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 24, 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.

Contents

Chapter 1

Introduction

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, usually a graph, by showing that the probability that a randomly chosen object has that property is greater than zero. In this case, we will apply the probabilistic method to numerical semigroups.

A numerical semigroups is a subset of $\mathbb{N}$ that is closed under addition (Definition 3.1.1). These objects are studied in the context of commutative algebra and algebraic geometry, and they have applications in integer programming, coding theory and cryptography [1]. There are numerical invariants that are used to study numerical semigroups, such as the embedding dimension, the genus and the Frobenius number (Definitions 3.2.2, 3.2.6 and 3.2.4). For example, the Frobenius number is defined as the maximum of the complement of the numerical semigroup over the integers.

The Erdös-Rényi (ER) model is a commonly used model of random graphs, where each edge is chosen with probability p, independently of the other edges (Definition 2.4.1). In contrast to graphs, the algebraic nature of numerical semigroups allows for a wider range of random models. This thesis investigates the average behavior random numerical semigroup invariants using a probabilistic model similar to the ER model (Definition 4.2.1).

Our central result is Theorem 6.1.1, which is similar to Theorem 4.2.1, the main result of

- 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.

Theorem 4.2.1 describes the behavior of the expected embedding dimension, genus, and Frobenius number of a random numerical semigroup, depending on the parameters of the model. It gives an explicit bound of the expected value of these invariants. On the other hand, Theorem 6.1.1 describes the behavior of the invariants almost surely, that is, with probability that tends to one as the parameters of the model converge to certain values. Our proof is more elementary and provides asymptotically tighter bounds of the behaviour of these invariants.

We used experiments to study the behavior of ER-type random numerical semigroups, which led to the proof of Theorem 6.1.1. For our experiments, we used `numsgps-sage` [3, O'Neill][4, Delgado] and for visualizations we used `IntPic` [5, Delgado]. We also implemented our own publicly available repository `randnumsgps` [6] for generating and visualizing random numerical semigroups in Python.

The structure of the thesis is as follows:

- Chapter 2 discusses the Probabilistic Method, based on the work of Noga Alon and Joel H. Spencer.
- Chapter 3 focuses on numerical semigroups, providing definitions, examples, and results necessary for understanding their structure.
- Chapter 4 introduces three models of random numerical semigroups, including our newly proposed model. We also present recent results in the field, including Theorem 4.2.1.
- Chapter 5 details the algorithms and experiments conducted.
- Chapter 6 presents the main results, including the proof of Theorem 6.1.1, its implications and its relation with Theorem 4.2.1.

The aim of this thesis is to extend the methods used in the study of random graphs to numerical semigroups. The research seeks to contribute to the understanding of random numerical semigroups through a probabilistic perspective.

To sum up, we provide a detailed study of random numerical semigroups using probabilistic models, experimental data and software tools.

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 [7].

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*.

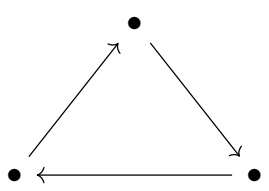

Figure 2.1: A tournament on 3 vertices with property S_1.

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 [7])**.** *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 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.0.1 and A.0.3, 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 $a \in \mathbb{R}$,

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

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 $\mathrm{Pr}[X = x] > 0$, and there is a point $x \leq \mathrm{E}[X]$ such that $\mathrm{Pr}[X = x] > 0$. 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 [7])**.** *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 [7])**.** $\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}$.

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

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j]. \tag{2.5}$$

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

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j] \leq \mathrm{E}[X] + \sum_{i \sim j} \Pr[A_i \wedge A_j] = \mathrm{E}[X] + \Delta. \tag{2.6}$$

The result follows from Theorem 2.3.2 and Corollary 2.3.1. □

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.

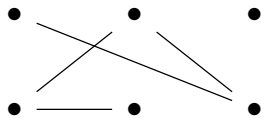

Figure 2.2: A graph with an isolated vertex.

2.4.1 A threshold function for isolated vertices

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,

   $$\sum_{i\neq j} \mathrm{E}[X_i X_j] = \sum_{i\neq j} \Pr[X_i = X_j = 1]$$

$$= n(n-1)(1-p)^{n-1}(1-p)^{n-2}$$
$$= n(n-1)(1-p)^{2n-3},$$

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 $\lim_{n\to\infty} p(n) = 0$,

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

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, a related function called the "expectation threshold" offers a simpler calculation alternative. In 2022, Park demonstrated that the threshold function closely aligns with the expectation threshold, within a logarithmic factor [8]. 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 [9] and [1].

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 $\mathbb{N}_0 \setminus \{2\}$.

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 [10]. 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

$$44 = 2 \cdot 20 + 2 \cdot 9 + 6, \qquad 45 = 5 \cdot 9,$$
$$46 = 2 \cdot 20 + 6, \qquad 47 = 20 + 3 \cdot 9,$$

$$48 = 8 \cdot 6, \qquad 49 = 2 \cdot 20 + 9.$$

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 addition. 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.

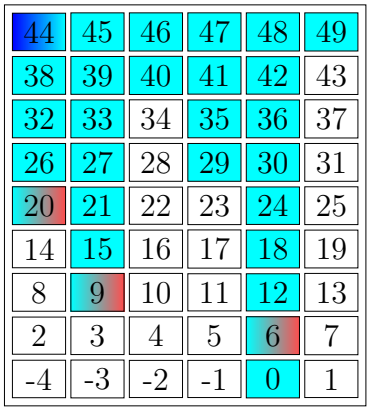

Figure 3.1: Visualization of the McNugget semigroup.

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 [11]. 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 [12].

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 [13]. 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* [14]. Note that, for S such that $g(S) > 0$,

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

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 shows the following for a numerical semigroup S with $g(S) > 0$.

Proposition 3.2.3. $g(S) \leq F(S) \leq 2g(S)$.

Finally, we define a class of semigroups which are used in the proof of Theorem 4.2.1.

Definition 3.2.7. A numerical semigroup S is *irreducible* if it is maximual with respect to inclusion among all numerical semigroups with the same Frobenius number.

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)**.** *[15] 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 [1]: 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 [16].

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 by the Erdös-Rényi approach to 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

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

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 $\mathcal{S}(T)$ is a probability space over the set of semigroups $\mathcal{S} = \langle \mathcal{A} \rangle$ 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 [17]. For any numerical semigroup, for $\mathcal{A} = \{a_1, \ldots, a_n\}$ such that $a_1 \leq \ldots \leq a_n$, Erdös and Graham [18] 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 [19] 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 [17] and [20], 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 [12] proved Arnold's conjecture, by showing the following theorem.

Theorem 4.1.1. *Let $n \geq 3$. Then, for every constant D,*

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

The statement of this theorem does not depend on T, it only depends on n. The proof is based on a discrete inverse arithmetic-geometric mean inequality.

4.2 ER-type model

In order to use methods such as those that apply to the Erdös-Rényi model for random graphs, a similar model for random numerical semigroups was proposed in [2].

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

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

with these events mutually independent.

Note that this definition does not require a numerical semigroup $\mathcal{S}$ to be co-finite. A semigroup $\mathcal{S}(M,p)$ is obtained by using the following procedure:

1. Initialize an empty set $\mathcal{A}$.
2. As i goes from 1 to M, add i to $\mathcal{A}$ with probability p, independently of the other steps.
3. Return the semigroup $\mathcal{S} = \langle\mathcal{A}\rangle$.

