On Sufficient Conditions for the Existence of Twin Values in Sieves over the Natural Numbers

by

Luke Szramowski

Submitted in Partial Fulfillment of the Requirements

For the Degree of

Masters of Science

in the

Mathematics

Program

Youngstown State University

May, 2020

On Sufficient Conditions for the Existence of Twin Values in Sieves over the Natural Numbers

Luke Szramowski

I hereby release this thesis to the public. I understand that this thesis will be made available from the OhioLINK ETD Center and the Maag Library Circulation Desk for public access. I also authorize the University or other individuals to make copies of this thesis as needed for scholarly research.

Signature: ______________________________

Luke Szramowski, Student   Date

Approvals: ______________________________

Dr. Eric Wingler, Thesis Advisor   Date

____

Dr. Thomas Wakefield, Committee Member   Date

____

Dr. Thomas Madsen, Committee Member   Date

____

Dr. Salvador A. Sanders, Dean of Graduate Studies   Date

Abstract

For many years, a major question in sieve theory has been determining whether or not a sieve produces infinitely many values which are exactly two apart. In this paper, we will discuss a new result in sieve theory, which will give sufficient conditions for the existence of values which are exactly two apart. We will also show a direct application of this theorem on an existing sieve as well as detailing attempts to apply the theorem to the Sieve of Eratosthenes.

Contents

1 Introduction

Sieve theory is a powerful tool that is used to analyze sets of numbers produced via a removal game. Start with a set of numbers and removed values which satisfy a certain set of criteria to produce a new set. If we continue this procedure for some time, we will be left with the sieved set at the end.

In general, it is a difficult problem determining the behavior of the sieved set; a classical example of this is the set of prime numbers, which are notoriously difficult to analyze. When sieving over the natural numbers, there are a variety of questions which can be asked about the sieved set, the most famous of which is; will the sieved set contain infinitely many values which are only 2 apart? In Section 4 of this paper, we will give sufficient conditions which, if met, will guarantee the existence of these so called twin values in a general way. The material in this thesis is original work and all definitions, unless otherwise stated, are likewise original.

2 Preliminary Material

Definition 2.1 *Admissible Sequence*

A sequence of subsets of the natural numbers $\{R_i\}_{i=0}^{\infty}$ *is called a admissible sequence if and only if:*

$$R_i \cap R_j = \emptyset \iff i \neq j$$

for all $i, j \in \mathbb{N}$.

The definition of the admissable sequence is important because later in the paper it will help with various counting arguments, since having an empty intersection will guarantee that no "double counting" of values will occur.

Definition 2.2 *General Natural Sieve*

Let S be a subset of the natural numbers. S is called the sieve set with respect to $\{R_i\}_{i=1}^{\infty}$ if:

1. $\{R_i\}_{i=0}^{\infty}$ is an admissible sequence.

2. There exists a sequence of sets $\{S_i\}_{i=0}^{\infty}$ where $S_0 = \mathbb{N}$ and $R_n \subseteq S_n$ for each n.

3. If $S_{n+1} = S_n - R_n$ then;

$$S = \lim_{n\to\infty} S_n$$

In the above definition, we will refer to the R_n sets as *removal sets*, and we will refer to the S_n sets as the *iteration* sets. For example, we may refer to the 3^{rd} iteration of the sieve, in which case we are referring to the set S_3.

Now, for the sake of convenience, we would like to index the values contained in the removal sets and the iteration sets. We will use the following notation:

$$S_n = \{s(1,n), s(2,n), ...\}$$

where $i < j \to s(i,n) < s(j,n)$. Similarly;

$$R_n = \{r(1,n), r(2,n)...\}$$

where $i < j \rightarrow r(i,n) < r(j,n)$. By the definition of our sieve process, we have that $R_n \subseteq S_n$. This leads us to our next definition.

Definition 2.3 *The Locator Function*

Let S be a sieved set with respect to $\{R_i\}_{i=0}^{\infty}$. For each $n \in \mathbb{N} \cup \{0\}$, define the function:

$$\Psi_n : R_n \rightarrow \mathbb{N}$$

where:

$$\Psi_n(r(i,n)) = j \;\Rightarrow\; r(i,n) = s(j,n)$$

The Locator Function helps us determine the location of the removed values relative to the iteration of the sieve.

Definition 2.4 *Order Invariant Sieves*

Let S be a sieved set with respect to $\{R_i\}_{i=0}^{\infty}$. This sieved set is called order invariant if, for any bijective function $f : \mathbb{N} \cup \{0\} \rightarrow \mathbb{N} \cup \{0\}$, if $\bar{S}$ is the sieved set with respect to $\{R_{f(i)}\}_{i=0}^{\infty}$, then $\bar{S} = S$.

We define the Order Invariance property because we will come across examples of sieves which are defined in such a way that if the order of removal is altered, the end result is not necessarily the same. This causes a natural question to arise; does changing the order of the removal sets change the core properties of the end sieve? Does this change necessarily removed the infinitude of certain values or are these properties preserved upon permuting the order of the removal values?

Definition 2.5 *Sieve Equivalence*

Let S be the sieved set with respect to $\{R_i\}_{i=0}^{\infty}$ and let T be the sieved set with respect to $\{P_i\}_{i=0}^{\infty}$. We say that the sieve process of S and T are equivalent if it is the case that $S = T$.

In some cases it may be advantageous to study one form of a sieve over another. Historically, there are multiple sieves which produce prime numbers, each sieve with very different processes. Given that the sieve processes are equivalent, any results we derive from one formulation of a sieve will apply to the others, which will increase our arsenal of results to tackle a particular problem.

