it has at least $(d-1)d/2$ boxs. Thus,

$$ij + \frac{(d-1)d}{2} \leq b.$$

So the range of j is

$$1 \leq j \leq \left\lfloor \frac{b-(d-1)d/2}{i} \right\rfloor.$$

Next, we show (7). The number $Y_{b,c,1}$ is the same as the number of rectangles with at most b boxes and at most c rows. Let i and j be the numbers of rows and columns, respectively, of such a rectangle. Clearly, $1 \leq i \leq c$. Since $ij \leq b$, we have $j \leq \left\lfloor \frac{b}{i} \right\rfloor$. □

3. CONCENTRATION

Recall that $N^R_{n-1-k,d}$ be the number of zero-sum free sequences of length $n-1-k$ having d distinct elements in a random subset R. From (3), recall that

$$\mathbb{E}(N^R_{n-1-k,d}) = p^d \varphi(n) X_{k,d-1}.$$

From now on, we consider a concentration of $N^R_{n-1-k,d}$ and N^R_{n-1-k} using a graph theoretical approach called the Kim–Vu polynomial concentration result.

3.1. **Kim–Vu polynomial concentration result.** Let $\mathcal{H} = (V, E)$ be a weighted hypergraph with $V = [n] := \{0, 1, \ldots, n-1\}$. Recall that R is a random subset of $[n]$ obtained by selecting each $v \in [n]$ independently with probability p. Let $\mathcal{H}[R]$ be the sub-hypergraph of $\mathcal{H}$ induced on R, and we let Z be the sum of weights of hyperedges in $\mathcal{H}[R]$. Kim and Vu [12] obtained a result that provides a concentration of Z around its mean $\mathbb{E}(Z)$ with high probability. For more details, see Alon and Spencer [1]. To state the result, we need some definitions.

Definition 8. Let ℓ be the maximum size of hyperedges in $\mathcal{H}$, and let $A \subset [n]$ be such that $|A| \leq \ell$. We let

- $Z_A :=$ the sum of weights of hyperedges in $\mathcal{H}[R]$ containing A,
- $\mathbb{E}_A := \mathbb{E}\left(Z_A \mid A \subset R\right)$,
- $\mathbb{E}_i :=$ the maximum of $\mathbb{E}_A$ for $A \subset [n]$ with $|A| = i$,
- $$\mathbb{E}' := \max_{1 \leq i \leq \ell} \mathbb{E}_i \quad \text{and} \quad \mathbb{E}^* := \max\{\mathbb{E}', \mathbb{E}(Z)\}.$$

The concentration result by Kim and Vu [12] is as follows.

Theorem 9 (**Kim–Vu polynomial concentration inequality**). *With the notation as above, we have that, for each $\lambda > 1$,*

$$\Pr\left[|Z - \mathbb{E}(Z)| > a_\ell \sqrt{\mathbb{E}' \cdot \mathbb{E}^*} \lambda^\ell\right] < 2e^{-\lambda+2} n^{\ell-1},$$

where $a_\ell = 8^\ell (\ell!)^{1/2}$.

3.2. **Hypergraph and example.** For a given positive integer k, we define the hypergraph $\mathcal{H}_{n-1-k} = \mathcal{H}_{n-1-k}(C_n) = ([n], E)$ such that $a_1 \cdot a_2 \cdot \ldots \cdot a_{n-1-k}$ is a zero-sum free sequence over C_n if and only if the corresponding set $\{b_1, b_2, \ldots, b_\ell\} = \{a_1, a_2, \ldots, a_{n-1-k}\}$, with $1 \leq \ell \leq n-1-k$, is contained in E. The weight of an hyperedge $\{b_1, b_2, \ldots, b_\ell\}$ of $\mathcal{H}_{n-1-k}$ is the number of zero-sum free sequences over C_n consisting of $b_1, b_2, \ldots, b_\ell$.