The main result of [2] is the following theorem.

Theorem 4.2.1. *Let $\mathcal{S} \sim \mathcal{S}(M,p)$, where $p = p(M)$ is a monotone decreasing function of M. Then,*

(a) If $p(M) \in o\left(\frac{1}{M}\right)$, then $\mathcal{S} = \{0\}$ almost always.

(b) If $\frac{1}{M} \in o(p(M))$ and $\lim_{M\to\infty} p(M) = 0$, then $\mathcal{S}$ is co-finite almost always and

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

(c) If $\lim_{M\to\infty} p(M) > 0$, then

$$\lim_{M\to\infty} \mathrm{E}[e(\mathcal{S})] < \infty, \quad \lim_{M\to\infty} \mathrm{E}[g(\mathcal{S})] < \infty \quad \text{and} \quad \lim_{M\to\infty} \mathrm{E}[F(\mathcal{S})] < \infty,$$

and each limit is bounded by explicit rational functions in p.

Note that this proves that $\frac{1}{M}$ is a threshold function for co-finiteness, if we extend the definition of threshold functions of random graphs (Definition 2.4.2) to ER-type random numerical semigroups. The proof of this part of the theorem follows from standard arguments using the probabilistic method (Theorem 4.2.3).

On the other hand, parts (b) and (c) follow from the construction of a shellable simplicial complex ([2, Definition 8]) whose facets are in bijection with irreducible numerical semigroups of a fixed Frobenius number n (Definition 3.2.7). It turns out that the faces of the $n-$th simplicial complex count the number of sets $\mathcal{A} \subset \{1, \ldots, n-1\}$ satisfying $n \notin \langle \mathcal{A} \rangle$ ([2, Proposition 23]). Thus, the expected value of the embedding dimension can be calculated from the entries of the h-vector (in the sense of algebraic combinatorics [21]) of this simplicial complex.

The authors also provide bounds for the expected value of the embedding dimension and the genus, when p is constant.

Theorem 4.2.2. *Let $\mathcal{S} \sim \mathcal{S}(M, p)$, where p is a constant. Then,*

$$\begin{aligned} \frac{6-8p+3p^2}{2-2p^2+p^3} &\leq \lim_{M\to\infty} \mathrm{E}[e(\mathcal{S})] \leq \frac{2-p^2}{p}, \\ \frac{6-14p+11p^2-3p^3}{2p-2p^3+p^4} &\leq \lim_{M\to\infty} \mathrm{E}[g(\mathcal{S})] \leq \frac{(1-p)(2-p^2)}{p^2}, \text{ and} \\ \frac{6-14p+11p^2-3p^3}{2p-2p^3+p^4} &\leq \lim_{M\to\infty} \mathrm{E}[f(\mathcal{S})] \leq \frac{2(1-p)(2-p^2)}{p^2}. \end{aligned}$$

As p tends to 0, this says that the limit of the expected value of the embedding dimension is between a function that tends to a constant and a function that is asymptotically linear in $\frac{1}{p}$. Also, the expected values of the genus and the Frobenius number are between a function that is asymptotically linear in $\frac{1}{p}$ and a function that is asymptotically quadratic in $\frac{1}{p}$. It is useful to think of these functions in terms of $\frac{1}{p}$, since it is the expected value of the gap between randomly selected elements of $\mathcal{A}$ in the ER-type model.

We now prove part (a) of Theorem 4.2.1.

Theorem 4.2.3 (Theorem 5 [2])**.** *$\frac{1}{M}$ is a threshold function for co-finiteness.*

Proof. We prove the two parts of the theorem separately.

Part 1

Suppose that $p \in o\left(\frac{1}{M}\right)$. As in Inequality 2.4,

$$\Pr[|\mathcal{A}| > 0] \leq \mathrm{E}[|\mathcal{A}|] = Mp.$$

Since $Mp \to 0$ as $M \to \infty$, we have that $\Pr[|\mathcal{A}| > 0] \to 0$ as $M \to \infty$. Thus, $\mathcal{S} = \{0\}$ almost always.

Part 2

Suppose that $\frac{1}{M} \in o(p)$ and $\lim_{M\to\infty} p = 0$. We prove that $\mathcal{S}$ is co-finite almost always, by showing that $\mathcal{A}$ almost always contains a co-prime pair of elements.

For each pair of distinct co-prime integers $i, j \in \{1, \ldots, m\}$, let $A_{i,j}$ be the event that $i, j \in \mathcal{A}$. Let $1_{i,j}$ be the indicator variable of event $A_{i,j}$ and let

$$X = \sum_{\substack{i<j \\ \gcd(i,j)=1}} 1_{i,j}.$$

We show that $X > 0$ almost always using the second moment method. First we calculate $\mathrm{E}[X]$. It is a well-established result in number theory [22, Theorem 332] that the limiting probability of two integers being co-prime is $\frac{6}{\pi^2}$. That is,

$$\lim_{M\to\infty} \frac{|\{(a,b) \in \{1,\ldots,M\}^2 : \gcd(a,b) = 1\}|}{M^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

Therefore, $E[X] \sim \frac{3}{\pi^2} M^2 p^2$. Now, $\{i, j\} \sim \{i', j'\}$ (see (2.5)) if $i = i'$ and $j \neq j'$. Thus, by varying over triples i, j, j', we obtain

$$\Delta = \sum_{\{i,j\}\sim\{i',j'\}} \Pr[A_{i,j} \wedge A_{i',j'}] \leq M^3 p^3.$$

Therefore, $\Delta \in o(\mathrm{E}[X]^2)$, since $\frac{1}{M} \in o(p)$. Thus, by the second moment method (Corollary 2.3.3), $X > 0$ almost always. This means that $\mathcal{A}$ almost always contains a co-prime pair of elements and so $\mathcal{S}$ is co-finite. □

In chapter 6, we prove part (b) of Theorem 4.2.1 using standard probabilistic methods. We also prove a result related to part (c) (Lemma 6.3.2) .

4.3 Downward model

We propose a new model for random numerical semigroups that fixes the Frobenius number, which could be helpful in the study of Wilf's conjecture. We have not yet proved any results about this model, but we present some experiments in Chapter 5.

Definition 4.3.1. Fix $f \in \mathbb{N}$, then the downward random numerical semigroup $\mathcal{S}_{\downarrow}(f, p)$ is defined by the following procedure:

1. Initialize a set $\mathcal{A} = \{f + 1, \ldots, 2f, 2f + 1\}$.
2. As i goes down from $f-2$ to 1. If $F(\langle A \cup \{i\}\rangle) = f$, add i to $\mathcal{A}$ with probability p. Otherwise, proceed to the next step.
3. Return the semigroup $\mathcal{S} = \langle \mathcal{A} \rangle$.

Step 1 initializes the minimal numerical semigroup (with respect to inclusion) such that $f(\mathcal{S}) = f$. Step 2 randomly adds elements to $\mathcal{A}$ in decreasing order, such that $F(\langle A \cup \{i\}\rangle) = f$. Note that in this case, the probability of adding an element to $\mathcal{A}$ is not independent of the other steps.

