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
November 12, 2023

Abstract

We investigate properties of random numerical semigroups using a probabilistic model based on the Erdös-Rényi model for random graphs and propose a new probabilistic model. We provide a new and more elementary proof of a lower bound of the expected embedding dimension, genus, and Frobenius number of a random semigroup, and provide a tighter probabilistic upper bound. Our results derive from the application of the Probabilistic Method to the generation of random numerical semigroups and observations about sums of uniformly random subsets of cyclic groups. We include experiments that motivated our results and offer conjectures for subsequent research.

Dedication

For my mother.

Acknowledgements

I would like to thank Tristram Bogart, Alexander Getmanenko (supervisor, family, research collaborators, anyone else who significantly helped you with this work). Christopher O'Neill, Johanna Franklin and the Polymath Jr. program.

Santiago Morales Duarte
November 12, 2023
Bogotá, Colombia

Contents

List of Figures

Chapter 1

Introduction

The start of the introduction provides some context and brief background. This is a test for github.

1.1 Research Objectives and Overview

The research question which this Thesis aims to answer is...

The specific research objectives of this Thesis are:

1. Objective 1
2. Objective 2

Chapter 2 provides an overview of the Probabilistic Method and introduces the concept of random graphs.

Chapter 3 introduces numerical semigroups and their properties.

Chapter 4 introduces the concept of random numerical semigroups and defines a new model for random numerical semigroups.

Chapter 5 presents the results of experiments performed on random numerical semigroups.

Chapter 6 proves

Chapter 2

The Probabilistic Method

2.1 Introduction

This chapter is based on the book *The Probabilistic Method* by Noga Alon and Joel H. Spencer [1].

The Probabilistic Method is a powerful tool, with applications in Combinatorics, Graph Theory, Number Theory and Computer Science. It is a nonconstructive method that proves the existence of an object with a certain property, by showing that the probability that a randomly chosen object has that property is greater than zero. The method requires an appropriate sample space and is best illustrated by an example:

Definition 2.1.1. A *tournament* is a directed graph T on n vertices such that for every pair of vertices $i, j \in V(T)$, exactly one of the edges (i, j) or (j, i) is in $E(T)$.

The name of a tournament comes from the fact that it can be thought of as a sports tournament where each vertex represents a team and each team plays every other team exactly once. The edge (i, j) represents a win for team i over team j. A tournament T has property S_k if for every subset $K \subseteq V(T)$ of size k, there is a vertex $v \in V(T)$ such that $(v, s) \in E(T)$ for all $s \in K$. That is, for every set of k teams there is a team that beats all of them. For example, the tournament in Figure 2.1 has property S_1 since every team is beaten by another team.

A natural question to ask is: for every k, is there a tournament with property S_k? The answer is yes. We will prove this using the Probabilistic Method. First we define a probability space over the set of tournaments on n vertices:

A *random* tournament on a set of n vertices is a tournament T such that for every pair of vertices $i, j \in V(T)$, the edge (i, j) is in $E(T)$ with probability $\frac{1}{2}$ and the edge (j, i) is in $E(T)$ with probability $\frac{1}{2}$, independently of all other edges. Thus, every tournament on n vertices has the same probability, which means that this probability space is *symmetric*.

The main idea is to show that for sufficiently large n as a function of k, such that the probability that a random tournament on n vertices has property S_k is greater than zero. This implies that there is at least one tournament with property S_k.

Theorem 2.1.1 (Theorem 1.2.1 [1])**.** *For every $k \in \mathbb{N}$, there is a tournament with property S_k.*

Proof. Fix a subset $K \subseteq V(T)$ of size k. Consider the event A_K that there is no vertex $v \in V(T)$ such that $(v, s) \in E(T)$ for all $s \in S$. For any vertex $v \in V(T) \setminus K$, the probability

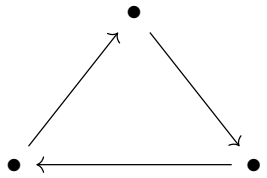

Figure 2.1: A tournament on 3 vertices with property S_1.

that $(v, s) \notin E(T)$ for all $s \in K$ is 2^{-k}. Thus,

$$\Pr[A_K] = \left(1 - 2^{-k}\right)^{n-k}.$$

Now, if we consider all subsets $K \subseteq V(T)$ of size k, then the probability that T does not have property S_k is the probability that at least one of the events A_K occurs. Since there are $\binom{n}{k}$ such subsets, by the union bound,

$$\Pr\left[\bigvee_{\substack{K \subseteq V(T) \\ |K|=k}} A_K\right] \leq \sum_{\substack{K \subseteq V(T) \\ |K|=k}} \Pr[A_K] = \binom{n}{k}\left(1 - 2^{-k}\right)^{n-k}.$$

We want to show that, for some n, the probability of this event is less than 1. Using Propositions A.1.1 and A.1.3, we have that

$$\Pr\left[\bigvee_{\substack{K \subseteq V(T) \\ |K|=k}} A_K\right] \leq \binom{n}{k}\left(1 - 2^{-k}\right)^{n-k} \tag{2.1}$$

$$\leq \left(\frac{en}{k}\right)^k \left(e^{-2^{-k}}\right)^{n-k} = e^{k\log\left(\frac{n}{k}\right) - \frac{n-k}{2^k}}. \tag{2.2}$$

Then, (2.2) is less than 1 if

$$k\log\left(\frac{n}{k}\right) - \frac{n-k}{2^k} < 0.$$

Which is true if

$$\begin{aligned} 0 > k\log n - \frac{n}{2^k} &> k\log n - k\log k + \frac{k}{2^k} - \frac{n}{2^k} \\ &= k\log\left(\frac{n}{k}\right) - \frac{n-k}{2^k}. \end{aligned}$$

Thus,

$$\frac{n}{\log n} > k2^k \implies \Pr\left[\bigvee_{\substack{K \subseteq V(T) \\ |K|=k}} A_K\right] < 1.$$

Hence, for sufficiently large n, the probability that a random tournament on n vertices does not have property S_k is less than one. Therefore, the probability that there exists a tournament

on n vertices with property S_k is greater than zero, which means that there exists at least one tournament with property S_k. □

We make two observations:

1. We used the *union bound*. The union bound is a common technique in the Probabilistic Method. It states that for any events $A_1, ..., A_n$,

   $$\Pr[A_1 \cup ... \cup A_n] \leq \Pr[A_1] + ... + \Pr[A_n].$$

   We will extensively use this technique in this thesis. In a measure space, the union bound is the same property as *subadditivity*.

2. The proof is nonconstructive. It does not give us a way to find a tournament with property S_k. It only shows that there is at least one. This is a common feature of the Probabilistic Method. However, in this case, we have that for large enough n, the probability that a random tournament on n vertices has property S_k is close to one. This means that we can find a tournament with property S_k by generating random tournaments until we find one with the desired property. If n is large enough, it will be highly probable, though harder to verify.

In this chapter, we will introduce some tools that are useful for applying the Probabilistic Method in discrete settings. We will also give some examples of the method in action.