Then $\mathbb{E}_{d,A}$ defined above is the expected number of zero-sum free sequences of length $n-1-k$ having d distinct elements that contains $A \subset C_n$ and is contained in R under the condition that $A \subset R$. Also, for $1 \leq i \leq d$, let

$$\mathbb{E}_{d,i} = \max\{\mathbb{E}_{d,A} \mid A \subset C_n \text{ with } |A| = i\}.$$

We will estimate $\mathbb{E}_{d,A}$ and $\mathbb{E}_{d,i}$.

For an easier understanding, we give an example in C_8 before estimating $\mathbb{E}_{d,A}$ and $\mathbb{E}_{d,i}$ in a general C_n. Let $C_8 = \{0, 1, 2, 3, 4, 5, 6, 7\}$ and we consider the case where $k = 2$. In this case, the length of zero-sum free sequences is $n - 1 - k = 8-1-2 = 5$. All generators in C_8 are $1, 3, 5, 7$, and all possible (x_1, x_2) in Theorem 1 are $(1, 1), (1, 2), (1, 3)$, and $(2, 2)$. Thus, Theorem 1 gives that all zero-sum free sequences of length 5 over C_8 are

$$\begin{array}{llll} 1\cdot1\cdot1\cdot1\cdot1 & 3\cdot3\cdot3\cdot3\cdot3 & 5\cdot5\cdot5\cdot5\cdot5 & 7\cdot7\cdot7\cdot7\cdot7 \\ 1\cdot1\cdot1\cdot1\cdot2 & 3\cdot3\cdot3\cdot3\cdot6 & 5\cdot5\cdot5\cdot5\cdot2 & 7\cdot7\cdot7\cdot7\cdot6 \\ 1\cdot1\cdot1\cdot1\cdot3 & 3\cdot3\cdot3\cdot3\cdot1 & 5\cdot5\cdot5\cdot5\cdot7 & 7\cdot7\cdot7\cdot7\cdot5 \\ 1\cdot1\cdot1\cdot2\cdot2 & 3\cdot3\cdot3\cdot6\cdot6 & 5\cdot5\cdot5\cdot2\cdot2 & 7\cdot7\cdot7\cdot6\cdot6. \end{array}$$

Hence, the hypergraph $\mathcal{H}_5(C_8)$ has hyperedges as follows:

Hyperedge	$\{1\}$	$\{3\}$	$\{5\}$	$\{7\}$	$\{1,2\}$	$\{1,3\}$	$\{3,6\}$	$\{5,2\}$	$\{5,7\}$	$\{7,6\}$
Weight	1	1	1	1	2	2	2	2	2	2

As an example, we estimate $\mathbb{E}_{2,1}$ by considering $\mathbb{E}_{2,\{a\}}$ for $a \in C_8$. First, let $a = 1$. Note that our goal here is not to get the exact value of $\mathbb{E}_{2,\{1\}}$ but to obtain a uniform upper bound of $\mathbb{E}_{2,\{a\}}$ for all $a \in C_8$. For a generator g, there are several cases we need to deal with:

• **Case 1 ($a = 1 = g$):** Trivially, $g = 1$. Since $(x'_1, x'_2) = (0,1), (0,2)$, or $(1,1)$, we have $(x_1, x_2) = (1,2), (1,3)$, or $(2,2)$, and hence, all zero-sum free sequences of this case in C_8 are

$$1\cdot1\cdot1\cdot1\cdot(2\cdot1) \qquad 1\cdot1\cdot1\cdot1\cdot(3\cdot1) \qquad 1\cdot1\cdot1\cdot(2\cdot1)\cdot(2\cdot1).$$

