Empirical Sieve Based on Cyclic Frequencies: A Deterministic Alternative for Prime Detection

Héctor Cárdenas Campos

April 20, 2025

Contents

1. Introduction

The study of prime numbers has been a recurring theme throughout the history of mathematics, due not only to their theoretical importance but also to their practical applications in fields such as cryptography, information theory, and computing. One of the oldest and best-known methods for generating prime numbers is the **Sieve of Eratosthenes**, which systematically eliminates multiples of known primes to identify the remaining primes within a given range.

Despite its effectiveness, this method presents limitations when the goal is to analyze the **internal structure** of prime numbers or to identify predictive patterns within their distribution. In this context, we present an alternative model, developed from empirical observations, which makes it possible to determine whether a number is prime **without prior knowledge of other primes** or the need for explicit factorization.

This approach is based on the fact that all prime numbers greater than 3 belong to the **arithmetic sequence** defined by the terms $6n+1$ and $6n-1$. From this property, a model is constructed that generates **cyclic frequencies** associated with each number in that sequence, and a mechanism of collision-based cancellation is applied, including an inverse formulation that enables direct evaluation of individual numbers without constructing full sequences.

The method is characterized by:

- Relying solely on basic arithmetic operations.
- Using deterministic functions to generate frequencies from an index called the "consecutive."
- Defining a **safe analysis range**, based on internal properties of the number being tested.
- Offering a theoretical formulation that enables prediction of cancellations without exhaustively traversing all sequences.

This document aims to formally present the mathematical structure of the model, precisely defining the functions used, their relationship to the $6n \pm 1$ sequence, and the criteria for determining whether a given number is prime.

1.1. Sequence of Prime Candidates

All prime numbers greater than 3 belong to the sequence defined by:

$$a_c = \begin{cases} 6n - 1, & \text{if } c \text{ is even} \\ 6n + 1, & \text{if } c \text{ is odd} \end{cases}$$

where $n \in \mathbb{N}$ and c is the index or "consecutive" within the sequence.

Alternatively, this sequence can be generated directly, without explicit reference to n, using the recursive formula:

$$a_0 = 5, \quad a_{c+1} = a_c + \begin{cases} 2, & \text{if } c \text{ is even} \\ 4, & \text{if } c \text{ is odd} \end{cases}$$

The terms generated in this way are:

$$5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 25,\ 29,\ 31,\ \ldots$$

1.2. Consecutive Index c

The number $c \in \mathbb{N}$ represents the position index within the alternating sequence $6n \pm 1$. From each value of c, three fundamental functions of the model are derived: the frequencies $f_1(c)$, $f_2(c)$, and the mean value $f_3(c)$, which will be described in the next section.

2. Generation of Frequencies

To each value of the consecutive index $c \in \mathbb{N}$, two frequencies f_1 and f_2 are associated, as well as a mean value f_3, all defined deterministically by the following expressions:

2.1. Generation Formulas

$$\begin{aligned} f_2(c) &= 2c + 3 \\ f_1(c) &= 2 \cdot f_2(c) + (-1)^c = 4c + 6 + (-1)^c \\ f_3(c) &= \frac{f_1(c) + f_2(c)}{2} \end{aligned}$$

These formulas generate the values that will participate in the cancellation process. It is observed that all generated values are positive integers.

2.2. Interpretation of f_1, f_2, and f_3

- $f_2(c)$: The smaller frequency of the pair, always odd. It defines one of the alternating steps.
- $f_1(c)$: The larger frequency, which complements f_2 to form an alternating sequence.
- $f_3(c)$: The arithmetic mean of both frequencies. It coincides with the value a_c of the sequence $6n \pm 1$ defined previously. That is,

$$f_3(c) = a_c \quad \text{for all } c \in \mathbb{N}$$

2.3. Example

For $c = 3$:

$$\begin{aligned} f_2(3) &= 2 \cdot 3 + 3 = 9 \\ f_1(3) &= 2 \cdot 9 + (-1)^3 = 18 - 1 = 17 \\ f_3(3) &= \frac{17 + 9}{2} = 13 \end{aligned}$$

And indeed, $a_3 = 13$, which is a number of the form $6n + 1$, located at the fourth position of the sequence $6n \pm 1$.

