

THE RULE BEHIND THE OCCURRENCE OF PRIME NUMBERS

Matteo Arpe

Claudia La Chioma¹

e-mail: matteo.arpe@satorgroup.com

claudlc@math.uio.no

Abstract. In this paper we present a methodology to identify the occurrence of prime numbers in their apparently random sequence. We prove that each prime number contains all the needed information to identify the subsequent primes. In particular:

1. every prime number p determines a vector $\mathbf{d}^{(p)}$ whose components are symmetric but a tail of three points;
2. the couple $(p, \mathbf{d}^{(p)})$ identifies uniquely the primes $\{p_i\}_{i=1}^n$ for some n , the related vectors $\{\mathbf{d}^{(p_i)}\}_{i=1}^n$ and the set of coprimes with q prime for all $q \leq p$;
3. dimension and mass of $\mathbf{d}^{(p)}$ are uniquely determined from dimension and mass of $\mathbf{d}^{(q)}$, $q = \sup\{a \in \mathbb{N} \text{ s.t. } a \text{ is prime and } a \leq p\}$;
4. symmetry and three–points tail are preserved by all the $\mathbf{d}^{(p)}$.

We underline that the methodology presented here is an inductive one and that its by-product is the complete determination of prime numbers by the only knowledge of a couple $(p, \mathbf{d}^{(p)})$.

Keywords: prime numbers, sieve of Erathostenes, totient function.

1. Introduction

All over the centuries a great number of Mathematicians has devoted its research to find out a rule hidden in the sequence of prime numbers.

The first important result about prime numbers is due to Euclid who proves the existence of infinitely many prime numbers, without any information about their distribution.

¹Corresponding author.

In the eighteenth century, Euler focused his attention on the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s \in \mathbb{N}.$$

He proved that, for every $s \in \mathbb{N}$, it is related to prime numbers by means of the following relation

$$\zeta(s) = \prod_{p \text{ is prime}} \frac{p^s}{p^s - 1}.$$

This function is called *zeta function* and it seems completely unrelated with primes except that by means of the previous equality. In his work, Euler conjectured that ζ satisfies a kind of symmetrical functional equation, and this important property was proved by Riemann, [6], who extended ζ to complex numbers: he solved the problem of studying prime numbers using Euler product formula; moreover he found a formula for counting the number of primes in a prescribed interval $[1, N]$ by means of a main term related to the statistical distribution of primes and a lower order oscillatory correction term, determined by the solutions of

$$\zeta(s) = 0, \quad s \in \mathbb{C}.$$

Riemann conjectured that all the solutions of this equation with $Re(s) > 0$ are of the form $s = \frac{1}{2} + bi$, $b \in \mathbb{R}$. Moreover there are infinitely many zeros with $Re(s) < 0$ and $Im(s) = 0$. This is the still unproved *Riemann hypothesis*. Several mathematicians tried to prove this conjecture giving only partial answers which contribute to give credit to its validity.

If it could be proved to be true, we would know only approximatively the distribution of prime numbers, which would be random. Otherwise, if it was proved to be false, we could say that the prime numbers are distributed with a law still to be found.

A complete collection of results in Number Theory can be found in Dickson [2].

In this paper we present a theory to derive the sequence of prime numbers explaining their random occurrence. It is based upon an important property of symmetry obtained after a clever remark on a filtering and sieving procedure of the set \mathbb{N} . It is well known (sieve of Erathostehes) that whenever we identify a prime number p , by definition none of its multiple can be prime, therefore it has to be erased from \mathbb{N} .

With this procedure we get a nested sequence of sets of the following form: given a prime number p define the set $\{a \in \mathbb{N} \text{ s. t. } \text{MCD}(a, q) = 1, \text{ for all } q \leq p, q \text{ prime}\}$, i.e. the set of numbers coprime with q for all $q \leq p$. This procedure is quite expensive: we show that the filtering procedure can be optimized looking only a finite number of multiples distributed periodically and symmetrically in the set itself.

More precisely, every set of the previous form can be associated to a characteristic vector $\mathbf{d}^{(p)}$, in which every component represents the relative distance between two subsequent numbers of the set itself, with a known dimension ℓ_p

and mass M_p . We note that the components of the vector are symmetric but a symmetric tail of three points and that the relative distances in the set itself are periodic w.r.t the vector $\mathbf{d}^{(p)}$. Moreover all the multiples of p can be located simply knowing the position of the first ℓ_p multiples and that the vector of the relative distances between the multiples is derived from $\mathbf{d}^{(p)}$ itself.

We point out that every vector $\mathbf{d}^{(p)}$ is obtained once we know p and $\mathbf{d}^{(q)}$, where $q = \sup\{a \in \mathbb{N} \text{ s. t. } a \text{ is prime and } a < p\}$.

The paper is organized as follows: in Section 2 we list some definition and quantities which will be useful in the following; Section 3 contains an intuitive description of the problem and the logical approach which leads to the theoretical result explained in Section 4. Section 5 contains a brief description of the algorithm used to select the sequence of prime numbers.

2. Definitions and Notation

We recall here the definitions and notations we will use throughout the paper.

Definition 2.1. We say that $p \in \mathbb{N}$, $p > 1$, is a *prime number* if it can be divided only by 1 and by p itself.

We say that $a, b \in \mathbb{N}$ are *coprime* if the only factor they have in common is 1.

If p is a prime number, then we say that $a \in \mathbb{N}$ is *coprime* with p if p does not divide a .

Let p, q be prime numbers, $p > q$; we call them *twin numbers* if $p - q = 2$, *cousin numbers* if $p - q = 4$.

In the sequel we will use the following notation: for any $a, b \in \mathbb{N}$

$$\text{MCD}(a, b) = (a, b).$$

For any $m \in \mathbb{N}$ we list in increasing order the integers a coprime with n , for all $n \leq m$:

$$(2.1) \quad \mathbb{P}_m = \{a = a^{(m)} \in \mathbb{N}, \text{ s. t. } (a^{(m)}, m!) = 1\} = \{a_1^{(m)} < a_2^{(m)} < \dots\}.$$

Using this notation we write $\{a_k^{(1)}\}_{k \geq 0} = \{n \in \mathbb{N}, n \geq 1\}$.