Thus, the expected number of all zero-sum free sequences of this case in R is

$$X_{2,1}\cdot p.$$

• **Case 2 ($a = 1 = 2g$):** There is no such g, but we go forward to get a uniform upper bound. Since $a = 2g$, we have $x'_\ell = 1$ for some ℓ. Hence, the number of all zero-sum free sequences of this case in C_8 is at most $X_{2-1,1} + X_{2-1,0}$, where the first term is from the situation when all other x' are different from x'_ℓ and the second term is from the other situation. Thus, the expected number of all zero-sum free sequences of this case in R is at most

$$(X_{2-1,1} + X_{2-1,0})\, p.$$

• **Case 3 ($a = 1 = 3g$):** We infer that $g = 3$ and $(x'_1, x'_2) = (0,2)$. Hence, every zero-sum free sequences of this case in C_8 is

$$3\cdot3\cdot3\cdot3\cdot(3\cdot3) = 3\cdot3\cdot3\cdot3\cdot1.$$

The number of all zero-sum free sequences of this case over C_8 is

$$X_{2-2,1} + X_{2-2,0} \le X_{2-1,1} + X_{2-1,0}.$$

Thus, the expected number of all zero-sum free sequences of this case in R is at most

$$(X_{2-1,1} + X_{2-1,0})\, p.$$

Therefore,

$$\mathbb{E}_{2,\{1\}} \leq (X_{2,1} + 2\,(X_{2-1,1} + X_{2-1,0}))\, p.$$

By the same argument, for every $a \in C_8$, we have that $\mathbb{E}_{2,\{a\}}$ has the same upper bound, and hence,

$$\mathbb{E}_{2,1} \leq (X_{2,1} + 2\,(X_{2-1,1} + X_{2-1,0}))\, p.$$

In a similar way, one can estimate $\mathbb{E}_2$ in C_8, which gives $\mathbb{E}'$ and $\mathbb{E}^*$.

3.3. **Estimating $\mathbb{E}_{d,i}$.** We are ready to estimate $\mathbb{E}_{d,i}$ in a general C_n. First, we consider the case where $i = 1$.

Lemma 10. *For $2 \leq d \leq \left\lfloor \frac{1+\sqrt{1+8k}}{2} \right\rfloor$, we have that*

$$\begin{aligned} \mathbb{E}_{d,1} &\leq p^{d-1}\,(X_{k,d-1} + k(X_{k-1,d-1} + X_{k-1,d-2})) \quad \textit{and} \\ \mathbb{E}_{1,1} &= 1. \end{aligned}$$

Proof. Fix $a \in C_n$. We estimate the expected number of zero-sum free sequences

$$g \boldsymbol{\cdot} \ldots \boldsymbol{\cdot} g \boldsymbol{\cdot} (x_1 g) \boldsymbol{\cdot} \ldots \boldsymbol{\cdot} (x_k g)$$

in R containing $\{a\}$ with two cases separately: for a generator g, the first case is when $a = g$, and the second case is when $a = jg$ for $2 \leq j \leq k+1$.

- **Case 1 ($a = g$):** The number of zero-sum free sequences over C_n containing $a = g$ is $X_{k,d-1}$. Hence, the expected number of zero-sum free sequences in R containing $a = g$ is

$$X_{k,d-1} \cdot p^{d-1}. \tag{8}$$

- **Case 2 ($a = jg$ for $2 \leq j \leq k+1$):** We first estimate the number of zero-sum free sequences over C_n containing $a = jg = x_\ell g$ for some ℓ. Since $x'_\ell = x_\ell - 1 \geq 1$, the remaining $x'_1, \ldots, x'_{\ell-1}, x'_{\ell+1}, \ldots, x'_k$ satisfy $\sum_{\substack{1 \leq i \leq k \\ i \neq \ell}} x'_i \leq k-1$. If $x'_1, \ldots, x'_{\ell-1}, x'_{\ell+1}, \ldots, x'_k$ are different from x'_ℓ, then the number of zero-sum free sequences over C_n is at most $X_{k-1,d-2}$. Otherwise, the number of zero-sum free sequences over C_n is at most $X_{k-1,d-1}$. Since $2 \leq j \leq k+1$, the expected number of zero-sum free sequences of this case in R is

$$k(X_{k-1,d-1} + X_{k-1,d-2})p^{d-1}. \tag{9}$$

From (8) and (9), we have that

$$\mathbb{E}_{d,1} \leq \max_{\{a\}} \mathbb{E}_{d,\{a\}} \leq p^{d-1}\,(X_{k,d-1} + k(X_{k-1,d-1} + X_{k-1,d-2}))\,,$$

which completes our proof of the lemma. □

Next, we consider a general i with $|A| = i$.

