Numerical analysis approach
to twin primes conjecture

Gabriele Di Pietro

Via Annunziata, 42, 64021, Giulianova (TE), Italy
e-mail: `petreius@libero.it`

Abstract: The purpose of this paper is to demonstrate how the modified Sieve of Eratosthenes is used to have an approach to twin prime conjecture. If the Sieve is used in its basic form, it does not produce anything new. If it is used through the numerical analysis method explained in this paper, we obtain a specific counting of twin primes. This counting is based on the false assumption that distribution of primes follows punctually the Logarithmic Integral function denoted as $\mathrm{Li}(x)$ (see [5] and [10], pp. 174–176). It may be a starting point for future research based on this numerical analysis method technique that can be used in other mathematical branches.
Keywords: Numerical analysis, Number theory, Sieves.
2020 Mathematics Subject Classification: 11N35.

1   Introduction

A number is considered as a *prime* when it can be divided by one or itself; *twin primes* are prime numbers that differ by 2 from the previous or following number of a sequence.

A famous algorithm used to create a table of primes is the Sieve of Eratosthenes (see [1, 7]): it consists writing down the integers from 2 to a number $x \in \mathbb{N}$ in a table and then striking out all number greater than 2 which are divisible by 2. The numbers that remain are all twins but only some of them will remain after the next deletions. As a result, the smallest remaining number greater than 2, which is 3, can be found. All numbers greater than 3 which are divisible by 3 are struck out and the remaining ones are twins of the $6k + 1$ and $6k - 1$ form ($\forall k \in \mathbb{N}$ and $k > 0$). This process can be carried out until $\lfloor \sqrt{x} \rfloor$ to ensure that the remaining numbers are primes; the floor function (see [11]) $\lfloor \sqrt{x} \rfloor$ gives the largest integer less than or equal to $\sqrt{x}$.

According to the twin prime conjecture, there are infinitely many twin primes. We cannot demonstrate it but the modified Sieve of Eratosthenes can be used to prove a small partial result based on the false assumption that the natural distribution of primes punctually follows the formula $\frac{x}{\ln x}$, discovered by Gauss. The last theorem is a subsequent extension of previous Gauss formula to other orders of Logarithmic Integral $\mathrm{Li}(x) = \int_2^x \frac{1}{\ln y} dy$.

The paper is organized as follows. Sections 1–4 give us an introduction to terminology, the method used and a clarification on Υ terms. Using the orders of Logarithmic Integral $\mathrm{Li}(x)$, the Sieve of Eratosthenes can be modified counting twin primes in Υ function in Section 5. Section 6 summarizes the previous explanations and gives us two theorems using the Logarithmic Integral orders and the modified Sieve of Eratosthenes.

2 Towards a table of twin primes

Firstly, it can be observed that if a number $x \in \mathbb{N}$ is prime, it will be sufficient to control that it is divisible by integers less than $\sqrt{x}$ (see [15]). So, if a sequence of primes from 1 to $x \in \mathbb{N}$ was made using the Sieve of Eratosthenes, it would be sufficient to cross out the multiples of primes less than to $\sqrt{x}$. The table from 1 to x to which the Sieve of Eratosthenes is applied can be referred to as $[1, x]$ *Sieve of Eratosthenes table*.

This table can be split in two parts, namely *eliminators zone* from 1 to $\sqrt{x}$ and *would-be twin primes zone* from $\sqrt{x}$ (not included) to x. The numbers of the $6k + 1$ and $6k - 1$ form ($\forall k \in \mathbb{N}$ and $k > 0$) included between x and its root can be referred to as *would-be twin primes*. Their amount can be discovered through the calculation

$$\left\lfloor \frac{2x}{6} \right\rfloor - \left\lfloor \frac{2\sqrt{x}}{6} \right\rfloor. \tag{1}$$

Using (1) there can be a miscalculation, specifically a maximum of two numbers in excess. To find the would-be twin prime couples, it will be sufficient to divide by two.

