symmetry

MDPI

Article

# An Investigation on the Prime and Twin Prime Number Functions by Periodical Binary Sequences and Symmetrical Runs in a Modified Sieve Procedure

**Bruno Aiazzi *, Stefano Baronti, Leonardo Santurri and Massimo Selva**

Institute of Applied Physics "Nello Carrara", IFAC-CNR, Research Area of Florence, 50019 Sesto Fiorentino, Italy; s.baronti@ifac.cnr.it (S.B.); l.santurri@ifac.cnr.it (L.S.); m.selva@ifac.cnr.it (M.S.)

* Correspondence: b.aiazzi@ifac.cnr.it; Tel.: +39-055-5226451

Received: 18 April 2019; Accepted: 4 June 2019; Published: 10 June 2019

**Abstract:** In this work, the Sieve of Eratosthenes procedure (in the following named Sieve procedure) is approached by a novel point of view, which is able to give a justification of the Prime Number Theorem (P.N.T.). Moreover, an extension of this procedure to the case of twin primes is formulated. The proposed investigation, which is named Limited INtervals into PEriodical Sequences (LINPES) relies on a set of binary periodical sequences that are evaluated in limited intervals of the prime characteristic function. These sequences are built by considering the ensemble of deleted (that is, 0) and undeleted (that is, 1) integers in a modified version of the Sieve procedure, in such a way a symmetric succession of runs of zeroes is found in correspondence of the gaps between the undeleted integers in each period. Such a formulation is able to estimate the prime number function in an equivalent way to the logarithmic integral function $\mathrm{Li}(x)$. The present analysis is then extended to the twin primes, by taking into account only the runs whose size is two. In this case, the proposed procedure gives an estimation of the twin prime function that is equivalent to the one of the logarithmic integral function $\mathrm{Li}_2(x)$. As a consequence, a possibility is investigated in order to count the twin primes in the same intervals found for the primes. Being that the bounds of these intervals are given by squares of primes, if such an inference were actually proved, then the twin primes could be estimated up to infinity, by strengthening the conjecture of their never-ending.

**Keywords:** prime numbers; Prime Number Theorem (P.N.T.); modified Sieve procedure; binary periodical sequences; prime number function; prime characteristic function; limited intervals; logarithmic integral estimations; twin prime numbers

## 1. Introduction

The Sieve procedure is able to achieve heuristic justifications of the Prime Number Theorem (P.N.T.) [1]. Such a theorem gives the asymptotic trend of the prime number function $\pi(x)$, where $\pi(x)$ denotes the quantity of prime numbers $p$ less or equal to $x \in \mathbb{R}$, that is,

$$\pi(x) = \text{ number of primes } p,\ p \leq x. \tag{1}$$

Let $\log(x)$ be the natural logarithm of $x$. If the real functions $A(x)$ and $B(x)$ are asympthotically equal, that is, $\lim_{x\to\infty} A(x)/B(x) = 1$, then we say that $A(x)$ and $B(x)$ are equivalent as $x \to \infty$, and we write $A(x) \sim B(x)$. Consequently, the P.N.T. can be written as

$$\pi(x) \sim x/\log(x). \tag{2}$$

After the infinitude of primes was recognized since ancient times, the estimation (2) was conjectured by Gauss [2] and Legendre [3] at the end of the $18th$ century. Gauss himself improved Equation (2), by considering the logarithmic integral function Li(x), which is defined as

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}. \tag{3}$$

Again, the function (3) is such that

$$\pi(x) \sim \mathrm{Li}(x) \tag{4}$$

but the approximation (4) is much more precise than (2). In fact, it can be demonstrated that the piece $x/\log(x)$ is only the first term of the series expansion of (3). The aim of this work is to introduce a novel heuristic procedure (LINPES, Limited INtervals into PEriodical Sequences) that is equivalent to the $\mathrm{Li}(x)$ approximation, in the sense of Equation (4), apart from a simple multiplicative constant, by exploiting some binary periodic sequences, and related symmetrical runs. Pieces of these sequences compose limited intervals of the prime characteristic function $\xi_p(n)$, which is defined as

$$\xi_p(n) = \begin{cases} 1 & \text{if } n \text{ is prime} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

As a matter of fact, a topic that is very much discussed nowadays in the literature just concerns the possible discovering of some regularities and periodicities in the distribution of the primes in certain intervals of the integer sequence [4]. In this work, the implications of the LINPES procedure are also investigated, in particular with an extension to the twin primes, whose distribution is given by a function known as twin prime function $\pi_2(x)$, which is similar to (1), that is,

$$\pi_2(x) = \text{number of pairs of twin primes } (p,\ p+2),\ \ p \le x. \tag{6}$$

Unlike the case of primes, the infinitude of twin primes is still unproved. However, analogously to the P.N.T., the density of the twin primes has been conjectured [5], by considering that the probability to be a prime of an integer $n$ is equal to $1/\log(n)$. Consequently, the probability that $n$ and $n+2$ are both prime can be computed, in such a way the strong twin prime conjecture[6] gives an equivalence between the twin prime function $\pi_2(x)$ and the logarithmic integral function $\mathrm{Li}_2(x)$, that is,

$$\pi_2(x) \sim C\,\mathrm{Li}_2(x) \tag{7}$$

where $\mathrm{Li}_2(x)$ is defined as

$$\mathrm{Li}_2(x) = \int_2^x \frac{dt}{(\log t)^2} \tag{8}$$

and $C = 2\,\Pi_2 \simeq 1.3203$ is a multiplicative constant that takes into account the statistical dependence of the primes $n$ and $n+2$ [5]. The related constant $\Pi_2 \simeq 0.6602$ is named twin prime constant, that is,

$$\Pi_2 = \prod_{p>2,\, p\,\text{prime}} \left(1 - \frac{1}{(p-1)^2}\right). \tag{9}$$

As it will be shown later, the proposed LINPES procedure is able to estimate the twin prime function in an equivalent way as the $\mathrm{Li}_2(x)$ function, apart from a multiplicative constant. However, this is made by admitting that a basic relation, which is true for the primes, is also valid for the twin primes. In this case, the contribution of the present work will be a more probable assertion of the infinitude of twin primes.

Before starting our discussion, we itemize the variables utilized in this paper

- $\pi(x)$: prime number function (1)
- $\mathrm{Li}(x)$: logarithmic integral function (3), which leads to an estimation of $\pi(x)$
- $\pi_2(x)$: twin prime number function (6)
- $\mathrm{Li}_2(x)$: logarithmic integral function (8), which leads to an estimation of $\pi_2(x)$
- $\pi(N)$: prime number function computed in the fixed integer $N$
- $p$: generic prime number
- $p(n)$: arithmetic function that gives the succession of primes
- $\xi_p(n)$: arithmetic function that gives the characteristic function of primes (5)
- $R_s(n)$: number of residual integers in the $n-th$ step of the Sieve procedure
- $\pi_R(N)$: estimation of $\pi(N)$ given by the heuristic method of Section 2
- $\xi(k,n)$: approximation of $\xi_p(n)$ after the $k-th$ step of the Sieve procedure
- $\psi(k,n)$: periodic binary sequence obtained in the $k-th$ step of the modified Sieve procedure
- $T(k)$: period of the periodic binary sequence $\psi(k,n)$
- $J(k,n)$: sliding interval whose size is the same of $I(k)$ and whose initial point is given by $n$
- $S(k)$: size of the interval $I(k)$
- $R(k)$: number of residual runs of zeroes in each period $T(k)$
- $L(m,k)$: size of the $m-th$ run of zeroes in each period $T(k)$
- $I(k) = [p(k)^2, p(k+1)^2)$: interval of $\xi_p(n)$ where a piece of $\psi(k,n)$ is stored
- $D(k,n)$: local density of the residual runs of zeroes by moving a sliding interval $J(k,n)$ in $T(k)$
- $\overline{D}(k)$: average density of the residual runs of zeroes in the period $T(k)$
- $P(k)$: estimated number of primes in the interval $I(k)$ by using the proposed procedure
- $L(k)$: estimated number of primes in $I(k)$ by using the logarithmic integral function $\mathrm{Li}(x)$
- $\pi(k)$: real number of primes in the interval $I(k)$
- $\pi_P(N)$: estimation of $\pi(N)$ by using the proposed procedure
- $\mathrm{Li}(N)$: estimation of $\pi(N)$ by using the logarithmic integral function $\mathrm{Li}(x)$
- $\widetilde{\pi}_P(N)$: corrected version of the estimation $\pi_P(N)$
- $R_2(k)$: number of runs sized $2$ in each period $T(k)$
- $\overline{D}_2(k)$: average density of the residual runs $2$ in the period $T(k)$
- $P_2(k)$: estimated number of twin primes in the interval $I(k)$ by using the proposed procedure
- $\pi_{2P}(N)$: estimation of $\pi_2(N)$ by using the proposed procedure
- $\mathrm{Li}_2(N)$: estimation of $\pi_2(N)$ by using the logarithmic integral function $\mathrm{Li}_2(x)$
- $L_2(k)$: estimated number of twin primes in $I(k)$ by using the logarithmic integral function $\mathrm{Li}_2(x)$
- $\widetilde{\pi}_{2P}(N)$: corrected version of the estimation $\pi_{2P}(N)$
- $\pi_2(k)$: real number of twin primes in the interval $I(k)$.

This paper is organized as follows: Section 2 reports a well-known heuristic method, which is able to estimate the prime number function $\pi(x)$ in the sense of (2), apart from a multiplicative constant. Section 3 shows instead how the LINPES procedure is able to obtain an estimation of $\pi(x)$ that is equivalent to the logarithmic-integral function $\mathrm{Li}(x)$. Section 4 extends the proposed procedure to the case of twin primes. Finally, future research and conclusive remarks are provided in Section 5.

## 2. A Heuristic Estimation of $\pi(x)$ Equivalent to the $x/\log(x)$ Function

In this section, a well-known heuristic method to justify the P.N.T. in a probabilistic way is briefly resumed, by starting from the Sieve procedure, which splits the primes from the composites in a list of integers up to a given number $N$. The Sieve procedure is the most common way to obtain the primes, and it is also presently a research topic in order to improve its efficiency [7]. Let $p(n)$ be the arithmetic function whose *n-th* element is the *n-th* prime, with $n \in \mathbb{N}$ [8,9]. The Sieve procedure can be summarized by the following steps:

- Step 1: List the integers in the interval $I_N = (1, N]$, with $N \in \mathbb{N}$, then put $n = 1$ and start from the lowest prime $p(n) = p(1) = 2$.
- Step 2: Cancel all the multiples of $p(n)$ not yet struck out, by starting from $p(n)^2$ up to $N$.
- Step 3: Go to the next remaining integer $q > p(n)$ in the list. If $q^2 > N$, the procedure ends, otherwise increase $n$ to $n+1$.
- Step 4: Put $p(n) = q$ and return to Step 2.

In order to directly compute the characteristic function of primes $\xi_p(n)$, we can memorize the status of each integer in a binary vector ranging from $1$ to $I_N$. In practice, we associate the value $0$ to an integer that has been struck out by the procedure, and the value $1$ otherwise. Such a vector is initialized by all $1$ values, because no integer is deleted when the procedure starts. Then, in each iteration of the Sieve procedure, a $0$ value is assigned to the cells that identify the deleted integers (that is, the composite integers). At the end of the procedure, only the cells related to the prime numbers will retain the initial $1$ value.