2.2 Linearity of Expectation

Let X be a discrete random variable, then the *expected value* of X is defined as

$$\mathrm{E}[X] = \sum_{x \in \mathrm{Rg}(X)} x\Pr[X = x].$$

Theorem 2.2.1 (Linearity of expectation)**.** $\mathrm{E}[X]$ *is linear.*

Proof. Let X and Y be discrete random variables. Then, the expected value of $X + Y$ is

$$\begin{aligned}
\mathrm{E}[X+Y] &= \sum_{x \in \mathrm{Rg}(X)} \sum_{y \in \mathrm{Rg}(Y)} (x+y)\Pr[X = x \wedge Y = y] \\
&= \sum_{x \in \mathrm{Rg}(X)} \sum_{y \in \mathrm{Rg}(Y)} x\Pr[X = x \wedge Y = y] + \sum_{x \in \mathrm{Rg}(X)} \sum_{y \in \mathrm{Rg}(Y)} y\Pr[X = x \wedge Y = y] \\
&= \sum_{x \in \mathrm{Rg}(X)} x \sum_{y \in \mathrm{Rg}(Y)} \Pr[X = x \wedge Y = y] + \sum_{y \in \mathrm{Rg}(Y)} y \sum_{x \in \mathrm{Rg}(X)} \Pr[X = x \wedge Y = y] \\
&= \sum_{x \in \mathrm{Rg}(X)} x\Pr[X = x] + \sum_{y \in \mathrm{Rg}(Y)} y\Pr[Y = y] \\
&= \mathrm{E}[X] + \mathrm{E}[Y].
\end{aligned}$$

Also, if $a \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{E}[aX] &= \sum_{x \in \mathrm{Rg}(X)} ax\Pr[X = x] \\
&= a \sum_{x \in \mathrm{Rg}(X)} x\Pr[X = x] \\
&= a\mathrm{E}[X].
\end{aligned}$$

Thus, the expected value is linear. □

Example 2.2.1. Let σ be a random permutation of $\{1, \ldots, n\}$ chosen uniformly at random. Let X_i be the indicator variable for the event that $\sigma(i) = i$. Then, $\mathrm{E}[X_i] = \frac{1}{n}$ since there are n possible values for $\sigma(i)$ and only one of them is i. Now, let $X = \sum_{i=1}^{n} X_i$. Then, X is the number of fixed points of σ. By the linearity of expectation,

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1. \quad \triangle$$

Note that there is a point x such that $x \geq \mathrm{E}[X]$ and $\Pr[X = x] > 0$, and there is a point $x \leq \mathrm{E}[X]$ such that $\Pr[X = x] > 0$. The following result by Szele (1943) is often considered as one of the first applications of the Probabilistic Method.

Theorem 2.2.2 (Theorem 2.1.1 [1])**.** *There is a tournament with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.*

Proof. Let X be the number of Hamiltonian paths in a random tournament. Let σ be a permutation, and let X_σ be the indicator variable for the event that σ is a Hamiltonian path of the random tournament. That is, σ is an ordering of the vertices such that $(\sigma(1), \sigma(2)), ..., (\sigma(n-1), \sigma(n))$ are edges of the tournament. Then, $X = \sum_\sigma X_\sigma$. By the linearity of expectation,

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{\sigma} X_\sigma\right] = \sum_{\sigma} \mathrm{E}[X_\sigma] = \sum_{\sigma} \frac{1}{2^{n-1}} = n!2^{-(n-1)}.$$

Therefore, there exists a tournament with at least $n!2^{-(n-1)}$ Hamiltonian paths. □

2.3 Second Moment Method

Just as we can use the linearity of expectation to prove results with the Probabilistic Method, we can also use the *second moment method*, which relies on the *variance* of a random variable. Let X be a random variable with expected value $\mathrm{E}[X]$. Then, the variance of X is defined as

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2].$$

By the linearity of expectation,

$$\mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X] - 2\mathrm{E}[X\mathrm{E}[X]] + \mathrm{E}[X]^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2.$$

The standard practice is to denote the expected value by μ and the variance by σ^2. The use of the following inequality is called the second moment method

Theorem 2.3.1 (Chebyshev's inequality)**.** *For $\lambda > 0$,*

$$\Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof.

$$\sigma^2 = \mathrm{Var}[X] = \mathrm{E}[(x - \mu)^2] \geq \lambda^2\sigma^2\Pr[|X - \mu| \geq \lambda\sigma]. \quad \square$$

If $X = X_1 + ... + X_n$, then, by the linearity of expectation,

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j],$$

where

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j].$$

Note that $\mathrm{Cov}[X_i, X_j] = 0$ if X_i and X_j are independent. Furthermore, if, for each i, X_i is an indicator variable of event A_i, that is, $X_i = 1$ if A_i occurs and $X_i = 0$ otherwise, then

$$\mathrm{Var}[X_i] = \Pr[A_i](1 - \Pr[A_i]) \leq \mathrm{E}[X_i],$$

and we have that

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j]. \tag{2.3}$$

Suppose that X only takes nonnegative integer values, and we are interested in bounding $\Pr[X = 0]$. First, note that

$$\Pr[X > 0] \leq \mathrm{E}[X]. \tag{2.4}$$

For a sequence of variables $X_1, X_2, ...$, we say that X satisfies a property A *almost always* if $\lim_{n\to\infty} \Pr[X_n \text{ satisfies } A] = 1$.

Thus, using (2.4), if $\mathrm{E}[X] \to 0$, then $X = 0$ almost always. On the other hand, if $\mathrm{E}[X] \to \infty$, it is not necessarily true that $X > 0$ almost always. For instance, consider an obviously imaginary game where you throw a coin until it lands heads up and you get paid 2^n dollars if it takes n throws. Then, $\mathrm{E}[X] = \infty$ but $X = 0$ with probability $\frac{1}{2}$. In some cases, we can use the second moment method to show that if $\mathrm{E}[X] \to \infty$ and we have more information about $\mathrm{Var}[X]$, then $X > 0$ almost always.

Theorem 2.3.2 (Theorem 4.3.1 [1])**.** $\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}$.

Proof. We apply Chebyschev's inequality 2.3.1 with $\lambda = \frac{\mu}{\sigma}$. Thus,

$$\Pr[X = 0] \leq \Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}. \quad \square$$

The following result is a direct consequence.