With these definitions set, let us look at a handful of examples of sieves; some of which are commonly known and others which are less so.

3 Sieves

In this section, we will detail the various sieves which we will come across. Some of the sieves live in relative obscurity, while others are relatively well known. Regardless, this section serves a dual purpose; it allows introduction of the sieves as well as giving us the opportunity to reword the sieves in terms of the definitions given in Section 2.

3.1 The Sieve of Eratosthenes

The Sieve of Eratosthenes, also known as the Prime Number sieve, is the process by which the prime numbers can be produced from the set of natural numbers. The way this is done is by listing out the natural numbers and procedurally removing all of the composite numbers, with the end result being a set which contains all primes numbers and 1. The way this is typically seen is as follows:

1. Begin by listing out all of the natural numbers:

$$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14...$$

2. Determine the smallest number greater than 1, which in this case is 2, and strike out all numbers on the list which are a positive integer multiple of 2 away from 2. This gives us:

$$1\ 2\ 3\ \cancel{4}\ 5\ \cancel{6}\ 7\ \cancel{8}\ 9\ \cancel{10}\ 11\ \cancel{12}\ 13\ \cancel{14}...$$

3. Now determine the smallest number greater than 2 which has not been struck out. In this case, it is 3. Now in the new list, strike out all values which are a positive integer

multiple 3 away from 3. This gives us:

$$1\ 2\ 3\ \cancel{4}\ 5\ \cancel{6}\ 7\ \cancel{8}\ \cancel{9}\ \cancel{10}\ 11\ \cancel{12}\ 13\ \cancel{14}\ldots$$

4. From here on, we continue the process by picking the smallest number not struck out by the sieve and larger than the number previously used. Doing this indefinitely, we will be left with the set of prime numbers and 1.

We can see that while the above is a fine example of a sieve, it does not fit our definition of a General Natural Sieve. To begin, the classical statement of the Sieve of Eratosthenes allows for the same value to be removed *multiple* times. We can see that in the first iteration of the sieve, the number 6 is struck out, but it is yet again struck out on the second iteration of the sieve.

As we wish to stay uniform in our definitions, let us rephrase the Sieve of Eratosthenes with the following definition:

Definition 3.1 *The Adjusted Sieve of Eratosthenes*

Define the sequence of removal sets as follows:

$$R_n = \{r \in S_n : p_{n+1} \mid r\ ,\ r > p_{n+1}\}$$

with $S_0 = \mathbb{N}$. *Then the set* S *which is sieved with respect to* $\{R_i\}_{i=0}^{\infty}$, *is equal to* $\mathbb{P} \cup \{1\}$, *the set of prime numbers with 1. We will refer to the iteration sets as* E_n *and the removal sets as* R_n *throughout this paper.*

While this is not a drastic change, the Adjusted Sieve is now a general number sieve, which will help us in our later analysis. Even the above Adjusted Sieve has its shortcomings; any analysis of the removal sets is hampered by their abnormal structure. For example, in the classical version of the sieve, we know exactly where the values which are struck out will be; we simply remove the multiples of the chosen *removal* value. In this new sieve, we do not have that same process. Since the removal sets are disjoint, we cannot simply remove multiples of some number starting at some point. Let us look at the some iterations of the adjusted sieve:

Iteration 1

Start with the set of natural numbers:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 ...

Now we will be removing all multiples of 2 greater than two:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 ...

In this case, all the values removed will be exactly a multiple of 2 away from 2.

Iteration 2

Start with the ending set from iteration one:

1 2 3 5 7 9 11 13 15 17 19 21 23 25 ...

Now we will remove all multiples of 3 greater than 3:

1 2 3 5 7 9 11 13 15 17 19 21 23 25 ...

We can notice that again, in this case, all the values removed will be a multiple of 3 away from 3. This changes with the next iteration.

Iteration 3

Begin with the ending set from iteration 2:

1 2 3 5 7 11 13 17 19 23 25 29 31 35 ...

We will remove all multiples of 5 greater than 5:

1 2 3 5 7 11 13 17 19 23 25 29 31 35 ...

We can see that in this case that the first removed value 25, is actually seven terms away from 5, and 35 is 10 terms away and so on. The gaps between the removed values is not uniformly 5, which is a shift from the previous examples.

This trend continues and only gets worse and more chaotic. The position of the removed values are not constant anymore and this will pose some difficulty in future analysis. In order to combat this, we can inspect different variations of this sieve which behave in a *similar* way.

3.2 The Block Sieve

Definition 3.2 *The Block Sieve*

Define R_n*, for each non-negative integer* n*, as:*

$$R_n = \{r \in \mathbb{N} : \gcd(r, p(n)) = 1 \ , \ p_{n+1}|r\}$$

where $p(n) = \prod_{i=1}^{n} p_i$ *and* $p(0) = 1$*.*

For this sieve, we define the iteration sets as T_n *and the removal sets as* Y_n *throughout this paper.*

We can see that the sieved set generated by the Block Sieve is simply the set $\{1\}$. It seems like the Block Sieve and the Sieve of Eratosthenes are generated by almost the same removal sets, however, we are removing the prime number in question as well.

This sieved set is of interest to us, not because of the end result, but rather because of the behavior of each of the iterations of the sieve. We can see that, when comparing the iterations of the Adjusted Sieve of Eratosthenes and those of the Block Sieve:

$$T_n \cup \{p_1, p_2, ..., p_n\} = E_n$$

And;

$$Y_n = R_n - \{p_1, p_2, ..., p_{n+1}\}$$

This is to say that the sieves on the same iteration will only differ by the prime values whose multiples have already been sieved out. Because of the above relation, it may be advantageous to study the Block Sieve, as the absence of these prime values allows us to exploit a special underlying structure.