Moreover, given the sequence $\{a_k^{(m)}\}_k$ and an integer $\ell = \ell_m$, which will be specified later, define the vector of differences $\mathbf{d}^{(m)} = (d_1^{(m)}, \dots, d_\ell^{(m)})$ as follows

$$(2.2) \quad \mathbf{d}^{(m)} \in \mathbb{N}^\ell, \quad d_i^{(m)} = a_i^{(m)} - a_{i-1}^{(m)}, \text{ for } i = 1, \dots, \ell.$$

It is clear that $d_i^{(m)} > 0$ for all $i = 1, \dots, \ell$.

Given a vector $v \in \mathbb{N}^\ell$ we define the **mass** of v to be the following positive quantity:

$$(2.3) \quad M = \sum_{i=1}^{\ell} v_i.$$

Given a finite sequence of prime numbers $\{p_i\}_{1 \leq i \leq n}$ define the function:

$$\begin{aligned} \! : \{p_i\}_{1 \leq i \leq n} &\longrightarrow \mathbb{N}, \\ p_n &\longrightarrow \! (p_n) = \prod_{i=1}^n p_i =: p_n!. \end{aligned}$$

3. Prime numbers and their characteristic vector: an intuitive approach

In this section we present the procedure of the identification of prime numbers based on the symmetrical vector $\mathbf{d}^{(p)}$ of the relative distances between the elements of \mathbb{P}_p for any p prime. Before giving an abstract result, we present an intuitive idea of the procedure.

By definition $a_0^{(1)} = 2$ is the first prime number, therefore none of its multiples can be prime. We focus then our attention to the set $\mathbb{N} \setminus \{2n\}_{n \in \mathbb{N}}$: it is the set of odd numbers and

$$\mathbb{P}_2 = \{a_n^{(2)} \in \mathbb{N}, \text{ s. t. } (a_n^{(2)}, 2) = 1\} = \{a_n^{(2)} = 2n + 1\}_{n \in \mathbb{N}}.$$

In order to find other prime numbers, we repeat the sieving procedure with respect to the first prime number in \mathbb{P}_2 , $a_0^{(2)} = 3$; in this case $d_n^{(2)} = 2$ for all n , therefore we pose $\ell = 1$ and $\mathbf{d}^{(2)} = (2)$.

Using the result of the sieve of Eratosthenes, it is well known that erasing all the multiples of the primes q such that $q \leq p$, we are sure of having selected all the prime numbers less than p^2 . In this case we can select all the primes less than $(a_0^{(2)})^2 = 9$.

The main difference between the sieve of Eratosthenes and our approach stands in the number of iterations needed to go further in the selection of primes: in this classical sieve, for every number $b \in \mathbb{P}_2$ we have to check if it is a multiple of $a_0^{(2)}$, while in our case we focus only on the relative distance from the previous number in the sequence, as it will be explained in the sequel. We point out that it suffices to know only a finite number of distances to get information on all the remaining numbers, since we note that the distances are periodic in \mathbb{P}_p , for every p and that those distances are symmetric but a tail of three points.

As a matter of fact, heuristically we notice that the multiples of $a_0^{(2)}$ occupy every third position in the sequence \mathbb{P}_2 ; erasing the elements using this criterion, we produce the set \mathbb{P}_3 .

If we look at the relative distances $\mathbf{d}^{(3)}$ in \mathbb{P}_3 compared to $\mathbf{d}^{(2)}$ in \mathbb{P}_2 we have the following situation:

\mathbb{P}_2	$a_0^{(2)}$	$a_1^{(2)}$	$a_2^{(2)}$	$a_3^{(2)}$	$a_4^{(2)}$	$a_5^{(2)}$	$a_6^{(2)}$	$a_7^{(2)}$	$a_8^{(2)}$	$a_9^{(2)}$	$a_{10}^{(2)}$	$a_{11}^{(2)}$...
	<u>3</u>	5	7	9	11	13	15	17	19	21	23	25	...
$\mathbf{d}^{(2)}$		2	2	2	2	2	2	2	2	2	2	2	...
\mathbb{P}_3		$a_0^{(3)}$	$a_1^{(3)}$		$a_2^{(3)}$	$a_3^{(3)}$		$a_4^{(3)}$	$a_5^{(3)}$		$a_6^{(3)}$	$a_7^{(3)}$...
		<u>5</u>	7		11	13		17	19		23	25	...
$\mathbf{d}^{(3)}$			2		4	2		4	2		4	2	...

TABLE T1: the sets \mathbb{P}_2 and \mathbb{P}_3 and the vectors $\mathbf{d}^{(2)}$ and $\mathbf{d}^{(3)}$ repeated periodically. The next iteration is done with respect to the underlined number, which is prime.

We note that the vector (2, 4) repeats itself periodically along all the numbers in \mathbb{P}_3 .

Define $\mathbf{d}^{(3)} = (2, 4) \in \mathbb{N}^{\ell_3}$; here $\ell_3 = 2$ is the number of elements in the fundamental period. We notice that now

$$\ell_3 = 2, M_3 = 6, \text{ while } \ell_2 = 1, M_2 = 2.$$

Recall that $[\cdot]$ represents the integer part function and $\chi_{x>b}$ is the Heaviside function of the set $(b, +\infty)$; for any $a_k^{(3)} \in \mathbb{P}_3$ define at first

$$C = \left[\frac{a_k^{(3)} - a_0^{(3)}}{M_3} \right],$$

then there exists $j = 0, \dots, \ell_3 - 1$ such that we have the following: by periodicity and summing up on the equations (2.2) for $i = 1, \dots, k$ we have

$$a_k^{(3)} = a_0^{(3)} + C \sum_{i=1}^{\ell_3} d_i^{(3)} + \sum_{i=1}^{\ell_3} \left(1 - \chi_{x>j(i)} \right) d_i^{(3)};$$

here the last term represents the remainder. Clearly $j = k - C \cdot \ell_3$.