Corollary 2.3.1. *If* $\mathrm{Var}[X] \in o(\mathrm{E}[X^2])$, $X > 0$ *asymptotically almost always.*

Let $\varepsilon > 0$, following the proof of Theorem 2.3.2, if $\lambda = \frac{\varepsilon\mu}{\sigma}$, then

$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2 \mathrm{E}[X]^2}.$$

Thus, we have a tighter result:

Corollary 2.3.2. *If* $\mathrm{Var}[X] \in o(\mathrm{E}[X]^2)$, *then* $X \sim \mathrm{E}[X]$ *almost always..*

Finally, if $X = X_1 + \cdots + X_n$, where each X_i is the indicator variable of event A_i. For indices i, j such that $i \neq j$, we say that $i \sim j$ if the events A_i and A_j are not independent. Let

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j].$$

Corollary 2.3.3. *If* $\mathrm{E}[X] \to \infty$ *and* $\Delta = o(\mathrm{E}[X^2])$*, then* $X > 0$ *almost always. Also,* $X \sim \mathrm{E}[X]$ *almost always.*

Proof. When $i \sim j$,

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j] \leq \mathrm{E}[X_i X_j] = \Pr[A_i \wedge A_j],$$

and so

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j] \leq \mathrm{E}[X] + \sum_{i \sim j} \Pr[A_i \wedge A_j] = \mathrm{E}[X] + \Delta. \quad \square$$

We are now ready to show an application of the second moment method.

2.4 Threshold Functions

Let $n \in \mathbb{N}$ and $0 \leq p \leq 1$.

Definition 2.4.1. The Erdös-Rényi model for random graphs $G(n, p)$ is a probability space over the set of graphs on n labeled vertices determined by

$$\Pr[\{i, j\} \in G] = p$$

with these events mutually independent.

Given a graph theoretic property A, there is a probability that $G(n, p)$ satisfies A, which we write as $\Pr[G(n, p) \vDash A]$. As n grows, we let p be a function of n, $p = p(n)$.

Definition 2.4.2. $r(n)$ is a threshold function for a graph theoretic property A if

1. When $p(n) \in o(r(n))$, $\lim_{n \to \infty} \Pr[G(n, p(n)) \vDash A] = 0$,
2. When $r(n) \in o(p(n))$, $\lim_{n \to \infty} \Pr[G(n, p(n)) \vDash A] = 1$,

or vice versa.

We give an example of a threshold function which illustrates a common method for proving that a function is a threshold.

2.4.1 A threshold function for isolated vertices

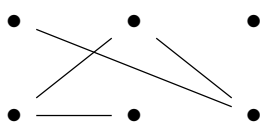

Figure 2.2: A graph with an isolated vertex.

Let G be a graph on n labeled vertices. An isolated vertex of G is a vertex which does not belong to any of the edges of G. Let A be the property that G contains an isolated vertex.

Theorem 2.4.1. $r(n) = \dfrac{\ln n}{n}$ *is a threshold for having isolated vertices.*

Proof. For each vertex i in G, let A_i be the event that i is an isolated vertex and define its indicator variable

$$X_i = \begin{cases} 1 & \text{if } i \text{ is an isolated vertex,} \\ 0 & \text{if } i \text{ is not an isolated vertex.} \end{cases}$$

Now, the probability that a vertex i is isolated is $(1-p)^{n-1}$,since it is the probability that none of the other $n-1$ vertices is connected to i. Let $X = \sum_{i=1}^{n} X_i$, then the expected number of isolated vertices is

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \Pr[A_i] = n(1-p)^{n-1}.$$

Let $p = k\dfrac{\ln n}{n}$ for $k \in \mathbb{R}_{>0}$. Then, since $1 - k\frac{\log n}{n} = e^{-k\frac{\log n}{n} \pm O\left(k^2 \frac{(\log n)^2}{n^2}\right)}$,

$$\begin{aligned} \lim_{n\to\infty} \mathrm{E}[X] &= \lim_{n\to\infty} n\left(1 - \frac{\ln n}{n}\right)^{n-1} \\ &= \lim_{n\to\infty} ne^{-k\ln n} = \lim_{n\to\infty} n^{1-k}. \end{aligned}$$

There are two cases:

1. If $k > 1$, $\lim_{n\to\infty} \mathrm{E}[X] = 0$. Since $\mathrm{E}[X] \geq \Pr[X > 0]$, we conclude that

   $$\lim_{n\to\infty} \Pr[G(n,p) \vDash A] = \lim_{n\to\infty} \Pr[X > 0] = 0.$$

   Thus, if $r(n) \in o(p(n))$, then $\lim_{n\to\infty} \Pr[G(n,p(n)) \vDash A] = 0$.

2. If $k < 1$, the fact that $\lim_{n\to\infty} \mathrm{E}[X] = \infty$ is not enough to conclude that

   $$\lim_{n\to\infty} \Pr[G(n,p) \vDash A] = 1.$$

   We have to use the second moment method. We will prove that $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$. First,

   $$\begin{aligned} \sum_{i\neq j} \mathrm{E}[X_i X_j] &= \sum_{i\neq j} \Pr[X_i = X_j = 1] \\ &= n(n-1)(1-p)^{n-1}(1-p)^{n-2} \\ &= n(n-1)(1-p)^{2n-3}, \end{aligned}$$

   for if i is an isolated vertex, then there is no edge between i and j so we only have to account for the remaining $n-2$ edges that contain j.

   Thus, since $\sum_{i=1}^{n} \mathrm{E}[X_i^2] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \mathrm{E}[X]$ and $\lim_{n\to\infty} p(n) = 0$,

$$\begin{aligned}\lim_{n\to\infty}\frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2} &= \lim_{n\to\infty}\frac{\mathrm{E}[X^2]-\mathrm{E}[X]^2}{\mathrm{E}[X]^2}\\ &= \lim_{n\to\infty}\frac{\sum_{i=1}^{n}\mathrm{E}[X_i^2]+\sum_{i\neq j}\mathrm{E}[X_iX_j]}{\mathrm{E}[X]^2}-1\\ &= \lim_{n\to\infty}\frac{\mathrm{E}[X]}{\mathrm{E}[X]^2}+\frac{n(n-1)(1-p)^{2n-3}}{n^2(1-p)^{2n-2}}-1\\ &= \lim_{n\to\infty}\frac{1}{1-p}-1=0.\end{aligned}$$

We conclude that $\mathrm{Var}[X]\in o(\mathrm{E}[X]^2)$ and so, by Corollary 2.3.1, if $k<1$,

$$\lim_{n\to\infty}\Pr[G(n,p)\vDash A]=\lim_{n\to\infty}\Pr[X>0]=1.$$

Therefore, if $p(n)\in o(r(n))$, then $\lim_{n\to\infty}\Pr[G(n,p(n))\vDash A]=1$. Furthermore, if $p(n)\in o(r(n))$, $X\sim\mathrm{E}[X]$ almost always as n tends to infinity.

We conclude that $r(n)=\frac{\ln n}{n}$ is a threshold function for property A. □

We could have used Corollary 2.3.3 since we are dealing with a sum of indicator variables. However, since we could show the result without using the upper bound on the covariance, we only needed Corollary 2.3.1. The definition of Δ will be useful in further results.

Also, note that we proved a stronger result than what we needed. We also showed that there are functions which are constant multiples of $r(n)$ such that the probability that $G(n,p)$ satisfies A is close to one or close to zero.