Proposition 3.1 *The Structure of T_n*

For any $n \in \mathbb{N} \cup \{0\}$;

$$T_n = \bigcup_{i=0}^{\infty} (i \cdot p(n) + U(p(n)))$$

where $U(p(n))$ is the set of units modulo $p(n)$.

This proposition shows us that each iteration of the Block Sieve produces a series of unit sets *stacked* on top of each other across the natural numbers. We can similarly characterize the *removal* sets:

Proposition 3.2 *The Structure of Y_n*

For any $n \in \mathbb{N} \cup \{0\}$;

$$Y_n = p_{n+1} \cdot T_n$$

[The proof for this should likewise be clear.]

With this, we can see that the sieved sets are recursively generated. Indeed, with knowledge of the prior set alone, we are able to generate the next set. We can see that with this set up, we do not even need to magically "know" the next prime; if we take the $\min(E_n - \{1\})$, that value is guaranteed to be our next prime number.

Because of the prior two propositions, we can see that the structure of the block sieve can be condensed down. Let us look at an example first:

For a non-trivial example, let us look at T_3. By Proposition 2.1;

$$T_3 = \bigcup_{i=0}^{\infty}(i \cdot p(3) + U(p(3))) = \bigcup_{i=0}^{\infty}(i \cdot 30 + \{1, 7, 11, 13, 17, 19, 23, 29\})$$

By organizing the values in this set in a particular way, certain peculiar patterns will become evident. These patterns are consistent and will allow us to generate yet another variation of the sieve, which will allow for easier inspection. We organize them by generating an array of values, which has infinitely many rows and finitely many columns. The j^{th} row will be the entire set $j \cdot p(3) + U(p(3))$. Below is a concrete example of the Block Sieve, where the values which are removed are contained in boxes.

1	$\boxed{7}$	11	13	17	19	23	29
31	37	41	43	47	$\boxed{49}$	53	59
61	67	71	73	$\boxed{77}$	79	83	89
$\boxed{91}$	97	101	103	107	109	113	$\boxed{119}$
121	127	131	$\boxed{133}$	137	139	143	149
151	157	$\boxed{161}$	163	167	169	173	179
181	187	191	193	197	199	$\boxed{203}$	209
211	$\boxed{217}$	221	223	227	229	233	239
241	247	251	253	257	$\boxed{259}$	263	269
271	277	281	283	$\boxed{287}$	289	293	299
$\boxed{301}$	307	311	313	317	319	323	$\boxed{329}$
331	337	341	$\boxed{343}$	347	349	353	359
361	367	$\boxed{371}$	373	377	379	383	389
391	397	401	403	407	409	$\boxed{413}$	419
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

We can notice that if we break up the array into blocks of 7 rows, that the values which are removed are done so in the same positions relative to those blocks. With this pattern, we can produce yet another variation of the sieve.

3.3 The Finite Block Sieve

From the previous sieve, we can see that the distances between the removed values in each “block” is the same. Due to this, we really only need to consider a single block from the prior sieve. This sieve process will be different from the others, in the sense that it will be a constructive sieve process: when we complete an iteration of the sieve, we will construct a new set out of the produced set and sieve over that. It will be constructed as follows:

Definition 3.3 *The Finite Block Sieve*

Let us initially define $F_0 = \{1\}$. We will produce F_{n+1} from the set F_n by:

$$F_{n+1} = \bigcup_{i=0}^{p_{n+1}} (i \cdot p(n) + F_n) - p_{n+1} \cdot F_n$$

This follows directly from truncating the Block Sieve definition. Iterations of this sieve can be found in Section 5 at the end of this paper. Due to its finite nature, analysis is significantly easier than in the finite case.

This sieve is important to study as it is directly related to the ordinary Block Sieve, and by extension, directly related to the Adjusted Sieve of Eratosthenes.

3.4 The Sieve of Joseph Flavius

The Sieve of Joseph Flavius, more commonly referred to as the Lucky Number sieve, is defined as follows:

1. Begin with the set of natural numbers.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 ...

2. Identify the smallest number greater than 1, in this case, 2, and removed every element which is a multiple of 2 away from 1. This leaves us with the set:

$$1 \quad 3 \quad 5 \quad 7 \quad 9 \quad 11 \quad 13 \quad 15 \quad 17 \quad 19 \quad 21 \quad 23 \quad 25 \quad 27 \quad 29 \quad ...$$

3. Again, we now identify the smallest element in the set greater than 2, which in this case is 3, and we will removed every element that is a multiple of 3 away from 1. This leaves us with the set:

$$1 \quad 3 \quad 7 \quad 9 \quad 13 \quad 15 \quad 19 \quad 21 \quad 25 \quad 27 \quad 31 \quad 33 \quad 37 \quad 39 \quad 43 \quad ...$$

We continue this processes moving forward and doing so will result in the set of Lucky Numbers.

Let us formalize this process in terms of our definition of a General Number Sieve:

Definition 3.4 *The Sieve of Joseph Flavius*

Define R_n, for each non-negative integer n, as:

$$R_n = \{s \in S_n \, : \, \ell_n \,|\, |S_n|_s|\}$$

Where $S_n|_n = \{k \in S_n : k \leq s\}$ and ℓ_n is the n^{th} lucky number and $\ell_0 = 2$.

This is probably one of the less well known explicit sieves, yet it is studied due to its connection to the Sieve of Eratosthenes. We can see that the processes are extraordinarily similar: both rely on choosing the smallest element in the set greater than 1 which was not chosen before, and removing values which are a multiple of that chosen value away from some starting point. Because of the similarities of the sieve processes, it is suspected that the sieved sets share many essential properties in common.