Primes greater than 3, less than $\sqrt{x}$, can be referred to as *eliminators* as they will cut out some of the would-be twin primes through the sieve. These numbers, being $6k + 1$ or $6k - 1$, have a period of 6; so

$$6k + 1 \equiv 1 \mod 6$$

$$6k - 1 \equiv -1 \mod 6$$

The following Figure 1 clarifies the *eliminators* terminology showing a y number and its multiples: a multiple will be *eliminated* from twin primes counting as we know through Sieve of Eratosthenes.

In Figure 1, we have y as an eliminator and $2y, 3y, \ldots$ as its multiples. Specifically the eliminator $6k + 1$ cuts out a $6k - 1$ number after 4 multiples and it cuts out a $6k + 1$ number after 6 multiples; in this case, the rule $6k - 1$ behaves in a specular manner.
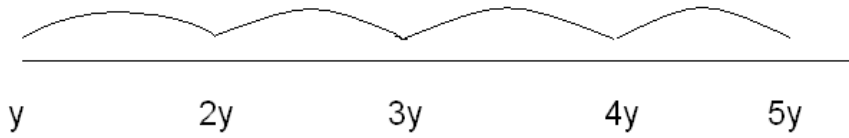
Figure 1. An eliminator y and its multiples.

So, if for every y eliminator we obtain x/y multiples, then for every eliminator of the $6k+1$ form $x/6y$ number of its own form will be deleted as well as $x/6y$ numbers of the opposite one ($6k-1$ form). An y eliminator will delete the amount of *would-be* twin primes:

$$\left\lfloor \frac{x}{6y} \right\rfloor + \left\lfloor \frac{x}{6y} \right\rfloor \sim \left\lfloor \frac{x}{3y} \right\rfloor \tag{2}$$

Among these, we can eliminate the multiples within the eliminators zone, so we only consider the would-be twin primes greater than $\sqrt{x}$. The formula (2) represents the total amount of would-be twin primes cut out within the eliminators zone

$$K = \left\lfloor \frac{x}{3y} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{3y} \right\rfloor. \tag{3}$$

This result does not avoid a double or multiple counting: in fact two or more eliminators could delete the same would-be twin prime but in this way we consider two or more times the same deleted number; in order to affirm that the first chosen eliminator multiple cross out a pair of twin primes, we have to refine the counting. However, it could delete a number which as already been deprived of its twin.

The formula (1) calculates the would-be twin prime numbers. The amount of couples of would-be twin prime numbers can be obtained by dividing it by two that is to say $\left\lfloor \frac{x}{6} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{6} \right\rfloor$. The final formula is

$$\left\lfloor \frac{x}{6} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{6} \right\rfloor - \sum_{y \in \pi(\sqrt{x})} K, \tag{4}$$

where K is (3) and $\pi(\sqrt{x})$ are primes less then $\sqrt{x}$. Affirming that (4) could verify the conjecture in Section 6 (i.e., it is positive) is a utopia, unless we presume that the distribution of eliminators is near $\sqrt{x}$. We only have to refine the estimate by removing repetitions and introducing new remarks.

3   The $w(z)$ function

The distinct prime factors of a positive integer $x \geq 2$ are defined as the $w(x)$ numbers $p_1, \ldots, p_{w(x)}$ in the prime factorization (Hardy and Wright in [9], p. 354). So the $w(x)$ function is the number of distinct prime factors of a number $x \in \mathbb{N}$. In this case we can avoid repetitions: all numbers hit by an eliminator are deleted necessarily by another eliminator because of factorization so we can easily halve the total number of multiples. Nevertheless, a would-be twin prime could be deleted by more than two numbers: for example, if all numbers were hit by three eliminators,

we should divide the total amount of multiples by three. All we need is a function that shows how many different primes factorize a number. Having $w(x)$ an irregular pointwise increase, an average increase is preferred and

$$L = \frac{1}{\#Z} \sum_{z \in [1,x]} w(z) \tag{5}$$

is the way in which the eliminator multiples should be divided in a $[1, x]$ Sieve of Eratosthenes table, where z is an integer number of $6k + 1$ or $6k - 1$ form in the $[1, x]$ Sieve of Eratosthenes table; Z is the set of integer numbers z and $\#Z$ is its cardinality.