3. Mechanism of Cancellation by Frequencies

Each pair of frequencies $(f_1(c), f_2(c))$ defines an alternating sequence constructed by interleaved summation of these frequencies, starting from the initial consecutive index c. This sequence generates what we call a **cancellation row**.

3.1. Definition of the Cancellation Row

Let $c \in \mathbb{N}$ be a given consecutive index. The alternating sequence r_n associated with the pair $(f_1(c), f_2(c))$ is defined as:

$$r_0 = c, \quad r_{n+1} = \begin{cases} r_n + f_1(c), & \text{if } n \text{ is even} \\ r_n + f_2(c), & \text{if } n \text{ is odd} \end{cases}$$

The resulting sequence $r_0, r_1, r_2, \ldots$ contains the consecutive indices that will be canceled by the frequencies generated from c.

3.2. Cancellation Criterion

Given a number X to be tested, its target consecutive c_X is first determined as the index such that $f_3(c_X) = X$, that is, the value of c for which $a_c = X$.

Then, the number X is considered **composite** if its target consecutive appears in the cancellation row of at least one previous value c, such that:

$$f_3(c) \leq \sqrt{X}$$

This condition restricts the evaluation to a finite set of frequencies, similar in logic to the Sieve of Eratosthenes, where only multiples of primes less than or equal to $\sqrt{X}$ are eliminated.

3.3. Interpretation

The model assumes that if a frequency originating from some value c reaches (through alternating summation) the target consecutive c_X, then the corresponding number cannot be prime, as it has been canceled by an internal regularity of the system. If no frequency reaches it, then the number is considered prime.

4. Full Example: Evaluation of Number 13

Below is a detailed example of the primality evaluation process using the frequency cancellation model.

4.1. Step 1: Determine the Consecutive Associated with the Number

The number $X = 13$ belongs to the sequence $6n \pm 1$. To find the corresponding consecutive c_X such that $f_3(c_X) = 13$, we evaluate successive values of $f_3(c)$ until a match is found:

$$
\begin{aligned}
&f_2(0) = 3, \quad && f_1(0) = 7, \quad && f_3(0) = \frac{3+7}{2} = 5 \\
&f_2(1) = 5, \quad && f_1(1) = 11, \quad && f_3(1) = \frac{5+11}{2} = 8 \\
&f_2(2) = 7, \quad && f_1(2) = 15, \quad && f_3(2) = \frac{7+15}{2} = 11 \\
&f_2(3) = 9, \quad && f_1(3) = 17, \quad && f_3(3) = \frac{9+17}{2} = 13
\end{aligned}
$$

Therefore, the associated consecutive is $c_X = 3$.

4.2. Step 2: Evaluate Previous Frequencies

We must consider all values of c such that $f_3(c) \leq \sqrt{X} \approx 3.605$. Evaluate:

- $f_3(0) = 5 \rightarrow$ already exceeds $\sqrt{13}$

There is no value of $c < 3$ for which $f_3(c) \leq \sqrt{13}$. Therefore, there are no earlier frequencies that could have generated a cancellation row reaching the target consecutive $c_X = 3$.

4.3. Step 3: Conclusion

Since no earlier row reaches the target consecutive c_X, the number 13 is considered **prime** according to the model.

This example illustrates how the model enables primality testing without factorizations or divisions, relying solely on arithmetic functions and position comparison within the generated sequence.

5. General Evaluation Criterion

Given a number $X > 3$, the model establishes that it will be considered composite if there exists a value $c \in \mathbb{N}$ such that:

- $f_3(c) \leq \sqrt{X}$
- The index c_X, associated with the number X, appears in the alternating sequence of indices generated by the frequencies $f_1(c)$ and $f_2(c)$, starting from c

Note: This sequence is constructed in the domain of the consecutive indices c, not in that of the values a_c. That is, the frequencies cancel positions within the $6n \pm 1$ sequence, not the numerical values themselves.