For example, in the paper *The Lucky Number Theorem* by *Hawkins and Briggs*, they were able to prove that if ℓ_n is the n^{th} Lucky Number, then:

$$\ell_n \approx n \cdot \ln(n)$$

We can recall that this is the same statement as the Prime Number Theorem, that the n^{th} prime number, p_n is approximately $n \cdot \ln(n)$. This result shows that the values that survive the sieve are approximately the same size.

It was further pointed out in the paper *On Certain Sequences of Integers Defined By Sieves* by *Gardnerer, Lazurus, Metropolis and Ulam* that gaps between consecutive prime numbers and consecutive lucky numbers are approximately the same. It also appears that beneath a given natural number, the number of twin primes and twin lucky numbers are approximately the same. While it is not necessarily the case that the infinitude of lucky twins would imply the infinitude of twin primes, it would certainly lend some empirical evidence to the assertion.

4 Analysis of Twin Values

Theorem 4.1 *The Twin Sieve Theorem*

Let S be a sieved set with respect to $\{R_i\}_{i=0}^{\infty}$. Suppose the following hold:

1. There exists an S_m such that, for all but finitely many values, if $s \in S_m$, then $s - 2 \in S_m$ or $s + 2 \in S_m$.

2. There exists a sequence of natural numbers $\{c_i\}_{i=m}^{\infty}$, such that:

i. $3 \leq c_m \leq c_{m+1} \leq c_{m+2} \leq \ldots$

ii. $\lim_{n\to\infty} c_n = \infty$

iii. For all $y \geq m$, and for all $r(k, y) \in R_y$;

$$\Psi_y(r(k, y)) \geq c_y \cdot k$$

Then it is the case that there are infinitely many elements $s \in S$ such that $s + 2 \in S$.

Proof

Let S be the sieved set with respect to $\{R_i\}_{i=0}^{\infty}$, and let us all assume all hypotheses. By Hypothesis 1, there exists a set S_m such that for all but finitely many values, if $s \in S_m$, then $s - 2 \in S_m$ or $s + 2 \in S_m$. Let us partition this set into two pieces;

$$S_m = K_m \cup P_m$$

where K_m is the set of all values $s \in S_m$ for which $s \pm 2 \notin S_m$ and $S_m - K_m = P_m$. It should be clear that by the hypothesis, K_m is a finite set and P_m is infinite.

Now, let us take P_m and generate a special set from it in the following way. We first index P_m:

$$P_m = \{p(1,m), p(2,m), p(3,m), ...\}$$

We must then filter this set via the following rule:

If it is the case that there are elements $p(i,m)$, $p(i+1,m)$ and $p(i+2,m)$ such that $p(i,m)+2 = p(i+1,m)$ and $p(i+1,m)+2 = p(i+2,m)$ and $p(i+2,m)+2 \neq p(i+3,m)$, then remove $p(i+2,m)$ from the set.

With the surviving elements, define the set Y_m as the collection of those surviving elements. We will index this set in the standard way as:

$$Y_m = \{y(1,m), y(2,m), y(3,m), ...\}$$

From Y_m, we produce the set H_m as:

$$H_m = \{h(1,m), h(2,m), h(3,m), ...\}$$

where each of the $h(i,m)$ will be a set which will contain two elements so that:

$$h(i,m) = \{y(2i-1,m), y(2i,m)\}$$

It should be clear from our filtering rule that each $h(i,m)$ contains precisely one pair of twin values, that is $y(2i-1,m)+2 = y(2i,m)$. With this new set, we will define an alternative sieving process on it, which will mirror the sieving process on the original set.

Definition 4.1 *The H Sieving Process*

Define the removal sets $V_k = \{h(i,k) \in H_k \,:\, \Psi_k(r(j,k)) = 2i-1 \text{ or } \Psi_k(r(j,k)) = 2i\}$

This removal process will function in the following way: In the original sieve, if we remove the k^{th} element in the sieve, then in the corresponding H set, we will remove the the set of values that would contain the k^{th} element. We can see that in doing this, we will be removing entire pairs instead of just one element of the pair.

Now, let us define two valuable functions we will use:

Definition 4.2 *The Delta Function*

For each $k \geq m$ define the function:

$$\Delta_2 : \mathbb{N}^2 \to \mathbb{N}$$

such that:

$$\Delta_2(k,x) = |\{s \in S_k \,:\; s-2 \in S_k \, and \, s < x\}|$$

Additionally, we define:

$$\Delta_2(x) = |\{s \in S \,:\, s - 2 \in S_k \,\wedge\, s < x\}|$$

Intuitively, this function will count the number of pairs in S_k which are less than or equal to x. It will count each pair as one object; that is, this function will not count each member of a pair individually.

Definition 4.3 *The Delta Bar Function*

For each $k \geq m$ define the function:

$$\bar{\Delta}_2 : \mathbb{N}^2 \to \mathbb{N}$$

such that:

$$\bar{\Delta}_2(k, x) = \{h \in H_k \,:\, \max(h) < x\}$$

Additionally, we define:

$$\bar{\Delta}_2(x) = |\{h \in H \,:\, max(h) < x\}|$$

This function will essentially do the same thing that the ordinary delta function will do, but it will count over our H sets instead. We cannot use the original delta function because while the S sets are sets of natural numbers, the H sets are set *of* sets of natural numbers.

With these two functions, we will prove our first supplementary lemma:

Lemma 4.1 *The Delta Inequality*

For all $k \geq m$ and for all $x \in \mathbb{N}$;

$$\Delta_2(k, x) \geq \bar{\Delta}_2(k, x)$$

Proof

To prove this, we shall proceed by induction.