For the interested reader, an important result concerning threshold functions has been proven recently. For certain properties, determining the threshold can be difficult. Nonetheless, the "expectation threshold" function offers a simpler calculation alternative. In 2022, Park demonstrated that the threshold function closely aligns with the expectation threshold, within a logarithmic factor [2]. This finding confirms a conjecture previously proposed by Kahn and Kalai in 2006.

Chapter 3

Numerical Semigroups

3.1 Introduction

So far we have only discussed graphs. In this chapter, we will introduce a new object which has a different structure, but for which the Probabilistic Method can be used to prove results. Definitions and results in this chapter can be found in [3] and [4].

Definition 3.1.1. A *numerical semigroup* is a subset $S \subseteq \mathbb{N}$ for which

1. $0 \in S$,
2. S is closed under addition, i.e. $a, b \in S$ implies $a + b \in S$, and
3. S has finite complement in $\mathbb{N}$.

Examples of numerical semigroups include $\mathbb{N}$ and $\mathbb{N} \setminus \{1\}$. Subsets of $\mathbb{N}$ which are not numerical semigroups include the set of even numbers, any finite set and $\mathbb{N}_0 \setminus \{2\}$. Numerical semigroups are studied in the context of commutative algebra and algebraic geometry, and they have applications in integer programming, coding theory and cryptography [4].

Example 3.1.1. The *McNugget Semigroup* is the set of all non-negative integers which can be expressed as a sum of non-negative multiples of 6, 9 and 20 (see Figure 3.1).

Suppose you are in the United Kingdom and you wish to order 43 McNuggets. The cashier will hesitate for a while before telling you that they do not sell 43 McNuggets, since there is no combination of boxes of 6, 9 and 20 McNuggets which add up to 43 [5]. However, if you order 44 McNuggets, one possibility is that you will receive one box of 20 McNuggets, two boxes of 9 McNuggets and one box of 6 McNuggets. This is because 44 can be expressed as a sum of non-negative multiples of 6, 9 and 20, namely $44 = 2 \cdot 20 + 2 \cdot 9 + 6$. In general, if you order more than 43 McNuggets, you will receive your order.

Let us see why the McNugget Semigroup is a numerical semigroup. First, we note that 0 can be expressed as a sum of non-negative multiples of 6, 9 and 20, namely $0 = 0 \cdot 6 + 0 \cdot 9 + 0 \cdot 20$. Next, we note that if a and b can be expressed as a sum of non-negative multiples of 6, 9 and 20, then so can $a + b$. Finally, we note that the complement of the McNugget Semigroup in $\mathbb{N}_0$

is finite, since

$$\begin{aligned} 44 &= 2 \cdot 20 + 2 \cdot 9 + 6, & 45 &= 5 \cdot 9, \\ 46 &= 2 \cdot 20 + 6, & 47 &= 20 + 3 \cdot 9, \\ 48 &= 8 \cdot 6, & 49 &= 2 \cdot 20 + 9. \end{aligned}$$

And every integer greater than 49 can be expressed as a sum of one of these numbers plus a multiple of 6.

The McNugget semigroup is an example of a numerical semigroup which is *finitely generated*. This means that there exists a finite set $A = \{a_1, \ldots, a_n\}$ such that $S = \langle A \rangle$, where

$$\langle A \rangle = \{c_1 a_1 + \cdots + c_n a_n : c_1, \ldots, c_n \in \mathbb{N}\}.$$

Theorem 3.1.1. *All numerical semigroups are finitely generated.*

Proof. Let S be a numerical semigroup. Let m be the first non-zero element of S. Let b_i be the first element of S such that $b_i \equiv i \pmod{m}$, which exists since S has a finite complement in $\mathbb{N}$. Let $A = \{m, b_1, \ldots, b_{m-1}\}$. Then $S = \langle A \rangle$, since every non-zero element of S can be expressed as a sum of an element of A plus a non-negative multiple of m. □

Also, note that $\gcd(\{6, 9, 20\}) = 1$.

Theorem 3.1.2. *Let $A \subseteq \mathbb{N}$ be a non-empty finite set. Then $\langle A \rangle$ is a numerical semigroup if and only if* $\gcd(A) = 1$.

Proof. Let $A = \{a_1 \ldots, a_n\}$, where a_1 is the first non-zero number in $S = \langle A \rangle$. Note that a_1 is in A, since it cannot be expressed as the sum of non-negative multiples of other elements in S.

If $\gcd(A) = d > 1$, then every element in S is divisible by d and so there are infinitely many numbers in $\mathbb{N}$ which are not in S.

Now, suppose that $\gcd(A) = 1$. If $a_1 = 1$, then $S = \mathbb{N}$ is a numerical semigroup. Suppose that $a_1 > 1$. By definition of generating set, $0 \in S$ and S is closed under addtion. Since $\gcd(A) = 1$ then there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{Z}$ such that

$$\lambda_1 a_1 + \ldots + \lambda_n a_n = 1.$$

Then,

$$k := \sum_{i=1}^{n} \lambda_i a_i + a_1 \sum_{i=1}^{n} |\lambda_i| a_i \equiv 1 \pmod{a)_1},$$

and k is a non-negative sum of multiples of elements of A:

$$k = \sum_{i=1}^{n} (\lambda_i + |\lambda_i| a_1) a_i,$$

and so $k \in S$. Thus, for $0 \leq j < a_1$, $jk \in S$ and jk is congruent with j modulo a_1. Therefore, every number greater than $(a_1 - 1)k$ belongs to S, since it can be expressed as the sum of a multiple of k plus a multiple of a_1. This means that the complement of S in $\mathbb{N}$ is finite and so S is a numerical semigroup. □

The McNugget semigroup example and the proofs of the previous theorems motivates the following invariants of numerical semigroups.

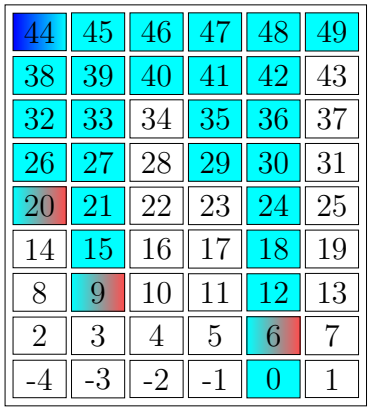

Figure 3.1: Visualization of the McNugget semigroup.

3.2 Invariants

Let S be a numerical semigroup.

Definition 3.2.1. The *multiplicity* of S, denoted by $m(S)$, is the smallest non-zero element of S.

For instance, the multiplicity of the McNugget Semigroup is 6.

Let A and B be non-empty finite subsets of $\mathbb{N}$. Then we denote by $A + B$ the set

$$\{a + b : a \in A, b \in B\}.$$

Theorem 3.2.1. *There exists a unique minimal generating set A with $S = \langle A \rangle$.*

Proof. Let $A = S \setminus (S + S)$. This means that every element in A is not the sum of two elements in S. First we prove that A generates S. Note that A generates 0. Suppose that $s \in S \setminus A$. Then $s = a + b$, such that a and b are in S and $a, b < s$. If we proceed recursively, in a finite number of steps we can express s as a sum of elements of A.