Otherwise, if no value of c satisfies both conditions simultaneously, the number X is considered prime.

5.1. Justification of the Bound $f_3(c) \leq \sqrt{X}$

This bound ensures that only frequencies generated from consecutive indices whose mean value $f_3(c)$ is less than or equal to the square root of the number under evaluation are considered. This is justified by analogy with the Sieve of Eratosthenes, where only multiples of primes less than or equal to $\sqrt{X}$ are removed.

In this model, $f_3(c)$ plays a similar role: it represents the numeric value from which the propagation of frequencies that could collide with X begins. Once $f_3(c) > \sqrt{X}$, any cancellation row generated from that value will only reach consecutive values greater than c_X, and its inclusion in the analysis becomes unnecessary.

5.2. Procedure Summary

To evaluate a number X, the following steps are followed:

1. Compute c_X, the consecutive such that $f_3(c_X) = X$
2. Compute $\sqrt{X}$
3. Evaluate all values $c \in \mathbb{N}$ such that $f_3(c) \leq \sqrt{X}$
4. For each c, generate the alternating sequence starting from c using increments $f_1(c)$ and $f_2(c)$
5. If c_X appears in any of them, then X is composite
6. If it does not appear in any, then X is prime

6. Inverse Formulas for Collision Detection

Given a consecutive index $c \in \mathbb{N}$, the cancellation row is defined as the alternating sequence generated from c using the frequencies $f_1(c)$ and $f_2(c)$. Instead of traversing this sequence step by step, it is possible to determine whether a target index c_X (associated with the number X) appears in it by means of inverse formulas.

6.1. Collision Model

The alternating sequence of indices generated from c can be expressed as:

$$r_0 = c, \quad r_1 = c + f_1(c), \quad r_2 = r_1 + f_2(c), \quad r_3 = r_2 + f_1(c), \quad \ldots$$

That is, a sequence defined by alternating sums of the two frequencies.

We wish to determine whether there exists some $n \in \mathbb{N}$ such that $r_n = c_X$.

6.2. Collision Formulas

Collision verification can be carried out by solving the following expressions:

$$r_1 = \frac{c_X - f_1(c) - c}{f_1(c) + f_2(c)}$$

$$r_2 = \frac{c_X - f_1(c) - f_2(c) - c}{f_1(c) + f_2(c)}$$

Where:

- r_1: index (integer) indicating whether c_X matches a **even-positioned** term in the sequence (after applying f_1, then alternating), - r_2: index (integer) indicating whether it matches an **odd-positioned** term (after applying both increments before alternating).

6.3. Cancellation Condition

If either of the expressions evaluates to an integer, then the target consecutive c_X belongs to the alternating sequence generated from c, and therefore the number X is considered composite.

Otherwise, X has not been canceled by that frequency.

6.4. Illustrative Example

Let $X = 25$. We know that $f_3(7) = 25$, so $c_X = 7$. We want to check whether this value is canceled by the frequency generated from $c = 2$.

We calculate:

$$f_1(2) = 15, \quad f_2(2) = 7$$

$$r_1 = \frac{7 - 15 - 2}{15 + 7} = \frac{-10}{22} \notin \mathbb{Z}, \quad r_2 = \frac{7 - 15 - 7 - 2}{15 + 7} = \frac{-17}{22} \notin \mathbb{Z}$$

Neither value is an integer; therefore, the consecutive $c = 2$ does not cancel $c_X = 7$, and the frequency generated from it does not affect the number 25.