We note that $a_0^{(3)}$ is the first prime number left after the sieving; at the next step, the new periodicity will appear after repeating the vector $\mathbf{d}^{(3)}$ for $a_0^{(3)}$ -times and deleting the first ℓ_3 multiples of $a_0^{(3)}$ which will occur. As a matter of fact, consider the vector $\mathbf{d}^{(3)} \in \mathbb{N}^{a_0^{(3)} \cdot \ell_3}$ obtained by repeating $a_0^{(3)}$ -times the vector $\mathbf{d}^{(3)}$: in the set of elements in $\mathbb{P}_{a_0^{(3)}}$ identified by $\mathbf{d}^{(3)}$ we recover all the first ℓ_3 multiples of $a_0^{(3)}$, starting from $(a_0^{(3)})^2$ and therefore select $\mathbb{P}_{a_0^{(3)}} = \mathbb{P}_5$.

The multiples to erase have the form

$$a_{k_j}^{(3)} = a_0^{(3)} \cdot \left[a_0^{(3)} + \sum_{i=1}^{\ell_3} \left(1 - \chi_{x>j(i)} \right) d_i^{(3)} \right] = a_0^{(3)} \cdot a_j^{(3)}, \quad j = 0, \dots, \ell_3 - 1.$$

The sequence of distances obtained from $\mathbf{d}^{(3)}$ once we have selected the multiples to erase gives the vector $\mathbf{d}^{(5)}$, as it can be seen from the following table:

\mathbb{P}_3	$a_0^{(3)}$	$a_1^{(3)}$	$a_2^{(3)}$	$a_3^{(3)}$	$a_4^{(3)}$	$a_5^{(3)}$	$a_6^{(3)}$	$a_7^{(3)}$	$a_8^{(3)}$	$a_9^{(3)}$	$a_{10}^{(3)}$	$a_{11}^{(3)}$	$a_{12}^{(3)}$
	<u>5</u>	7	11	13	17	19	23	25	29	31	35	37	41
$d_n^{(3)}$		2	4	2	4	2	4	2	4	2	4	2	4
$D_n^{(3)}$		$\mathbf{d}^{(3)}$										$(D_1^{(3)}, D_2^{(3)})$	
\mathbb{P}_5		$a_0^{(5)}$	$a_1^{(5)}$	$a_2^{(5)}$	$a_3^{(5)}$	$a_4^{(5)}$	$a_5^{(5)}$		$a_6^{(5)}$	$a_7^{(5)}$		$a_8^{(5)}$	$a_9^{(5)}$
		<u>7</u>	11	13	17	19	23		29	31		37	41
$d_n^{(5)}$			4	2	4	2	4		6	2		6	4
\mathbb{P}_3		$a_{13}^{(3)}$	$a_{14}^{(3)}$	$a_{15}^{(3)}$	$a_{16}^{(3)}$	$a_{17}^{(3)}$	$a_{18}^{(3)}$	$a_{19}^{(3)}$	$a_{20}^{(3)}$	$a_{21}^{(3)}$	$a_{22}^{(3)}$	$a_{23}^{(3)}$...
		43	47	49	53	55	59	61	65	67	71	73	
		2	4	2	4	2	4	2	4	2	4	2	
		$(D_3^{(3)}, \dots, D_{10}^{(3)})$...	
\mathbb{P}_5		$a_{10}^{(5)}$	$a_{11}^{(5)}$	$a_{12}^{(5)}$	$a_{13}^{(5)}$		$a_{14}^{(5)}$	$a_{15}^{(5)}$		$a_{16}^{(5)}$	$a_{17}^{(5)}$	$a_{18}^{(5)}$...
		43	47	49	53		59	61		67	71	73	
		2	4	2	4		6	2		6	4	2	

TABLE T2: \mathbb{P}_3 and \mathbb{P}_5 with their vectors $\mathbf{d}^{(3)}$, $\mathbf{d}^{(5)}$ and \mathbf{d}^3 with the previous period repeated.

In analogy with this procedure, we give another instance, to better illustrate the recurrence.

By construction we have $\mathbf{d}^{(5)} = (4, 2, 4, 2, 4, 6, 2, 6) \in \mathbb{N}^{\ell_5}$ with

$$\ell_5 = 8 = \ell_3 \cdot (5 - 1), \text{ and } M_5 = 30 = M_3 \cdot 5.$$

The vector $\mathbf{d}^{(5)}$ is clearly symmetric but the tail, $(6, 2, 6)$, which is symmetric as well.

Moreover, for all $a_k^{(5)} \in \mathbb{P}_5$ there exists $j = 0, \dots, \ell_5 - 1$ such that it can be written as

$$a_k^{(5)} = a_0^{(5)} + C \sum_{i=1}^{\ell_5} d_i^{(5)} + \sum_{i=1}^{\ell_5} \left(1 - \chi_{x>j}(i)\right) d_i^{(5)},$$

with C , χ and j as before.

Reiterating this procedure we notice that whenever we pass from \mathbb{P}_{p_n} to $\mathbb{P}_{p_{n+1}}$ we erase the multiples of $a_0^{(p_n)} = p_{n+1}$ of the form $p_{n+1} \cdot a_j^{(p_n)}$, $j = 0, \dots, \ell_{p_n} - 1$

$$(3.1) \quad a_{k_j}^{(p_n)} = p_{n+1} \cdot \left[p_{n+1} + \sum_{i=1}^{\ell_{p_n}} \left(1 - \chi_{x>j}(i)\right) d_i^{(p_n)} \right], \text{ for some } k_j.$$

Let ℓ_{p_n} be the number of elements in the fundamental period, this means that we look at the first ℓ_{p_n} elements in \mathbb{P}_{p_n} that are those identified by $\mathbf{d}^{(p_n)}$ except the final point (we start considering the multiples from $(p_{n+1})^2$). Since this vector is symmetric but a symmetric tail, the distances between two subsequent multiples are symmetric as well along the elements identified by $\mathbf{d}^{(p_n)}$.

