Random numerical semigroups and iterated sumsets modulo p

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
October 29, 2023

Abstract

This is the text of the abstract, providing a brief summary of the context, research problem, main contributions and conclusions of this work.

Dedication

For my mother. (name of person - this is optional!)

Acknowledgements

I would like to thank (supervisor, family, research collaborators, anyone else who significantly helped you with this work).

Santiago Morales Duarte
October 29, 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 **??** provides a comprehensive review of literature which is relevant to the overall aim. This includes ...

Chapter 4 aims to ...

This chapter resulted in the following publications:

- D. R. Franklin and K. J. Wilson, "A LaTeX Thesis Template for the School of Electrical and Data Engineering," *IEEE Transactions on LaTeX Thesis Templates*, vol. 1, no. 1, Oct. 2021
- D. R. Franklin and K. J. Wilson, "A LaTeX Thesis Template for the School of Electrical and Data Engineering," *IEEE Transactions on LaTeX Thesis Templates*, vol. 1, no. 1, Oct. 2021

Chapter 5 aims to ...

This chapter resulted in the following publications:

- D. R. Franklin and K. J. Wilson, "A LaTeX Thesis Template for the School of Electrical and Data Engineering," *IEEE Transactions on LaTeX Thesis Templates*, vol. 1, no. 1, Oct. 2021

Chapter 6 aims to ...

This chapter resulted in the following publications:

- D. R. Franklin and K. J. Wilson, “A LaTeX Thesis Template for the School of Electrical and Data Engineering,” *IEEE Transactions on LaTeX Thesis Templates*, vol. 1, no. 1, Oct. 2021

Finally, Chapter 7 summarises the results and implications of this work, and provides recommended directions for continuation of this work in the future.

1.1.1 Additional Research Contributions

A number of additional research publications and presentations are listed below:

- xxx

Test [1]

Chapter 2

The Probabilistic Method

2.1 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, by showing that the probability that a randomly chosen object has that property is greater than zero. The method requires an appropiate 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 $E(T)$. [2]

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 $s \in K$ [2]. 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: is there a tournament with property S_k for every 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*.

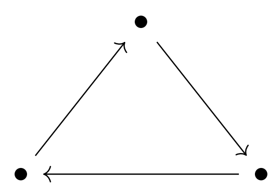

Figure 2.1: A tournament on 3 vertices with property S_1.

The main idea is to show that there is a 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. *For every $k \in \mathbb{N}$, there is a tournament with property S_k. [2]*

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 any 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, this 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 one if

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

which holds if and only if

$$\frac{n}{\log n} > k2^k.$$

Now, if $n = k^2 2^k \log 2$, then

$$\lim_{n\to\infty} \frac{n}{k2^k \log n} = \lim_{k\to\infty} \frac{k\log 2}{2\log k + k\log 2 + \log\log 2} = 1.$$

Hence, we conclude that there exists $n = (\log 2)k^2 2^k(1 + o(1))$, such that 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. $\square$

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, this will not take too long.

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

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

We show that the expected value is linear.

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

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

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

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

This result is known as the *linearity of expectation.*

Example 2.2.1. [2] 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, $E[X_i] = \frac{1}{n}$ 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,

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

Note that there is a point in the probability space $x \geq E[X]$ such that $\Pr[X = x] > 0$, and there is a point $x \leq E[X]$ 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 [2].

Theorem 2.2.1. *There is a tournament with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths. [2]*

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,

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

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 $E[X]$. Then, the variance of X is defined as

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

By the linearity of expectation,

$$E[(X - E[X])^2] = E[X] - 2E[XE[X]] + E[X]^2 = E[X^2] - E[X]^2.$$

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

Theorem 2.3.1 (Chebyschev's inequality)**.** *[2] For $\lambda \geq 0$,*

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

Proof.

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

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

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

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

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

and we have that

$$\operatorname{Var}[X] \le E[X] + \sum_{i \ne j} \operatorname{Cov}[X_i, X_j]. \tag{2.3}$$

«««< HEAD If X only takes nonnegative integer values, we can bound $\Pr[X = 0]$ ======= Suppose that X only takes nonnegative integer values, we are interested in bounding $\Pr[X = 0]$. First, note that

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

Thus, if $E[X] \to 0$, then $X = 0$ almost always. On the other hand, if $E[X] \to \infty$ does not necessarily imply 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, $E[X] = \infty$ but $X = 0$ with probability $\frac{1}{2}$. In some cases, we can use the second moment method to show that if $E[X] \to \infty$ and we have more information about $\operatorname{Var}[X]$, then $X > 0$ almost always.

»»»> 6bea5c627ef1f7570a7203298c5128bebb69a164

Theorem 2.3.2. *$Pr[X = 0] \le \frac{Var[X]}{E[X]^2}$.[2]*

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

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

In asymptotic terms,

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

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

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

Thus, we have a tighter result:

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

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 $i \ne j$, 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 $E[X] \to \infty$ and $\Delta = o(E[X^2])$, then $X > 0$ almost always. Also, $X \sim E[X]$ almost always. [2]*

Proof. When $i \sim j$,

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

and so

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

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

2.4 Threshold Functions

Let $n \in \mathbb{N}$ and $0 \leq p \leq 1$. The random graph $G(n, p)$ 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 [2]. 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 2.4.1. $r(n)$ 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. [2]

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

2.4.1 Threshold function for having isolated vertices

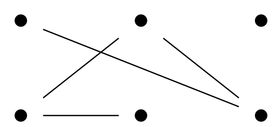

Figure 2.2: A graph with an isolated vertex.

