Random numerical semigroups and sums of subsets of cyclic groups

by **Santiago Morales Duarte**

Thesis submitted in fulfilment of the requirements for the degree of
Bachelor of Science
under the supervision of Tristram Bogart

Department of Mathematics
Faculty of Science
Universidad de los Andes
May 7, 2024

Contents

Chapter 1

Expected value proof

1.1 Upper bound

Before proving part (b) of the main theorem, we will prove a lemma that shows that a cyclic group of prime order is covered by the sums of a random subset of logarithmic size almost alwasys.

Lemma 1.1.1. *Let q be a prime number and $\mathcal{A}$ be a random subset of $\mathbb{Z}_q$ of size $4\lfloor 3\log_2 q\rfloor$. As q tends to infinity, $2\lfloor 3\log_2 q\rfloor\mathcal{A}$ covers $\mathbb{Z}_q$ almost always.*

Proof. Let $s \in \mathbb{N}$ such that $s \leq q$. Let $\mathcal{A}$ be a uniformly random subset of $\mathbb{Z}_q$ of size s, that is,

$$\Pr(\mathcal{A}) = \frac{1}{\binom{q}{s}}.$$

For a given $z \in \mathbb{Z}_q$ and $k \in \mathbb{N}$ for which $k \leq s/2$, let

$$N_z^k := \left\{ K \subseteq \mathbb{Z}_q : |K| = k, \sum_{t\in K} t = z \right\}.$$

Note that $|N_z^k| = \frac{1}{q}\binom{q}{k}$, since $K \in N_0^k$ if and only if $K + k^{-1}z \in N_z^k$ for every $z \in \mathbb{Z}_q$.

For $K \in N_z^k$, let E_K be the event that $K \subset \mathcal{A}$. Let X_K be the indicator variable of E_K. We define the random variable

$$X_z = \sum_{K\in N_z^k} X_K.$$

Note that X_z counts the number of sets of size k which add up to z. We now find $\mathrm{E}[X_z]$. Since the sum of every subset $K \subset S$ is in $\mathbb{Z}_q$,

$$\sum_{z\in Z_q} X_z = \binom{s}{k},$$

and so

$$\binom{s}{k} = E\left[\sum_{z\in Z_q} X_z\right] = \sum_{z\in Z_q} E[X_z].$$

As in the argument for finding $|N_z^k|$, for every $z \in \mathbb{Z}_q$,

$$\mathrm{E}[X_0] = \sum_{K \in N_0^k} \mathrm{E}[X_K] = \sum_{K \in N_0^k} \mathrm{E}[X_{K+k^{-1}z}] = \sum_{K \in N_z^k} \mathrm{E}[X_K] = \mathrm{E}[X_z].$$

Therefore, we have that

$$E[X_z] = \frac{1}{q}\binom{s}{k}. \tag{1.1}$$

Now, for $K, L \in N_z^k$, let $j \in \mathbb{N}$ such that $j \le k$ and define

$$\Delta_j := \sum_{|K \cap L| = j} \Pr[E_K \wedge E_L].$$

If $|K \cap L| = j$,

$$\Pr[E_K \wedge E_L] = \frac{\binom{q-2k+j}{s-2k+j}}{\binom{q}{s}}.$$

We can bound the number of events for which $|K \cap L| = j$. First we choose K as any set in N_z^k and then we choose the remaining $k - j$ elements as any subset of $\mathbb{Z}_q \setminus K$ with size $k - j$. Thus,

$$\Delta_j \le \frac{1}{q}\binom{q}{k}\binom{q-k}{k-j}\frac{\binom{q-2k+j}{s-2k+j}}{\binom{q}{s}}.$$

This implies that, using 1.1,

$$\begin{aligned}
\frac{\Delta_j}{\mathrm{E}[X_z]^2} &\le \frac{\binom{q}{k}\binom{q-k}{k-j}\binom{q-2k+j}{s-2k+j}}{\frac{1}{q}\binom{s}{k}\frac{1}{q}\binom{s}{k}q\binom{q}{s}} \\
&= \frac{\frac{q!}{(q-k)!k!}\frac{(p-k)!}{(k-j)!(q-2k+k)!}\frac{(q-2k+j)!}{(s-2k+j)!(q-s)!}}{\frac{1}{q}\binom{s}{k}\frac{s!}{(s-k)!k!}\frac{q!}{(q-s)!s!}} \\
&= \frac{q\binom{s-k}{k-j}}{\binom{s}{k}}.
\end{aligned}$$