The Sieve procedure is able to obtain heuristic justifications of the relation (2) by considering purely probabilistic considerations [10]. To show this, let be $N$ an integer whose order of magnitude is large enough to allow sufficiently robust statistics. In the first step ($n = 1$), the multiples of $p(1) = 2$ are struck out, starting from $p(1)^2 = 4$, and the number of deleted integers is approximately given by

$$\left\lfloor \frac{N}{2} \right\rfloor - 1 \simeq \frac{N}{2}. \tag{10}$$

Therefore, the quantity of residual integers is about $R_s(1) \simeq N/2$. In the following step ($n = 2$), the multiples of $p(2) = 3$ are struck out. Given the independence of the congruences modulo $p$, where $p$ is a prime, about $1/3$ of the residual integers will be deleted (for the Chinese Remainder Theorem [9]). The updated number of the residual integers $R_s(2)$ will be given by

$$R_s(2) \simeq \left(1 - \frac{1}{2}\right) \times \left(1 - \frac{1}{3}\right) \times N. \tag{11}$$

In general, about $1 - 1/p(k)$ of the residual integers will be struck out in the $k-th$ step of the Sieve procedure. The procedure ends when the greatest prime number not exceeding $N^{1/2}$ is reached, that is, $p(K)$, where $K$ is such that $p(K)^2$ is the greatest prime square lower than $N$. At this point, we obtain an estimation $\pi_R(N)$ of the number of residual integers $R_s(K)$, and consequently of the quantity of primes $\pi(N)$, that is,

$$\pi_R(N) = \left(1 - \frac{1}{2}\right) \times \left(1 - \frac{1}{3}\right) \times \left(1 - \frac{1}{p(K)}\right) \times N = N \times \prod_{k=1}^{K} \left(1 - \frac{1}{p(k)}\right) = N \times \prod_{k=1}^{K} \frac{p(k) - 1}{p(k)}. \tag{12}$$

Let us apply the Merten's Third Theorem [11] to the reciprocal of the product structure (12), by taking the limit as $N \to \infty$, that is, as $K \to \infty$. We obtain

$$\lim_{K \to \infty} \prod_{k=1}^{K} \frac{p(k)}{p(k) - 1} \times \frac{1}{\log\left(p(K)^2\right)} = \frac{1}{2} \times e^{\gamma} \simeq \frac{1}{2} \times 1.7811 \simeq 0.8905 \tag{13}$$

where $\gamma$ is the *Eulero-Mascheroni constant*. Consequently, we can get the limit of $\pi_R(N)$ as $N \to \infty$, that is, an approximation of the limit of $\pi(N)$, by considering

$$\lim_{N \to \infty} \pi_R(N) = \lim_{N \to \infty} N \times \prod_{k=1}^{K} \frac{p(k) - 1}{p(k)} = \lim_{N \to \infty} N \times \frac{c}{\log N} = \lim_{N \to \infty} \frac{c\,N}{\log N} \tag{14}$$

that is, $\pi_R(N) \sim \frac{cN}{\log N}$, with $c = 2e^{-\gamma} \simeq 1/0.8905 \simeq 1.1229$, and being $\lim_{N\to\infty} N = \lim_{K\to\infty} p(K)^2$. Noticeably, from the relations (2) and (14), the real quantity of prime numbers in the interval $I_N = [1, N]$, is overestimated, as $N \to \infty$, by a factor $c$, due to the previous approximations.

As a conclusion, this heuristic procedure gives a justification of the P.N.T. that is equivalent to the relation (2), except for the $c$ constant [10,12]. In Section 3, the proposed LINPES procedure will be described, which gives a justification of the P.N.T. that is instead equivalent to the more precise estimation (4), by means of a procedure that is not purely probabilistic, but that is also featured by analytic considerations, which can be shared with other scientific sectors.

## 3. The LINPES Estimation of $\pi(x)$ Equivalent to the $\mathrm{Li}(x)$ Function

In this section, the novel heuristic LINPES procedure is described, by showing that it can give an estimation of the prime number function $\pi(x)$. To this end, an ensemble of periodic binary sequences will be considered in limited intervals of the prime characteristic function $\xi_p(n)$. Such a topic is of a great interest because the distribution of primes in short intervals has been deeply investigated in literature, up to the present [13,14]. The proposed procedure is also able to provide useful insights into the estimation of the trend of the twin prime number function $\pi_2(x)$. In this analysis, we denote in the following $p(0) = 1$ for convenience, even if the integer $1$ is not considered to be a prime.

### *3.1. Periodic Binary Sequences Inside the Prime Characteristic Function $\xi_p(n)$*

The occurrence of pieces of periodic binary sequences inside the prime characteristic function $\xi_p(n)$ is discussed here. To this end, both the Sieve procedure and a modified version of it are investigated step-by-step, where each step is labelled with the progressive index $k$, with $k = 0$ denoting the beginning of the two procedures. The difference between the modified and the true Sieve procedure is simply that in the Sieve procedure, in each step $k \geq 1$, only the multiplies of the prime $p(k)$ are struck out, but not the prime itself, whereas in the modified Sieve procedure the prime itself is also deleted. As previously stated, the status of each integer (0→deleted, 1→undeleted) is stored in a $N$-size vector, which is initialized with all $1$ values. The outputs of the Sieve procedure and its modified version are denoted as $\xi(k, n)$ and $\psi(k, n)$, respectively, for each step $k > 0$. Consequently, the deletion of an integer from the true or the modified Sieve procedure simply means that a $0$ value replaces a $1$ value in the two previous sequences. In the case of the Sieve procedure, the sequence $\xi(k, n)$ is an approximation at the step $k$ of the prime characteristic function $\xi_p(n)$.

At the beginning of the procedures ($k = 0$), we have two equal periodic sequences of all $1$ values, that is, $\xi(0, n)$ and $\psi(0, n)$, whose period is $T(0) = 1$. In the first step of the modified Sieve procedure ($k = 1$), the multiples of $p(1) = 2$ are struck out, including $p(1)$ itself. Consequently, we obtain a sequence $\psi(1, n)$, which is still periodic, with alternating $1$ and $0$ symbols. The period of $\psi(1, n)$ is given by the prime value $p(1)$ itself, that is, $T(1) = 2$. In the following, $T(k)$ will denote the period of the sequence $\psi(k, n)$. Conversely, in the Sieve procedure, the prime $p(1)$ is not deleted. In this case, the output sequence $\xi(1, n)$ is not periodic, but includes a piece of the periodic sequence $\psi(1, n)$, by starting from the square $p(1)^2 = 4$. Before such a value, the previous sequence $\xi(0, n)$ is preserved, which coincides with $\psi(0, n)$. It follows that $\xi(1, n)$ is a mixed sequence, being composed by pieces of both $\psi(0, n)$ and $\psi(1, n)$, that is,

$$\xi(1, n) = \begin{cases} \psi(0, n) & \text{if } p(0)^2 \leq n < p(1)^2 \\ \psi(1, n) & \text{if } n \geq p(1)^2. \end{cases} \tag{15}$$

Similarly, in the second step of the modified Sieve procedure ($k = 2$), every multiple of $p(2) = 3$, which is not yet struck out, is deleted, including the prime itself, to give the new sequence $\psi(2, n)$. Therefore, this sequence comes from the deletion of all the multiplies of the primes $p(1)$ and $p(2)$, including the primes themselves. It follows that the sequence $\psi(2, n)$ is periodic, with a period equal to the product of $p(1)$ and $p(2)$, as it will be demonstrated in Theorem 1. If we consider the second

step of the Sieve procedure, where the primes $p(1)$ and $p(2)$ have not been deleted, we obtain the sequence $\xi(2,n)$. This is again a mixed sequence, where a piece of the periodic sequence $\psi(2,n)$ is introduced, by starting from the square $p(2)^2 = 9$, whereas the previous binary values are saved before this square. Consequently, we have

$$\xi(2,n) = \begin{cases} \psi(0,n) & \text{if } p(0)^2 \leq n < p(1)^2 \\ \psi(1,n) & \text{if } p(1)^2 \leq n < p(2)^2 \\ \psi(2,n) & \text{if } n \geq p(2)^2. \end{cases} \tag{16}$$

In general, the multiples of the prime $p(k)$, which are not yet struck out in the previous steps, are deleted in the *k-th* step of the modified Sieve procedure, including the prime $p(k)$ itself. Consequently, after performing all the first $k$ steps, we obtain the periodic sequence $\psi(k,n)$, as shown in Theorem 1. In the case of the original Sieve procedure, after the *k-th* step, we obtain the sequence $\xi(k,n)$, which is an approximation of the prime characteristic function until the prime $p(k)$. Such an approximation differs from the previous one $\xi(k-1,n)$, only by starting from the square $p(k)^2$. In fact, after this point, a piece of the periodic sequence $\psi(k,n)$ is recognizable. It follows that $\xi(k,n)$ can be eventually written as a mixed sequence, which is a generalization of Equations (15) and (16), that is,

$$\xi(k,n) = \begin{cases} \psi(0,n) & \text{if } p(0)^2 \leq n < p(1)^2 \\ \psi(1,n) & \text{if } p(1)^2 \leq n < p(2)^2 \\ \dots & \\ \psi(k-1,n) & \text{if } p(k-1)^2 \leq n < p(k)^2 \\ \psi(k,n) & \text{if } n \geq p(k)^2. \end{cases} \tag{17}$$

By evaluating the expression (17), we can recognize that subsets of the periodic binary sequences $\psi(k,n)$ are present, for each $k$, in the related intervals $I(k) = [p(k)^2, p(k+1)^2)$ of the prime characteristic function. This happens until the end of the Sieve procedure, because each $k-th$ interval is not influenced by the deletions done in the following steps. We now show that the sequences $\psi(k,n)$ are periodic and that their periods are given by the product of all the primes up to $p(k)$.

**Theorem 1.** *Let be given the binary sequences $\psi(k,n)$, which are generated by the deletion of the multiplies of all the primes up to $p(k)$, including the primes themselves. Then, the sequences $\psi(k,n)$ are periodic, and their periods $T(k)$ are given by the product of all the primes up to $p(k)$, that is,*

$$T(k) = \prod_{i=1}^{k} p(i) \tag{18}$$

**Proof.** The deletion of the multiplies of all the primes up to $p(k)$ gives all the sets, as a function of $k$, of reduced residue systems modulo $T(k)$, where $T(k)$ is given by Equation (18). Each set is composed by all the positive integers relatively prime to $T(k)$, that is, by all the numbers such that $gcd\,(n, T(k)) = 1$. The quantity of integers in each set is given by the Euler phi function $\phi(T(k))$, which computes the number of positive integers less than $T(k)$ and relatively prime to $T(k)$. However, the sets of reduced residue systems are abelian groups, so that each of them is associated to a principal Dirichlet character function. This is an arithmetical function $\chi_1(k,n)$, which is nothing but $\psi(k,n)$, being defined as

$$\chi_1(k,n) = \begin{cases} 1 & \text{if } gcd(n, T(k)) = 1 \\ 0 & \text{if } gcd(n, T(k)) > 1. \end{cases} \tag{19}$$

In [8], it is proven that $\chi_1(k,n)$ is a periodic sequence, and in particular that

$$\chi_1(k, n + T(k)) = \chi_1(k, n) \qquad \forall n \tag{20}$$

This completes the proof. □

Table 1 reports the periods $T(k)$ of the sequences $\psi(k,n)$, $k = 0, \ldots, 7$, in comparison with the sizes $S(k) = p(k+1)^2 - p(k)^2$ of the intervals $I(k)$, where subsets of each $\psi(k,n)$ are recognizable. The pseudo-prime $p(0) = 1$ is put in brackets.