Now, we show that A is minimal. If $S = \langle A' \rangle$, then for $a \in A$, if a is a sum of non-negative multiples of elements of A', then, since a is not the sum of two elements in S, a must be an element of A'. □

Since the minimal generating set is unique, we can define the following invariant.

Definition 3.2.2. The *embedding dimension* of S, denoted by $e(S)$, is the cardinality of the minimal generating set of S.

Corollary 3.2.1. $e(S) \leq m(S)$.

Proof. Apply Theorem 3.2.1 and the proof of Theorem 3.1.1. □

Let n be a non-zero element of S.

Definition 3.2.3. The *Apéry set* of n in S is the set

$$\mathrm{Ap}(S, n) = \{s \in S : s - n \notin S\}$$

With the observation that $s - n \notin S$ if and only if s is the first element in its congruence class modulo n, the Apéry set can also be defined as

$$\mathrm{Ap}(S, n) = \{0, b_1, \ldots, b_{n-1}\},$$

where b_i is the first element of S such that $b_i \equiv i \pmod{n}$. For instance,

$$\mathrm{Ap}(\langle 6, 9, 20\rangle, 6) = \{0, 49, 20, 9, 40, 29\}.$$

.

Proposition 3.2.1. *Each element of* $\mathrm{Ap}(S, n)$ *is either an element of the minimal generating set or a sum of two elements of* $\mathrm{Ap}(S, n)$.

Proof. First, $0 = 0 + 0$. If $0 < s \in \mathrm{Ap}(S, n)$, the other option is that $s = a + b$ such that $a \in S \setminus \mathrm{Ap}(S, n)$ and $b \in S$. But this is not possible, since that means that $a - n \in S$, and so $a - n + b \equiv s \pmod{n}$, which contradicts that s is the first element of its congruence class modulo n. □

For example, in the case of the McNugget semigroup,

$$29 = 20 + 9, \quad 40 = 20 + 20 \quad \text{and} \quad 49 = 40 + 9.$$

Definition 3.2.4. The *Frobenius number* of S, denoted by $F(S)$, is the largest element of $(\mathbb{N} \cup \{-1\}) \setminus S$.

The Frobenius number of the McNugget semigroup is 43. For $n \in S$, using the definition of the Frobenius number, we have that

$$F(S) = \max(\mathrm{Ap}(S, n)) - n, \tag{3.1}$$

since any number larger than this will belong to S. It has been proven that finding the Frobenius number from a variable number of generators is an NP-hard problem [6]. On the other hand, integer programming can be used to find the Frobenius number of a numerical semigroup and, vice versa, the Frobenius number plays an important role in the analysis of integer programming algorithms [7].

Theorem 3.2.2. *Let* $a, b \in \mathbb{N}$ *such that* $\gcd(a, b) = 1$. *Then*

$$F(\langle a, b\rangle) = ab - a - b.$$

Proof. If a and b are relatively prime, then b modulo a generates $\mathbb{Z}_a$ and $(a - 1)b$ is the maximum element of $\mathrm{Ap}(\langle a, b\rangle, a)$. Apply equation 3.1. □

Definition 3.2.5. The *conductor* of S, denoted by $c(S)$, is the first element of S greater than $F(S)$, so $c(S) = F(S) + 1$.

In Figure 3.1, the elements of the McNugget semigroup are colored with light blue, the elements of the minimal generating set are shown in red and the conductor is shown in dark blue. The lowest element of each column is the first element of its congruence class modulo 6, which is the Apéry set of 6 in the McNugget semigroup.

In general, the Apéry set can be used in to calculate many properties of numerical semigroups [8]. For instance, it can be used to calculate the cardinality of the set of gaps $G(S) = \mathbb{N} \setminus S$.

Definition 3.2.6. The *genus* of S, denoted by $g(S)$, is the cardinality of $G(S)$.

Proposition 3.2.2. *Let n be a non-zero element of S.*

$$g(S) = \frac{1}{n}\left(\sum_{s \in Ap(S,n)} s\right) - \frac{n-1}{2}. \tag{3.2}$$

Proof. First, if we divide the complement of S in $\mathbb{N}$ into congruence classes modulo n, we can find the number of gaps in congruence class i by counting the number of elements in that class before the first element of S in that class. Thus, if $b_i \in \mathrm{Ap}(S,n)$ and $b_i \equiv i \pmod{n}$, we can write $b_i = k_i n + i$, and the number of gaps in congruence class i is k_i. Therefore,

$$\begin{aligned} g(S) = \sum_{i=1}^{n-1} k_i &= \frac{1}{n}\left(n\sum_{i=1}^{n-1} k_i + \frac{n(n-1)}{2}\right) - \frac{n-1}{2} \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} k_i n + i\right) - \frac{n-1}{2} \\ &= \frac{1}{n}\left(\sum_{s \in Ap(S,n)} s\right) - \frac{n-1}{2}. \quad \square \end{aligned}$$

Equations 3.1 and 3.2 are known as the *Selmer formulas* [9]. Note that

$$g(S) \leq F(S),$$

since $F(S)$ is the largest element of $\mathbb{N} \setminus S$. Also,

$$g(S) \geq \frac{c(S)}{2} = \frac{F(S)+1}{2},$$

since $s \in S$ implies that $F(S) - s \notin S$, which means that at least half of the elements which are less than $c(S)$ do not belong to S. This also shows that

$$g(S) \leq F(S) \leq 2g(S). \tag{3.3}$$

3.3 Wilf's Conjecture

Now, we present a dangerous problem, in the sense that it is easy to state, but it has not been solved yet. Let $n(S)$ be the number of elements of S which are less than $c(S)$, so that $n(S) + g(S) = c(S)$.

Conjecture 3.3.1 (Wilf, 1978)**.** *[10] For any numerical semigroup S,*

$$\frac{n(S)}{c(S)} \geq \frac{1}{e(S)}.$$

Wilf's conjecture states that the density of the elements of a numerical semigroup which are less than the conductor (also called the *small elements*) is bounded below by the inverse of the embedding dimension. This conjecture has been verified for specific classes of numerical semigroups [4]: for a numerical semigroup S, Wilf's conjecture holds whenever one of these conditions is satisfied:

- $e(S) \leq 3$,
- $e(S) = m(S)$,
- $F(S) - 1 + g(S) \leq 4$,
- $4g(S) \leq 3c(S)$,
- $n(S) \leq 4$,
- $4n(S) \geq c(S)$,
- $F(S) \leq 20$.

This is not an exhaustive list. As there are several published papers on partial solutions, a recent survey on Wilf's conjecture can be found in [11].

Chapter 4

Random Numerical Semigroups

We present three distinct models of random numerical semigroups. The first model employs a uniform distribution for its generators and is the most extensively studied. The second model is inspired in the Erdös-Rényi approach for random graphs. The third model is a new approach we introduce, characterized by a fixed Frobenius number.