From this construction we can easily derive the vector $\mathbf{d}^{(p_{n+1})}$ and we have that

$$\ell_{p_{n+1}} = \ell_{p_n} \cdot p_{n+1} - \ell_{p_n} = \ell_{p_n} \cdot (p_{n+1} - 1),$$

then we find

$$\ell_{p_{n+1}} = \prod_{i \leq n+1, p_i \text{ prime}} (p_i - 1);$$

and

$$M_{p_{n+1}} = p_{n+1} \cdot M_{p_n},$$

therefore $M_{p_{n+1}} = p_{n+1}!$.

Moreover, for every $a_k^{(p_n)} \in \mathbb{P}_{p_n}$ there exists $j = 0, \dots, \ell_{p_n} - 1$ such that

$$(3.2) \quad a_k^{(p_n)} = a_0^{(p_n)} + C \sum_{i=1}^{\ell_{p_n}} d_i^{(p_n)} + \sum_{i=1}^{\ell_{p_n}} \left(1 - \chi_{x>j}(i)\right) d_i^{(p_n)},$$

where $C = \left[\frac{a_k^{(p_n)} - a_0^{(p_n)}}{M_{p_n}} \right]$ and $\chi_{x>b}$ the Heaviside function on the set $(b, +\infty)$.

Remark 3.1. This construction gives an evidence of the following important result: every prime number p determines a symmetric frequency $\mathbf{d}^{(p)}$ along the natural numbers \mathbb{N} which generates \mathbb{P}_p . Moreover it identifies another symmetric frequency, $\mathbf{d}_{mult}^{(p)}$, which interacts with the first originating a new symmetric frequency $\mathbf{d}^{(q)}$, with $q = \inf\{a > p \text{ s.t. } a \text{ is prime}\}$ and then \mathbb{P}_q .

The superposition of all these symmetric frequencies generates the distribution of primes.

3.1. Historical remarks

The function defining the length ℓ_{p_n} of the vector $\mathbf{d}^{(p_n)}$ has already been studied, see Euler, [4], [3], [5], [2], and it is known as *totient or Euler function*. It is defined as follows.

Let $\{p_i\}_{i=1, \dots, n}$ be a sequence of prime numbers in increasing order and M_{p_n} defined as before; then the number of integers $m \in \mathbb{N}$ such that $m \leq M_{p_n}$ and $(m, M_{p_n}) = 1$ is given as

$$(3.3) \quad \phi(M_{p_n}) := \prod_{p_i} (p_i - 1) = \ell_{p_n}.$$

This function could also be expressed in a different way: for any $n \in \mathbb{N}$ we have

$$\phi(n) = \#\{m \in \mathbb{N} \text{ s. t. } m \leq n, (m, n) = 1\};$$

then

$$\phi(M_{p_n}) = \#\{m \in \mathbb{P}_{p_n}, m \leq M_{p_n}\}.$$

This set includes integers which are either primes greater than p_n or have all their prime factors greater than p_n .

We can notice that for any p prime, $\phi(p) = p - 1$ since all $n \in \mathbb{N}$, $n < p$ are such that $(n, p) = 1$.

In 1849, A. De Polignac [1] observed similar properties of periodicity in the subsets of \mathbb{N} obtained after the sieving procedure: erasing all the multiples of 2 and 3 he obtained the so called “table p_2 ”

$$(0) 1 (2) (3) (4) 5 (6) 7 (8) (9) (10) 11 \dots$$

If we count the number of terms in every group of consecutive deleted numbers we get the sequence

$$1, 3, 1, 3, 1, \dots$$

which forms the *diatomic series of 3*.

In a similar way, after deleting all the multiples of the first n primes we get the table p_n and the diatomic series of the n -th prime p_n .

We can observe that the terms after 1 of the period are symmetrical distributed while the middle term is 3. De Polignac indicated with

$$\pi_n = p_n !,$$

the product of the first n primes, $\phi(\pi_n)$ the number of terms in the period and observed that their sum is $\pi_n - \phi(\pi_n)$: using our notation

$M_n - \ell_n$ = sum of the terms in the fundamental period of the diatomic series of p_n ;

moreover this number indicates the number of integers less than π_n which are divisible by one or more primes p_i , $i \leq n$.

Under these premises, we formalize the procedure with an algorithm and we derive our main result.

4. Algorithm

In this section we briefly describe the algorithm to produce the vector $\mathbf{d}^{(p)}$ from $\mathbf{d}^{(q)}$, with $q = \sup\{a < p \text{ s.t. } a \text{ is prime}\}$. The purpose of this section is to give the hint of the iterative procedure we are presenting. Theoretical results and rigorous proofs are postponed in Section 5.

In Section 3 we have seen that, given the couple $(p, \mathbf{d}^{(p)})$, p prime, we can completely determine the set \mathbb{P}_p . The procedure to obtain $\mathbf{d}^{(p_{n+1})}$ is quite simple and straightforward: starting from $a_0^{(p_n)} = p_{n+1}$, we repeat the vector $\mathbf{d}^{(p_n)}$ p_{n+1} -times, obtaining $\mathbf{d}^{(p_n)}$

$$\mathbf{d}^{(p_n)} = \underbrace{(\mathbf{d}^{(p_n)}, \mathbf{d}^{(p_n)}, \dots, \mathbf{d}^{(p_n)})}_{p_{n+1} \text{ - times}}.$$