Chapter 5

Experiments

5.1 ER-type model experiments

In this section, we will present the experiments we conducted to study the behavior of ER-type random numerical semigroups. We used the following algorithm for ER-type random numerical semigroup generation (see Definition 2.4.1). The `NumericalSemigroup` class is part of the `numsgps-sage` package [3].

```
    # Generates random sample using Erdös–Rényi model
    def generate_sample(M, p, sample_size):
        semigroups = []
        for j in range(sample_size):
            generators = []
            for i in range(M):
                if random.random() < p:
                    generators.append(i)
            if gcd(generators) != 1:
                continue
            semigroups.append(NumericalSemigroup(generators))
        return semigroups
```

For the experiments, we generated random numerical semigroups for several values of $1/p$ in the range $[4, 1000]$. For each $1/p$, the sample size was 1000 and we calculated the average embedding dimension (Table 5.1) and the average Frobenius number (Table 5.2). As shown in the experiments done in [2], the bounds found in Theorem 4.2.2 are not tight.

$1/p$	Lower Bound	$\overline{e(\mathcal{S})}$	Upper bound
4.00	2.21	2.79	7.75
75.14	2.95	9.05	150.27
146.29	2.97	10.67	292.56
217.43	2.98	12.28	434.85
288.57	2.99	12.94	577.14
359.71	2.99	13.65	719.43
430.86	2.99	14.14	861.71
502.00	2.99	14.73	1,004.00
573.14	2.99	15.29	1,146.28
644.29	2.99	16.17	1,288.57
715.43	2.99	16.07	1,430.86
786.57	2.99	16.34	1,573.14
857.71	3.00	16.61	1,715.43
928.86	3.00	17.39	1,857.71
1,000.00	3.00	17.22	2,000.00

Table 5.1: Average embedding dimension of random numerical semigroups generated using the ER-type model (15 samples of 1000 random numerical semigroups).

$1/p$	Lower Bound	$\overline{F(\mathcal{S})}$	Upper bound
4.00	7.26	13.96	46.50
75.14	218.56	1,088.82	22,283.25
146.29	431.93	2,483.26	85,010.91
217.43	645.33	4,174.94	188,229.03
288.57	858.75	5,859.29	331,937.60
359.71	1,072.17	7,794.18	516,136.62
430.86	1,285.59	9,594.56	740,826.09
502.00	1,499.02	11,533.38	1,006,006.00
573.14	1,712.45	13,765.73	1,311,676.37
644.29	1,925.87	16,239.19	1,657,837.19
715.43	2,139.30	17,769.34	2,044,488.45
786.57	2,352.73	19,806.19	2,471,630.17
857.71	2,566.15	22,157.78	2,939,262.33
928.86	2,779.58	25,079.10	3,447,384.94
1,000.00	2,993.01	26,637.46	3,995,998.00

Table 5.2: Average Frobenius number of random numerical semigroups generated using the ER-type model (15 samples of 1000 random numerical semigroups).

We plot the average embedding dimension and the average Frobenius number found in the experiments in Figures 5.1 and 5.2. These experiments led us to conjecture that the average embedding dimension grows as $\log(1/p)$ and the average Frobenius number grows as $(1/p)\log(1/p)$. Although we did not prove this conjecture, it led us in the right direction to prove Theorem 6.1.1.

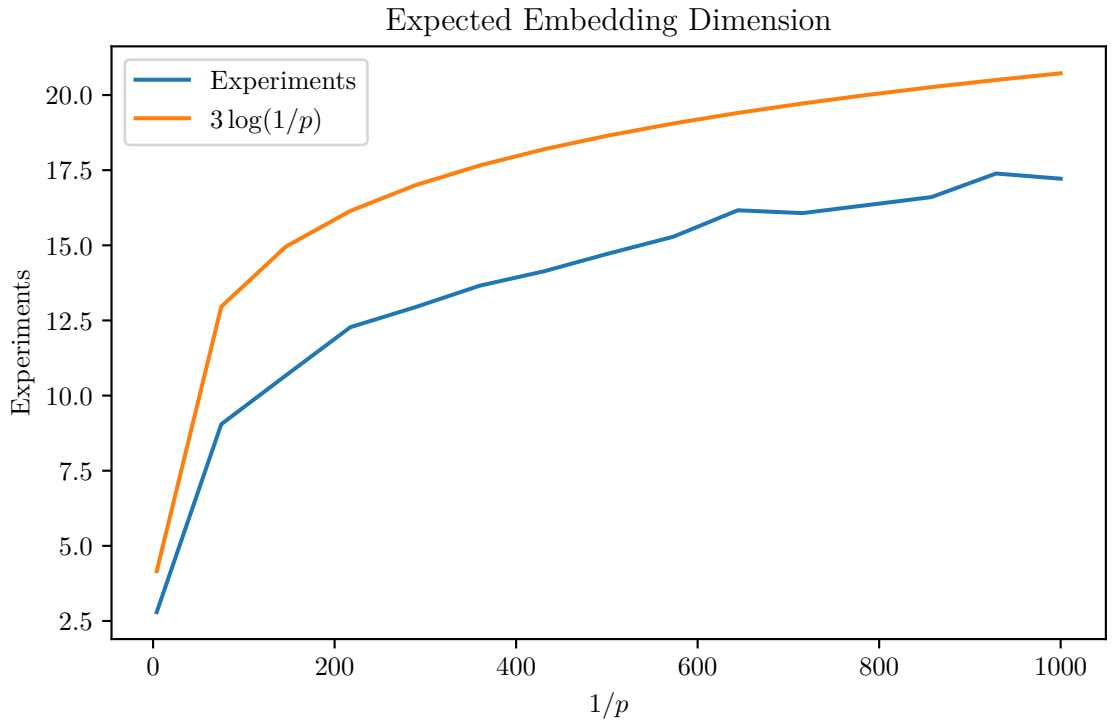

Figure 5.1: Average embedding dimension of random numerical semigroups generated using the ER-type model.

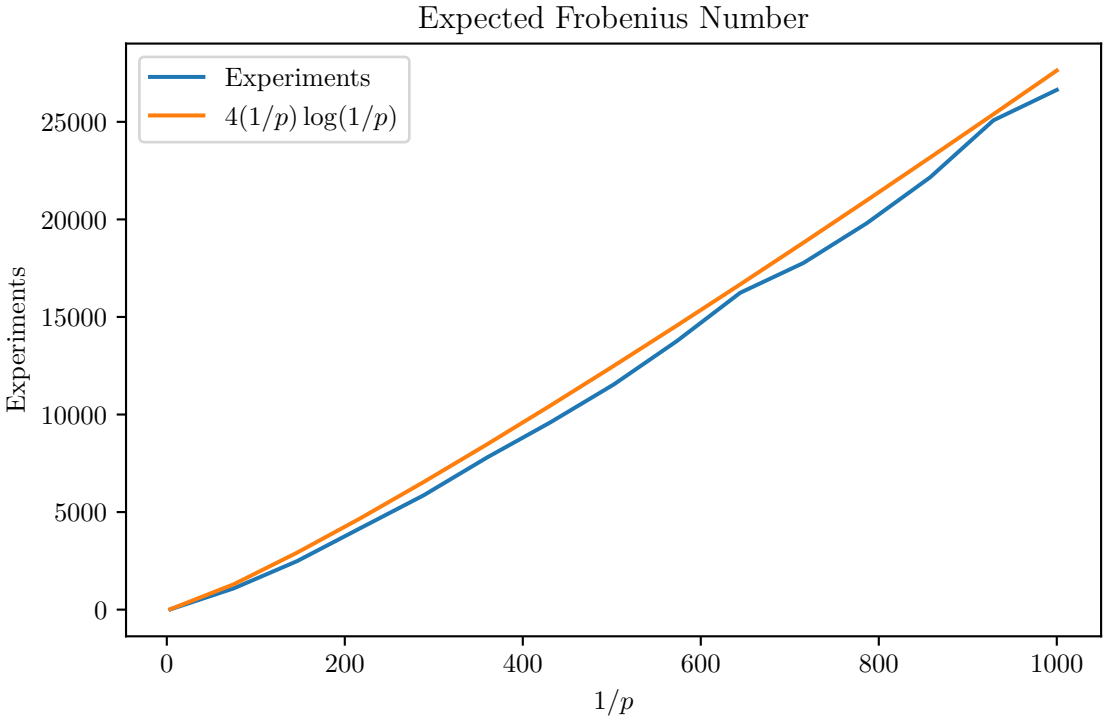

Figure 5.2: Average Frobenius number of random numerical semigroups generated using the ER-type model vs $4(1/p)\log(1/p)$