Let $s = 4\lfloor 3\log_2 q \rfloor$ and $k = 2\lfloor 3\log_2 q \rfloor$. Using that $\binom{s-k}{k-j}$ is maximized at $k - j = \lfloor (s-k)/2 \rfloor$,

$$\frac{\Delta_j}{\mathrm{E}[X_z]^2} \le \frac{q\binom{2\lfloor 3\log_2 q \rfloor}{\lfloor 3\log_2 q \rfloor}}{\binom{4\lfloor 3\log_2 q \rfloor}{2\lfloor 3\log_2 q \rfloor}} \le \frac{q}{\binom{2\lfloor 3\log_2 q \rfloor}{\lfloor 3\log_2 q \rfloor}} \le \frac{q}{2^{\lfloor 3\log_2 q \rfloor}} \sim \frac{1}{q^2},$$

since $\binom{2\lfloor \log_2 q \rfloor}{\lfloor 3\log_2 q \rfloor}^2 \le \binom{4\lfloor 3\log_2 q \rfloor}{2\lfloor 3\log_2 q \rfloor}$ (Proposition A.0.4).

Hence, by (**??**) and Theorem **??**,

$$\Pr[X_z = 0] \le \frac{\mathrm{E}[X_z] + \Delta}{\mathrm{E}[X_z]^2} = \frac{1}{E[X_z]} + \sum_{j=0}^{k} \frac{\Delta_j}{\mathrm{E}[X_z]^2}$$

$$\leq \frac{1}{E[X_z]} + \frac{(k+1)}{q^2} = \frac{1}{E[X_z]} + \frac{2\lfloor 3\log_2 q\rfloor + 1}{q^2}.$$

Therefore, by the union bound and since $q \to \infty$ as $p \to 0$,

$$\Pr\left[\bigvee_{z\in\mathbb{Z}_q} X_z = 0\right] \leq \frac{q}{E[X_z]} + \frac{2\lfloor 3\log_2 q\rfloor + 1}{q^2} = o(1).$$

We conclude that $X_z > 0$ for every $z \in \mathbb{Z}_q$ almost always. Thus, for every $z \in \mathbb{Z}_q$, there exists $K \in N_z^k$ such that $K \subset \mathcal{A}$ almost always. This means that $2\lfloor 3\log_2 q\rfloor\mathcal{A}$ covers $\mathbb{Z}_q$ almost always. □

1.1.1 Proof of the upper bound

Lemma 1.1.2. *Let $\psi(x)$ be a function for which $x(\log x)^2 \in o(\psi(x))$ .Then*

$$\lim_{p\to 0} \Pr\left[F(\mathcal{S}) \leq \psi\left(\frac{1}{p}\right)\right] = 1.$$

The proof of this theorem consists of several parts. The strategy is to prove that the Ápery set of a subsemigroup of S is completed before step $\psi\left(\frac{1}{p}\right)$ with high probability, since $F(\mathcal{S})$ is less than the maximum element of this Ápery set. The proof has the following structure:

1. First, we will find a step for which a prime q is chosen with high probability (E_1).
2. Then, in the spirit of Lemma 1.1.1 we will find a step such that a set $\mathcal{A}$ of s elements which are different modulo q are chosen with high probability (E_2).
3. Finally, we will apply Lemma 1.1.1 to $\mathrm{Ap}(\langle \mathcal{A} \cup \{q\}\rangle, q)$.

Proof.

Part 1

Let $h(x)$ be a function such that $h(x) \in o(x(\log x)^2)$ and $x\log x \in o(h(x))$. Let $t(x) = 20x\log x$. Consider the event E_1 that there exists a prime $q \in \mathcal{S}$, such that

$$t\left(\frac{1}{p}\right) \leq q \leq h\left(\frac{1}{p}\right).$$

Let q_n be the n-th prime number. By the prime number theorem [1, Theorem 8],

$$q_n \sim n\log n. \tag{1.2}$$

Let $k(x)$ be the number of primes between $t(x)$ and $h(x)$. Now, for sufficiently large n, $t(n) \leq q_{20n}$. Also, for every $c \in \mathbb{R}^+$, $q_{cn} \in o(h(n))$ since $cn\log cn \in o(h(n))$. Thus, for sufficiently large x and every $c \in \mathbb{R}^+$, $k(x) > cx$ and we get that

$$\lim_{p\to 0} \Pr[\neg E_1] \leq \lim_{p\to 0}(1-p)^{k(1/p)} \leq \lim_{p\to 0}(1-p)^{\frac{c}{p}} = e^{-c}.$$