Then we identify the multiples to erase: they are given by (3.1),

$$a_{k_j}^{(p_n)} = a_0^{(p_n)} \cdot \left(a_0^{(p_n)} + \sum_{i=1}^{\ell_{p_n}} (1 - \chi_{x>j}(i)) d_i^{(p_n)} \right), \text{ for all } j = 0, \dots, \ell_{p_n} - 1.$$

The vector $\mathbf{d}^{(p_{n+1})}$ is obtained from $\mathbf{d}^{(p_n)}$ modifying the values corresponding to the positions k_j . Let

$$(4.1) \quad \begin{aligned} \mathcal{T}_n^{n+1} : \mathbb{N}^{p_{n+1} \cdot \ell_{p_n}} &\longrightarrow \mathbb{N}^{p_{n+1}!}, \\ \mathbf{d}^{(p_n)} &\longrightarrow \mathcal{T}_n^{n+1} \mathbf{d}^{(p_n)} := \mathbf{d}^{(p_{n+1})}, \end{aligned}$$

the *transition operator* defined by the following relations:

- for $1 \leq i \leq k_0$, $d_i^{(p_{n+1})} = d_{i+1}^{(p_n)}$,
- for $i = k_j$, $j = 0, \dots, \ell_{p_n} - 1$, $d_i^{(p_{n+1})} = d_{i+j+1}^{(p_n)} + d_{i+j+2}^{(p_n)}$,
- for $k_j < i < k_{j+1}$, $j = 0, \dots, \ell_{p_n} - 1$, $d_i^{(p_{n+1})} = d_{i+j+2}^{(p_n)}$.

From equation (3.1) we derive the vector of distances between the multiples, $\mathbf{d}_{mult}^{(p_n)}$. It is easy to see that this vector is symmetric as well with a symmetric tail by definition:

$$d_{mult,j}^{(p_n)} = a_{k_j}^{(p_n)} - a_{k_{j-1}}^{(p_n)} = a_0^{(p_n)} \cdot d_j^{(p_n)}.$$

5. Theoretical results

Theorem 5.1. *Let $\{p_n\}_n$ be the sequence of prime numbers arranged in increasing order, \mathbb{P}_{p_n} defined as in (2.1), $\mathbf{d}^{(p_n)}$ the vector of the distances associated to p_n . Then*

$$(5.1) \quad \ell_{p_n} = \prod_{i \leq n, p_i \text{ prime}} (p_i - 1), \quad M_{p_n} = p_n!.$$

Proof. This result follows easily by construction.

We notice from the previous examples that the following relations hold:

$$\begin{aligned} p_1 = 2, \quad M_2 = 2, & & \ell_2 = 1, \\ p_2 = 3, \quad M_3 = 6 = 3 \cdot 2, & & \ell_3 = 2 = (3 - 1) \cdot (2 - 1), \\ p_3 = 5, \quad M_5 = 30 = 5 \cdot 3 \cdot 2, & & \ell_5 = 8 = (5 - 1) \cdot (3 - 1) \cdot (2 - 1). \end{aligned}$$

Given the set \mathbb{P}_{p_n} , we construct the set $\mathbb{P}_{p_{n+1}}$ identifying all the multiples of $a_0^{(p_n)} = p_{n+1}$. We have seen that it suffices to find the first ℓ_{p_n} multiples starting from $(p_{n+1})^2$: they are located along the sequence identified by repeating the vector $\mathbf{d}^{(p_n)}$ for p_{n+1} -times, therefore

$$\ell_{p_{n+1}} = \prod_{i \leq n+1, p_i \text{ prime}} (p_i - 1);$$

Observing that M_{p_n} is the sum of the components of the vector $\mathbf{d}^{(p_n)}$, reiterating this procedure we get

$$M_{p_{n+1}} = p_{n+1} \cdot M_{p_n} = \dots = p_n! . \quad \blacksquare$$

Theorem 5.2. *Let $\mathbf{d}^{(p_n)}$ be the vector associated to \mathbb{P}_{p_n} , p_n prime. Then $\mathbf{d}^{(p_n)}$ is symmetric but a tail of three points, that is*

$$(5.2) \quad d_i^{(p_n)} = d_{\ell_{p_n}-2-i}^{(p_n)}, \quad \text{for all } i = 1, \dots, \ell_{p_n} - 3.$$

Moreover

$$(5.3) \quad (d_{\ell_{p_n}-2}^{(p_n)}, d_{\ell_{p_n}-1}^{(p_n)}, d_{\ell_{p_n}}^{(p_n)}) = (a_0^{(p_n)} - 1, 2, a_0^{(p_n)} - 1).$$

Proof. The proof of this result follows by induction.

Consider $p_3 = 5$. From Table T2 we have $\mathbf{d}^{(5)} = (4, 2, 4, 2, 4, 6, 2, 6)$, therefore it verifies the statement. Suppose now that the property is true for the set \mathbb{P}_{p_n} ; we prove that it is true for $\mathbb{P}_{p_{n+1}}$.

The property of symmetry for $\mathbf{d}^{(p_{n+1})}$ follows easily from the symmetry of $\mathbf{d}^{(p_n)}$ and the symmetry of the vector of distances between the multiples. We focus now on the tail.

By assumption $\mathbf{d}^{(p_n)} = (d_1^{(p_n)}, \dots, d_{\ell_{p_n}-3}^{(p_n)}, p_{n+1} - 1, 2, p_{n+1} - 1)$. The tail part of the vector $\mathbf{d}^{(p_{n+1})}$ comes from $\mathbf{d}^{(p_n)}$ once we have erased the last two multiples of p_{n+1} , corresponding to $d_{\ell_{p_n}-2}^{(p_n)}$ and $d_{\ell_{p_n}-1}^{(p_n)}$. By assumption

$$d_{\ell_{p_n}-2}^{(p_n)} = p_{n+1} - 1, \quad d_{\ell_{p_n}-1}^{(p_n)} = 2.$$

Let $k_{\ell_{p_n}-2}$ and $k_{\ell_{p_n}-1}$ be the index of the multiple to be erased. By definition (3.1) we have