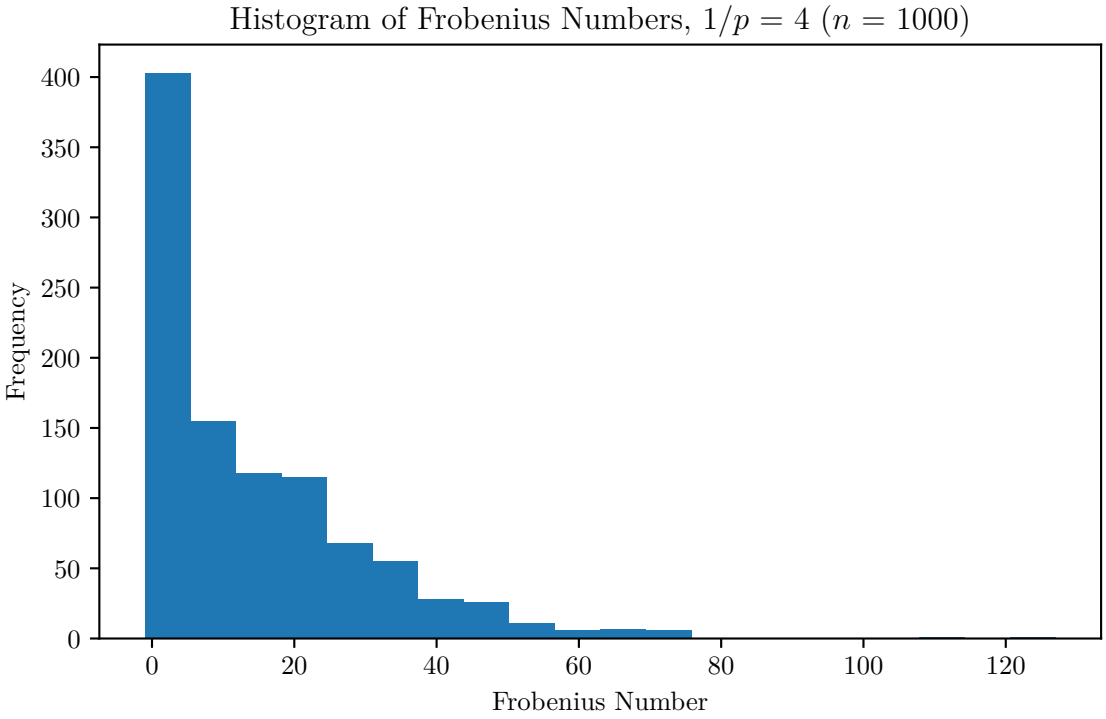

Figure 5.3: Histogram of the Frobenius number of random numerical semigroups generated using the ER-type model, for $1/p = 4$.

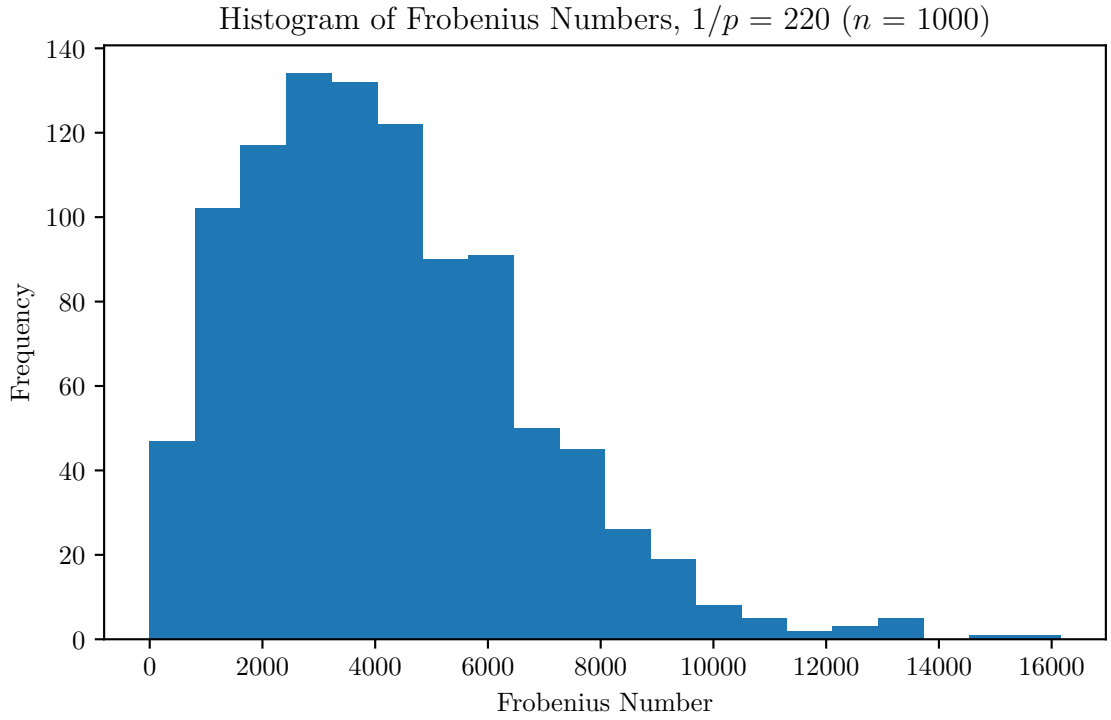

Figure 5.4: Histogram of the Frobenius number of random numerical semigroups generated using the ER-type model, for $1/p = 220$.

We also plot the histograms of the embedding dimension and the Frobenius number for $1/p = 4, 220, 1000$ (Figures 5.3-5.8). These histograms hint that the distribution of the embedding dimension and the Frobenius number of ER-type random numerical semigroups may converge to a known distribution, which is a topic for future research.