4.1 Box Model

Let $n, T \in \mathbb{N}$. We consider the set of points

$$G(T) = \{\mathcal{A} \in \mathbb{N}^n : \gcd(\mathcal{A}) = 1,\ |\mathcal{A}|_\infty \leq T\}.$$

This is the set of possible sets of n generators of numerical semigroups such that each generator is at most T.

Definition 4.1.1. A Box model random numerical semigroup $S(T)$ is a probability space over the set of semigroups $S = \langle \mathcal{A} \rangle$ with $\mathcal{A} \in G(T)$, determined by

$$\Pr[\mathcal{A}] = \frac{1}{|G(T)|}.$$

In other words, a point in $G(T)$ is chosen uniformly at random and the corresponding semigroup is returned. V. I. Arnold was the first to study this model [12]. Assuming that $\mathcal{A} = \{a_1, \ldots, a_n\}$ and $a_1 \leq \ldots \leq a_n$, Erdös and Graham [13] proved that:

$$F(\langle \mathcal{A} \rangle) \leq 2a_n \left\lceil \frac{a_1}{n} \right\rceil - a_1.$$

On the other hand, Aliev and Gruber [14] proved an optimal lower bound for the Frobenius number, namely:

$$F(\langle \mathcal{A} \rangle) > (n-1)!^{\frac{1}{n-1}}(a_1 \cdots a_n)^{\frac{1}{n-1}} - (a_1 + \cdots + a_n).$$

In [12] and [15], Arnold conjectured that the average behavior of the Frobenius number is, up to a constant, given by the lower bound, i.e.:

$$F(\langle \mathcal{A} \rangle) \sim (n-1)!^{\frac{1}{n-1}}(a_1 \cdots a_n)^{\frac{1}{n-1}}.$$

In 2009, Aliev, Henk and Hindrichs [7] proved Arnold's conjecture, by showing the following theorem.

Theorem 4.1.1. *Let $n \geq 3$. Then, as n grows,*

$$\Pr\left[\frac{F(\langle\mathcal{A}\rangle)}{(a_1\cdots a_n)^{\frac{1}{n-1}}} \geq D\right] = o(D^{-2\frac{n-1}{n+1}}).$$

The proof of this theorem is based on a discrete inverse arithmetic-geometric mean inequality.

4.2 ER-type model

In order to use similar methods that can be applied to the Erdös-Rényi model for random graphs, a similar model for random numerical semigroups was proposed in [16].

Definition 4.2.1. For $p \in [0,1]$ and $M \in \mathbb{N}$, an ER-type random numerical semigroup $S(M,p)$ is a probability space over the set of semigroups $S = \langle\mathcal{A}\rangle$ with $\mathcal{A} \subseteq \{1,...,M\}$, determined by

$$\Pr[n \in \mathcal{A}] = p,$$

with these events mutually independent.

In other words, a semigroup $S(M,p)$ is obtained by using the following procedure:

1. Initialize a set $\mathcal{A} = \{0\}$.
2. From step 1 to M, add i to $\mathcal{A}$ with probability p, independently of the other steps.
3. Return the semigroup $S = \langle\mathcal{A}\rangle$.

The main result of [16] is the following theorem.

Theorem 4.2.1.

They also provide the following bounds.

- Test.

We prove the threshold function for co-finiteness.

Theorem 4.2.2.

When p is a constant and let M tend to infinity, we think of the ER-type model as

4.3 Downward model

In relation with Wilf's conjecture 3.3.1, we propose a new model for random numerical semigroups that fixes the Frobenius number.

The main tool for proving the other parts of Theorem 4.2.1 is a correspondence between irreducible numerical semigropus with a simplicial complex.

Definition 4.3.1.

Chapter 5

Experiments

5.1 ER-type model experiments

We used the following algorithm for random numerical semigroup generation.

Algorithm 1 An algorithm with caption

Require: $n \geq 0$
Ensure: $y = x^n$
  $y \leftarrow 1$
  $X \leftarrow x$
  $N \leftarrow n$
  while $N \neq 0$ **do**
    if N is even **then**
      $X \leftarrow X \times X$
      $N \leftarrow \frac{N}{2}$ ▷ This is a comment
    else if N is odd **then**
      $y \leftarrow y \times X$
      $N \leftarrow N - 1$
    end if
  end while

For random sums of cyclic groups, we used the following algorithm.

In this Chapter, XXX is presented. Include pseudocode. ..

5.2 Downward model experiments

Etc. etc.

Chapter 6

Results

In this chapter, we present the main results of this Thesis. We start by proving a Theorem found in [16] without the use of the simplicial complex.

6.1 Lower Bound

We prove a Theorem found in [16] without the use of the simplicial complex.

Theorem 6.1.1. *Let $S \sim S(M,p)$, where $p = p(M)$ is a monotone decreasing function of M. If $\frac{1}{M} \ll p \ll 1$, then S is cofinite, i.e., the set of gaps is finite, a.a.s and*

$$\lim_{M\to\infty} \mathrm{E}[e(S)] = \lim_{M\to\infty} \mathrm{E}[g(S)] = \lim_{M\to\infty} \mathrm{E}[F(S)] = \infty.$$

Proof. Let $X := \min(S \setminus \{0\})$ be a random variable. Then, for $0 < n \leq M$,

$$\Pr[X = n] = p(1-p)^{n-1},$$

and so

$$\begin{aligned}\mathrm{E}[X] &= \sum_{n=0}^{\infty} n\Pr[X=n] = \sum_{n=0}^{M} np(1-p)^{n-1} = p\frac{d}{dp}\left[-\sum_{n=0}^{M}(1-p)^n\right] \\ &= p\frac{d}{dp}\frac{(1-p)^{M+1}-1}{p} = p\frac{1-(1-p)^{M+1}-(M+1)(1-p)^M p}{p^2} \\ &= \frac{1-(1-p)^M - M(1-p)^M p}{p} \geq \frac{1-e^{-Mp}-Mpe^{-Mp}}{p}.\end{aligned}$$

Thus, since $\lim_{M\to\infty} Mp = \infty$, then $\lim_{M\to\infty} Mpe^{-Mp} = \lim_{M\to\infty} e^{-Mp} = 0$, which implies that

$$\lim_{M\to\infty} \mathrm{E}[X] = \lim_{M\to\infty} \frac{1-e^{-Mp}-Mpe^{-Mp}}{p} = \infty.$$

Also, note that if $p = \frac{c}{M}$, $c \in \mathbb{R}_+$ $(0 < e^{-c} + ce^{-c} < 1)$,

$$\lim_{M\to\infty} \mathrm{E}[X] = \lim_{M\to\infty} \frac{1-e^{-c}-ce^{-c}}{p} = \infty.$$

Proof. Fix $a \in \mathbb{N}$ such that $a > 11$ and let $A = \{1, \ldots, \lfloor \frac{a}{p} \rfloor\}$. Since $\frac{1}{M} \ll p$, we have that $\lfloor \frac{a}{p} \rfloor \leq M$ for large enough M. Consider the following events:

- E_1: No generator selected is less than $\frac{1}{ap}$.