Lemma 11. *For* $1 \le i < d \le \left\lfloor \frac{1+\sqrt{1+8k}}{2} \right\rfloor$, *we have that*

$$\begin{aligned} \mathbb{E}_{d,i} &\le p^{d-i}\left[i\binom{k}{i-1}+\binom{k}{i}\right]\left(\sum_{j=0}^{i} X_{k-i+1,d-1-j}\right) \quad \textit{and} \\ \mathbb{E}_{d,d} &= 1. \end{aligned}$$

Proof. Fix $a_1, a_2, \dots, a_i \in C_n$. We estimate the number of zero-sum free sequences

$$g \cdot \ldots \cdot g \cdot (x_1 g) \cdot \ldots \cdot (x_k g)$$

containing $\{a_1, a_2, \dots, a_i\}$ with two cases separately: for a generator g, the first case is when $g = a_\ell$ for some ℓ, and the second case is when $g \neq a_\ell$ for all ℓ.

• **Case 1 (**$a_1 = g$ **and** $\{a_2, \dots, a_i\} = \{j_2 g, \dots, j_i g\}$ **for** $2 \le j_2 < \cdots < j_i \le k+1$**):** For fixed g and $j_2, \dots, j_i$, the number of zero-sum free sequences over C_n containing $\{a_1, a_2, \dots, a_i\} = \{g, j_2 g, \dots, j_i g\}$ is at most $X_{k-i+1,d-1} + \cdots + X_{k-i+1,d-i}$. The number of choices $(g, j_2, \dots, j_i)$ such that $g = a_\ell$ for some ℓ and $2 \le j_2 < \cdots < j_i \le k+1$ is at most $i\binom{k}{i-1}$. Hence, the expected number of zero-sum free sequences in R containing $\{a_1, a_2, \dots, a_i\} = \{g, j_2 g, \dots, j_i g\}$ is

$$i\binom{k}{i-1}(X_{k-i+1,d-1} + \cdots + X_{k-i+1,d-i})p^{d-i}. \tag{10}$$

• **Case 2 (**$\{a_1, a_2, \dots, a_i\} = \{j_1 g, j_2 g, \dots, j_i g\}$ **for** $2 \le j_1 < \cdots < j_i \le k+1$**):** For fixed $j_1 < \cdots < j_i$, we first consider the number of zero-sum free sequences over C_n containing $\{a_1, a_2, \dots, a_i\} = \{j_1 g, j_2 g, \dots, j_i g\}$. Without loss of generality, we let $x_1 = j_1, \dots, x_i = j_i$. Since $x'_\ell = x_\ell - 1 \ge 1$, the remaining $x'_{i+1}, \dots, x'_k$ satisfy $\sum_{i+1 \le \ell \le k} x'_\ell \le k - i$. The number of distinct $x'_{i+1}, \dots x'_k$ from $x'_1, \dots, x'_i$ are possibly $d-1$, $d-2, \dots,$ or $d-1-i$, and hence, the number of zero-sum free sequences over C_n containing $\{a_1, a_2, \dots, a_i\} = \{j_1 g, j_2 g, \dots, j_i g\}$ is at most $X_{k-i,d-1} + X_{k-i,d-2} + \cdots + X_{k-i,d-1-i}$. From the choices of $2 \le j_1 < \cdots < j_i \le k+1$, the expected number of zero-sum free sequences in R containing $\{a_1, \dots, a_i\}$ is at most

$$\binom{k}{i}(X_{k-i,d-1} + X_{k-i,d-2} + \cdots + X_{k-i,d-1-i})p^{d-i}. \tag{11}$$