6.5. Derivation of the Inverse Formulas

To derive the previous expressions, we begin with the alternating sequence generated from a value $c \in \mathbb{N}$, using increments $f_1(c)$ and $f_2(c)$. The sequence is constructed as follows:

$$
\begin{aligned}
r_0 &= c \\
r_1 &= c + f_1(c) \\
r_2 &= c + f_1(c) + f_2(c) \\
r_3 &= c + f_1(c) + f_2(c) + f_1(c) \\
&= c + f_1(c) + f_2(c) + f_1(c) \\
&= c + f_1(c) + f_2(c) + f_1(c) + f_2(c) + \cdots
\end{aligned}
$$

Generalizing, for $k \in \mathbb{N}$, the elements of the sequence alternate between:

$$
\begin{aligned}
r_{\text{even}} &= c + f_1(c) + k \cdot (f_1(c) + f_2(c)) \\
r_{\text{odd}} &= c + f_1(c) + f_2(c) + k \cdot (f_1(c) + f_2(c))
\end{aligned}
$$

To check whether a given target consecutive c_X appears in this sequence, we set up the equations:

$$
\begin{aligned}
c_X &= c + f_1(c) + k \cdot (f_1(c) + f_2(c)) \quad \text{(even position)} \\
c_X &= c + f_1(c) + f_2(c) + k \cdot (f_1(c) + f_2(c)) \quad \text{(odd position)}
\end{aligned}
$$

Solving for k, we obtain:

$$
\begin{aligned}
k &= \frac{c_X - f_1(c) - c}{f_1(c) + f_2(c)} \quad \text{(even position)} \\
k &= \frac{c_X - f_1(c) - f_2(c) - c}{f_1(c) + f_2(c)} \quad \text{(odd position)}
\end{aligned}
$$

These are the exact formulas that allow us to verify whether a frequency hits the target consecutive c_X, without building the entire sequence.

7. Optimization of the Analysis Range

One of the main benefits of this model is that it does not require evaluating all possible frequencies, but only those whose mean value $f_3(c)$ is less than or equal to a threshold defined by the target number X.

7.1. Growth of the Functions f_1, f_2, f_3

Given that:

$$f_2(c) = 2c + 3$$
$$f_1(c) = 4c + 6 + (-1)^c$$
$$f_3(c) = \frac{f_1(c) + f_2(c)}{2}$$

then the function $f_3(c)$ grows linearly with respect to c, with small oscillations due to the term $(-1)^c$ in $f_1(c)$. Specifically:

$$f_3(c) = \frac{(4c + 6 + (-1)^c) + (2c + 3)}{2} = 3c + \frac{9 + (-1)^c}{2}$$

This expression allows the value of $f_3(c)$ to be estimated directly for any c, and consequently, it determines whether that consecutive index should be included in the analysis for a given number X.

7.2. Cutoff Condition

The model establishes that only those values of c satisfying:

$$f_3(c) \leq \sqrt{X}$$

should be considered, since any frequency generated from a consecutive c with $f_3(c) > \sqrt{X}$ will never reach c_X, due to the linear growth of the alternating sequences, which prevents them from going backward.

7.3. Practical Application

This condition significantly reduces the set of frequencies to be evaluated for a given number. Instead of performing a full search, it is sufficient to determine the maximum value of c such that $f_3(c) \leq \sqrt{X}$, and to evaluate only those consecutive indices within that range.

In computational terms, this enables an efficient implementation of the model, with much lower cost than classical methods of factorization or trial division.

8. Predictive Model for Individual Evaluation

8.1. Direct use of the inverse formula

The inverse formula for collision detection, derived in the previous section, allows the evaluation of a number X through its associated index c_X, without the need to generate or traverse any alternating sequence. This makes it suitable as a predictive model on its own, especially when analyzing isolated values.

The process consists of identifying whether c_X is part of the cancellation row generated from a particular value c. If either of the following expressions yields an integer, then the number is considered composite:

$$r_1 = \frac{c_X - f_1(c) - c}{f_1(c) + f_2(c)}, \qquad r_2 = \frac{c_X - f_1(c) - f_2(c) - c}{f_1(c) + f_2(c)}$$

8.2. Example: Evaluating $X = 25$

We know that $f_3(7) = 25$, so the associated consecutive is $c_X = 7$. Let us test whether this value is canceled by the frequencies generated from $c = 0$:

$$f_2(0) = 3, \quad f_1(0) = 7$$
$$r_1 = \frac{7 - 7 - 0}{7 + 3} = \frac{0}{10} = 0 \in \mathbb{Z}$$
$$\Rightarrow \text{The value is canceled by } c = 0$$

Therefore, 25 is correctly identified as composite using only the inverse formula, without building any sequence.