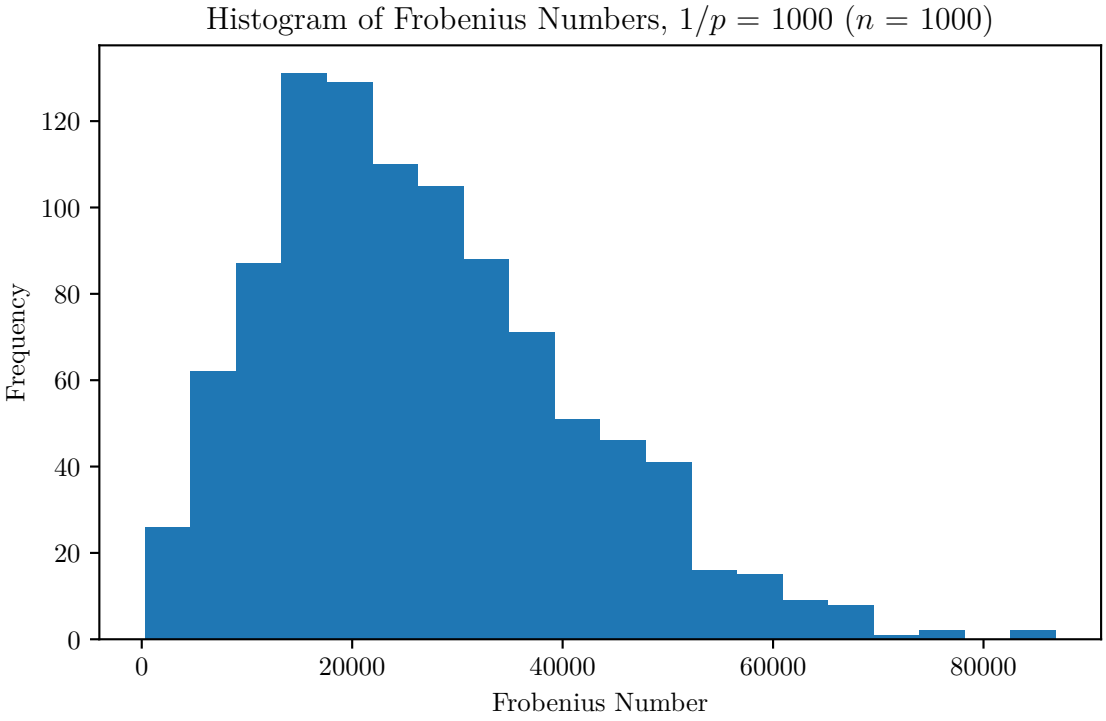

Figure 5.5: Histogram of the Frobenius number of random numerical semigroups generated using the ER-type model, for $1/p = 1000$.

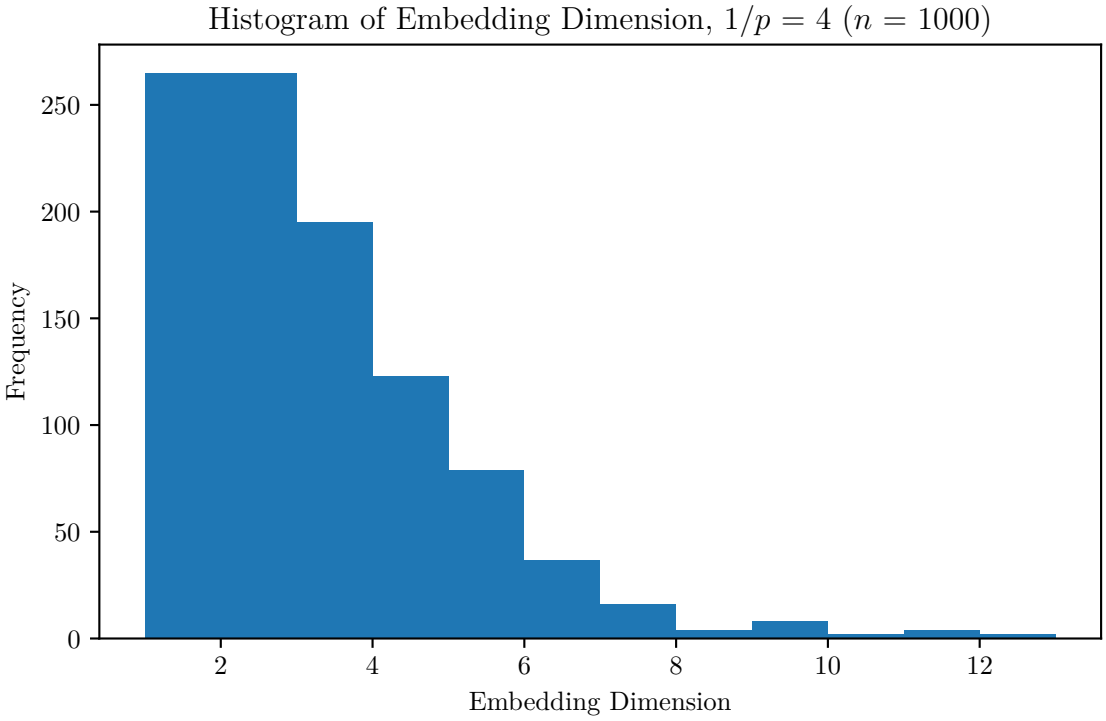

Figure 5.6: Histogram of the embedding dimension of random numerical semigroups generated using the ER-type model, for $1/p = 4$.

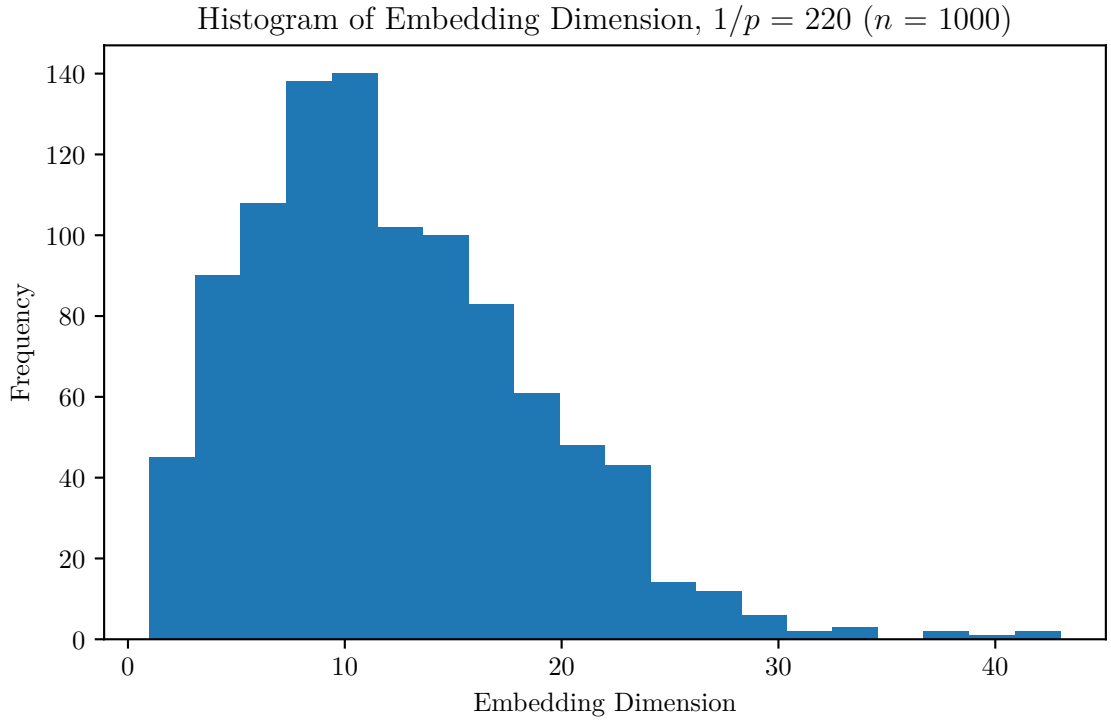

Figure 5.7: Histogram of the embedding dimension of random numerical semigroups generated using the ER-type model, for $1/p = 220$.