4 Completely deleted twins

Using the prime number theorem (see [2, 8]),

$$\lim_{x \to \infty} \frac{\pi(x)}{x / \ln(x)} = 1$$

and according to Chebyschev's inequality (see [10], p. 186) for $x \to \infty$

$$0.92 \ldots \frac{x}{\ln(x)} \leq \pi(x) \leq 1.105 \ldots \frac{x}{\ln(x)},$$

if we have x numbers (considering that $\frac{x}{\ln(x)}$ are primes), we can choose as average distance among two near primes $\ln(x)$ but it is a false assumption (see [4, 6, 14]) and we only use it to explain the method in this paper. In fact, this is a weak form (first order expansion) of Logarithmic Integral $\mathrm{Li}(x)$ (see [3], pp. 116–117):

$$\mathrm{Li}(x) = \frac{x}{\ln x} \sum_{k=0}^{\infty} \frac{k!}{(\ln x)^k}$$

Littlewood in [12] demonstrated that $\mathrm{Li}(x) - \pi(x)$ changes sign infinite times so $\mathrm{Li}(x)$ is a good would-be mean. In order to simplify the calculation, $\frac{x}{\ln(x)}$ (first order approximation) must be used mentioning the upper orders in the final theorem. According to the previous estimate, the distance between two primes is

$$\mu - x = \ln(x) \tag{6}$$

and it increases as far as it becomes bigger than 8. The result is that between two primes there is a completely deleted couple of would-be twin primes (remembering that the period is 6) and the eliminator divides a prime number and its twin. This is a good thing because, due to this, the remaining couples of primes will prove our conjecture. As a matter of fact, if a couple is completely erased, we will use two multiples to eliminate it. So we take into account the multiples that will be numbers deprived of their twin from the whole number of multiples. These multiples, in fact, are irrelevant to count the would-be couples of twin that will be deleted (we can consider a couple of twin already deleted, since it misses one of the two members).

Mathematically, we can write

$$T = \sum_{p \in [\sqrt{x}, x]} \left\lfloor \frac{\mathrm{dist}(p) - 3}{6} \right\rfloor, \tag{7}$$

where $\mathrm{dist}(p)$ gives distance function between p (prime) and the following prime number. Now

$$\tilde{T} = \sum_{p=\sqrt{x}}^{x} \left(\left\lfloor \left| \frac{\ln(p) - 3}{6} \right| \right\rfloor \right)^{+} \tag{8}$$

is the number of multiples needed to completely erase the twin couples between $\sqrt{x}$ and x. Every x in (6) is, in fact, a prime. Hence, if we develop the formula above, we obtain

$$\sum_{p=\sqrt{x}}^{x} \left\lfloor \frac{\ln(x)}{6} \right\rfloor - \left\lfloor \frac{3}{6} \left(\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) \right\rfloor \sim \sum_{p=\sqrt{x}}^{x} \frac{\ln(p)}{6} - \frac{5.5}{6} \left(\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) \tag{9}$$

In this case we have 5.5 terms in approximation as we added the floor function

$$\sum_{p=\sqrt{x}}^{x} \left\lfloor \frac{\ln(p)}{6} \right\rfloor \tag{10}$$

so we have to take away at least a number of elements equal to decimals it may be obtained from the formula above after we pass in following approximation using

$$\frac{\ln(p)}{6}. \tag{11}$$

In fact, when we have a numerator divided by 6, the following decimals can be obtained: $\frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}$; using an extreme approximation, we obtain 2.5 by adding these numbers. We thus have to add to the numerator $2.5 \cdot \left(\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right)$ for the same reason: this clarifies 5.5.

5 Alternative approach to the would-be twin prime couples function

The explanations contained in Section 3 can be avoided by using the following approach.