  Let X_1 be the number of generators selected from $\{1, \ldots, \lfloor \frac{1}{ap} \rfloor\}$. Then

$$\Pr[\overline{E_1}] = \Pr[X_1 > 0] \leq \mathrm{E}[X_1] \leq p \cdot \frac{1}{ap} = \frac{1}{a}.$$

- E_2 : At most $\frac{3a}{2}$ generators are selected from A.

  Let X_2 be the number of generators selected in A, then X_2 is a binomial random variable with $n = \frac{a}{p}$ and we can use the bound (Feller I can add this to the appendix)

$$\Pr[\overline{E_2}] = \Pr\left[X_2 > \frac{3a}{2}\right] \leq \frac{\frac{3a}{2}(1-p)}{(\frac{3a}{2} - a)^2} \leq \frac{6}{a}.$$

  Also, note that by the union bound

$$\Pr[E_1 \wedge E_2] \leq 1 - \frac{1}{a} - \frac{6}{a} = 1 - \frac{7}{a}.$$

- E_3 : At least $\frac{a}{2}$ generators are selected from A.

  Similarly, we can use the bound for the other tail of the distribution so that

$$\Pr[\overline{E_3}] = \Pr\left[X_2 < \frac{a}{2}\right] \leq \frac{(n - \frac{a}{2})p}{(np - \frac{a}{2})^2} = \frac{a - (\frac{a}{2})p}{(\frac{a}{2})^2} \leq \frac{4}{a}.$$

- E_4 : The generators selected from A are minimal.

  Let $Y_{(1)}, Y_{(2)}, \ldots, Y_{(k)}$ denote the first k generators selected in A. Assume E_1 and E_2. We have that E_1 implies $Y_{(1)} \geq \frac{1}{ap}$ and E_2 implies $k \leq \frac{3a}{2}$.

  First we bound for the probability that, given E_1 and E_2, $b \in A$ is selected as a generator. By conditional probability

$$\begin{aligned}\Pr[b \text{ is selected}] = &\Pr[b \text{ is selected} | E_1 \wedge E_2] \Pr[E_1 \wedge E_2] \\ &+ \Pr[b \text{ is selected} | \overline{E_1 \wedge E_2}] \Pr[\overline{E_1 \wedge E_2}],\end{aligned}$$

  and so

$$\Pr[b \text{ is selected} | E_1 \wedge E_2] \leq \frac{\Pr[b \text{ is selected}]}{\Pr[E_1 \wedge E_2]} \leq \frac{p}{1 - \frac{7}{a}}.$$

  Now, note that $Y_{(2)}$ is not minimal if a multiple of $Y_{(1)}$ is selected in A. Thus, if we fix $Y_{(1)} = y_1 \geq \frac{1}{ap}$, $Y_{(1)}$ is not minimal if $b \in \{2y_1, 3y_1, \ldots, c_1 y_1\}$ is selected, where $c_1 y_1$ is the largest multiple of y_1 which does not exceed $\frac{a}{p}$. Since $y_1 \geq \frac{1}{ap}$, we have that $c_1 \leq a^2$. Then, using the union bound,

$$\Pr[Y_{(2)} \text{ is not minimal} | E_1 \wedge E_2 \wedge Y_{(1)} = y_1] \leq \frac{pa^2}{1 - \frac{7}{a}}.$$

  If we sum over all possible y_1, we get that

$$\Pr[Y_{(2)} \text{ is not minimal} | E_1 \wedge E_2] \leq \frac{pa^2}{1 - \frac{7}{a}}.$$

Similarly, for $2 \leq t \leq k$ and fixed $Y_{(1)} = y_1, \ldots, Y_{(t-1)} = y_{t-1}$, $Y_{(t)}$ is not minimal if the first $t-1$ numbers selected from A can generate $Y_{(t)}$. For the possible numbers generated by the first t numbers selected, there are at most a^2 choices for each coefficient, so there are at most a^{2t} such linear combinations. Then

$$\Pr[Y_{(t)} \text{ is not minimal} | E_1 \wedge E_2] \leq \frac{pa^{2t}}{1 - \frac{7}{a}}.$$

Therefore, since $Y_{(1)}$ is always minimal, we can use the union bound and $k \leq \frac{3a}{2}$ to conclude that

$$\Pr[E_4 | E_1 \wedge E_2] \geq 1 - \frac{p}{1 - \frac{7}{a}} \sum_{t=1}^{\frac{3a}{2}-1} a^{2t} = 1 - o(1).$$

Thus,

$$\Pr[E_4] = \Pr[E_4 | E_1 \wedge E_2] \Pr[E_1 \wedge E_2] \geq 1 - \frac{7}{a} - o(1).$$

Finally, note that by union bound,

$$\Pr[E_4 \wedge E_3] \geq 1 - \frac{11}{a} - o(1).$$

Therefore, for every $N \in \mathbb{N}$ and $\varepsilon > 0$, there exists K such that $M \geq K$ implies

$$\Pr[f(S) > N],\ \Pr[g(S) > N],\ \Pr[e(S) > N] > 1 - \varepsilon.$$

.

6.2 Upper bound

I conjecture that the hypothesis that q is prime can be relaxed.

6.2.1 Iteraded sumsets in cyclic groups

Lemma 6.2.1. *Let q be a prime number and S be a random subset of $\mathbb{Z}_q$ of size $4\lfloor 3 \log_2 q \rfloor$. As q tends to infinity, $2\lfloor 3 \log_2 q \rfloor S$ covers $\mathbb{Z}_q$ almost always.*

Let q be a prime number and let $s \in \mathbb{N}$ such that $s \leq q$. Let S be a uniformly random subset of $\mathbb{Z}_q$ of size s, that is,

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

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 A_K be the event that $K \subset S$. Let X_K be the indicator variable of A_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 provide two ways of finding $\mathrm{E}[X_z]$. The first one uses that, for any $K \subset N_z^k$,

$$\mathrm{E}[X_K] = \Pr[A_K] = \frac{\binom{q-k}{s-k}}{\binom{q}{s}},$$

and so we get that

$$\mathrm{E}[X_z] = \sum_{K \in N_z^k} \mathrm{E}[X_K] = |N_z^k| \mathrm{E}[X_K] = \frac{1}{q}\binom{q}{k}\frac{\binom{q-k}{s-k}}{\binom{q}{s}} = \frac{1}{q}\binom{s}{k}.$$

This motivates the second way, for we know that

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

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 also find that

$$E[X_z] = \frac{1}{q}\binom{s}{k}. \tag{6.1}$$

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

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

Fix $j \leq k$, then

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

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,

$$\Delta_j \leq \frac{\binom{p}{k}\binom{q-k}{k-j}\binom{q-2k+j}{s-2k+j}}{q\binom{q}{s}}.$$

This implies that, using 6.1,

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

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

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

since $\binom{2\lfloor q^\alpha\rfloor}{\lfloor 3\log_2 q\rfloor}^2 \leq \binom{4\lfloor 3\log_2 q\rfloor}{2\lfloor 3\log_2 q\rfloor}$ (I can prove this in a lemma or in the appendix).