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

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

Let $p = k\dfrac{\ln n}{n}$ for $k \in \mathbb{R}_{>0}$. Then

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

There are two cases:

1. If $k > 1$, $\lim_{n\to\infty} E[X] = 0$. Since $E[X] \geq \Pr[X > 0]$, 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 $\lim_{n\to\infty} E[X] = \infty$ 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 $\mathrm{Var}[X] = o(E[X]^2)$. First,

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

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

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

   We conclude that $\mathrm{Var}[X] \in o(E[X]^2)$ 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 $p(n) \in o(r(n))$, $X \sim E[X]$ almost always as n tends to infinity.

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

We could have used Corollary 2.3.3 since we are dealing with a sum of indicator variables. However, since we could show 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.

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.

Definition 3.1.1. [3] A *numerical semigroup* is a subset $S \subseteq \mathbb{N}_0$ 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 $\mathbb{N}_0$.

Examples of numerical semigroups include $\mathbb{N}_0$,

3.2 Invariants

Definition 3.2.1 (Multiplicity)**.**

Definition 3.2.2 (Embedding Dimension)**.**

Definition 3.2.3 (Apéry Set)**.**

Definition 3.2.4 (Frobenius Number)**.**

Definition 3.2.5 (Genus)**.**

3.3 Wilf's Conjecture

Etc. etc.

Chapter 4

Random Numerical Semigroups

4.1 Box Model

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

4.1.1 Results

4.1.2 Subtopic C

4.2 ER-type model

We generate a random numerical semigroup with a model similar to the Ërdos-Renyi model for random graphs.

Definition 4.2.1. For $p \in [0, 1]$ and $M \in \mathbb{N}$, a 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.

4.3 Downward model

Etc. etc.

4.4 Conclusion

Chapter 5

Experiments

5.1 ER-type model experiments

In this Chapter, XXX is presented. Include pseudocode.

Section 5.2 discusses Theme 1. Section 5.3 discusses Theme 2....

5.2 Downward model experiments

5.2.1 Subtopic A

5.2.2 Subtopic B

5.2.3 Subtopic C

5.3 Theme 2

Etc. etc.

Chapter 6

Results

6.1 Introduction

In this Chapter, XXX is presented.

Section 6.2 discusses Theme 1. Section 6.4 discusses Theme 2....

6.2 Lower Bound

6.3 Expected Frobenius Number

We prove a Theorem found in [4] without the use of the simplicial complex.

Theorem 6.3.1. *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*

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

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

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

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

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

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

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

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

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

- 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 $b \in \{2y_1, 3y_1, \ldots, c_1y_1\}$ is selected, where c_1y_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

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

.

6.3.1 Subtopic A

6.3.2 Subtopic B

6.3.3 Subtopic C

6.4 Lower Bound result

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

6.4.1 Lemma

- 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 $E[X_z]$. The first one uses that, for any $K \subset N_z^k$,

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

and so we get that

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

This motivates the second way, for we know that

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

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

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

Therefore, we also find that

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

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 (1),

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

Let $s = 4\lfloor 3\log_2 q\rfloor$ and $k = 2\lfloor 3\log_2 q\rfloor$, where $\alpha \in (0,1)$. Using that $\binom{s-k}{t}$ is maximized at $t = \lfloor (s-k)/2\rfloor$,

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

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

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

Therefore, by the union bound,

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

6.5 Theorem

- 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,

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

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

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

6.6 Conclusion

Chapter 7

Conclusions and Future Work

7.1 Summary of Outcomes

7.2 Recommendations & Future Work

7.3 Concluding Remarks

In summary, ...

Bibliography

[1] D. R. Franklin and K. J. Wilson, “A LaTeX Thesis Template for the School of Electrical and Data Engineering,” *IEEE Transactions on LaTeX Thesis Templates*, vol. 1, no. 1, Oct. 2021.

[2] N. Alon and J. H. Spencer, *The probabilistic method.* John Wiley & Sons, 2016.

[3] J. C. Rosales, P. A. García-Sánchez, *et al.*, *Numerical semigroups.* Springer, 2009, vol. 20.

[4] J. De Loera, C. O’Neill, and D. Wilburne, “Random numerical semigroups and a simplicial complex of irreducible semigroups,” *arXiv preprint arXiv:1710.00979*, 2017.

[5] K. Frankston, J. Kahn, B. Narayanan, and J. Park, “Thresholds versus fractional expectation-thresholds,” *Annals of Mathematics*, vol. 194, no. 2, pp. 475–495, 2021.

[6] J. Park and H. T. Pham, “A proof of the kahn-kalai conjecture,” *arXiv e-prints*, arXiv–2203, 2022.

[7] S. Chapman, R. Garcia, and C. O’Neill, “Beyond coins, stamps, and chicken mcnuggets: An invitation to numerical semigroups,” *A Project-Based Guide to Undergraduate Research in Mathematics: Starting and Sustaining Accessible Undergraduate Research*, pp. 177–202, 2020.

[8] M. Delgado, “Conjecture of wilf: A survey,” *Numerical Semigroups: IMNS 2018*, pp. 39–62, 2020.

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.

Proposition A.1.1. $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ *for* $1 \leq k \leq n$.

Proposition A.1.2. $\left(\frac{n}{k}\right)^k \leq \binom{n}{k}$ *for* $1 \leq k \leq n$.

Proposition A.1.3. $(1-p) \leq e^{-p}$ *for* $0 \leq p \leq 1$.

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

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

Proposition A.1.4. $\binom{n}{k}^2 \leq \binom{2n}{2k}$ *for* $n \geq 1$.

Proof. We have that

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:

`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