Therefore,

$$\lim_{p\to 0} \Pr[E_1] = 1. \tag{1.3}$$

Part 2

Now, assume E_1. Then $\mathcal{S}$ contains a prime number q for which

$$t\left(\frac{1}{p}\right) \leq q \leq h\left(\frac{1}{p}\right).$$

Let q be such a prime. Let $s = 4\lfloor 3\log_2 q\rfloor$, as in Lemma 1.1.1. Let $T = \{1, \ldots, q\}$. Consider the event E_2 that at least s generators are selected in T. Let X_1 be the number of generators selected in T, then $X_1 \sim \mathrm{Bin}(q, p)$. We first show that for sufficiently small p, $qp > s$ in order to use a bound of the left tail of the binomial distribution (Proposition A.0.6).

Since

$$q \geq t\left(\frac{1}{p}\right) = \frac{20}{p}\log\frac{1}{p},$$

then

$$qp \geq 20\log\frac{1}{p}.$$

Also, since

$$q \leq h\left(\frac{1}{p}\right) \leq \frac{1}{p}\left(\log\frac{1}{p}\right)^2,$$

then

$$s = 4\lfloor 3\log_2 q\rfloor \leq 4\left\lfloor 3\log_2\frac{1}{p}\left(\log\frac{1}{p}\right)^2\right\rfloor = 4\left\lfloor 3\log_2\frac{1}{p} + 6\log_2\log\frac{1}{p}\right\rfloor.$$

Thus, for sufficiently small p, $qp > s$ and we can use Proposition A.0.6 with $r = s$ to show that

$$\Pr[\overline{E_2}|E_1] = \Pr\left[X_1 < s\right] \leq \frac{(q-s)p}{(qp-s)^2}.$$

Thus, bounding by the worst case asymptotically,

$$\begin{aligned}
\lim_{p\to 0} P[\overline{E_2}|E_1] &\leq \lim_{p\to 0}\frac{\left(h\left(\frac{1}{p}\right) - 4\left\lfloor 3\log_2 t\left(\frac{1}{p}\right)\right\rfloor\right)p}{\left(t\left(\frac{1}{p}\right)p - 4\left\lfloor 3\log_2 h\left(\frac{1}{p}\right)\right\rfloor\right)^2} \\
&\leq \lim_{p\to 0}\frac{\left(h\left(\frac{1}{p}\right) - 4\left\lfloor 3\log_2\frac{20}{p}\log\frac{1}{p}\right\rfloor\right)p}{\left(20\log\frac{1}{p} - 4\left\lfloor 3\log_2\frac{1}{p}\left(\log\frac{1}{p}\right)^2\right\rfloor\right)^2} \\
&= \lim_{p\to 0}\frac{o\left(\frac{1}{p}\left(\log\frac{1}{p}\right)^2\right)p}{\left(20\log\frac{1}{p} - 4\left\lfloor 3\log_2\frac{1}{p}\left(\log\frac{1}{p}\right)^2\right\rfloor\right)^2}
\end{aligned}$$

$$= \lim_{p\to 0} \frac{o\left(\left(\log \frac{1}{p}\right)^2\right)}{\left(20\log\frac{1}{p} - 4\left\lfloor 3\log_2 \frac{1}{p}\left(\log\frac{1}{p}\right)^2\right\rfloor\right)^2} = 0.$$

We conclude that

$$\lim_{p\to 0} \Pr[E_2|E_1] = 1,$$

and so, using (1.3),

$$\lim_{p\to 0} \Pr[E_1 \wedge E_2] = \lim_{p\to 0} \Pr[E_2|E_1]\Pr[E_1] = 1.$$

Part 3

Finally, assume E_1 and E_2. Let $\mathcal{A} = \{Y_1, \ldots, Y_s\}$ be a randomly selected subset of size s of the generators selected in T. Since the generators are chosen randomly and $|T| = q$, we can apply Lemma 1.1.1 to $\mathbb{Z}_q \cong \mathrm{Ap}(\langle \mathcal{A} \cup \{q\}\rangle, q)$ to get that it will be completed before step

$$qs \leq h\left(\frac{1}{p}\right) 2\left\lfloor 3\log_2 h\left(\frac{1}{p}\right)\right\rfloor \in O\left(h\left(\frac{1}{p}\right)\log\frac{1}{p}\right),$$

almost always as $p \to 0$.