From (10) and (11), we have that

$$\begin{aligned} \mathbb{E}_{d,i} &\le \max_{\{a_1,\dots,a_i\}} \mathbb{E}_{d,\{a_1,\dots,a_i\}} \\ &\le \left[i\binom{k}{i-1}+\binom{k}{i}\right]\left(\sum_{j=0}^{i} X_{k-i+1,d-1-j}\right) p^{d-i}, \end{aligned}$$

which completes our proof of the lemma. ☐

3.4. **Proofs of Theorems 4 and 5.**

Proof of Theorem 4. Let $X = N^R_{n-1-k}$. Under the assumption that k is fixed and $p \ll 1$, Corollary 3 gives that

$$\mathbb{E}(X) = p\varphi(n)\left(1 + O_k(p)\right).$$

Since k is fixed, Lemmas 10 and 11 yield that

$$\begin{aligned}
\mathbb{E}_1 &= \mathbb{E}_{1,1} + \mathbb{E}_{2,1} + \cdots + \mathbb{E}_{D,1} = O_k(1),\\
\mathbb{E}_2 &= \mathbb{E}_{2,2} + \mathbb{E}_{3,2} + \cdots + \mathbb{E}_{D,2} = O_k(1),\\
&\vdots\\
\mathbb{E}_D &= \mathbb{E}_{D,D} = 1.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}' &= \max_{1\le i\le D}\{\mathbb{E}_i\} = O_k(1) \text{ and}\\
\mathbb{E}^* &= \max\{\mathbb{E}', \mathbb{E}\} = p\varphi(n),
\end{aligned}$$

provided that $p\varphi(n) \gg 1$, i.e., $p \gg \frac{\log\log n}{n}$.

Set $\lambda = d\log n$, then $e^{-\lambda}n^{d-1} = 1/n = o(1)$, and hence, the Kim–Vu polynomial concentration result (Theorem 9) gives that a.a.s.

$$|X - \mathbb{E}(X)| = O_k\left(\sqrt{p\varphi(n)}(\log n)^d\right),$$

that is,

$$X = p\varphi(n) + O_k\left(p^2\varphi(n) + \sqrt{p\varphi(n)}(\log n)^d\right).$$

Note that $p\varphi(n) \gg \sqrt{p\varphi(n)}(\log n)^d$ is equivalent to $p \gg \frac{(\log n)^{2d}\log\log n}{n}$, and hence, our assumption on p is

$$\frac{(\log n)^{2d}\log\log n}{n} \ll p \ll 1.$$

Thus, we complete the proof of Theorem 4. ☐

For the proof of Theorem 5, we use the following version of Chernoff's bound.

Lemma 12 (**Chernoff's bound**, Corollary 4.6 in [14])**.** *Let X_i be independent random variables such that*

$$\Pr[X_i = 1] = p_i \quad \textit{and} \quad \Pr[X_i = 0] = 1 - p_i,$$

and let $X = \sum_{i=1}^n X_i$. For $0 < \lambda < 1$,

$$\Pr\Big[|X - \mathbb{E}(X)| \ge \lambda\mathbb{E}(X)\Big] \le 2\exp\Big(-\frac{\lambda^2}{3}\mathbb{E}(X)\Big).$$

We are ready to prove Theorem 5.

Proof of Theorem 5. Let $X = N^R_{n-1-k,d}$. First, we consider the case where $d = 1$. Observe that $X \sim Bin(\varphi(n), p)$, and hence, Chernoff's bound with $\lambda = \omega/\sqrt{p\varphi(n)}$ implies that a.a.s.

$$|X - p\varphi(n)| < \omega(p\varphi(n))^{1/2},$$

provided that $p\varphi(n) \gg 1$, i.e., $p \gg \frac{\log\log n}{n}$, where ω tends to ∞ arbitrarily slowly as $n \to \infty$.

Next we consider the case when $d \geq 2$. It follows from (3) that

$$\mathbb{E}(X) = p^d\varphi(n)X_{k,d-1}.$$

Lemma 11 gives that for a fixed k,

$$\begin{aligned} \mathbb{E}_{d,i} &= O_k(p^{d-i}) \text{ for } 1 \leq i \leq d-1 \text{ and} \\ \mathbb{E}_{d,d} &= 1. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}' &= \max_{1\leq i\leq d} \mathbb{E}_{d,i} = O_k(1) \text{ and} \\ \mathbb{E}^* &= O_k(\max\{1, p^d\varphi(n)\}) = O_k(p^d\varphi(n)), \end{aligned}$$

provided that $p^d\varphi(n) \gg 1$, i.e., $p \gg \left(\frac{\log\log n}{n}\right)^{1/d}$.