**Table 1.** Periods $T(k)$ of the sequences $\psi(k,n)$, for primes $p(k) \leq p(7)$, in comparison with the sizes $S(k)$ of the intervals $I(k)$. The ratios $S(k)/T(k)$ are rapidly decreasing as the prime $p(k)$ grows.

| $k$ | $p(k)$ | $p(k+1)$ | $I(k)$ | $S(k)$ | $T(k)$ | $S(k)/T(k)$ |
|---|---|---|---|---|---|---|
| (0) | (1) | 2 | $[1,4)$ | 3 | 1 | 3.000000 |
| 1 | 2 | 3 | $[4,9)$ | 5 | 2 | 2.500000 |
| 2 | 3 | 5 | $[9,25)$ | 16 | 6 | 2.666667 |
| 3 | 5 | 7 | $[25,49)$ | 24 | 30 | 0.800000 |
| 4 | 7 | 11 | $[49,121)$ | 72 | 210 | 0.342857 |
| 5 | 11 | 13 | $[121,169)$ | 48 | 2 310 | 0.020779 |
| 6 | 13 | 17 | $[169,289)$ | 120 | 30 030 | 0.003996 |
| 7 | 17 | 19 | $[289,361)$ | 72 | 510 510 | 0.000141 |

By considering the ratios $S(k)/T(k)$, it is evident that the periods $T(k)$ increase much faster than the width of the intervals $S(k)$. This makes sense because the periodicity of the sequences $\psi(k,n)$ is hardly recognizable by simply investigating the subsets of each $\psi(k,n)$ in the intervals $I(k)$.

### *3.2. The Symmetric Sequences of the Runs of Zeroes in the Periods $T(k)$*

In Section 3.1, the prime distribution has been represented as the intersection of an endless number of periodic binary sequences $\psi(k,n)$, whose periods $T(k)$ rapidly grow, and such that subsets of these sequences can be found in limited intervals $I(k)$ of the prime characteristic function $\xi_p(n)$. In particular, each of these intervals ranges between the squares of a prime $p(k)$ and of the successive $p(k+1)$. Consequently, the real primes in each interval $I(k)$ are given by the $1$ values of the correspondent sequence $\psi(k,n)$. In order to complete this analysis, we now consider the gaps between these primes, by following an established trend in literature. In particular, we are interested to investigate the distributions of the runs of zeros $R(k)$ in each period $T(k)$, being the binary sequences $\psi(k,n)$ composed by isolated ones followed by strings, more or less large, of zeroes. It follows that the quantity $R(k)$ also gives the number of undeleted integers (i.e., isolated ones) in each period $T(k)$, because the quantity $T(k)$, for $k \geq 1$, is an even number, so that the last digit of each period is a zero.

Let us consider the Sieve procedure described step-by-step in Section 3.1 and the number of runs of zeroes $R(k)$ in each period $T(k)$ of the binary sequences $\psi(k,n)$. For $k = 0, 1$, we have only one run ($R(0) = R(1) = 1$), whose sizes are $L(1,0) = 1$ and $L(1,1) = 2$, respectively. For $k = 2$, the deletion of both the multiples of $p(1)$ and $p(2)$ give two runs ($R(2) = 2$) in the period $T(2) = 6$, whose sizes are $L(1,2) = 4$ and $L(2,2) = 2$, respectively, and so on. Table 2 reports the number of runs $R(k)$ and their sizes $L(m,k)$, for $k \leq 4$, where the index $m$ identifies the specific run and $k$ gives the step of the Sieve procedure. Noticeably, the runs of each period $T(k)$ are symmetrical around a symmetry center given by a run sized $4$, except for a final run that is sized $2$. Such a trend is expected to be a rule also for the successive steps.

**Table 2.** Runs of zeroes in the periods $T(k)$ of the sequences $\psi(k,n)$, for primes $p(k) \leq p(4)$. For each $k$, the number of runs $R(k)$ and their sizes $L(m,k)$ are reported, with $m = 1,\ldots,R(k)$. Let us notice the symmetry of the runs in each period $T(k)$. By starting from $k = 2$, the symmetry center is given by a run of length 4, whereas the final run of length 2 is out of symmetry.

| $k$ | $p(k)$ | $T(k)$ | $R(k)$ | $L(m,k)$ |
|---|---|---|---|---|
| (0) | (1) | 1 | 1 | 1 |
| 1 | 2 | 2 | 1 | 2 |
| 2 | 3 | 6 | 2 | 4 2 |
| 3 | 5 | 30 | 8 | 6 4 2 4<br>2 4 6 2 |
| 4 | 7 | 210 | 48 | 10 2 4 2 4 6 2 6 4 2 4 6 6 2 6 4 2 6 4 6 8 4 2 4<br>2 4 8 6 4 6 2 4 6 2 6 6 4 2 4 6 2 6 4 2 4 2 10 2 |

### 3.3. The Relation Between the Primes in an Interval $I(k)$ and the Runs in a Period $T(k)$

For evidencing the relation between each period $T(k)$ and the correspondent number of runs of zeroes $R(k)$, we report in Table 3 the scores of $R(k)$ for $k \leq 7$.

**Table 3.** Periods $T(k)$ and related runs of zeroes $R(k)$ for the primes $p(k) \leq p(7)$. The special prime $p[0] = 1$ is put in round brackets.

| $k$ | $p(k)$ | $T(k)$ | $R(k)$ |
|---|---|---|---|
| (0) | (1) | 1 | 1 |
| 1 | 2 | 2 | 1 |
| 2 | 3 | 6 | 2 |
| 3 | 5 | 30 | 8 |
| 4 | 7 | 210 | 48 |
| 5 | 11 | 2310 | 480 |
| 6 | 13 | 30,030 | 5760 |
| 7 | 17 | 510,510 | 92,160 |

Such scores also give the number of the integers that have not been struck out by the modified Sieve procedure in the period $T(k)$, which in turn can be related to the number of undeleted integers (and consequently of the primes) in the correspondent interval $I(k)$. We will show in Theorem 2 that a correlation exists between $T(k)$ and $R(k)$, in such a way the number of primes in each interval $I(k)$ can be inferred. According on the theory of congruences, Theorem 2 gives the quantity of the integers that have not been struck out (i.e., $R(k)$) in each period $T(k)$, that is,

**Theorem 2.** *Let be given the periodic binary sequences $\psi(k,n)$ defined in Theorem 1, and whose periods are $T(k) = \prod_{i=1}^{k} p(i)$. Then, the number of undeleted integers, that is, the number of runs of zeroes $R(k)$, in a period $T(k)$, for $k \geq 1$, is given by*

$$R(k) = \prod_{i=1}^{k} (p(i) - 1), \qquad k \geq 1 \tag{21}$$

**Proof.** The number of undeleted integers in each period $T(k)$ is given by the number of integers in the reduced residue systems modulo $T(k)$, that is, the number of positive integers less than $T(k)$ and relatively prime to $T(k)$. Such a value is given by the Euler phi function $\phi(T(k))$, once computed in $T(k)$, that is [8]

$$\phi(T(k)) = T(k) \cdot \prod_{p|T(k)} \left(1 - \frac{1}{p}\right) = T(k) \cdot \prod_{p|T(k)} \left(\frac{p-1}{p}\right) = T(k) \cdot \frac{\prod_{i=1}^{k} (p(i)-1)}{\prod_{i=1}^{k} p(i)} = \prod_{i=1}^{k} (p(i)-1) \tag{22}$$

where $p(i)$, $i = 1, \ldots, k$, are the primes dividing $T(k)$. □

By starting from $p(4) = 7$, Table 1 shows that the interval $I(k)$ is included in the first period of the sequence $\psi(k, n)$. Consequently, a subset of the undeleted integers $R(k)$ in each period $T(k)$ lies in the correspondent interval $I(k)$, where they are just primes. Therefore, we can infer the quantity of primes $P(k)$ in each $I(k)$, by starting from the quantity $R(k)$ in the correspondent period $T(k)$. As a first approximation, a simple proportional relationship is investigated. Let us consider the local density $D(k, n)$ of the undeleted integers in the period $T(k)$, where $D(k, n)$ is computed in sliding intervals $J(k, n)$ whose size is the same of $I(k) = [p(k)^2, p(k+1)^2)$, that is, $p(k+1)^2 - p(k)^2$. In this context, the index $n$ represents the starting point of each $J(k, n)$. If such intervals span the whole period $T(k)$, we assume that the density $D(k, n)$ is not a function of $n$. In this case, it is equal to the average density $\overline{D}(k)$ over $T(k)$, and we have

$$D(k, n) = \overline{D}(k) = \frac{R(k)}{T(k)} = \frac{\prod_{i=1}^{k} (p(i) - 1)}{\prod_{i=1}^{k} p(i)} = \prod_{i=1}^{k} \frac{p(i) - 1}{p(i)}, \qquad k \geq 1 \tag{23}$$

It is noteworthy that the product structure in Equation (23) is the same as in Equation (12). Let us suppose that the previous assumption holds. Then, an estimation of the local density $D(k, n)$ in each interval $I(k)$ (that is, for $n = p(k)^2$), will be just the average density $\overline{D}(k)$ over the period $T(k)$. Consequently, we can write

$$D\left(k, p(k)^2\right) \simeq \overline{D}(k), \qquad k \geq 1. \tag{24}$$

Therefore, by starting from Equation (23), we can estimate the quantity of primes $P(k)$ in each interval $I(k)$, for $k \geq 1$. To this end, the average density $\overline{D}(k)$ is multiplied by the size $S(k) = p(k+1)^2 - p(k)^2$, that is,

$$P(k) = \overline{D}(k) \cdot S(k) = (p(k+1)^2 - p(k)^2) \cdot \prod_{i=1}^{k} \frac{p(i) - 1}{p(i)}, \qquad k \geq 1. \tag{25}$$

Evidently, Equation (25) is analogous to Equation (12), apart from the size $N$ of the global interval $I_N$, where $N \in I_K = [p(K)^2, p(K+1)^2)$, that is changed into the size $p(k+1)^2 - p(k)^2$ of the local interval $I(k)$.

### *3.4. The Novel LINPES Estimation of the Prime Number Function $\pi(x)$*

Equation (25) gives a succession of estimations $P(k)$ of the real number of primes $\pi(k)$ in each interval $I(k) = [p(k)^2, p(k+1)^2)$. Therefore, the next step will be to blend all these scores to compute a global estimation $\pi_P(N)$ of the quantity of the primes up to $N$, where $N \in I(K)$, analogously to Equation (12). In theory, $\pi_P(N)$ is simply computable by adding all the contributions $P(k)$ of Equation (25), for $k = 1, \ldots, K$, where $p(K)$ is the greatest prime number not exceeding $N^{1/2}$. However, such a procedure includes the term $p(K+1)$, which is unknown. In order to overcome this issue, the computation of $\pi_P(N)$ has to involve only the terms up to $P(K-1)$, plus a final term $P(K, N)$, where the interval $I_K$ is only partially considered. Consequently, we obtain

$$\pi_P(N) = \sum_{k=0}^{K-1} P(k) + P(K, N) = P(0) + \sum_{k=1}^{K-1} \left[ \left(p(k+1)^2 - p(k)^2\right) \cdot \prod_{i=1}^{k} \frac{p(i) - 1}{p(i)} \right] + P(K, N) \tag{26}$$

where $P(0) = p(1)^2 - p(0)^2$, and $P(K,N) = \left(N - p(K)^2\right) \cdot \prod_{i=1}^{K} \frac{p(i)-1}{p(i)}$. Let us notice that Equation (26) includes as many contributions as the primes are, where each term is given by a relation similar to Equation (12), with the global size $N$ that is replaced by the size of the interval $I(k)$. Each contribution includes an average number of primes that is given by $\prod_{i=1}^{k} \frac{p(i)-1}{p(i)}$, so that the average distance $p(k+1) - p(k)$ between two consecutive primes is $\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}$, which is of the order of magnitude of $\log(p(k))$. For the Cramér conjecture [15], this average distance is $p(k+1) - p(k) = \mathcal{O}(\log^2(p(k))$. Another conjecture by Cramér, by starting from the Riemann's hypothesis, was $p(k+1) - p(k) = \mathcal{O}(\sqrt{p(k)}\,\log(p(k))$ [12,16]. Consequently, the error given by neglecting the partial term $P(K,N)$ is smaller than the loading term of the Cramér conjectures, so that the partial term $P(K,N)$ could be omitted.