$$(5.4) \quad \begin{aligned} a_{k_{\ell_{p_n}-2}}^{(p_n)} &= p_{n+1} \cdot \left(p_{n+1} + \sum_{i=1}^{\ell_{p_n}-2} d_i^{(p_n)} \right), \\ a_{k_{\ell_{p_n}-1}}^{(p_n)} &= p_{n+1} \cdot \left(p_{n+1} + \sum_{i=1}^{\ell_{p_n}-1} d_i^{(p_n)} \right); \end{aligned}$$

on the other hand, using equation (3.2) we can write

$$(5.5) \quad \begin{aligned} a_{k_{\ell_{p_n}-2}}^{(p_n)} &= p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^*} d_i^{(p_n)}, \\ a_{k_{\ell_{p_n}-1}}^{(p_n)} &= p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^{**}} d_i^{(p_n)}, \end{aligned}$$

where j^* and j^{**} are the corresponding indexes. We notice that

$$a_{k_{\ell_{p_n}-1}}^{(p_n)} = p_{n+1} + p_{n+1} \cdot M_{p_n},$$

therefore, comparing with (5.5) we get $j^{**} = \ell_{p_n}$.

We evaluate now the distance between $a_{k\ell_{p_n-1}}^{(p_n)}$ and $a_{k\ell_{p_n-2}}^{(p_n)}$, using equations (5.5) and (5.4) we get

$$\sum_{i=j^{*}+1}^{\ell_{p_n}} d_i^{(p_n)} = p_{n+1} \cdot d_{\ell_{p_n-1}}^{(p_n)} = 2p_{n+1}.$$

From the inductive assumption on the tail of $\mathbf{d}^{(p_n)}$

$$2p_{n+1} = d_{\ell_{p_n-2}}^{(p_n)} + d_{\ell_{p_n-1}}^{(p_n)} + d_{\ell_{p_n}}^{(p_n)},$$

therefore, comparing with the previous equality, we derive $j^* = \ell_{p_n} - 3$.

We can now evaluate the tail of $\mathbf{d}^{(p_{n+1})}$. By definition, the $(\ell_{p_{n+1}} - 2)$ -th element of the vector is given as the distance between the corresponding elements in \mathbb{P}_{p_n} , once we have erased the $(\ell_{p_{n+1}} - 2)$ -th multiple of p_{n+1} , therefore

$$\begin{aligned} d_{\ell_{p_{n+1}-2}}^{(p_{n+1})} &= a_{k\ell_{p_n-2}+1}^{(p_n)} - a_{k\ell_{p_n-2}-1}^{(p_n)} = \\ &= p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^*+1} d_i^{(p_n)} - \\ &\quad - \left[p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^*-1} d_i^{(p_n)} \right] = \\ &= d_{j^*+1}^{(p_n)} + d_{j^*}^{(p_n)} = d_{\ell_{p_n-2}}^{(p_n)} + d_{\ell_{p_n-3}}^{(p_n)}. \end{aligned}$$

Using now the inductive symmetry of $\mathbf{d}^{(p_n)}$ and the assumption on the tail we have

$$d_{\ell_{p_{n+1}-2}}^{(p_{n+1})} = p_{n+1} - 1 + d_1^{(p_n)} = p_{n+2} - 1.$$

Analogously we have

$$\begin{aligned} d_{\ell_{p_{n+1}-1}}^{(p_{n+1})} &= a_{k\ell_{p_n-2}+2}^{(p_n)} - a_{k\ell_{p_n-2}+1}^{(p_n)} = \\ &= p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^*+2} d_i^{(p_n)} - \\ &\quad - \left[p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^*+1} d_i^{(p_n)} \right] = \\ &= d_{j^*+2}^{(p_n)} = d_{\ell_{p_n-1}}^{(p_n)} = 2, \end{aligned}$$

and

$$\begin{aligned}
d_{\ell_{p_{n+1}}}^{(p_{n+1})} &= a_{k_{\ell_{p_n-1}+1}}^{(p_n)} - a_{k_{\ell_{p_n-1}-1}}^{(p_n)} = \\
&= p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^{**}+1} d_i^{(p_n)} - \\
&\quad - \left[p_{n+1} + (p_{n+1} - 1)M_{p_n} + \sum_{i=1}^{j^{**}-1} d_i^{(p_n)} \right] = \\
&= \sum_{i=j^{**}}^{j^{**}+1} d_i^{(p_n)} = d_{\ell_{p_n}}^{(p_n)} + d_1^{(p_n)} = p_{n+2} - 1,
\end{aligned}$$

where we have used that “ $d_{\ell_{p_n+1}}^{(p_n)}$ ” = $d_1^{(p_n)}$. ■

Remark 5.3. The symmetry of the vector $\mathbf{d}^{(p)}$ and of its tail is presented in Figures 1, 2 and 3 for the case of $p = 5, 7, 11$ respectively.

Corollary 5.4. Let $\mathbf{d}^{(p_n)}$ be the symmetric vector associated to \mathbb{P}_{p_n} , p_n prime. The center of the first symmetric part (that is, omitting the tail of three entries) is

$$d_{k_n^*}^{(p_n)} = 4, \quad k_n^* = \frac{\ell_{p_n} - 2}{2},$$

and it corresponds to the point

$$(5.6) \quad a_{k_n^*}^{(p_n)} = c^{(p_n)} = \frac{p_n!}{2} + 2.$$

Proof. In order to prove the first part we focus our attention on the point $\frac{p_n!}{2}$: as a matter of fact, it is the middle point of the segment, in the continuous line, identified by the first symmetric part of $\mathbf{d}^{(p_n)}$, $[p_{n+1}, p_n! - p_{n+1}]$. We notice that $\frac{p_n!}{2}$ is an odd number, by construction; moreover it is a multiple of p_n therefore it does not belong to \mathbb{P}_{p_n} . The numbers $\frac{p_n!}{2} \pm 1$ are even numbers by construction, therefore they do not belong to \mathbb{P}_{p_n} either.