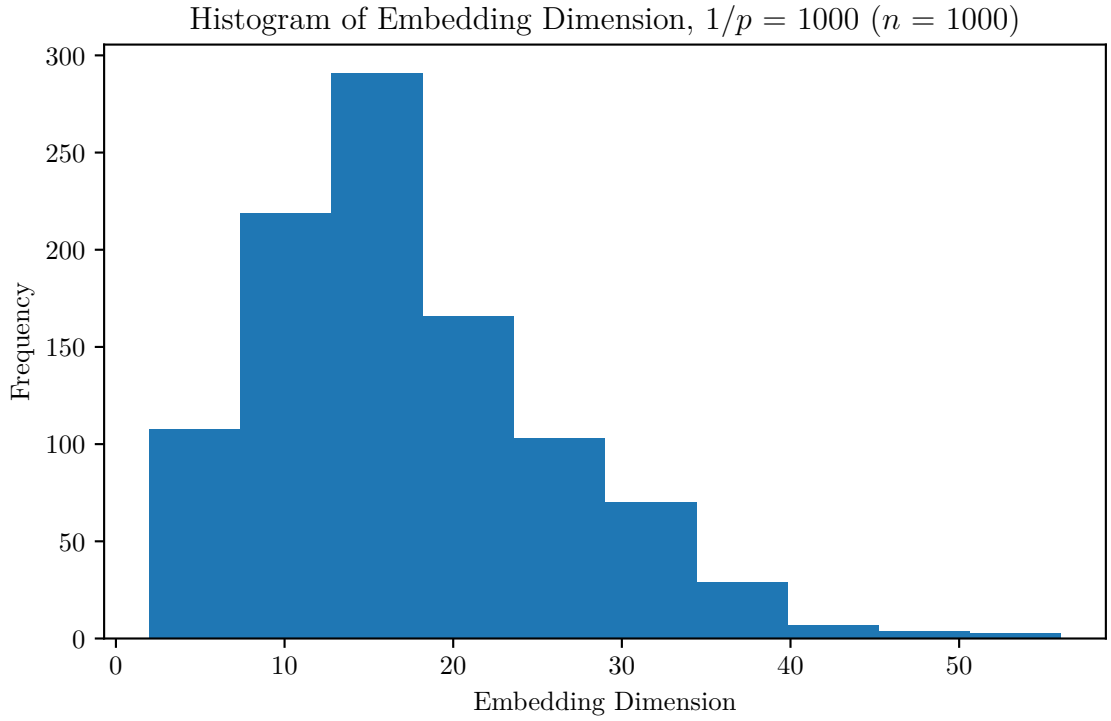

Figure 5.8: Histogram of the embedding dimension of random numerical semigroups generated using the ER-type model, for $1/p = 1000$.

Finally, we show some visualizations of ER-type random numerical semigroups, using the algorithm below for the generation of $\mathrm{Ap}(\mathcal{S}, m(\mathcal{S}))$, with M, p and $m(\mathcal{S})$ as parameters.

```python
    # Function to generate Ap(S, m(S))
    def generate_apery_set(M, p, first):
        # Initialize apery set
        apery_set = [first]

        # Initialize lists (new elements are elements chosen with probability p,
        #  they belong to the minimal generating set)
        generated = []
        new_elements = [first]
        non_apery = []

        # Generate apery set
        for i in range(first + 1, M + 1):
            # Check if a representative of a mod first is already in the apery set
            for a in apery_set:
                if (i - a) % first == 0:
                    non_apery.append(i)
                    break
            # If not add it to the apery set
            else:
                # Check if element can be generated by the apery set
                for a in apery_set:
                    if (i - a) in apery_set:
                        apery_set.append(i)
                        generated.append(i)
                        break
                # If not, add it to the generators with probability p
                else:
                    if random.random() < p:
                        apery_set.append(i)
                        new_elements.append(i)
                # Break when the apery_set is complete
                if len(apery_set) == first:
                    break
        return apery_set, generated, new_elements, non_apery
```

To generate a visualization, run the script `smgps_visualizer.py` found in the repository [6] and introduce a number n. The script will generate a ER-type random numerical semigroup with multiplicity n and $p = 1/n$, and will plot a visualization.

This visualizations are useful to understand the structure of ER-type random numerical semigroups. For example, a number is less likely to be in the minimal generating set as it grows. This is because there are more elements that can be in the span by the minimal generating set as the generation of the numerical semigroup progresses.

A related observation is that there are patterns in the density of the numerical semigroup, in the density of the minimal generating set and in the density of the Apéry set. One could think of this patterns as phase transitions. A topic for further research would be to formalize a definition of this density and to study its behavior.

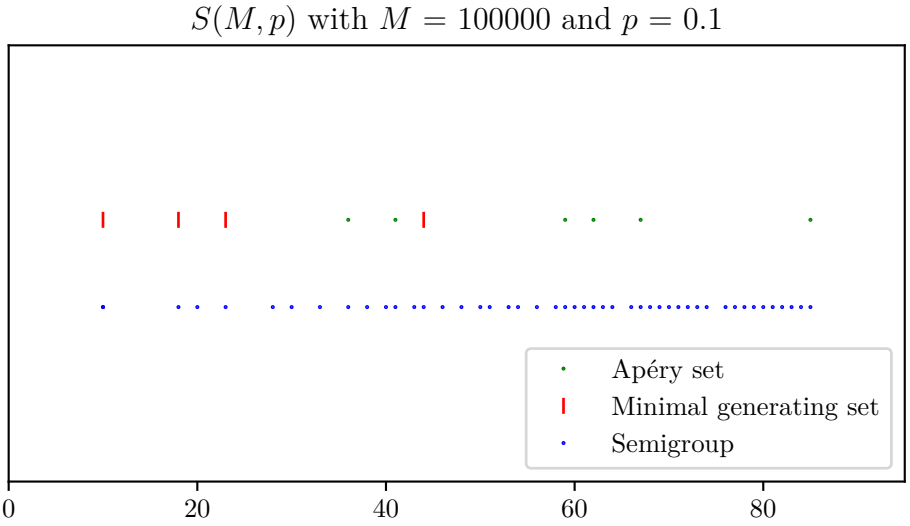

Figure 5.9: ER-type random numerical semigroup, $1/p = 10$.

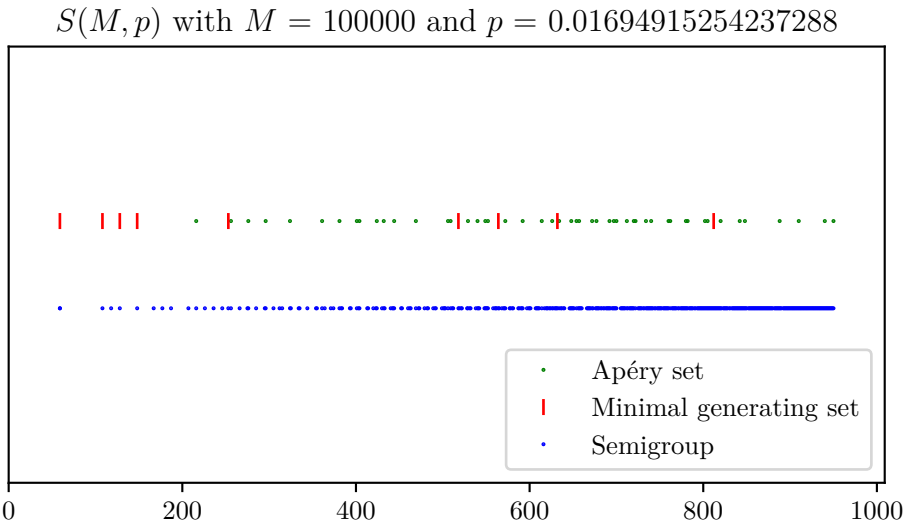

Figure 5.10: ER-type random numerical semigroup, $1/p = 59$.