### *3.5. The Corrected LINPES Estimation by Using the Equivalence with the $Li(x)$ Function*

We want now to show that Equations (3) and (26) are related. To this end, we write the logarithmic integral function $\mathrm{Li}(N)$ as a summation of integrals, each of them is computed in the interval $I(k) = [p(k)^2, p(k+1)^2)$, that is,

$$\mathrm{Li}(N) = \int_2^{p(1)^2} \frac{dt}{\log t} + \sum_{k=1}^{K-1} \int_{p(k)^2}^{p(k+1)^2} \frac{dt}{\log t} + \int_{p(K)^2}^{N} \frac{dt}{\log t}, \tag{27}$$

where the first term starts from $2$ to cope with a possible improper integral, and $p(K)^2$ is the greatest square of a prime less than $N$. Consequently, the $\mathrm{Li}(N)$ function is expressed by Equation (27) as a succession of estimations $L(k)$, in a similar way to Equation (26), that is,

$$\mathrm{Li}(N) = L(0) + \sum_{k=1}^{K-1} L(k) + L(K,N), \tag{28}$$

where $L(0) = \int_2^{p(1)^2} \frac{dt}{\log t}$, $L(K,N) = \int_{p(K)^2}^{N} \frac{dt}{\log t}$, and

$$L(k) = \int_{p(k)^2}^{p(k+1)^2} \frac{dt}{\log t}. \tag{29}$$

We now apply the Mean Value Theorem to each interval $I(k)$ in Equation (27), that is,

$$\mathrm{Li}(N) = \frac{p(1)^2 - 2}{\log(\varsigma_0)} + \sum_{k=1}^{K-1} \frac{p(k+1)^2 - p(k)^2}{\log(\varsigma(k))} + \frac{N - p(K)^2}{\log(\varsigma_K)}, \tag{30}$$

where $\varsigma_0 \in I(0)$, $I(0) = [p(0)^2, p(1)^2)$, $\varsigma(k) \in I(k)$, $k = 1, \ldots, K-1$, and $\varsigma_K \in I(K,N)$, $I(K,N) = [p(K)^2, N)$. In order to show the equivalence between the Equations (26) and (30), we also consider the lower bound $p(k)^2$ of the interval $I(k)$. By taking, in the two summations, the ratio between the two terms multiplying the interval size $S(k) = p(k+1)^2 - p(k)^2$, we can write

$$\frac{\prod_{i=1}^{k} \frac{p(i)-1}{p(i)}}{\frac{1}{\log(\varsigma(k))}} = \left( \frac{\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}}{\log(\varsigma(k))} \right)^{-1} \tag{31}$$

From Equation (13), we have

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}}{\log(\varsigma(k))} = \lim_{k\to\infty} \left[ \frac{\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}}{\log(p(k)^2)} \times \frac{\log(p(k)^2)}{\log(\varsigma(k))} \right] = \frac{1}{2} \times e^{\gamma} \times \lim_{k\to\infty} \frac{\log(p(k)^2)}{\log(\varsigma(k))} \tag{32}$$

where $\varsigma(k) \in I(k) = [p(k)^2, p(k+1)^2)$, so that its maximum distance from $p(k)^2$ is $p(k+1)^2 - p(k)^2$. However, we know that the $k-th$ prime $p(k)$ is given asymptotically by $p(k) \sim k\,log(k)$ [9]. Therefore, $p(k)^2 \sim k^2 \cdot log(k)^2$ and $p(k+1)^2 \sim (k+1)^2 \cdot log(k+1)^2 \sim k^2\,log(k)^2$, so that for each point $\varsigma(k) \in [p(k)^2, p(k+1)^2)$ we have $\varsigma(k) \sim k^2\,log(k)^2$. It follows that

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}}{\log(\varsigma(k))} = \frac{1}{2} \times e^{\gamma} \times \lim_{k\to\infty} \frac{\log\left(p(k)^2\right)}{\log(\varsigma(k))} = \frac{1}{2} \times e^{\gamma} = \frac{1}{c} \simeq 0.8905 \tag{33}$$

and consequently Equation (31) gives, for each fixed $k$,

$$\frac{\prod_{i=1}^{k} \frac{p(i)-1}{p(i)}}{\frac{1}{\log(\varsigma(k))}} = c_I(k) \qquad \text{where } \lim_{k\to\infty} c_I(k) = c = 2 \times e^{-\gamma} \simeq 1.1229. \tag{34}$$

It follows that the trends of the two estimations (26) and (30) are the same as $k \to \infty$, apart from the constant coefficient $c$. Due to this multiplicative factor, the proposed estimation (26) overestimates the prime number function $\pi(N)$ with respect to Equation (30), and in this sense it is similar to the heuristic procedure described in Section 2. However, it has to be noticed that this last one is completely probabilistic, whereas the proposed method is also based on an analytical procedure, that is, the recognition of an infinite number of binary periodical sequences and related intervals of the prime characteristic function. In order to correct this discrepancy, we relax the conjecture of Section 3.3, in such a way the trend of the local density $D(k,n)$ becomes a function of $n$. Experimentally, the values of the local density $D(k, p_k^2)$ in the interval $I(k)$ are lower than those of the average density $\overline{D}(k)$. The following conjecture is then proposed, which links $D(k, p_k^2)$ and $\overline{D}(k)$ by means of the constant $c$ of the Third Mertens' Theorem [11].

**Conjecture 1.** *The local density $D(k,n)$ of the undeleted integers in the period $T(k)$, if computed in sliding intervals whose size is the same of $I(k) = [p(k)^2, p(k+1)^2)$, is a function of the starting point $n$ of the sliding interval. In particular, the average density $\overline{D}(k)$ is greater than the local density $D\left(k, p(k)^2\right)$ in the interval $I(k)$, in such a way the succession $c_I(k)$ of their ratios exceeds the unity. Moreover, the limit value as $k \to \infty$ of $c_I(k)$ is equal to the constant $c = 2 \cdot e^{-\gamma} \simeq 1.1229$ of the Third Mertens' Theorem, that is,*

$$\lim_{k\to\infty} \frac{\overline{D(k)}}{D\left(k, p_k^2\right)} = c. \tag{35}$$

The typical trend of $D(k,n) = D(16,n) = D(n)$, for $k = 16$ and varying $n$, is plotted in Figure 1, together with the average density $\overline{D}(k) = \overline{D}(16) = \overline{D}$ in the period $T(k) = T(16)$. Let us notice that, as it will be discussed in the following, such a trend is less appreciable for small values of the primes.

Figure 1 can be explained as follows. Let us consider the sequences $\psi(k,n)$ defined in Section 3.1, where the multiples of the primes up to $p(k)$ have been struck out, included the primes themselves. In each of these sequences, all the undeleted integers are just primes in the range $[p(k+1), p(k+1)^2]$, whereas the undeleted integers greater than $p(k+1)^2$ can be indifferently primes or composites, because the multiples of the primes greater than $p(k)$ have not yet been struck out.

At the beginning of the modified Sieve procedure ($k = 0$), the local density $D(k,n)$ of the undeleted integers is not a function of $n$, because no integer has been still struck out. In the first step ($k = 1$), only the even integers (i.e., the multiplies of $p(1) = 2$) have been struck out, so that $D(k,n)$ is still a constant value up to infinity. Noticeably, the multipliers (i.e. the integers multiplying $p(1)$ to give the deleted multiplies) are equal to the undeleted integers when the procedure starts (i.e., all the integers). This rule also holds for the following steps, that is, the multipliers of the prime $p(k)$ in the $k-th$ step of the modified Sieve procedure are equal to the undeleted integers in the previous $(k-1)-th$ step. It follows that the multipliers of $p(2) = 3$ are all the odd integers, whose distribution is again uniform. Some of these multipliers (that is, $3, 5, 7$) are just primes in the interval

$[p(2),\ p(2)^2)$, but they can also be composites beyond $p(2)^2$. In this case, the distribution of the composite multipliers exactly compensate the decreasing trend of the distribution of the multipliers that are also prime numbers. If the primes $p(k)$ are sufficiently small, such a compensation happens quickly, because it starts from $p(k)^2$. In these cases, the distribution of the local density $D(k,n)$ is still approximately uniform. However, as $p(k)$ grows, a transient state is noticeable, because, for such values of $k$ and small values of $n$, the local density $D(k,n)$ is greater than the average density $\overline{D}(k)$. In fact, for such $n$ values, only a portion of the multiples of the primes $p(i), i = 1,\ldots,k$, have been struck out, because the deletion of the multiples of the prime $p(i)$, $i < k$, starts only from $p(i)^2$, apart from the prime $p(i)$ itself. This means that the deletion of the multiples of $p(i), i = 1,\ldots,k$, is completed only at the lower bound of the interval $I(k)$, that is, $p(k)^2$. Consequently, after this point, the transient state ends and the stationary state begins, where the local density $D(k,n)$ fluctuates around the average density $\overline{D}(k)$.

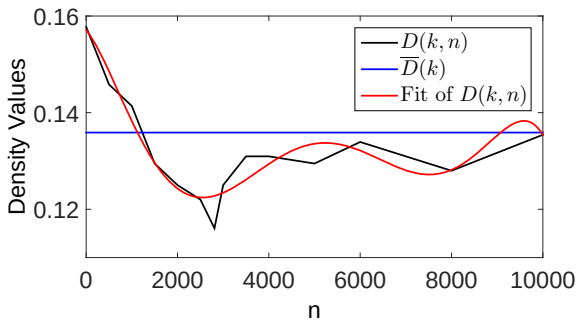


**Figure 1.** Typical trend (in black), with $k = 16$, $p(16) = 53$ and $p(17) = 59$, of the local density of the non-deleted integers $D(n)$ by varying $n$ in sliding intervals whose size is $S(16) = 3481 - 2809 = 672$. Notice that it is shown only the initial part of the period $T(16)$, whose order of magnitude is $10^{19}$, in such a way the symmetrical trend of the period falls outside the figure. The red line reports a polynomial fitting of the density $D(k,n)$, whereas the blue line concerns the average density $\overline{D}(k)$ in the period $T(k)$. The minimum value of the local density is just reached at the lower bound of the interval $I(k)$, that is, $p(16)^2 = 2809$.

Figure 1 shows the trend of the local density $D(k,n)$ in the case of $p(k) = 16$. Starting approximately from this value of $k$, we can notice a minimum value $D(k, p(k)^2)$ for the distribution of $D(k,n)$, which is located immediately after the transient state, that is, at the lower bound of the interval $I(k)$. Such a minimum value is about a 10 percent lower than the average density $\overline{D}(k)$. In fact, as previously explained, the multipliers of the prime $p(k)$ are just primes up to $p(k)^2$, whereupon they can be even composites. It follows that the distribution of the composite multipliers compensate the decreasing distribution of the multipliers that are prime numbers only starting from the multiple $p(k)^3 = p(k)^2 \cdot p(k)$. Therefore, as $k \to \infty$, such a compensation is delaying, in such a way the ratio between $\overline{D}(k)$ and $D(k,n)$ more and more grows up to the $c$ value of Equation (35). As a matter of fact, if all the multipliers were primes, their distribution would decrease by following a logarithmic trend, so that $D(k,n)$ would augment with the same trend, by starting from the minimum value in the interval $I(k)$. In the real case, however, the compensation given by the composite multipliers has the effect that the local density does not grow indefinitely, but tends to the limit value $c \cdot D(k, p(k)^2)$. Let us notice that, if we stop the procedure to a finite value of $k$, the ratio between $\overline{D}(k)$ and $D(k,n)$ is $c_I(k) \cdot D(k, p(k)^2)$, where the succession $c_I(k)$ is increasing and tends to the limit value $c$ as $k \to \infty$.