We focus now on the numbers $\frac{p_n!}{2} \pm 2$: we claim that they belong to \mathbb{P}_{p_n} , that is they are coprime with $p_n!$. We proceed by contradiction.

Assume that they are not coprime, then there exist $q \leq p_n$, q prime, such that

$$q \text{ divides } p_n! \quad \text{and} \quad q \text{ divides } \frac{p_n!}{2} \pm 2.$$

From the first we get $q = p_j$ for some $j = 0, \dots, n$, while from the second we get $q \neq p_j$ for all $j = 1, \dots, n$, therefore $q = p_0 = 2$. This leads to a contradiction since $\frac{p_n!}{2} \pm 2$ is an odd number, therefore $\frac{p_n!}{2} \pm 2 \in \mathbb{P}_{p_n}$.

We have already shown that $\frac{p_n!}{2} \pm 2$ are consecutive numbers in \mathbb{P}_{p_n} , therefore there exists $k_n^* \in \mathbb{N}$ such that

$$a_{k_n^*-1}^{(p_n)} = \frac{p_n!}{2} - 2, \quad a_{k_n^*}^{(p_n)} = \frac{p_n!}{2} + 2,$$

then

$$d_{k_n^*}^{(p_n)} = 4.$$

The remaining properties follow easily from the results of Theorems 5.1 and 5.2: the mass of the tail is

$$M_{p_n, tail} = 2(p_{n+1} - 1) + 2 = 2p_{n+1},$$

therefore we can derive the mass of every symmetric part:

$$M_{p_n, symm} = \frac{1}{2} (M_{p_n} - 2p_{n+1} - 4);$$

the center is given as

$$a_{k_n^*}^{(p_n)} = c^{(p_n)} = p_{n+1} + M_{p_n, symm} + d_{k_n^*}^{(p_n)} = \frac{p_n!}{2} + 2. \quad \blacksquare$$

Remark 5.5. The vector $\mathbf{d}^{(p_n)}$ is fundamental in finding twin and cousin numbers.

By definition, for every i , $d_i^{(p_n)}$ gives the distance between $a_i^{(p_n)}$ and $a_{i-1}^{(p_n)}$.

If $a_i^{(p_n)}, a_{i-1}^{(p_n)} < \min((p_{n+1})^2, c^{(p_n)})$ and $d_i^{(p_n)} = 2$ or $d_i^{(p_n)} = 4$ then $a_i^{(p_n)}$ and $a_{i-1}^{(p_n)}$ are respectively twin and cousin numbers. Moreover, because of the property of symmetry of $\mathbf{d}^{(p_n)}$ we have a necessary condition on where we can find twin and cousin numbers $a_j^{(p_n)}, a_{j-1}^{(p_n)} > \min((p_{n+1})^2, c^{(p_n)})$, $d_j^{(p_n)}$ symmetric with $d_i^{(p_n)}$.

As a matter of fact, at this stage, we cannot say if $a_j^{(p_n)}, a_{j-1}^{(p_n)}$ are primes, and probably they are not, but this approach gives a first indication on where we cannot find for sure twins and cousin numbers: if $d_i^{(p_n)} > 2$, then the corresponding symmetric points, even if prime, are not twins numbers.

Corollary 5.6. Let p_n be a prime number and $\mathbf{d}^{(p_n)} \in \mathbb{N}^{\ell_{p_n}}$ its characteristic vector. For every $a_j^{(p_n)} \in \mathbb{P}_{p_n}$, $j = 1, \dots, k_n^* - 1$ its symmetric point is $a_{\ell_{p_n}-j-2}^{(p_n)}$ and it is given by

$$(5.7) \quad a_{\ell_{p_n}-j-2}^{(p_n)} = a_0^{(p_n)} + M_{p_n, symm} + d_{k_n^*}^{(p_n)} + \sum_{h=j}^{k_n^*-1} d_h^{(p_n)}.$$

Proof. The proof of this result is quite easy and straightforward. We have observed that the symmetry property of the vector $\mathbf{d}^{(p_n)}$ gives that for any $j=1, \dots, k_n^* - 1$

$$d_j^{(p_n)} = d_{\ell_{p_n}-2-j}^{(p_n)};$$

moreover, by definition

$$a_j^{(p_n)} - a_{j-1}^{(p_n)} = d_j^{(p_n)},$$

therefore it follows that

$$a_j^{(p_n)} - a_{j-1}^{(p_n)} = a_{\ell_{p_n}-j-2}^{(p_n)} - a_{\ell_{p_n}-j-3}^{(p_n)}.$$

This relation leads to the thesis: $a_{\ell_{p_n}-j-2}^{(p_n)}$ is symmetric to $a_j^{(p_n)}$. Using equation (3.2) and the results of Corollary 5.4, we have

$$a_{\ell_{p_n}-2-j}^{(p_n)} = a_0^{(p_n)} + M_{p_n, symm} + d_{k_n^*}^{(p_n)} + \sum_{h=k_n^*+1}^{\ell_{p_n}-2-j} d_h^{(p_n)};$$

using the symmetry property of the vector $\mathbf{d}^{(p_n)}$ we get the thesis. ■

6. Conclusion

The results presented in this paper give us the knowledge on the distribution of prime numbers and on how to find them. In particular, we have shown that every prime number p is associated to a vector $\mathbf{d}^{(p)} \in \mathbb{N}^{\ell_p}$ which is symmetric but a symmetric tail of three points. Once we know a sequence of prime numbers up to a prime p_n we know the (infinite) set \mathbb{P}_{p_n} of coprimes with p_n only by periodical repetition of the fundamental vector $\mathbf{d}^{(p_n)}$. Using this approach we can reduce our search of prime numbers to the study of finite sets, because all the properties of the set \mathbb{P}_{p_n} are contained in the vector $\mathbf{d}^{(p_n)}$. To be more precise, the knowledge of the couple $(p, \mathbf{d}^{(p)})$, p prime, is sufficient to identify all the other primes.