Base Case

We wish to show that:

$$\Delta_2(m, x) \geq \bar{\Delta}_2(m, x)$$

however, this immediately follows by the definitions of the set. We can see that by the process which we defined H_m;

$$\bigcup H_m = \bigcup_{i=0}^{\infty} h(i, m) \subseteq S_m$$

So every pair that is in H_m is also in S_m, thus, it is the case that our inequality holds.

Inductive Step

Let us assume that;

$$\Delta_2(n-1,x) \geq \bar{\Delta}_2(n-1,x)$$

for all $x \in \mathbb{N}$.

Since this holds for all $x \in \mathbb{N}$, let us fix some arbitrary $y \in \mathbb{N}$ and consider:

$$S_{n-1}|_y = \{s \in S_{n-1} : s < y\}$$

$$H_{n-1}|_y = \{h \in H_{n-1} : \max(h) < y\}$$

which we call the restrictions of S_{n-1} and H_{n-1} with respect to y.

By our inductive hypothesis, we know that:

$$\Delta_2(n-1,y) \geq \bar{\Delta}_2(n-1,y)$$

Now let us consider that we implement the sieve on these two restrictions. This will produce the sets $S_n|_y$ and $H_n|_y$. We know that during the sieve process, we will have values removed from each set. The *number* of values removed from $S_n|_y$ will be $a_{n,y}$ and the number of values removed from $H_n|_y$ will be $b_{n,y}$. By the definition of our sieving process, $a_{n,y} = b_{n,y}$.

Since every value removed from $H_{n-1}|_y$ to produce $H_n|_y$ contributes to the count of the adjusted delta function:

$$\bar{\Delta}_2(n,y) = \bar{\Delta}_2(n-1,y) - b_{n,y}$$

However, this is not necessarily the case for the S set. We can consider that not all values removed from $S_{n-1}|_y$ are elements in a pair, so we have that:

$$\Delta_2(n, y) \geq \Delta_2(n-1, y) - a_{n,y}$$

Thus we have;

$$\bar{\Delta}_2(n, y) = \bar{\Delta}_2(n-1, y) - b_{n,y} = \bar{\Delta}_2(n-1, y) - a_{n,y} \leq \Delta_2(n-1, y) - a_{n,y}$$

Since it is the case that:

$$\Delta_2(n-1, y) - a_{n,y} \leq \Delta_2(n, y) \Rightarrow \bar{\Delta}_2(n, y) \leq \Delta_2(n, y)$$

As desired. ■

With this lemma, we can see that the adjusted delta function serves as a lower bound for the number of pairs that appear in the original sieved set. Additionally, it should be clear that, for all $x \in \mathbb{N}$:

$$\Delta_2(m, x) \geq \Delta_2(m+1, x) \geq ... \geq \Delta_2(x)$$

$$\bar{\Delta}_2(m, x) \geq \bar{\Delta}_2(m+1, x) \geq ... \geq \bar{\Delta}_2(x)$$

With the above lemma, we also have that:

$$\Delta_2(x) \geq \bar{\Delta}_2(x)$$

Now, we have that the number of twin values to appear in the original sieve is at least the number of elements which survive in our new sieve. Now, let us prove our last lemma:

Lemma 4.2 *The Infinitude of H*

$$|H| = \aleph_0$$

Proof

In order to show that H is infinite, let us invoke the hypotheses:

1. For all $y \geq m$, and for all $r(k, y) \in R_y$;

$$\Psi_y(r(k, y)) \geq c_y \cdot k$$

2.

$$3 \leq c_m \leq c_{m+1} \leq ...$$

3.

$$\lim_{n \to \infty} c_n = \infty$$

We can see that on the first iteration of the sieve on H_m, the smallest removed value must be greater than $h(1, m)$. We achieve this by our sieve process over the H sets. Recall that the removal set for H_m defined as:

$$V_m = \{h(i, m) \in H_m \,:\, \Psi_m(r(j, m)) = 2i - 1 \text{ } or \text{ } \Psi_m(r(j, m)) = 2i\}$$

Since $\Psi_m(r(j,m)) \geq 3$, this implies that the smallest $h(i,m)$ in V_m has $i \geq 1$. This implies that H is non-empty. By the second condition, any value less than the smallest removed value cannot be removed by future iterations on the sieve. Thus, $h(1,m) \in H$. Now we can invoke the third condition. Since $\lim_{n\to\infty} c_n = \infty$, there exists a $c_{n_1} \geq 5$. Now we can see that the smallest removed value from H_{n_1} is greater than or equal to $h(3,n_1)$, which implies that $h(2,n_1) \in H$.

With this process in mind, for each $2k+1 \in \mathbb{N}$, define $c_{n_k} \geq 2k+1$. We can see that on the n_k^{th} iteration of the sieve, there are k elements in the sieve which are less than the smallest removed value in the sieve. Subsequently, on the n_k^{th} iteration, we are guarenteed to have k elements in H. Since $\lim_{n\to\infty} c_n = \infty$, it must be the case that there are infinitely many values in H, as desired. ■

With this last lemma in hand, we can complete our proof. Since H is infinite, we have that:

$$\lim_{x\to\infty} \bar{\Delta}_2(x) = \infty$$

Since $\bar{\Delta}_2(x) \leq \Delta_2(x)$, it must be the case that:

$$\lim_{n\to\infty} \Delta_2(x) = \infty$$

As desired. ■

With this theorem, we are provided with sufficient conditions for the existence of twin values within a sieve. Before we move on to applications of this theorem, we have two corollaries that may be useful.