In order to evaluate the effect of the compensation delay for the small primes $p(k), k = 1,\ldots,7$, in comparison with the case of $p(16) = 53$, Table 4 reports: a) the multipliers $f_I$ such that the multiples $f_I \cdot p(k)$ lie in the interval $I(k) = [p(k)^2, p(k+1)^2)$, and b) the first multiplier that is a composite number, that is, $f_c = p(k)^2$, whose correspondent multiple is $p(k)^2 \cdot p(k) = p(k)^3$. Evidently, as $k$

grows, the difference between the upper bound $p(k+1)^2$ of $I(k)$ and $p(k)^3$ becomes so large that the compensation effect of the composite multipliers is no longer noticeable in the interval itself.

**Table 4.** Prime numbers $p(k)$, $k = 1, \ldots, 7$, and $k = 16$, and the related intervals $I(k)$, together with: a) the multipliers $f_I$ such that the multiples $f_I \cdot p(k)$ lie inside the intervals $I(k)$; b) the first multiplier $f_c$ that is a composite number. Let us notice that the difference between $f_c$ and the multipliers $f_I$ rapidly grows, so that the distance between the multiple $f_c \cdot p(k)$ and the upper bound of the interval $I(k)$ becomes larger and larger.

| $k$ | $p(k)$ | $I(k)$ | $f_I \mid (f_I \cdot p(k)) \in I(k)$ | $f_c$ | $f_c \cdot p(k)$ |
|---|---|---|---|---|---|
| 1 | 2 | $[4, 9)$ | 2; 3 | 4 | 8 |
| 2 | 3 | $[9, 25)$ | 3; 5; 7 | 9 | 27 |
| 3 | 5 | $[25, 49)$ | 5; 7 | 25 | 125 |
| 4 | 7 | $[49, 121)$ | 7; 11; 13; 17 | 49 | 343 |
| 5 | 11 | $[121, 169)$ | 11; 13 | 121 | 1331 |
| 6 | 13 | $[169, 289)$ | 13; 17; 19 | 169 | 2197 |
| 7 | 17 | $[289, 361)$ | 17; 19 | 289 | 4913 |
| 16 | 53 | $[2809, 3481)$ | 53; 59; 61 | 2809 | 148,877 |

Figure 2 shows the trend of the succession $c_I(k)$, as $k$ approaches infinity. Evidently, such a succession tends to the constant value $c$. The x-axis is in a logarithmic scale, in such a way the values of $p(k)^2$ can be visualized up to $10^{15}$.

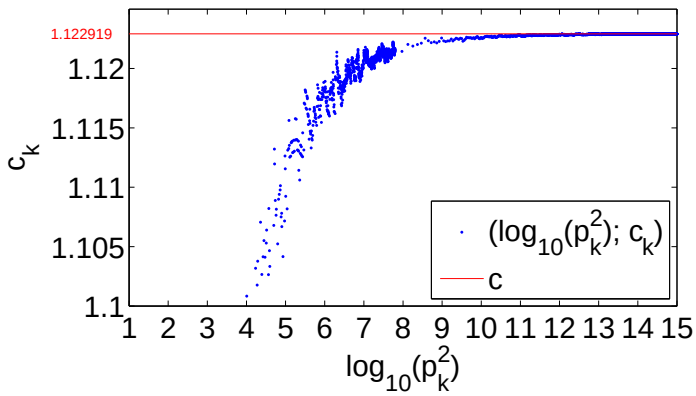


**Figure 2.** Trend of the succession $c_I(k)$ whose elements are the ratios between the average densities $\overline{D}(k)$ in the period $T(k)$ and the local densities $D(k, n)$ in the correspondent interval $I(k)$. For $k \to \infty$, such a succession asymptotically approximates the constant $c$. In the x-axis, a base-10 logarithmic scale has been chosen for a better visualization.

Finally, Table 5 highlights the equivalence between the proposed estimation (26) and the logarithmic-integral one (3). To this end, a number of linear regressions have been computed between the occurrences $P(k)$ (25) in each interval $I(k)$ of the proposed estimation versus the correspondent ones $L(k)$ (29) of the integral-logarithmic function. Each row of Table 5 is referred to the prime squares $p(k)^2$ ranging from a power-of-ten to the following one, except the first raw, which includes all the squares lower than $10^6$, in order to elaborate a sufficient number of points. For each of these ranges, we report the coefficients $m_1$ and $q_1$ of the linear regressions $y_i = m_1 x_i + q_1$, together with the coefficient of determination $R_1^2$, which is a measure of the fitting between the two estimations. Evidently, the coefficient of determination tends very fast to its optimal value, that is 1, despite that the number of observations has increased. Let us notice that the intercept $q_1$ is practically negligible with respect to the full-scale level, whereas the slope $m_1$ is approaching the constant value $1/c$.

For comparison, Table 5 also reports the parameters and the coefficient of determination in the case of the linear regressions $y_i = m_2\, x_i + q_2$ concerning the occurrences $P(k)$ versus the targets $\pi(k)$. These scores are defined as the number of primes in each interval $I(k)$. Even in this case, the fitting between $P(k)$ and $\pi(k)$ is impressive, as shown by the coefficient of determination $R_2^2$. Noticeably, the slope $m_2$ still approaches the value $1/c$, because the P.N.T. guarantees that the logarithmic-integral function and the prime number function goes to infinity in the same way.

**Table 5.** Parameters and coefficients of determination of the linear regressions $y_i = m_1\, x_i + q_1$ of the proposed estimations $P(k)$ versus the logarithmic-integral ones $L(k)$, together with the parameters and coefficients of determination of the linear regressions $y_i = m_2\, x_i + q_2$ of $P(k)$ versus the true number of primes $\pi(k)$. Each point is computed in an interval $I(k)$.

| $k$ | $p(k)^2$ | $m_1$ | $q_1$ | $R_1^2$ | $m_2$ | $q_2$ | $R_2^2$ |
|---|---|---|---|---|---|---|---|
| $[1, 168]$ | $(1, 10^6)$ | 0.894209 | 1.2846 | 0.9999932989 | 0.894649 | 0.2597 | 0.9996747582 |
| $[169, 446]$ | $(10^6, 10^7)$ | 0.892762 | 0.7754 | 0.9999985384 | 0.894052 | −2.5697 | 0.9998452835 |
| $[447, 1229]$ | $(10^7, 10^8)$ | 0.891565 | 1.0381 | 0.9999997462 | 0.891906 | −2.2200 | 0.9999418064 |
| $[1230, 3401]$ | $(10^8, 10^9)$ | 0.891025 | 2.1044 | 0.9999999196 | 0.891016 | 2.0534 | 0.9999821107 |
| $[3402, 9592]$ | $(10^9, 10^{10})$ | 0.890801 | 2.2963 | 0.9999999842 | 0.890751 | 5.6943 | 0.9999941659 |
| [9593, 27,293] | $(10^{10}, 10^{11})$ | 0.890657 | 4.9719 | 0.9999999945 | 0.890664 | 2.8478 | 0.9999981622 |
| [27,294, 78,498] | $(10^{11}, 10^{12})$ | 0.890606 | 5.7853 | 0.9999999989 | 0.890606 | 5.6440 | 0.9999993974 |
| [78,499, 227,647] | $(10^{12}, 10^{13})$ | 0.890570 | 10.3672 | 0.9999999997 | 0.890569 | 13.0142 | 0.9999998112 |
| [227,648, 664,579] | $(10^{13}, 10^{14})$ | 0.890555 | 14.8795 | 0.9999999999 | 0.890555 | 13.7581 | 0.9999999398 |
| [664,580, 1,951,957] | $(10^{14}, 10^{15})$ | 0.890546 | 20.1618 | 1.0000000000 | 0.890546 | 27.3660 | 0.9999999808 |

From the previous analysis, it follows that, for a given $N$, the proposed approximation $\pi_P(N)$ overestimates the prime number function $\pi(N)$ by a factor $c_N$, which can be computed by considering that we have an overestimation for each interval $I(k)$ that can be computed by considering a factor in the finite set $c_I(k)$, $k = 1, \ldots, K$, where $K$ is such that $N \simeq p(K)^2$ (see Equation (34)). If $N \to \infty$, the overestimation factor $c_N$ tends to the constant $c$. Being $c_N$ unknown, an adjusted version (36) of (26) can be defined by means of the correction factor $1/c$, that is,

$$\begin{aligned} \widetilde{\pi}_P(N) &= \frac{1}{c} \cdot \left( P_0 + \sum_{k=1}^{K-1} P(k) + P_{K,N} \right) = \\ &= \frac{1}{c} \cdot \left( p(1)^2 - p(0)^2 \right) + \frac{1}{c} \cdot \sum_{k=1}^{K-1} \left[ \left( p(k+1)^2 - p(k)^2 \right) \cdot \prod_{i=1}^{k} \frac{p(i)-1}{p(i)} \right] + \frac{1}{c} \cdot \left( N - p(K)^2 \right) \cdot \prod_{i=1}^{K} \frac{p(i)-1}{p(i)}. \end{aligned} \tag{36}$$

Clearly, the corrected version $\widetilde{\pi}_P(N) = \frac{1}{c} \cdot \pi_P(N)$ is able to give better estimations than $\pi_P(N)$ as $N$ approaches infinity. In order to give a quantitative assessment, Table 6 reports the scores of $\pi_P(N)$ (26) and of its adjusted version $\widetilde{\pi}_P(N)$ (36), in comparison with the logarithmic integral estimation $\mathrm{Li}(N)$ (27), and with the prime number function $\pi(N)$. The range of each row of Table 6 starts from a power-of-ten and ends to the following one up to $10^{15}$.

It can be noticed that the scores of $\widetilde{\pi}_P(N)$ slightly underestimate both the true number of primes $\pi(N)$ and the logarithmic integral function $\mathrm{Li}(N)$, which, in turn, is such that the sign of its difference with $\widetilde{\pi}_P(N)$ changes infinitely many times [17,18], by showing some irregularities in the distribution of the primes [19], which have been investigated by considering differences in some subsets of the primes themselves [20]. Concerning the previous underestimation, this is due to the fact that the limit value $c$ is an upper bound for the succession $c_I(k)$. Evidently, $\widetilde{\pi}_P(N)$ would be perfectly accurate if the terms $c_I(k)$ were available for the computation of (36), by considering the real number of primes in each interval $I(k)$.