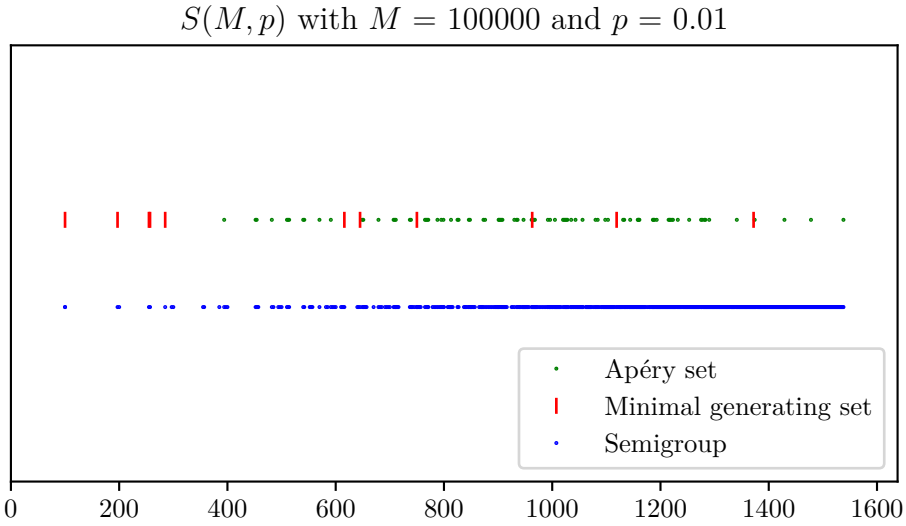

Figure 5.11: ER-type random numerical semigroup, $1/p = 100$.

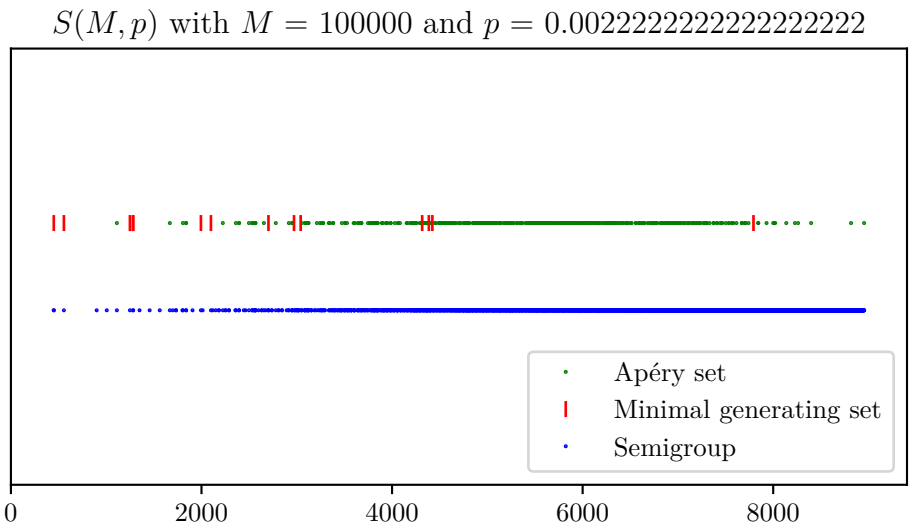

Figure 5.12: ER-type random numerical semigroup, $1/p = 450$.

Chapter 6

Results

6.1 Introduction

In this chapter, we present the main results of this thesis. We will prove a theorem similar to parts (b) and (c) of Theorem 4.2.1 using standard probabilistic arguments.

Theorem 6.1.1. *Let $\mathcal{S} \sim \mathcal{S}(M,p)$, where $p = p(M)$ is a monotone decreasing function of M and $\frac{1}{M} \in o(p(M))$. Then,*

(a) If $\lim_{M\to\infty} p(M) = 0$, then for every $K \in \mathbb{N}$,

$$\lim_{M\to\infty} \Pr[e(\mathcal{S}) > K] = \lim_{M\to\infty} \Pr[g(\mathcal{S}) > K] = \lim_{M\to\infty} \Pr[F(\mathcal{S}) > K] = 1.$$

(b) If $\lim_{M\to\infty} p(M) > 0$, then $e(\mathcal{S})$, $g(\mathcal{S})$ and $F(\mathcal{S})$ are bounded in probability, i.e., for every $\varepsilon > 0$, there exists K_ε such that

$$\Pr[e(\mathcal{S}) < K_\varepsilon] > 1-\varepsilon, \quad \Pr[g(\mathcal{S}) < K_\varepsilon] > 1-\varepsilon \quad \textit{and} \quad \Pr[F(\mathcal{S}) < K_\varepsilon] > 1-\varepsilon.$$

Furthermore, for every φ, ψ such that $(\log x)^2 \in o(\varphi(x))$, $x(\log x)^2 \in o(\psi(x))$,

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

Note that part (a) is *stronger* than part (b) of Theorem 4.2.1. However, part (b) of this theorem does not imply part (c) of Theorem 4.2.1. The proof of part (b) of this theorem is based on Lemma 6.3.1, which is a result on sums of random subsets of cyclic groups.

6.2 Lower Bound

We first prove part (a) of Theorem 6.1.1. We will show that, for each fixed number of generators a, there is a high probability that at least a minimal generators are chosen as $p \to 0$.

Proof. Fix $a \in \mathbb{N}$ such that $a > 11$ and let $T = \{1, \ldots, \lfloor \frac{a}{p} \rfloor\}$. Since $\frac{1}{M} \in o(p)$, 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

$$\Pr[\neg E_1] = \Pr[X_1 > 0] \leq \mathrm{E}[X_1] \leq p \cdot \frac{1}{ap} = \frac{1}{a}. \tag{6.1}$$

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

  Let X_2 be the number of generators selected in T, then $X_2 \sim \mathrm{Bin}(\frac{a}{p}, p)$ and we can use the bound in Proposition A.0.5 with $r = \frac{3a}{2}$ to get that

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

  Also, by the union bound and (6.1),

$$\Pr[E_1 \wedge E_2] \geq 1 - \frac{1}{a} - \frac{6}{a} = 1 - \frac{7}{a}. \tag{6.2}$$

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

  Similarly, we can use the bound for the left tail (Proposition A.0.6) with $r = \frac{a}{2}$ to get that

$$\Pr[\neg E_3] = \Pr\left[X_2 < \frac{a}{2}\right] \leq \frac{(n - \frac{a}{2})p}{(np - \frac{a}{2})^2} = \frac{a - (\frac{a}{2})p}{(\frac{a}{2})^2} \leq \frac{4}{a}. \tag{6.3}$$

- E_4 : The generators selected from T are all minimal.