Corollary 4.1 *The First Restatement of the Twin Sieve Theorem*

Let S be a sieved set with respect to $\{R_i\}_{i=0}^{\infty}$, suppose that the following hold:

1. There exists an S_m such that, for all but finitely many values, if $s \in S_m$, then $s - 2 \in S_m$ or $s + 2 \in S_m$.

2. There exists a sequence of natural numbers $\{c_i\}_{i=m}^{\infty}$ such that:

i. For all $i \geq m$, $c_i \geq 3$

ii. $\liminf_{n\to\infty} c_n = \infty$

iii. For all $y \geq m$, and for all $r(k, y) \in R_y$;

$$\Psi_y(r(k, y)) \geq c_y \cdot k$$

Then it is the case that there are infinitely many elements $s \in S$ such that $s + 2 \in S$.

In this case, we can remove the condition that there need to be an ascending chain of inequalities for the values of c_n. This proof follows exactly the same way as the original theorem, except for the last step in proving the infinitude of the H set.

Corollary 4.2 *The Second Restatement of the Twin Sieve Theorem*

Let S be a sieved set with respect to $\{R_i\}_{i=0}^{\infty}$, suppose that the following hold:

1. There exists an S_m such that, for all but finitely many values, if $s \in S_m$, then $s-2 \in S_m$ or $s+2 \in S_m$.

2. There exists a real valued function $f(x)$ such that:

i. For all $x \in \mathbb{N}$ where $x > 3$

$$f(x) \geq 3$$

ii. $\liminf_{x\to\infty} f(x) = \infty$

iii. For all $y \geq m$, and for all $r(k, y) \in R_y$;

$$\Psi_y(r(k, y)) \geq f(y) \cdot k$$

Then it is the case that there are infinitely many elements $s \in S$ such that $s + 2 \in S$.

In this case, we have that instead of searching for a sequence of natural numbers, we are able to instead find a function which bounds it from below. This restatement follows immediately from the first restatement, and can be clearly seen if we take:

$$c_n = \lfloor f(n) \rfloor$$

which then forces the real valued function to take on only integer values, which gives us the exact statement of Corollary 3.1.

Equipped with this theorem and its variations, let us move to the next section, where we will attempt to apply it to the various sieve processes detailed in Section 2.

5 Applications of the Twin Sieve Theorem

5.1 The Lucky Numbers

Equipped with the above theorem, let us show a direct application of it to the Sieve of Joseph Flavius.

Theorem 5.1

The set of Lucky Numbers contains infinitely many values ℓ such that $\ell + 2$ is also a Lucky number.

Proof

In order to prove this, we must guarantee that the hypotheses of the Twin Sieve Theorem are satisfied.

Condition 1

The first condition we must satisfy is that there is an iterated set which contains all but finitely many values which are not twins. Let us consider S_1, that is, the first iteration of the Lucky Sieve.

$$S_1 = \{1, 3, 5, 7, 9, 11, 13, 15, 17, ...\}$$

Which is the set of all odd integers. Clearly each element is a member of at least one pair, thus, our first condition is satisfied.

Condition 2

For this, we must show that there exists our desired sequence of constant values which bound the Locator function. Let us define the sequence $\{c_n\}_{n=1}^{\infty}$ so that $c_k = \ell_k$, which is the k^{th} Lucky number, with $\ell_0 = 1$.

$\boxed{\textit{Part 1}}$

Show that:

$$3 \leq c_m \leq c_{m+1} \leq c_{m+2} \leq ...$$

We can see that the sequence of Lucky numbers is a strictly increasing sequence, with $\ell_1 = 3$. Thus, it is clear that:

$$3 \leq \ell_1 < \ell_2 < \ell_3 < ...$$

Which satisfies part 1.

$\boxed{\textit{Part 2}}$

Show that:

$$\lim_{n \to \infty} c_n = \infty$$

Since $\ell_k \approx k \cdot \ln(k)$, for the paper *The Lucky Number Theorem*;

$$\lim_{n \to \infty} \ell_n = \lim_{n \to \infty\infty} n \cdot \ln(n) = \infty$$

Thus, the second part holds.

$\boxed{\textit{Part 3}}$

Show that, for all $y \geq m$, and for all $r(k, y) \in R_y$;

$$\Psi_y(r(k, y)) \geq c_y \cdot k$$

By definition, the Ψ function gives us the location of the removed values relative to the set they are removed from. However, by the definition of the Sieve of Joseph Flavius, on the y^{th} iteration of the sieve, every value removed will be exactly a multiple of ℓ_y away from 1, thus;

$$\Psi_y(r(k, y)) = \ell_y \cdot k \Rightarrow \Psi_y(r(k, y)) \geq \ell_y \cdot k$$

Thus, part 3 holds.

With all conditions of the Twin Sieve Theorem satisfied, it must be the case that the set of Lucky numbers must contain an infinite number of twin values, as desired. ■

The Sieve of Joseph Flavius is extremely similar to the Sieve of Eratosthenes, but it does have one very striking difference: The Sieve of Eratosthenes is an order invariant sieve, yet the sieve of Joseph Flavius is not. Because of this, changing the order of the removal sets or removal values will change the end set produced by the sieve.

For the sieve of Eratosthenes, it does not matter if we removed the multiples of 2 first, or if we removed the multiples of 5 first; the end result will always produce the set of prime numbers. However, let us consider that we start with the set of Lucky Numbers $\mathbb{L}$. If we start with the regular process, with the set of odd numbers, but we chose a different Lucky Number than 3 to remove elements by, we will be left with a considerably different set of numbers.