**Table 6.** The proposed estimation $\pi_P(N)$ and its adjusted version $\widetilde{\pi}_P(N)$ in comparison with the logarithmic integral estimation $\mathrm{Li}(N)$, and the prime number function $\pi(N)$. The scores of $\mathrm{Li}(N)$ have been computed by using the MATLAB® toolbox. The scores of $\pi_P(N)$ and $\widetilde{\pi}_P(N)$ have been rounded to the nearest integer.

| $N = 10^i$ | $\pi(N)$ | $\mathrm{Li}(N)$ | $\pi_P(N)$ | $\widetilde{\pi}_P(N)$ |
|---|---|---|---|---|
| $10^1$ | 4 | 6 | 4 | 4 |
| $10^2$ | 25 | 30 | 27 | 24 |
| $10^3$ | 168 | 178 | 181 | 161 |
| $10^4$ | 1229 | 1246 | 1348 | 1 200 |
| $10^5$ | 9592 | 9630 | 10,639 | 9 474 |
| $10^6$ | 78,498 | 78,628 | 87,688 | 78 090 |
| $10^7$ | 664,579 | 664,918 | 744,175 | 662,715 |
| $10^8$ | 5,761,455 | 5,762,209 | 6,460,497 | 5,753,306 |
| $10^9$ | 50,847,534 | 50,849,235 | 57,056,721 | 50,811,064 |
| $10^{10}$ | 455,052,511 | 455,055,615 | 510,796,987 | 454,883,106 |
| $10^{11}$ | 4,118,054,813 | 4,118,066,401 | 4,623,402,885 | 4,117,306,712 |
| $10^{12}$ | 37,607,912,018 | 37,607,950,281 | 42,226,535,908 | 37,604,250,381 |
| $10^{13}$ | 346,065,536,839 | 346,065,645,810 | 388,584,655,120 | 346,048,624,432 |
| $10^{14}$ | 3,204,941,750,802 | 3,204,942,065,692 | 3,598,796,310,868 | 3,204,857,671,495 |
| $10^{15}$ | 29,844,570,422,669 | 29,844,571,475,288 | 33,512,578,849,645 | 29,844,157,918,447 |

## 4. An Extension of the Procedure to the Twin Prime Numbers

### 4.1. Preliminary Concepts

Two prime numbers $p$ and $q$ are twin primes if $|p-q| = 2$, which is the lowest possible distance between primes, apart from $p = 2$ and $q = 3$, where $|p-q| = 1$. Let us note that two consecutive pairs of twin primes do not ever occur, apart from the case $\{3,5\}$ and $\{5,7\}$. In fact, one number in the sequence $\{n, n+2, n+4\}$ is certainly a multiple of 3. The gaps between consecutive primes have been extensively investigated in literature [13,15,21]. However, differently from the primes, *it is presently unknown whether there are infinitely many pairs of twin primes*. In any case, a preliminary counting shows that the twin primes are relatively abundant into the sequence of primes, and, consequently, it is reasonable to infer the so-called twin prime conjecture, which states that *there are infinitely many pairs of twin primes*. This conjecture is strengthened by the fact that the distribution of the primes does not change abruptly. Recently, significant progress has been made by showing that $\liminf_{k\to\infty} [p(k+1) - p(k)] = \ell < \infty$, that is, a finite upper bound exists for the limit inferior of the difference between consecutive primes. In particular, Zhang found that $\ell \leq 7 \cdot 10^7$ [22], and this bound has been successively improved by Maynard to $\ell \leq 600$ [23]. Finally, the Polymath's project, whose aim is to collect all the various efforts that try to put the bound lower as much as possible, has reached the value of $\ell \leq 246$ [24]. Evidently, in order to demonstrate the twin prime conjecture, a bound of $\ell = 2$ should be obtained. In this work, we try to give a contribution to the discussion of this conjecture, by following a different strategy, that is, by exploiting the concepts previously introduced for the primes. Consequently, as for the primes, the approach is not merely probabilistic, but also analytic, so constituting a possible significant step for further advancements, as in the case of approaches based on periodic functions [25]. The distribution of the twin primes is commonly characterized by using the twin prime function $\pi_2(x)$ (6). Such a distribution decays more rapidly than the distribution of the primes. In fact, Brun demonstrated in 1919 [26] that, if $S_T$ is the set of twin primes given by $S_T = \{p : p \text{ prime and } p+2 \text{ prime}\}$, the related series of the reciprocals converges to the finite limit $B \simeq 1.9022$ [1], that is,

$$\sum_{p \in S_T} \left( \frac{1}{p} + \frac{1}{p+2} \right) = B \tag{37}$$

regardless of the fact of whether the number of summation terms is infinite or not, whereas the same summation instead diverges for the primes.

Analogously to the P.N.T., a possible function for approximating the twin prime function $\pi_2(x)$ has been proposed [5] as the logarithmic integral function $\mathrm{Li}_2(x)$ (8). As for the primes, we want to obtain an equivalent procedure and investigate possible consequences.

### *4.2. A Possible Relation Between the Twin Primes in the Intervals and the Undeleted Integers in the Periods*

In Section 3.2, the distribution of the runs into each period $T(k)$ has been investigated. In the present analysis, the same investigation can be made for the particular case in which the size of the runs is 2. Evidently, such an investigation can potentially give an estimation of the quantity of twin primes, similarly to the one given by the Equation (26) for the primes. In fact, we will suggest that the number of the runs sized 2 in the interval $I(k)$ is equal to the quantity of twin primes in the same interval. Such a number is equal to the number of $\{101\}$ sequences, if the sequence $\{10\}$ is completely included in the interval. However, such a sequence cannot occur across two intervals, because each interval, apart from the first one, ends with an even number (that is, a 0), because it is followed by a square of an odd prime (that is, another 0), which is an odd number. For the sake of clarity, in the following we denote the runs sized 2 as runs 2. Let us notice that this procedure can be extended to run-lengths of whatever size, by following the *Hardy-Littlewood conjecture B* [6]. Such a topic will be the object of future explorations.

Table 7 reports the number $R_2(k)$ of the runs 2 in each period $T(k)$ for $p(k)$, $k = 0, \ldots, 7$. As for the total number of runs $R(k)$ (21) in the same period, a correlation can be found between $R_2(k)$ and the prime number $p(k)$. In particular, the scores of Table 7 suggest the following conjecture for $R_2(k)$

$$R_2(k) = \prod_{i=2}^{k} (p(i) - 2), \qquad k \geq 2. \tag{38}$$

**Table 7.** Number of runs 2, denoted as $R_2(k)$, that are included in the periods $T(k)$, for $p(k)$, $k = 0, \ldots, 7$. These scores are compared with the total number of runs $R(k)$. The special prime $p(0) = 1$ is put in round brackets.

| $k$ | $p(k)$ | $T(k)$ | $R(k)$ | $R_2(k)$ |
|---|---|---|---|---|
| (0) | (1) | 1 | 1 | 0 |
| 1 | 2 | 2 | 1 | 1 |
| 2 | 3 | 6 | 2 | 1 |
| 3 | 5 | 30 | 8 | 3 |
| 4 | 7 | 210 | 48 | 15 |
| 5 | 11 | 2310 | 480 | 135 |
| 6 | 13 | 30,030 | 5760 | 1485 |
| 7 | 17 | 510,510 | 92,160 | 22,275 |

Equation (38) can be investigated by taking the modified Sieve procedure. At the start of the procedure ($k = 0$), we have no run 2. In the first step ($k = 1$), the multiples of $p(1) = 2$ are struck out, in such a way the sequence $\psi(1, n)$ is made by runs 2 only. In particular, a single run 2 is included in the period $T(1) = 2$, so that $R_2(1) = 1$. For $k = 2$, we delete the multiples of $p(2) = 3$, so that the period $T(2) = 6$ becomes three times greater. This implies that the number of runs 2 could increase from 1 to 3, but the deletion in the point $n = 3$ vanishes two of these runs. Let us notice that the cancellation of one multiple vanishes two runs 2 only in this step, being all the runs 2 consecutive, but this does not happen in the following steps, where only one run 2, or even none, is deleted at the time. It follows that $R_2(2) = 1$, as in the previous step. On the whole, we obtain that the deleted runs 2 in the period $T(2)$ are a fraction $2/3 = 2/p(2)$ of the total number of runs 2 in the same period if no cancellations were made.

Similarly, for $k = 3$, the multiples of $p(3) = 5$ are struck out, so that the period $T(3)$ becomes five times greater. It follows that the number of runs 2 would grow from 1 to 5, but two cancellations (for $n = 5, 25$) vanish two of the five runs 2. Consequently, we obtain $R_2(3) = 3$ and the fraction of the deleted runs 2 is $2/5 = 2/p(3)$ of the total runs in this period if no cancellation were made. In this step, all the cancellations imply the deletion of one run 2, but this will not also be a rule for the following steps. In fact, for $k = 4$, we have eight cancellations in the period $T(4)$, but only six of them stroke out a run 2. However, the fraction of the deleted runs 2 in the period is still given by $6/21 = 2/7 = 2/p(4)$ of the pre-existing ones before the cancellations, being $R_2(4) = 3 \cdot 7 - 6 = 15$.

In the case of primes, it follows from the relation (21) that we struck out, in each step, a fraction $1/p(k)$ of the total number of runs in the period $T(k)$ if no cancellations were made, which is given by the product of the prime $p(k)$ by the actual number of runs in the previous period $T(k-1)$. By considering the scores of Table 7, a similar relation can be conjectured for the runs 2 in the case of twin primes, in order to link the number of cancelled runs 2 and the total number of runs 2 in the period $T(k)$ if no cancellations were made. Unfortunately, in general, the actual number of the deleted runs 2 is not easily computable, by starting from the total number of cancellations in $T(k)$. However, in the same way of the primes, our conjecture is that the deletion of the multiples of $p(k)$ has the effect to exactly cancel a fraction $2/p(k)$ of the runs 2 in the period $T(k)$.

If this conjecture holds, Equation (38) follows by induction. In fact, it is true for $p(2) = 3$. Let us suppose that Equation (38) holds for $p(k-1)$ and show that it is also true for $p(k)$. By the induction hypothesis, the number of runs 2 in the period $T(k-1)$ is given by $R_2(k-1) = \prod_{i=2}^{k-1}(p(i)-2)$. We must show that the number of runs 2 in the period $T(k)$ is $R_2(k) = \prod_{i=2}^{k}(p(i)-2)$. Given $R_2(k-1)$, the number of runs 2 in the new period $T(k)$ becomes $p(k) \cdot R_2(k-1)$, because $T(k)$ is $p(k)$ times greater than $T(k-1)$. By taking the previous conjecture, a fraction $2/p(k)$ of the runs 2 is struck out, in such a way we have a fraction of residual runs 2 given by $(p(k)-2)/p(k) \cdot R_2(k-1) = (p(k)-2)/p(k) \cdot \prod_{i=2}^{k-1}(p(i)-2) = \prod_{i=2}^{k}(p(i)-2) = R_2(k)$.

### *4.3. A Heuristic Estimation of $\pi_2(x)$ Equivalent to the $Li_2(x)$ Approximation*

From Equation (38), we can give an estimation $\pi_{2P}(N)$ of the twin prime function $\pi_2(x)$, which is equivalent to the approximation given by the $\mathrm{Li}_2(x)$ function (8). Such an estimation can be viewed as a generalization of Equation (26) to the case of the twin primes. To this end, analogously to Equation (23) for the primes, we compute the average density $\overline{D_2}(k)$ of the number of runs 2 in a period $T(k)$. By starting from the total number of runs 2 $R_2(k)$ in the period $T(k)$, the average density $\overline{D_2}(k)$ is given by the relation

$$\overline{D_2}(k) = \frac{R_2(k)}{T(k)} = \frac{\prod_{i=2}^{k}(p(i)-2)}{\prod_{i=1}^{k} p(i)} = \frac{1}{2} \times \prod_{i=2}^{k} \frac{p(i)-2}{p(i)}, \qquad k \geq 2. \tag{39}$$

As for the primes, we can initially approximate the local density $D_2(k,n)$ in the interval $I(k)$ as the average density $\overline{D_2}(k)$, that is, $D_2\left(k, p(k)^2\right) \simeq \overline{D_2}(k)$. In this case, the estimated number of twin primes $P_2(k)$ in $I(k)$, for $k \geq 2$, is given by

$$P_2(k) = \overline{D_2}(k) \times S(k) = (p(k+1)^2 - p(k)^2) \times \frac{1}{2} \times \prod_{i=2}^{k} \frac{p(i)-2}{p(i)}, \qquad k \geq 2 \tag{40}$$

The total estimation $\pi_{2P}(N)$ is then obtained by adding all the contributions $P_2(k)$, that is,

$$\pi_{2P}(N) = \sum_{k=0}^{K-1} P_2(k) + P_2(K,N) = P_2(0) + P_2(1) + \sum_{k=2}^{K-1} \left[ \left(p(k+1)^2 - p(k)^2\right) \cdot \tfrac{1}{2} \cdot \prod_{i=2}^{k} \tfrac{p(i)-2}{p(i)} \right] + P_2(K,N) \tag{41}$$

where $P_2(0) = p(1)^2 - p(0)^2$, $P_2(1) = \frac{1}{2} \cdot (p(2)^2 - p(1)^2)$, $P_2(K,N) = \left(N - p(K)^2\right) \cdot \frac{1}{2} \cdot \prod_{i=2}^{K} \frac{p(i)-2}{p(i)}$, and $K$ is the greatest prime number not exceeding $N^{1/2}$. As for the primes, Equation (41) overestimates

the true $\pi_2(N)$ scores, because the local density $D_2(k,n)$ is not actually constant in the period $T(k)$, but it is a function of $n$. However, the offset of the local density in the interval $I(k)$ with respect to the average density is greater than for the primes. Experimentally, each $P_2(k)$ value (40) overtakes the true quantity of twin primes computed in $I(k)$ of about 20%, that is, more or less a double of the percentage previously found for the primes, and reported in Figure 1, even if the trends of the local densities are similar. Quantitatively, the ratio between the average density $\overline{D_2}(n)$ and the local density $D_2\left(k, p(k)^2\right)$ seems to approximate the constant $c^2$ as $k \to \infty$, that is, the square of $c$.