Set $\lambda = d\log n$, then $e^{-\lambda}n^{d-1} = 1/n = o(1)$, and hence, the Kim–Vu polynomial concentration result (Theorem 9) implies that a.a.s.

$$|X - \mathbb{E}(X)| \leq a_d(\mathbb{E}'\mathbb{E}^*)^{1/2}\lambda^d = O_k\left(p^{d/2}\varphi(n)^{1/2}(\log n)^d\right).$$

Note that $p^d\varphi(n) \gg \sqrt{p^d\varphi(n)}(\log n)^d$ is equivalent to $p \gg \frac{(\log n)^2(\log\log n)^{1/d}}{n^{1/d}}$, which is our assumption on p. This completes our proof of Theorem 5. □

Acknowledgement. The authors thank Myungho Kim (Kyung Hee Univ.) for valuable comments.

References

[1] N. Alon and J. H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016. 3.1

[2] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), no. 2, 367–454. 1

[3] K. Cziszter and M. Domokos, *The Noether number for the groups with a cyclic subgroup of index two*, J. Algebra **399** (2014), 546–560. 1

[4] K. Cziszter, M. Domokos, and A. Geroldinger, *The interplay of invariant theory with multiplicative ideal theory and with arithmetic combinatorics*, Multiplicative ideal theory and factorization theory, Springer Proc. Math. Stat., vol. 170, Springer, [Cham], 2016, pp. 43–95. 1

[5] K. Cziszter, M. Domokos, and I. Szöllősi, *The Noether numbers and the Davenport constants of the groups of order less than 32*, J. Algebra **510** (2018), 513–541. 1
[6] P. Erdös, A. Ginzburg, and A. Ziv, *Theorem in the additive number theory*, Bull. Res. Council Israel Sect. F **10F** (1961), no. 1, 41–43. 1
[7] W. D. Gao, *Zero sums in finite cyclic groups*, Integers **0** (2000), A12, 7. 1, 1
[8] W. Gao and A. Geroldinger, *Zero-sum problems in finite abelian groups: a survey*, Expo. Math. **24** (2006), no. 4, 337–369. 1, 1, 1
[9] A. Geroldinger, D. J. Grynkiewicz, J. S. Oh, and Q. Zhong, *On product-one sequences over Dihedral groups*, submitted, https://arxiv.org/abs/1910.12484. 1
[10] W. M. Y. Goh and E. Schmutz, *The number of distinct part sizes in a random integer partition*, J. Combin. Theory Ser. A **69** (1995), no. 1, 149–158. 2.1
[11] B. Kim, *On the number of partitions of n into k different parts*, J. Number Theory **132** (2012), no. 6, 1306–1313. 2.1
[12] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434. 3.1, 3.1
[13] Y. Kohayakawa, T. Ł uczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), no. 2, 133–163. 1
[14] M. Mitzenmacher and E. Upfal, *Probability and computing*, Cambridge University Press, Cambridge, 2005, Randomized algorithms and probabilistic analysis. 12
[15] V. Ponomarenko, *Minimal zero sequences of finite cyclic groups*, Integers **4** (2004), A24, 6. 1
[16] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. 1
[17] S. Savchev and F. Chen, *Long zero-free sequences in finite cyclic groups*, Discrete Math. **307** (2007), no. 22, 2671–2679. 1
[18] M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), no. 2, 333–365. 1

Department of Mathematics, Kyung Hee University, Seoul 02447, South Korea
E-mail address: `sjlee242@khu.ac.kr, sjlee242@gmail.com`

Research Institute of Basic Sciences, Incheon National University, Incheon 22012, South Korea
E-mail address: `junseok1.oh@gmail.com`