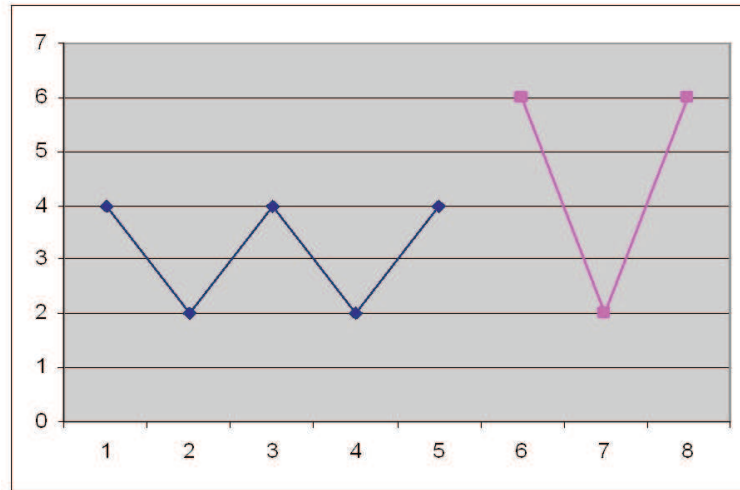


Figure 1: the vector $\mathbf{d}^{(5)}$. On the x -axis we write the index of coordinate i , on the y -axis the corresponding value $d_i^{(5)}$. It is easy to recognize the symmetric part and the tail.

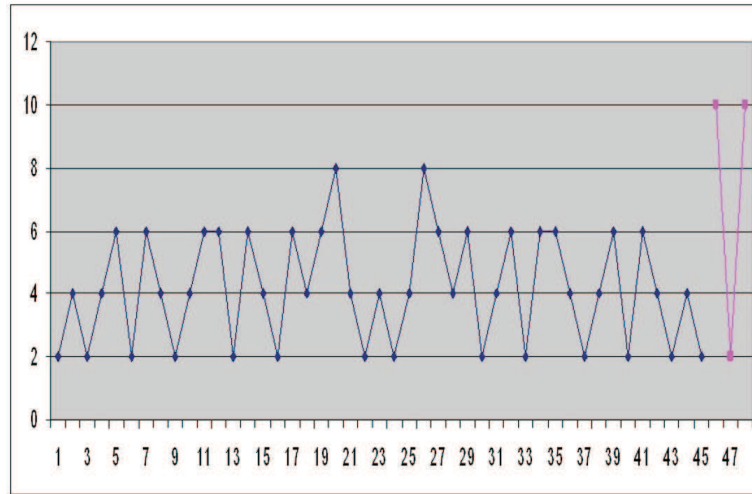


Figure 2: the vector $\mathbf{d}^{(7)}$ as a function of the coordinate i . The symmetric part and the tail are preserved.

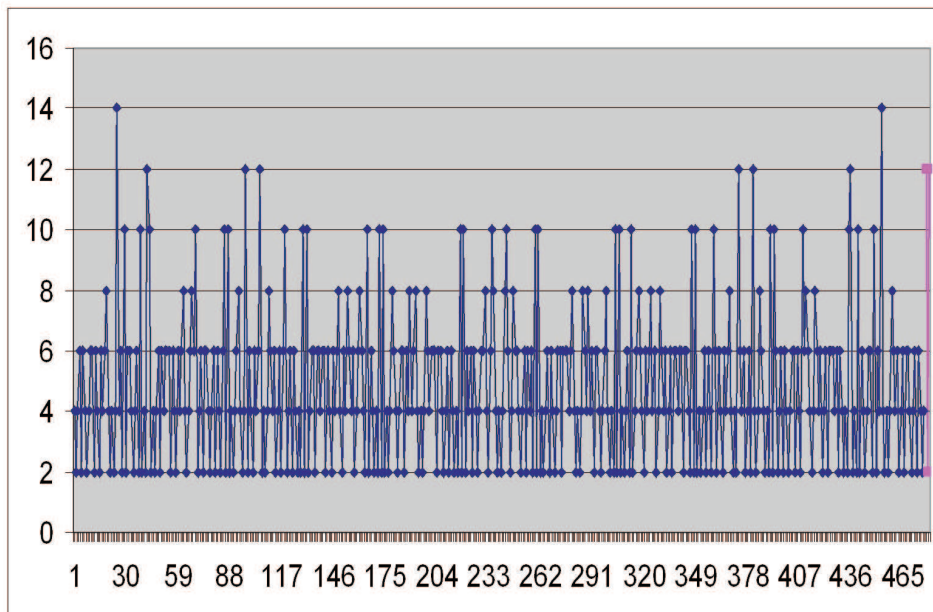


Figure 3: the vector $\mathbf{d}^{(11)}$. The x -axis is the index of coordinate i , the y -axis the value of the corresponding $d_i^{(11)}$. The dimension of the vector is 480, but we can still notice the symmetric part and the three points tail.

References

- [1] DE POLIGNAC, A., *Recherches nouvelles sur les nombres premiers*, C. R. Acad. Sci. Paris Math., 29 (1849), 397–401 and 738–739.
- [2] DICKSON, L.E., *History of the Theory of Numbers*, vol. I: *Divisibility and Primality*, Carnegie Institute of Washington, Publication No. 256, 1919. Reprinted by Chelsea, New York, 1971.
- [3] EULER, L., *Opera postuma*.
- [4] EULER, L., *Variae observationes circa series infinitas*, St. Petersburg Acad., Russia, 1737.
- [5] EULER, L., *Posthumous paper*, Comm. Arith. Coll., 2(512) (1862), 134–136.
- [6] RIEMANN, B., *Uebere die anzahl der primzahlen unter einer gegebenen grösse*, Gesammelte math. Werke und wissensch. Nachlass, 2, Aufl. 1892.

Accepted: 10.01.2007