Thus, if

$$\psi(x) = h\left(x\right) 2\left\lfloor 3\log_2 x\right\rfloor,$$

we have that $x(\log x)^2 \in o(\psi(x))$ and

$$\lim_{p\to 0} \Pr\left[F(\langle \mathcal{A} \cup \{q\}\rangle) \leq \psi\left(\frac{1}{p}\right)\right] = 1.$$

Since $F(\mathcal{S}) \leq F(\langle \mathcal{A} \cup \{q\}\rangle)$, we conclude that

$$\lim_{p\to 0} \Pr\left[F(\mathcal{S}) \leq \psi\left(\frac{1}{p}\right)\right] = 1.$$

Since the constraints on $h(x)$ are independent of multiplication by constants, the result is true for any function ψ such that $x(\log x)^2 \in \psi(g(x))$. □

The bound on the Frobenius number also implies bounds on the genus and the embedding dimension.

Corollary 1.1.1. *Let $\psi(x)$ be a function for which $x(\log x)^2 \in o(\psi(x))$ .Then*

$$\lim_{p\to 0} \Pr\left[g(\mathcal{S}) \leq \psi\left(\frac{1}{p}\right)\right] = 1.$$

Proof. Use Proposition **??**. □

Corollary 1.1.2. *Let $\varphi(x)$ be a function for which $(\log x)^2 \in o(\varphi(x))$ .Then*

$$\lim_{p\to 0} \Pr\left[e(\mathcal{S}) \leq \varphi\left(\frac{1}{p}\right)\right] = 1.$$

Proof. Since
$$\lim_{p\to 0} \Pr\left[F(\mathcal{S}) \le \psi\left(\frac{1}{p}\right)\right] = 1,$$
and the maximal element of the minimal generating set is at most $2F(\mathcal{S})$, the elements of the minimal generating set are chosen before step $2\psi\left(\frac{1}{p}\right)$ with high probability. Since
$$\left|\mathcal{A} \cap \left\{1, \ldots, \left\lfloor 2\psi\left(\frac{1}{p}\right)\right\rfloor\right\}\right| \sim \mathrm{Bin}\left(\left\lfloor 2\psi\left(\frac{1}{p}\right)\right\rfloor, p\right),$$
by the bound on the right tail of the binomial distribution (Proposition A.0.5), we have that
$$\lim_{p\to 0} \Pr\left[e(\mathcal{S}) \le (3p)\psi\left(\frac{1}{p}\right)\right] = 1.$$
Thus, if $\varphi(x) = \frac{3}{x}\psi(x)$, then $(\log x)^2 \in \varphi(x)$ and
$$\lim_{p\to 0} \Pr\left[e(\mathcal{S}) \le \varphi\left(\frac{1}{p}\right)\right] = 1. \quad \square$$

Bibliography

[1] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers.* Oxford university press, 1979.

[2] W. Feller, *An introduction to probability theory and its applications.* John Wiley & Sons, 1971, vol. 1.

[3] N. Alon and J. H. Spencer, *The Probabilistic Method.* John Wiley & Sons, 2016.

[4] J. Park and H. Pham, "A proof of the Kahn–Kalai conjecture," *Journal of the American Mathematical Society*, 2023.

[5] J. De Loera, C. O'Neill, and D. Wilburne, "Random numerical semigroups and a simplicial complex of irreducible semigroups," *The Electronic Journal of Combinatorics*, P4–37, 2018.

[6] J. C. Rosales, P. A. García-Sánchez, *et al.*, *Numerical semigroups.* Springer, 2009.

[7] M. Delgado, "Conjecture of Wilf: A survey," *Numerical Semigroups: IMNS 2018*, pp. 39–62, 2020.

[8] I. Aliev, M. Henk, and A. Hinrichs, "Expected Frobenius numbers," *Journal of Combinatorial Theory, Series A*, vol. 118, no. 2, pp. 525–531, 2011.

[9] J. Grime. "How to order 43 mcnuggets - numberphile," Youtube. (2012), [Online]. Available: `https://www.youtube.com/watch?v=vNTSugyS038&ab_channel=Numberphile`.