8.3. Interpretation

This mechanism constitutes a single-number predictive version of the model. It is mathematically equivalent to the full sequence-based cancellation but enables direct analysis with minimal computation. Its structure originates from the same algebraic reasoning that governs the cancellation rows, now distilled into a direct decision rule.

9. Empirical Validation with Numerical Examples

The following examples illustrate how the model operates in concrete cases, demonstrating the primality verification process using frequency-based cancellation.

Example 1: Prime Number $X = 13$

As shown previously, $f_3(3) = 13 \Rightarrow c_X = 3$. No frequency generated from any c such that $f_3(c) \leq \sqrt{13} \approx 3.605$ cancels this value.

Result: Prime

Example 2: Composite Number $X = 25$

We know that $f_3(7) = 25 \Rightarrow c_X = 7$.

We evaluate possible collisions from values of c such that $f_3(c) \leq \sqrt{25} = 5$, i.e., $c = 0, 1, 2$.

- $c = 2$: $f_1 = 15$, $f_2 = 7$ Alternating sequence: $2, 17, 24, 39, \ldots$ $c_X = 7$ does not appear.
- $c = 1$: $f_1 = 11$, $f_2 = 5$ Alternating sequence: $1, 12, 17, 28, 39, 50, \ldots$ $c_X = 7$ does not appear.

- $c = 0$: $f_1 = 7,\ f_2 = 3$ Alternating sequence: $0, 7, 10, 17, 20, 27, \ldots$ $c_X = 7$ appears.

Result: Composite

Example 3: Prime Number $X = 17$

We know that $f_3(4) = 17 \Rightarrow c_X = 4$.
We test values $c = 0, 1, 2, 3$. No frequency generates a sequence that includes $c_X = 4$.
Result: Prime

Example 4: Composite Number $X = 35$

We know that $f_3(10) = 35 \Rightarrow c_X = 10$.
We test several earlier frequencies:

- $c = 3$: $f_1 = 17,\ f_2 = 9$ Sequence: $3, 20, 29, 46, \ldots$ $c_X = 10$ does not appear.
- $c = 5$: $f_1 = 27,\ f_2 = 13$ Sequence: $5, 32, 45, 72, \ldots$ $c_X = 10$ does not appear.
- $c = 0$: $f_1 = 7,\ f_2 = 3$ Sequence: $0, 7, 10, 17, 20, 27, \ldots$ $c_X = 10$ appears.

Result: Composite

10. Conclusions

This document has formalized an alternative model for detecting prime numbers based on the generation of cyclic frequencies associated with the sequence $6n \pm 1$. From simple arithmetic functions defined by an index called the consecutive, frequency pairs are built whose interaction generates alternating cancellation sequences.

The central criterion of the model establishes that a number X is composite if the index c_X associated with it appears in one of these sequences generated from values c such that $f_3(c) \leq \sqrt{X}$. Otherwise, it is considered prime.

Inverse expressions have been derived and formalized into a predictive tool that enables the analysis of individual numbers without constructing full sequences. This predictive formulation is fully integrated into the model and shares the same structure as the original cancellation mechanism, allowing collision detection without constructing full sequences, which adds both efficiency and structural clarity to the method. These formulas have been algebraically justified based on the general form of the progressions involved.

Furthermore, it has been demonstrated that the mean value $f_3(c)$ grows linearly with c, allowing the analysis range to be limited and significantly reducing the computational cost of the procedure. This property gives the model a predictable and scalable behavior.

Through numerical examples, the correct functioning of the system has been validated, demonstrating its ability to distinguish between primes and composites by means of a purely deterministic approach without requiring factorization.

Therefore, the model constitutes a complete and self-consistent proposal, with potential for future theoretical extensions — including the characterization of its performance over large numerical ranges or the formalization of more refined exclusion criteria.