This proves that

$$\Pr[X_z = 0] \leq \frac{\Delta}{\mathrm{E}[X_z]^2} = \sum_{j=0}^{k}\frac{\Delta_j}{\mathrm{E}[X_z]^2} \leq \frac{(k+1)}{q^2}.$$

Therefore, by the union bound,

$$\Pr[\bigvee_{z\in\mathbb{Z}_q} X_z = 0] \leq \frac{(k+1)}{q}.$$

6.2.2 Proof of the upper bound

Theorem 6.2.1. *Let $g(x)$ be a function for which $x(\log x)^2 \in o(g(x))$. Then*

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

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 $g\left(\frac{1}{p}\right)$ with high probability, since $F(S)$ 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 (E1).
2. Then, in the spirit of the Lemma, we will find a step such that s elements, which are different modulo q, are chosen with high probability (E2).
3. Finally, we will apply the Lemma to the Ápery set of a subsemigroup of S generated by the subset in part 2.

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 $q \in S$, 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 and let k_x be the number of primes between $20x \log x$ and $h(x)$. For $n \geq 6$, by the Prime Number Theorem,

$$n(\log n + \log\log n - 1) < q_n < n(\log n + \log\log n) = o(h(n)).$$

Thus, $n = o(k_n)$ (I can prove this if it is not clear) and, for every $c > 0$,

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

Therefore,

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

Part 2

Now, assume E_1. Then S contains a prime number q for which

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

Let $s = 4\lfloor 3\log_2 q \rfloor$, as in the Lemma.

Let $A := \{q+1, q+2, \ldots, 2q\}$. Consider the event **E2** that at least s generators are selected in A. Let X_1 be the number of generators selected in A, then X_1 is a binomial random variable with parameters $n = q$ and p. Then, in a similar way to $E2$ in Theorem 1, we use the Binomial Distribution Tail Bound to show that, assuming that p is small enough so that $qp > s$ for all possible q,

$$\Pr[\overline{E_2}|E_1] = \Pr[X_1 < s] \leq \Pr[X_2 < s] \leq \frac{(n-s)p}{(np-s)^2} = \frac{(q-s)p}{(qp-s)^2}.$$

Thus, bounding by the worst case asymptotically, (needs to be explained better)

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

We conclude that

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

and so

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

Part 3

Finally, assume E_1 and E_2. Let $B = \{Y_1, \ldots, Y_s\}$ be a randomly selected subset of size s of the generators selected in E_2. Since the generators are chosen randomly and $|A| = q$, we can apply the Lemma to the Ápery set of the subsemigroup generated by B, denoted by $G(B)$, and conclude that the Ápery set of $G(B)$ will be completed before step $h\left(\frac{1}{p}\right) 2\left\lfloor 3\log_2 h\left(\frac{1}{p}\right)\right\rfloor$ with high probability as $p \to 0$.

Thus, if $g(x)$ be a function for which $x(\log x)^2 \in o(g(x))$ (Probably needs to be explained better),

$$\lim_{p\to 0} \Pr\left[F(G(B)) \leq g\left(\frac{1}{p}\right)\right] = 1.$$

Since $F(S) \leq F(G(B))$, we conclude that

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

6.3 Conclusions and Further work

Bibliography

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

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

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

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

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

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

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

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

[9] E. S. Selmer, "On the linear diophantine problem of Frobenius," 1977.

[10] 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.

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

[12] 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.

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

[14] 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.

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

[16] 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.

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

[18] B. Rosser, "Explicit bounds for some functions of prime numbers," *American Journal of Mathematics*, vol. 63, no. 1, pp. 211–232, 1941.

[19] P. Dusart, "The k^{th} prime is greater than $k(\log k + \log\log k - 1)$ for $k \geq 2$," *Mathematics of Computation*, vol. 68, no. 225, pp. 411–415, 1999.

Appendix A

Example Appendix

Here you might present some additional results, derivations, proofs etc. that were not included in the main text.

A.1 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.1.1. $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ *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.1.2. $\left(\frac{n}{k}\right)^k \geq \binom{n}{k}$ *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.1.3. $(1-p) \leq e^{-p}$ *for* $0 \leq p \leq 1$.

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

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

We prove a bound on the tails of the binomial distribution found in [17].

We also give a combinatorial proof of the following result.

Proposition A.1.4. $\binom{2n}{k}^2 \leq \binom{4n}{2k}$ *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 $\binom{2n}{k}^2$. □

Appendix B

Software Documentation

Here's an example source code listing, where the code is read in from an external file:

```
% Function to create a nice rotating animated GIF of 3D volumetric data V

function animation (V)

h = volshow (V, 'BackgroundColor', [0 0 0], 'Renderer', 'MaximumIntensityProjection', 'CameraPosition', [2 2 0], 'CameraUpVector', [1 0 0], 'ColorMap', jet);

camproj ('perspective');

N = 500;

filename = 'animation.gif';
vec = linspace(0, 4 * pi(), N)';
myPosition = 2 * [zeros(size(vec)) cos(vec) sin(vec)];

for idx = 1:N
% Update current view.
    h.CameraPosition = myPosition(idx, :);
% Use getframe to capture image.
    I = getframe(gcf);

    [indI, cm] = rgb2ind (I.cdata,256);
% Write frame to the GIF File.
    if idx == 1
        imwrite(indI, cm, filename, 'gif', 'Loopcount', inf, 'DelayTime', 0.05);
    else
        imwrite(indI, cm, filename, 'gif', 'WriteMode', 'append', 'DelayTime', 0.05);
    end
end
```

B.1 Code Availability

All scripts and source code used for simulation and analysis of the ... are available here

```
% Function to create a nice rotating animated GIF of 3D volumetric data V

function animation (V)

h = volshow (V, 'BackgroundColor', [0 0 0], 'Renderer', 'MaximumIntensityProjection', 'CameraPosition', [2 2 0], 'CameraUpVector', [1 0 0], 'ColorMap', jet);

camproj ('perspective');

N = 500;

filename = 'animation.gif';
vec = linspace(0, 4 * pi(), N)';
myPosition = 2 * [zeros(size(vec)) cos(vec) sin(vec)];

for idx = 1:N
% Update current view.
    h.CameraPosition = myPosition(idx, :);
% Use getframe to capture image.
    I = getframe(gcf);

    [indI, cm] = rgb2ind (I.cdata,256);
```

```
% Write frame to the GIF File.
    if idx == 1
        imwrite(indI, cm, filename, 'gif', 'Loopcount', inf, 'DelayTime', 0.05);
    else
        imwrite(indI, cm, filename, 'gif', 'WriteMode', 'append', 'DelayTime', 0.05);
    end
end
```

`https://bitbucket.org/username/gitrepo.git`

B.2 Software Requirements

- MATLAB code is confirmed working with version XXXX;
- Simulations require the use of gcc version XXX or llvm/clang version YYYY

B.3 Simulation Code - How to Run