For example:

Let us suppose that we remove by 2 and then remove by 7. On iteration 1, we have the set of odd numbers:

$$1 \quad 3 \quad 5 \quad 7 \quad 9 \quad 11 \quad 13 \quad 15 \quad 17 \quad 19 \quad ...$$

On the second iteration, we will be left with:

$$1 \quad 3 \quad 5 \quad 7 \quad 9 \quad 11 \quad 15 \quad 17 \quad 19 \quad 23 \quad ...$$

Now let us removed by 3, leaving us with:

$$1 \quad 3 \quad 7 \quad 9 \quad 15 \quad 17 \quad 23 \quad 25 \; ...$$

In this case 23 will survive the sieve, where in the original sieve, 23 was removed. Thus, by altering the order by which removal sets are implemented, a new set will be produced. However, the question is, does changing the order of removal change the existence of infinitely many twins in the end set?

Theorem 5.2 *The Lucky Number Consistency Theorem*

Let S be the sieved set with respect to $\{R_i\}_{i=0}^{\infty}$, where:

$$R_n = \{s \in S_n \, : \, \ell_{f(n)} \, | \, |S_n|_s|\}$$

where $f : \mathbb{N} \to \mathbb{N}$ and f is a bijection, and $\ell_{f(n)}$ is the $f(n)^{th}$ Lucky number.

Then the set S contains infinitely many twin values.

Proof

To prove this, we will use the First Restatement of the Twin Sieve Theorem. We can recall that if we define the sequence of natural numbers to be $\{\ell_{f(n)}\}_{n=1}^{\infty}$, we can immediately come to our conclusion.

By Theorem 4.1, we know that since $R_0 = 2\mathbb{N}$, S_1 is the set of odd numbers, which certainly contains only pairs. Thus, our first condition remains satisfied. Now, we also know that the last condition will be satisfied as well, since clearly we have equality:

$$\Psi_y(r(k, y)) = \ell_{f(n)} \cdot k$$

Thus, our third condition is satisfied. Lastly, we must show that:

$$\liminf_{n\to\infty} c_n = \infty$$

Since $\lim_{n\to\infty} \ell_n = \infty$, as the Lucky numbers are an increasing sequence of natural numbers, it is clear that any rearrangement of them must have a limit infimum of ∞, as desired. ■

This shows us, at least in the case of the Sieve of Joseph Flavius, that rearrangement of the removal sets does not change the end result. This may lead to a more general result, which we will attempt to achieve at a later time.

This is the current extent of the Twin Sieve Theorem as applied to the Lucky Numbers and its variations. After successfully applying the theorem to this sieve, we can attempt to apply it to the Sieve of Eratosthenes.

5.2 The Prime Numbers

Unfortunately, applying the Twin Sieve Theorem to the Sieve of Eratosthenes is not nearly as straight forward as it was with the Lucky Numbers. This is not a surprise however, since an easy application would imply the existence of infinitely many twin prime, which is known to be a difficult problem.

We can see that the first condition of the theorem and each of its restatements, is satisfied almost immediately. If we look at E_1;

$$E_1 = \{1, 2, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, ...\}$$

Which is the set of odd numbers, with 2 included. Every element in the set, with the exception of 2, is an element of a pair, so our first condition holds.

When attempting to apply the theorem to the Sieve of Eratosthenes, we will be using the Second Reformulation of the Twin Sieve Theorem. To satisfy this, we must find a function which bounds the Locator function from below. To do so, let us consider the following:

Definition 5.1 *The Legendre Function*

For each $x \in \mathbb{N}$ and each $k \in \mathbb{N}$, define the function:

$$\lambda(k, x) = \sum_{d|rad(k)} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor$$

> *Where $rad(k)$ is the radical of a natural number, defined as:*
>
> $$rad(k) = \prod_{p|k} p$$

We can see that this function is more or less a direct application of the inclusion-exclusion principle. Regardless, this function serves an important purpose; if we let $k = p(n)$, which is the product of the first n primes, $\lambda(p(n), x)$ counts the number of values less than x which are relatively prime to the first n primes. If we let our x range over the removal values, we have the equality:

$$\lambda(p(n), r(i, n)) + n = \Psi_n(r(i, n))$$

Where we add an additional n values since the Legendre function does not count the primes which are still in the set. So, we initially start with an explicit version of the Locator function, however, it is not extraordinarily useful. Due to the floor function, we cannot easily bound it from below. This is where the difficulty of the application is contained. However, with this function, we are able to consider what would happen if we take conservative estimates.

Estimate 1

Let us consider the function defined as:

$$\bar{\lambda}(p(n), x) = \sum_{d|p(n)} \frac{x}{d}$$

It is known that:

$$\sum_{d|p(n)} \mu(d)\frac{x}{d} = x \cdot \frac{\phi(p(n))}{p(n)}$$

Where ϕ is the Euler Totient Function.

If we assume that this function is a good estimate for the Legendre function, then we can consider the following:

$$\lambda(p(n), r(i,n)) \approx r(i,n) \cdot \frac{\phi(p(n))}{p(n)}$$

Now let us recall that all the removal values must be multiples of p_{n+1}, we can rewrite $r(i,n)$ as:

$$r(i,n) = p_{n+1} \cdot \bar{r}(i,n)$$

Where $\bar{r}(i,n) \in \mathbb{N}$. Now we can consider the quantity;

$$\frac{p_{n+1}\phi(p(n))}{p(n)}$$

We would like to know how this value behaves, particularly how it behaves as it approaches infinity. We can see that:

$$\frac{\phi(p(n))}{p(n)} = \prod_{i=1}^{n} \left(\frac{p-1}{p}\right) = \prod_{i=1}^{n} \left(1 - \frac{1}{p}\right)$$