To evidence this statement, let us consider the estimation given by the $\mathrm{Li}_2(x)$ function, that is, $C\,\mathrm{Li}_2(x)$, for $x = N$, from Equation (8), that is, $C\,\mathrm{Li}_2(N)$, as a summation of integrals, each of them is computed in the interval $I(k) = [p(k)^2, p(k+1)^2)$

$$C\,\mathrm{Li}_2(N) = C\int_2^{p(1)^2} \frac{dt}{\log^2 t} + C\sum_{k=1}^{K-1}\int_{p(k)^2}^{p(k+1)^2} \frac{dt}{\log^2 t} + C\int_{p(K)^2}^{N} \frac{dt}{\log^2 t} \tag{42}$$

being $p(K)^2$ the greatest square of a prime less than $N$. Similarly to Equation (28), we can write Equation (42) as a succession of estimations $L_2(k)$ in each interval $I(k)$, that is,

$$C\,\mathrm{Li}_2(N) = C\,L_2(0) + C\sum_{k=1}^{K-1} L_2(k) + C\,L_2(K,N), \tag{43}$$

where $L_2(0) = \int_2^{p(1)^2} \frac{dt}{\log^2 t}$, $L_2(K,N) = \int_{p(K)^2}^{N} \frac{dt}{\log^2 t}$ and

$$L_2(k) = \int_{p(k)^2}^{p(k+1)^2} \frac{dt}{\log^2 t}. \tag{44}$$

Then, we apply the Mean Value Theorem for Integrals to Equation (42) in each interval $I(k)$

$$C\,\mathrm{Li}_2(N) = C\,\frac{p(1)^2 - 2}{\log^2(\varsigma_0)} + C\sum_{k=1}^{K-1}\frac{p(k+1)^2 - p(k)^2}{\log^2(\varsigma(k))} + C\,\frac{N - p(K)^2}{\log^2(\varsigma_K)}, \tag{45}$$

where the point $\varsigma_0$ belongs to the interval $I(0) = [p(0)^2, p(1)^2)$, $\varsigma(k)$ belongs to the interval $I(k)$, $k = 1, \ldots, K-1$, and $\varsigma_K$ belongs to the interval $I(K,N) = [p(K)^2, N)$. As for the primes, we have to consider the lower bound $p(k)^2$ of the interval $I(k)$. Let us take the ratio between the two terms multiplying the size $S(k) = p(k+1)^2 - p(k)^2$, in the summations of the Equations (41) and (45), so that we obtain

$$\frac{\frac{1}{2}\cdot\prod_{i=2}^{k}\frac{p(i)-2}{p(i)}}{\frac{C}{\log^2(\varsigma(k))}} = \left(\frac{2C\cdot\prod_{i=2}^{k}\frac{p(i)}{p(i)-2}}{\log^2(\varsigma(k))}\right)^{-1} \tag{46}$$

If we consider the lower bound $p(k)^2$ of the interval $I(k)$, we have

$$\frac{2C\cdot\prod_{i=2}^{k}\frac{p(i)}{p(i)-2}}{\log^2(\varsigma(k))} = \frac{2C\cdot\prod_{i=2}^{k}\frac{p(i)}{p(i)-2}}{\log^2(p(k)^2)}\cdot\frac{\log^2(p(k)^2)}{\log^2(\varsigma(k))} \tag{47}$$

Let us notice that the ratio $\frac{p(i)-2}{p(i)}$ can be split as

$$\begin{aligned}\frac{p(i)-2}{p(i)} &= \frac{p(i)-2}{(p(i)-1)^2}\times\frac{(p(i)-1)^2}{p(i)} = \frac{p(i)^2 - 2\,p(i)}{(p(i)-1)^2}\times\frac{(p(i)-1)^2}{p(i)^2} = \frac{(p(i)-1)^2-1}{(p(i)-1)^2}\times\frac{(p(i)-1)^2}{p(i)^2}\\ &\Longrightarrow \frac{p(i)-2}{p(i)} = \frac{p(i)-1}{p(i)}\times\frac{p(i)-1}{p(i)}\times\left(1-\frac{1}{(p(i)-1)^2}\right)\end{aligned} \tag{48}$$

Consequently, we obtain

$$\prod_{i=2}^{k} \frac{p(i)}{p(i)-2} = \prod_{i=2}^{k} \left[ \frac{p(i)}{p(i)-1} \times \frac{p(i)}{p(i)-1} \times \frac{1}{1-\frac{1}{(p(i)-1)^2}} \right] \tag{49}$$

Then, we define

$$\begin{cases} C(k) = 2 \times \prod_{i=2}^{k} \left(1 - \frac{1}{(p(i)-1)^2}\right), & k \geq 2 \\ C(1) = C(0) = 1. \end{cases} \tag{50}$$

From Equation (49) and considering that $\lim_{k\to\infty} \frac{\log^2(p(k)^2)}{\log^2(\varsigma(k))} = 1$ (see Section 3.5), the limit, as $k \to \infty$, of the ratio (47) is given by

$$\lim_{k\to\infty} \frac{2C \times \prod_{i=2}^{k} \frac{p(i)}{p(i)-2}}{\log^2(p(k)^2)} \times \frac{\log^2(p(k)^2)}{\log^2(\varsigma(k))} = \lim_{k\to\infty} \frac{2C \times \prod_{i=2}^{k} \frac{p(i)}{p(i)-2}}{\log^2(p(k)^2)} = \lim_{k\to\infty} \frac{2C \times \frac{2}{C(k)} \times \prod_{i=2}^{k} \left[\frac{p(i)}{p(i)-1} \times \frac{p(i)}{p(i)-1}\right]}{\log^2(p(k)^2)}. \tag{51}$$

We noticed in the Equation (33) that

$$\lim_{k\to\infty} \frac{\prod_{i=1}^{k} \frac{p(i)}{p(i)-1}}{\log\left(p(k)^2\right)} = \frac{1}{2} \times e^{\gamma} \simeq \frac{1}{c} \simeq 0.8905. \tag{52}$$

Evidently, we have

$$\lim_{k\to\infty} \prod_{i=2}^{k} \frac{p(i)}{p(i)-1} = \frac{1}{2} \times \lim_{k\to\infty} \prod_{i=1}^{k} \frac{p(i)}{p(i)-1} \tag{53}$$

and, consequently, from Equation (9),

$$\lim_{k\to\infty} C(k) = C. \tag{54}$$

Finally, from Equation (51), we obtain the limit of the ratio (47)

$$\lim_{k\to\infty} \frac{4 \times \prod_{i=2}^{k} \left[\frac{p(i)}{p(i)-1} \times \frac{p(i)}{p(i)-1}\right]}{\log^2\left(p(k)^2\right)} = 4 \times \left(\frac{1}{2c}\right)^2 = \frac{1}{c^2} \simeq 0.8905^2 = 0.7931 \tag{55}$$

and Equation (46) gives

$$\frac{\frac{1}{2} \times \prod_{i=2}^{k} \frac{p(i)-2}{p(i)}}{\frac{C}{\log^2(\varsigma(k))}} = c_{2I}(k) \qquad \text{where } \lim_{k\to\infty} c_{2I}(k) = c^2 \simeq 1.2609. \tag{56}$$

For a given $N$, the proposed approximation $\pi_{2P}(N)$ overestimates the twin prime number function $\pi_2(N)$ by a factor $c_{2N}$, which can be computed by considering that we have an overestimation for each interval $I(k)$ that can be computed by considering a factor in the finite set $c_{2I}(k)$, $k = 1, \ldots, K$, where $K$ is such that $N \simeq p(K)^2$. Equations (55) and (56) show that the succession $c_{2I}(k)$ tends to the constant $c^2$ as $N \to \infty$. Consequently, we can define a corrected version $\widetilde{\pi}_{2P}(N)$ (57) of the proposed estimation $\pi_{2P}(N)$, by multiplying Equation (41) by the factor $1/c^2 \simeq 0.7931$, that is,

$$\begin{aligned} \widetilde{\pi}_{2P}(N) &= \frac{1}{c^2} \times \left(P_2(0) + P_2(1) + \sum_{k=2}^{K-1} P_2(k) + P_2(K,N)\right) = \frac{1}{c^2} \times \left(p(1)^2 - p(0)^2\right) + \frac{1}{2c^2} \times \left(p(2)^2 - p(1)^2\right) + \\ &+ \frac{1}{2c^2} \times \sum_{k=2}^{K-1} \left[\left(p(k+1)^2 - p(k)^2\right) \times \prod_{i=2}^{k} \frac{p(i)-2}{p(i)}\right] + \frac{1}{2c^2} \times \left(N - p(K)^2\right) \times \prod_{i=2}^{K} \frac{p(i)-2}{p(i)}. \end{aligned} \tag{57}$$

As for the primes, Equation (57) is expected to improve the estimation of $\pi_2(N)$ as $N$ approaches infinity. This is evidenced in the scores of Table 8, where a comparison is made between the proposed

estimation $\pi_{2P}(N)$ and its adjusted version $\widetilde{\pi}_{2P}(N)$ with the estimation $C\,\mathrm{Li}_2(N)$ given by the logarithmic integral function (8) and the twin prime number function $\pi_2(N)$. The ranges of $N$ are the same as Table 6.