[10] A. Assi, M. D'Anna, and P. A. García-Sánchez, *Numerical semigroups and applications.* Springer Nature, 2020, vol. 3.

[11] R. Apéry, "Sur les branches superlinéaires des courbes algébriques," *CR Acad. Sci. Paris*, vol. 222, no. 1198, p. 2000, 1946.

[12] E. S. Selmer, "On the linear Diophantine problem of Frobenius," 1977.

[13] H. S. Wilf, "A circle-of-lights algorithm for the "money-changing problem"," *The American Mathematical Monthly*, vol. 85, no. 7, pp. 562–565, 1978.

[14] J. L. Ramírez-Alfonsín, "Complexity of the Frobenius problem," *Combinatorica*, vol. 16, pp. 143–147, 1996.

[15] V. I. Arnold, "Weak asymptotics for the numbers of solutions of Diophantine problems," *Functional Analysis and Its Applications*, vol. 33, no. 4, pp. 292–293, 1999.

[16] P. Erdös and R. Graham, "On a linear Diophantine problem of Frobenius," *Acta Arithmetica*, vol. 1, no. 21, pp. 399–408, 1972.

[17] I. M. Aliev and P. M. Gruber, "An optimal lower bound for the Frobenius problem," *Journal of Number Theory*, vol. 123, no. 1, pp. 71–79, 2007.

[18] V. I. Arnold, *Arnold's problems.* Springer, 2004.

[19] R. P. Stanley, *Combinatorics and commutative algebra.* Springer Science & Business Media, 2007, vol. 41.
[20] M. Delgado, "Intpic," *a GAP package for drawing integers, Available via http://www. fc. up. pt/cmup/mdelgado/software*, 2013.
[21] M. Delgado, P. Garcıa-Sánchez, and J. Morais, "Numericalsgps," *A GAP package for numerical semigroups. Available via http://www. gap-system. org*, 2015.
[22] C. O'Neill, *Numsgps-sage*, `https://github.com/coneill-math/numsgps-sage`, 2013.
[23] S. Morales, *Randnumsgps*, `https://github.com/smoralesduarte/randnumsgps`, 2023.

Appendix A

Useful Bounds

We include some bounds that are useful in the proofs of the main results. By Stirling's Formula, we have that

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k. \tag{A.1}$$

Proposition A.0.1. $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ *for* $1 \leq k \leq n$.

Proof. Using (A.1), we have that, for $k \geq 1$,

$$k! \geq \left(\frac{k}{e}\right)^k.$$

Then

$$\binom{n}{k} \leq \frac{n^k}{\left(\frac{k}{e}\right)^k} = \left(\frac{en}{k}\right)^k. \quad \square$$

Proposition A.0.2. $\left(\frac{n}{k}\right)^k \geq \binom{n}{k}$ *for* $1 \leq k \leq n$.

Proof.

$$\binom{n}{k} = \prod_{i=0}^{k-1} \frac{n-i}{k-i} \geq \left(\frac{n}{k}\right)^k. \quad \square$$

Proposition A.0.3. $(1-p) \leq e^{-p}$ *for* $0 \leq p \leq 1$.

Proof. The Taylor series of e^{-p} is alternating with a decreasing sequence, so

$$e^{-p} = 1 - p + \frac{p^2}{2!} - \frac{p^3}{3!} + ... \geq 1 - p. \quad \square$$

We also give a combinatorial proof of the following result.

Proposition A.0.4. $\binom{2n}{k}^2 \leq \binom{4n}{2k}$ *for* $n \geq 1$.

Proof. The number of subsets of size $2k$ of a set of size $4n$ is $\binom{4n}{2k}$. This is greater than the number of subsets that can be expressed as the product of two subsets of size k of a set of size $2n$, which is $\binom{2n}{k}^2$. $\square$

The proof of the following bound can be found in [2, Section 6.3].

Proposition A.0.5. *Let* $X \sim \mathrm{Bin}(n, p)$*. If* $r > np$*,*

$$\Pr[X \geq r] \leq \frac{r(1-p)}{(r-np)^2}.$$

Since the binomial distribution is symmetric, we also have the following.

Proposition A.0.6. *Let* $X \sim \mathrm{Bin}(n, p)$*. If* $r < np$*,*

$$\Pr[X \leq r] \leq \frac{(n-r)p}{(np-r)^2}.$$