Considering would-be twin primes between x and its root, we find that these numbers are $\left\lfloor \frac{2x}{6} \right\rfloor - \left\lfloor \frac{2\sqrt{x}}{6} \right\rfloor$ according to (1). If we cut out from this formula the primes up to $[\sqrt{x}, x]$ we get the primes erased by the sieve without repetitions. Therefore the amount of eliminations of twin primes within the would-be twin primes zone is expressed as follows:

$$\frac{1}{L} \sum_{y \in \pi(\sqrt{x})} K = \left\lfloor \frac{x}{3} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{3} \right\rfloor - \left[\pi(x) - \pi(\sqrt{x}) \right] \tag{12}$$

without computing $w(n)$ for each would-be prime number. Naturally, using the prime number theorem we have:

$$\pi(x) - \pi\sqrt{x} \sim \left[\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln(\sqrt{x})} \right]. \tag{13}$$

Through the use of these approximations, the twin prime counting function can be written in the following way:

$$\Upsilon(x) = \left\lfloor \frac{x}{6} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{6} \right\rfloor - \frac{1}{L} \sum_{y \in \pi(\sqrt{x})} K + T, \tag{14}$$

where

$$\frac{1}{L} \sum_{y \in \pi(\sqrt{x})} K = \left\lfloor \frac{x}{3} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{3} \right\rfloor - \left\lfloor \frac{x}{\ln x} - \frac{\sqrt{x}}{\ln(\sqrt{x})} \right\rfloor \qquad (15)$$

and

$$T = \sum_{p \in [\sqrt{x}, x]} \left\lfloor \frac{dist(p) - 3}{6} \right\rfloor . \qquad (16)$$

$\Upsilon(x)$ represents the amount of twin prime couples included between $\sqrt{x}$ and x; if this amount is bigger than zero, the following conjecture in Section 6 (affirming that *there are infinite twin prime couples supposing a punctually* $\mathrm{Li}(x)$ *function distribution of primes*) is proved.

Figure 2 was built through Matlab code [13]. It represents a proof of our numerical analysis technique applied to twin prime conjecture. The blue function (our continuous function using the first order of Li as an average) fits the red one (non continuous real counting twin primes).
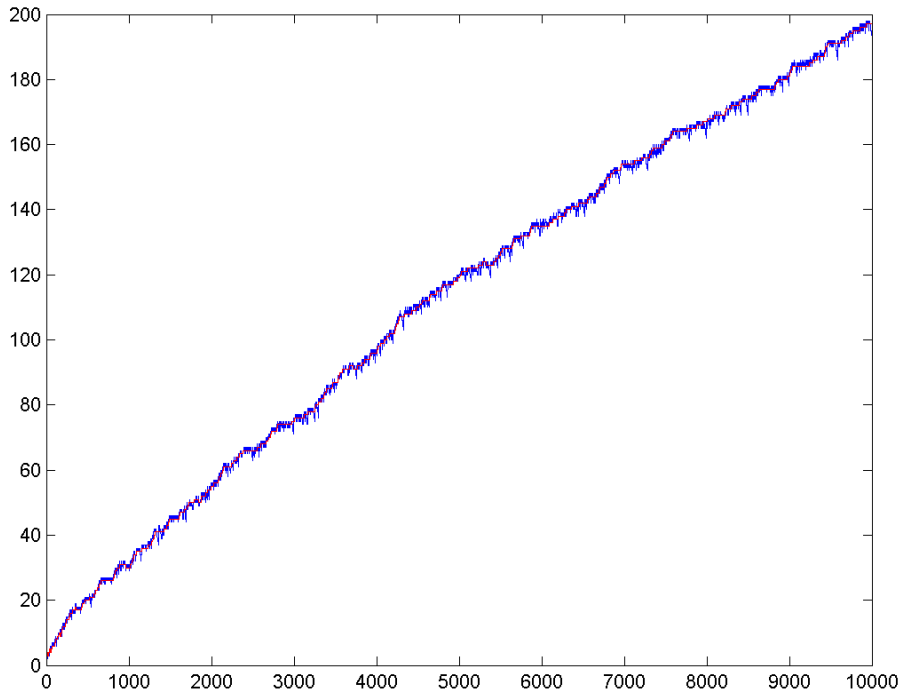

Figure 2. Red $\rightarrow$ exact twin primes function into $[\sqrt{x}, x]$; Blue $\rightarrow \Upsilon(x)$.