**Table 8.** The proposed estimation $\pi_{2P}(N)$ and its adjusted version $\widetilde{\pi}_{2P}(N)$ in comparison with the logarithmic integral estimation $C\,\mathrm{Li}_2(N)$ and the prime number function $\pi_2(N)$. The scores of the logarithmic integer function have been computed by using the MATLAB® toolbox. The scores of $\pi_{2P}(N)$ and $\widetilde{\pi}_{2P}(N)$ have been rounded to the nearest integer.

| $N = 10^i$ | $\pi_2(N)$ | $C\,\mathrm{Li}_2(N)$ | $\pi_{2P}(N)$ | $\widetilde{\pi}_{2P}(N)$ |
|---|---|---|---|---|
| $10^1$ | 2 | 2 | 4 | 3 |
| $10^2$ | 8 | 11 | 12 | 10 |
| $10^3$ | 35 | 43 | 48 | 38 |
| $10^4$ | 205 | 212 | 250 | 198 |
| $10^5$ | 1224 | 1246 | 1522 | 1207 |
| $10^6$ | 8169 | 8246 | 10,252 | 8131 |
| $10^7$ | 58,980 | 58,751 | 73,579 | 58,353 |
| $10^8$ | 440,312 | 440,365 | 553,514 | 438,977 |
| $10^9$ | 3,424,506 | 3,425,306 | 4,312,478 | 3,420,314 |
| $10^{10}$ | 27,412,679 | 27,411,414 | 34,537,569 | 27,390,848 |
| $10^{11}$ | 224,376,048 | 224,368,862 | 282,810,653 | 224,289,776 |
| $10^{12}$ | 1,870,585,220 | 1,870,559,864 | 2,358,205,655 | 1,870,231,592 |
| $10^{13}$ | 15,834,664,872 | 15,834,598,303 | 19,964,600,235 | 15,833,405,367 |
| $10^{14}$ | 135,780,321,665 | 135,780,264,892 | 171,202,650,560 | 135,776,370,890 |
| $10^{15}$ | 1,177,209,242,304 | 1,177,208,491,858 | 1,484,356,543,022 | 1,177,204,581,001 |

The connection between the $\pi_{2P}(N)$ estimation (41) and the $C\,\mathrm{Li}_2(N)$ estimation (42) is investigated in Table 9, by considering the parameters and the coefficient of determination of the linear regressions $y_i = m_1\,x_i + q_1$ between the occurrences of $P_2(k)$ (40) versus those of $C\,L_2(k)$, where $L_2(k)$ is given by (44), in each interval $I(k)$. As for the primes, an excellent fitting is given by the linear relationship between $P_2(k)$ and $C\,L_2(k)$. This is confirmed by the coefficient of determination $R_1^2$, which rapidly tends to 1 as $k$ grows. On the other hand, the intercept $q_1$ is negligible, whilst the slope $m_1$ approaches the limit value $1/c^2$.

The fitting of the linear regressions $y_i = m_2\,x_i + q_2$ between the occurrences of $P_2(k)$ (40) versus those of the twin prime number function $\pi_2(k)$, if computed in the same interval $I(k)$, is also reported in Table 9. Even if less impressive than in the case of Table 5 for the primes, the goodness of the fitting is clearly shown by the coefficient of determination $R_2^2$, which is practically at its best value. As for $m_1$, the slope $m_2$ seems to approximate the limit value $1/c^2$.

**Table 9.** Parameters and coefficients of determination of the linear regressions $y_i = m_1\,x_i + q_1$ of the proposed estimations for the twin primes $P_2(k)$ versus the logarithmic-integral ones $C\,L_2(k)$, together with the parameters and coefficients of determination of the linear regressions $y_i = m_2\,x_i + q_2$ of $P_2(k)$ versus the true number of twin primes $\pi_2(k)$. Each point is computed in an interval $I(k)$.

| $p(k)^2$ | $m_1$ | $q_1$ | $R_1^2$ | $m_2$ | $q_2$ | $R_2^2$ |
|---|---|---|---|---|---|---|
| $(1, 10^6)$ | 0.799120 | 0.3205 | 0.9999525356 | 0.784981 | 0.6900 | 0.9819305687 |
| $(10^6, 10^7)$ | 0.797052 | 0.1191 | 0.9999935927 | 0.807148 | −0.8767 | 0.9947532680 |
| $(10^7, 10^8)$ | 0.794901 | 0.1435 | 0.9999989157 | 0.792818 | 0.7199 | 0.9983424066 |
| $(10^8, 10^9)$ | 0.793935 | 0.2629 | 0.9999996567 | 0.794232 | −0.5788 | 0.9992892724 |
| $(10^9, 10^{10})$ | 0.793529 | 0.2602 | 0.9999999326 | 0.793309 | 1.2422 | 0.9998094846 |
| $(10^{10}, 10^{11})$ | 0.793273 | 0.5082 | 0.9999999770 | 0.793336 | −0.0376 | 0.9999152638 |
| $(10^{11}, 10^{12})$ | 0.793180 | 0.5523 | 0.9999999955 | 0.793186 | 0.6827 | 0.9999711368 |
| $(10^{12}, 10^{13})$ | 0.793115 | 0.9036 | 0.9999999988 | 0.793125 | 0.0544 | 0.9999902660 |
| $(10^{13}, 10^{14})$ | 0.793088 | 1.2179 | 0.9999999997 | 0.793089 | 0.9592 | 0.9999967269 |
| $(10^{14}, 10^{15})$ | 0.793072 | 1.5449 | 0.9999999999 | 0.793072 | 2.4024 | 0.9999988979 |

In summary, the proposed approach estimates the true number of twin primes by considering the number of runs $2$ in each interval $I(k) = \left[ p(k)^2, p(k+1)^2 \right)$, in such a way each estimation $P_2(k)$ fits the correspondent one given by $C\,L_2(k)$. Consequently, in the case the conjecture (38) holds, we can infer that the distribution of the twin primes follows the same trend in all the intervals $I(k)$. Because these intervals are a function of the squares of both the prime $p(k)$ and its successive one, it follows that, *being the primes are a never-ending succession, the unproved hypothesis of the infinitude of the twin primes would be further strengthened.*

## 5. Conclusions and Future Developments

In this work, an original heuristic procedure in order to obtain the distribution of the prime number function $\pi(x)$ is proposed and investigated, which gives estimations of the scores of $\pi(x)$ equivalently to the logarithmic integral function $\mathrm{Li}(x)$. However, this approach is not fully probabilistic, but it is also based on analytical concepts, that is, a set of infinitely many binary periodic sequences is found by means of a modified Sieve procedure, whose periods have a subset that is included in limited and disjoint intervals $I(k)$ of the prime characteristic function. In each period $T(k)$, these binary sequences define a succession of $1$ values, which are separated by runs of consecutive zeroes. Starting from the number of runs of zeroes in a period $T(k)$, an estimation of the total number of primes can be found, which is linked to the logarithmic integral estimation by the constant $c$ of the Third Mertens' Theorem. Noticeably, the succession of the runs of zeroes, whose elements are the gaps between two consecutive primes, is symmetric in each period $T(k)$. As a result, the proposed LINPES procedure estimates the prime number function in each interval $I(k)$, whose bounds are the squares of a prime number and of the successive one. As a particular case, this procedure is also specialized to the case of the twin primes, in such a way only the runs sized $2$ are considered in each period. Consequently, a heuristic relation for the number of these runs in a period $T(k)$ is formulated, whose trend is linked to the relation previously found for the total number of runs in the case of primes. Therefore, such a relation gives an estimation of the twin prime number function $\pi_2(x)$ in each interval $I(k)$, which is equivalent to the estimation of the logarithmic integral function $\mathrm{Li}_2(x)$, by means of the square of the constant $c$. Being the bounds of these intervals given by squares of primes, their number is infinite. As a consequence, the proposed procedure could give a contribution to the presumed infinity of the succession of the twin primes. Future developments will further investigate the relation of the number of runs $2$ in a period $T(k)$, together with the symmetry of the succession of the runs of zeroes.

**Author Contributions:** Conceptualization, B.A.; Methodology, B.A., S.B., L.S. and M.S.; Formal Analysis, L.S.; Investigation, B.A., S.B., L.S. and M.S.; Data Curation, B.A., L.S. and M.S.; Writing—Original Draft Preparation, B.A.; Writing—Review & Editing, B.A. and M.S.; Visualization, M.S.; Supervision, S.B.

**Funding:** This research received no external funding.

**Acknowledgments:** The author would thank the site https://primes.utm.edu/lists/small/millions/ for providing the prime numbers that have been used for the computations in this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Crandall, R.; Pomerance, C. *Prime Numbers: A Computational Perspective*; Springer: New York, NY, USA, 2001.
2. Gauss, C.F. Letter to Encke, dated 24 December (1849). *Werke Kng. Ges. Wiss. Gottingen* **1863**, *2*, 444–447.
3. Legendre, A.M. *Essai sur la thèorie des Nombres*; Duprat: Paris, France, 1798.
4. Torquato, S.; Zhang, G.; de Courcy-Ireland, M. Uncovering multiscale order in the prime numbers via scattering. *J. Stat. Mech.* **2018**. [CrossRef]
5. Goldston, D.A. Are There Infinitely Many Twin Primes? Available online: http://arxiv.org/pdf/0710.2123.pdf (accessed on 17 April 2019).
6. Hardy, G.H.; Littlewood, J.E. Some problems of "Partito Numerorum", III: On the expression of a number as a sum of primes. *Acta Math.* **1923**, *44*, 1–70. [CrossRef]

7. Helfgott, H.A. An improved Sieve of Eratosthenes. Available online: https://arxiv.org/abs/1712.09130 (accessed on 17 April 2019).
8. Apostol, T. M. *Introduction to Analytic Number Theory*; Springer: New York, NY, USA, 1976.
9. Fine, B.; Rosenberger, G. *Number Theory: An Introduction via the Distribution of Primes*; Birkhäuser: Boston, MA, USA, 2007.
10. Montgomery, H. A heuristic for the Prime Number Theorem. *Math. Intell.* **2006**, *28*, 6–9. [CrossRef]
11. Mertens, F. Ein Beitrag zur analytischen Zahlentheorie. *J. Reine Angew. Math.* **1874**, *78*, 46–62.
12. Granville, A. Harald Cramér and the Distribution of Prime Numbers. Available online: https://www.dartmouth.edu/~chance/chance_news/for_chance_news/Riemann/cramer.pdf (accessed on 17 April 2019).
13. Selberg, A. On the normal density of primes in small intervals, and the difference between consecutive primes. *Arch. Math. Naturvid.* **1943**, *47*, 87–105.
14. Languasco, A.; Zaccagnini, A. Short intervals asymptotic formulae for binary problems with prime powers. *J. Théorie Nombres Bordeaux* **2018**, *30*, 609–635. [CrossRef]
15. Cramer, H. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.* **1936**, *2*, 23–46. [CrossRef]
16. Cramer, H. On the distribution of primes. *Proc. Camb. Phil. Soc.* **1920**, *20*, 272–280.
17. Bays, C.; Hudson, R. H. A new bound for the smallest $x$ with $\pi(x) > \mathrm{Li}(x)$. *Math. Comput.* **2000**, *69*, 43–56. [CrossRef]
18. Saouter, Y.; Demichel, P. A sharp region where $\pi(x) - \mathrm{Li}(x)$ is positive. *Math. Comput.* **2010**, *79*, 2395–2405. [CrossRef]
19. Bays, C.; Hudson, R.H. Zeroes of the Dirichlet L-functions and irregularities in the distribution of primes. *Math. Comput.* **1999**, *69*, 861–866. [CrossRef]
20. Granville, A.; Martin, G. Prime number races. *Am. Math. Mon.* **2006**, *113*, 1–33. [CrossRef]
21. Pintz, J. Very large gaps between consecutive primes. *J. Num. Theor.* **1997**, *63*, 286–301. [CrossRef]
22. Zhang, Y. Bounded gaps between primes. *Ann. Math.* **2014**, *179*, 1121–1174. [CrossRef]
23. Maynard, J. Small gaps between primes. *Ann. Math.* **2015**, *181*, 383–413. [CrossRef]
24. Polymath, D.H.J. The "Bounded Gaps between Primes" Polymath Project: A Retrospective. Available online: https://arxiv.org/abs/1409.8361 (accessed on 17 April 2019).
25. Bagchi, B. A promising approach to the twin prime problem. *Reson* **2003**, *8*, 26–31. [CrossRef]
26. Brun, V. La série $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \ldots$ où les dénominateurs sont "nombres premiers jumeaux" est convergent ou finie. *Bull. Sci. Math.* **1919**, *43*, 100–104.