  Let $A_T = \{Y_{(1)}, Y_{(2)}, \ldots, Y_{(k)}\}$ be the generators selected in T in order. 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 the probability that $b \in T$ is selected as a generator. Using conditional probability, we have that

$$\Pr[b \in \mathcal{A}_T | E_1 \wedge E_2] = \frac{\Pr[(b \in \mathcal{A}_T) \wedge E_1 \wedge E_2]}{\Pr[E_1 \wedge E_2]} \tag{6.4}$$

$$\leq \frac{\Pr[b \in \mathcal{A}_T]}{\Pr[E_1 \wedge E_2]} = \frac{p}{1 - \frac{7}{a}}. \tag{6.5}$$

  Now, if no multiple of $Y_{(1)}$ is selected in T, then $Y_{(2)}$ is minimal. Thus, $Y_{(2)}$ is minimal if $\langle y_1 \rangle \cap \mathcal{A}_T = \{y_1\}$. Also, note that, since the generators are chosen independently,

$$P[b \in \mathcal{A}_T] = P[b \in \mathcal{A}_T | Y_{(1)} = y_1] \quad \text{if} \quad b > y_1.$$

  Since $y_1 \geq \frac{1}{ap}$, we have that

$$|\langle y_1 \rangle \cap \mathcal{A}_T| \leq |\langle y_1 \rangle \cap T| \leq a^2.$$

Then, using the union bound and (6.5),

$$\Pr[Y_{(2)} \text{ is not minimal} | E_1 \wedge E_2 \wedge Y_{(1)} = y_1] \leq \sum_{b \in \langle y_1 \rangle \cap T} \Pr[b \in \mathcal{A}_T | E_1 \wedge E_2 \wedge Y_{(1)} = y_1]$$
$$\leq \sum_{b \in \langle y_1 \rangle \cap T} \frac{p}{1 - \frac{7}{a}} \leq \frac{pa^2}{1 - \frac{7}{a}}.$$

If this bound is independent of y_1, we get that

$$\Pr[Y_{(2)} \text{ is not minimal} | E_1 \wedge E_2] \leq \frac{pa^2}{1 - \frac{7}{a}}. \tag{6.6}$$

Similarly, for $2 \leq t \leq k$ and fixed $Y_{(1)} = y_1, \ldots, Y_{(t-1)} = y_{t-1}$, $Y_{(t)}$ is minimal if $\langle y_1, \ldots, y_{t-1} \rangle \cap \mathcal{A}_T = \{y_1, \ldots, y_{t-1}\}$. To bound $|\langle y_1, \ldots, y_{t-1} \rangle \cap \mathcal{A}_T|$, there are at most a^2 choices for each multiple of y_i, so there are at most $a^{2(t-1)}$ such linear combinations and

$$|\langle y_1, \ldots, y_{t-1} \rangle \cap \mathcal{A}_T| \leq |\langle y_1, \ldots, y_{t-1} \rangle \cap T| \leq a^{2(t-1)}.$$

Also,

$$P[b \in \mathcal{A}_T] = P[b \in \mathcal{A}_T | Y_{(1)} = y_1 \wedge \cdots \wedge Y_{(t-1)} = y_{t-1}] \quad \text{if} \quad b > y_{t-1} > \ldots > y_1,$$

and so

$$\Pr[b \in \mathcal{A}_T | E_1 \wedge E_2 \wedge Y_{(1)} = y_1 \wedge \cdots \wedge Y_{(t-1)} = y_{t-1}] \leq \frac{p}{1 - \frac{7}{a}}.$$

Then, as in (6.6),

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

Therefore, we can use the union bound and $k \leq \frac{3a}{2}$ to conclude that

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

since a is constant and $p \to 0$ as $M \to \infty$. Thus,

$$\Pr[E_4] = \Pr[E_4 | E_1 \wedge E_2] \Pr[E_1 \wedge E_2] \geq 1 - \frac{7}{a} - o(1). \tag{6.7}$$

Finally, note that by the union bound, (6.3) and (6.7),

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

This means that for $\varepsilon > 0$ and sufficiently large a and M, the probability that at least $\frac{a}{2}$ minimal generators are selected is greater than $1 - \varepsilon$. Threfore, for every $K \in \mathbb{N}$,

$$\lim_{M \to \infty} \Pr[e(\mathcal{S}) > K] = 1.$$

Now, using Corollary 3.2.1 and Proposition 3.2.3, we have that

$$e(S) \leq m(S) \leq g(S) + 1 \leq F(S) + 1.$$

We conclude that, for every $K \in \mathbb{N}$,

$$\lim_{M \to \infty} \Pr[g(\mathcal{S}) > K] = \lim_{M \to \infty} \Pr[F(\mathcal{S}) > K] = 1. \quad \square$$

6.3 Upper bound

Before proving part (b) of theorem 6.1.1, 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.

Lemma 6.3.1. *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

$$E[X_z] = \frac{1}{q}\binom{s}{k}. \tag{6.8}$$

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

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

This implies that, using 6.8,

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

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

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

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

Hence, by (2.6) and Theorem 2.3.2,

$$\begin{aligned}
\Pr[X_z = 0] \leq \frac{\mathrm{E}[X_z] + \Delta}{\mathrm{E}[X_z]^2} &= \frac{1}{E[X_z]} + \sum_{j=0}^{k}\frac{\Delta_j}{\mathrm{E}[X_z]^2} \\
&\leq \frac{1}{E[X_z]} + \frac{(k+1)}{q^2} = \frac{1}{E[X_z]} + \frac{2\lfloor 3\log_2 q\rfloor + 1}{q^2}.
\end{aligned}$$

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. □

6.3.1 Proof of the upper bound

Lemma 6.3.2. *Let $\psi(x)$ be a function for which $x(\log x)^2 \in o(\psi(x))$ .Then*

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

Since $h(x)$ is a function for which $x(\log x) \in o(h(x))$, we have that

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 6.3.1 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 6.3.1 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 [22, Theorem 8],

$$q_n \sim n \log n. \tag{6.9}$$

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,

$$\lim_{p\to 0} \Pr[E_1] = 1. \tag{6.10}$$

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 6.3.1. 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

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

Thus, bounding by the worst case asymptotically,

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

We conclude that

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

and so, using (6.10),

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

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 6.3.1 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

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

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 6.3.1. *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 3.2.3. □

Corollary 6.3.2. *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

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

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

$$\lim_{p\to 0} \Pr\left[e(\mathcal{S}) \leq (3p)\psi\left(\frac{1}{p}\right)\right] = 1.$$

Thus, if $\varphi(x) = \frac{3}{x}\psi(x)$, then $(\log x)^2 \in \varphi(x)$ and

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

Bibliography

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

[2] 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.

[3] C. O'Neill, *Numsgps-sage*, `https://github.com/coneill-math/numsgps-sage`, 2013.

[4] M. Delgado, P. Garcıa-Sánchez, and J. Morais, "Numericalsgps," *A GAP package for numerical semigroups. Available via http://www. gap-system. org*, 2015.

[5] M. Delgado, "Intpic," *a GAP package for drawing integers, Available via http://www. fc. up. pt/cmup/mdelgado/software*, 2013.

[6] S. Morales, *Randnumsgps*, `https://github.com/smoralesduarte/randnumsgps`, 2023.

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

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

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

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

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

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

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

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

[15] 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.

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

[17] 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.

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

[19] 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.
[20] V. I. Arnold, *Arnold's problems*. Springer, 2004.
[21] R. P. Stanley, *Combinatorics and commutative algebra*. Springer Science & Business Media, 2007, vol. 41.
[22] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*. Oxford university press, 1979.
[23] W. Feller, *An introduction to probability theory and its applications*. John Wiley & Sons, 1971, vol. 1.

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 [23, 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* $r < np$*,*

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