We can recognize the end product as an Euler product. We know from result by Euler that:

$$\prod_{i=1}^{\infty}\left(1-\frac{1}{p_i^n}\right)=\zeta(n)$$

Where ζ is the Riemann Zeta function. The product of interest "evaluated" at infinity is equal to $\zeta(1)$ which we know diverges to infinity. We also know that:

$$\zeta(1)=\sum_{n=1}^{\infty}\frac{1}{n}$$

We know that while the harmonic series diverges, it does so very slowly, at a natural logarithmic rate, giving us the approximation:

$$\frac{\phi(p(n))}{p(n)}\approx\frac{1}{\ln(n)}$$

By the prime number theorem, we know:

$$p_{n+1}\approx(n+1)\cdot\ln(n+1)$$

Putting these two results together, we have:

$$p_{n+1}\cdot\frac{\phi(p(n))}{p(n)}\approx\frac{(n+1)\ln(n+1)}{\ln(n)}\approx n+1$$

If this were the case, it would imply that:

$$\Psi_n(r(i,n))\approx(n+1)\cdot\bar{r}(i,n)+n$$

If we take our constant sequence to be $c_n = n + 1$, we could conclude by this approximation that our theorem will hold over the Sieve of Eratosthenes.

However, the issue lies in the rigor of the approximation in question. This problem is one that has been widely studied in Sieve Theory and is historically is a difficult approximation to justify. The difficulty comes from the potential error accrued by disregarding the floor function. If the error is small, then the above estimate will hold and the theorem can be readily applied.

Showing that the error is consistently small is something that is currently, even with modern techniques, out of reach. The estimate we had gone over previously leads us to believe that, for each $r(k, n)$;

$$\lambda(p(n), r(k, n)) \geq (n + 1) \cdot k$$

This is perhaps a very strong bound, in the sense that it works under the assumption that the error term is small. Luckily, we do not need to have this bound be quite that strong. By the Second Reformulation of the Twin Sieve Theorem, we need to find *any* increasing function to bound our removal values by. To do so, let us try to build up a bit of an inventory of propositions to find such a function.

5.3 Bounding the Locator Function

In all iterations of the sieve, we know exactly what the smallest element to be removed is. On the n^{th} iteration of the sieve, we remove all multiples of p_n which remain in the set. It should then be clear that the smallest element removed by the sieve, that is, the first multiple of p_n removed, must be p_n^2. All elements removed by the sieve are of the form $p_n \cdot k$, for some natural number k, and if $k < p_n$, then k possesses a prime factor which is *strictly less* than p_n, which is impossible, since all such numbers have been removed already by previous iterations of the sieve. This leads us to our first result in this section:

Proposition 5.1 *For each* $n \in \mathbb{N}$*;*

$$r(1, n) = p_n^2$$

If we consider $\lambda(p(n), p_n^2)$, by definition, it must count all values which are relatively prime to the first n primes which are less than or equal to p_n^2. However, beneath p_n^2, the only values to survive must be prime. This gives us the following proposition:

Proposition 5.2 *For each* $n \in \mathbb{N}$*;*

$$\lambda(p(n), r(1, n)) = \pi(p_n^2) - n + 1$$

Additionally,

$$\Psi_n(p_n^2) = \pi(p_n^2) + 1$$

Where π *is the prime number counting function.*

This proposition gives us our ”jumping off” point. It lets us know that the smallest value removed will not be too small. More importantly, from the prime number theorem, we know:

$$\pi(x) \approx \frac{x}{\ln(x)} \Rightarrow \pi(p_n^2) \approx \frac{p_n^2}{\ln(p_n^2)}$$

Since p_n goes to infinity as n goes to infinity, it implies that:

$$\lim_{n\to\infty} \Psi_n(r(1,n)) = \infty$$

So the issue is not with the smallest value removed, but rather the values that are closely clustered together. Thus, it would be sufficient to show that the removal values are spread out enough so that we may find a function which bounds the Locator function from below. However, such an analysis is beyond the scope of this work, as all techniques to be used rely heavily on powerful results in analytic number theory and abstract algebra.

In conclusion, in this paper, we have established sufficient conditions for the existence of Twin Values in Sieves over the Natural Numbers, as well as showing its successful application to a known sieve, and its potential application to the Sieve of Eratosthenes.

6 Future Work

We have the following conjecture relating to the Sieve of Eratosthenes:

Conjecture 1 *The Eratosthenes Bounding Function*

For each $i \in \mathbb{N}$ *and for each* $n \in \mathbb{N}$ *which is sufficiently large:*

$$\lambda(p(n), r(i, n)) \geq \ln(\ln(n)) \cdot i$$

In future work, we plan on verifying the inequality in the conjecture. Coupled with the Twin Sieve Theorem, we will be left with a proof for the twin prime conjecture. The above inequality will no doubt pose significant difficulty in its proof, but at the very least, the above conjecture provides a restatement of the twin prime conjecture in simpler terms.

Additional research will be into extending the Twin Sieve Theorem to more general settings, in particular, we will attempt to answer the question:

In a General Natural Sieve, are there infinitely many values which are exactly k apart, where $k \in \mathbb{N}$?

7 Works Cited

Hawkins, D., and W. E. Briggs. "*The Lucky Number Theorem.*" Mathematics Magazine, vol. 31, no. 2, 1957, pp. 81–84. JSTOR, www.jstor.org/stable/3029213.

Gardiner, Verna, et al. "*On Certain Sequences of Integers Defined by Sieves.*" Mathematics Magazine, vol. 29, no. 3, 1956, pp. 117–122. JSTOR, www.jstor.org/stable/3029719.