6   The approximation of twin prime couples function

Considering the previous conclusions, we reach the following approximation

$$\tilde{\Upsilon}(x) = \left\lfloor \frac{x}{6} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{6} \right\rfloor - \frac{1}{\tilde{L}} \sum_{\tilde{y} \in \pi(\sqrt{x})} \tilde{K} + \tilde{T}, \qquad (17)$$

where

$$\frac{1}{\tilde{L}} \sum_{\tilde{y} \in \pi(\sqrt{x})} \tilde{K} = \left\lfloor \frac{x}{3} \right\rfloor - \left\lfloor \frac{\sqrt{x}}{3} \right\rfloor - \left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) \qquad (18)$$

180

and

$$\tilde{T} = \frac{(\ln(x) - 5.5)}{6} \left(\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right). \tag{19}$$

Theorem 1. *Supposing a punctually* $\mathrm{Li}(x)$ *function on first order, the function* $\tilde{\Upsilon}(x)$*, defined by* (17)–(19)*, is bigger than zero for each* $x > 141.83$*. This function is an approximation of* $\Upsilon(x)$*, the numerical twin prime couples function defined by* (14)–(16)*.*

Proof.

$$\frac{\sqrt{x} - x}{6} + \left(\frac{x}{\ln(x)} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) + \frac{(\ln x - 5.5)}{6} \left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) \geq 0$$

$$\frac{\sqrt{x} - x}{6} + \frac{\ln(x)}{6} \cdot \left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) + \frac{0.5}{6} \cdot \left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln \sqrt{x}} \right) \geq 0$$

using properties of logarithms

$$\frac{\sqrt{x} - x}{6} + \left(\frac{x}{6} - \frac{2\sqrt{x}}{6} \right) + \frac{0.5}{6} \cdot \left(\frac{x}{\ln x} - \frac{2\sqrt{x}}{\ln x} \right) \geq 0$$

$$\frac{0.5}{6} \cdot \left(\frac{x - 2\sqrt{x}}{\ln x} \right) - \frac{\sqrt{x}}{6} \geq 0$$

computing the least common multiple

$$\frac{\sqrt{x} \cdot (0.5 \cdot \sqrt{x} - 1 - \ln x)}{\ln x} \geq 0.$$

It follows that the denominator is bigger than zero for each $x > 1$, and the numerator is a logarithmic equation having as solutions 0.52 and 141.83. This equation is positive for external values of the following range: $[0.52, 141.83]$. Hence $\tilde{\Upsilon}(x)$ is positive for each $x \in [0.52, 1] \cup [141.83, \infty)$ proving our theorem. $\qquad \square$

For each $x < 20$, the twin primes may be counted by hand. The theorem claims that our approximated conjecture is proved: for each new x, twin prime couples amount will increase more and more, assuming that the function is always positive. This assumption guarantees that twin primes are infinite using our hypothesis based on the error of Numerical Analysis method and the false assumption about prime counting function using $\mathrm{Li}(x)$.

The very basic computing hypothesis to demonstrate the theorem is that the distance among primes is, in average, bigger or equal to $\ln(x)$. In fact, if we assumed that the distance among primes may be $k \ln(x)$, for $0 < k < 1$ according to the Gaussian approximation, the theorem wouldn't be proved. For instance, considering $k = 0.7851$ we would have:

$$\frac{\sqrt{x} \cdot [0.5 \cdot \sqrt{x} - 1 - 0.5702 \cdot \ln x - 0.2149 \cdot \ln(x)\sqrt{x}]}{\ln x} \geq 0,$$

$$\frac{\sqrt{x} \left[\sqrt{x} \cdot (0.5 - 0.2149 \ln x) - 1 - 0.5702 \cdot \ln x \right]}{\ln x} \geq 0.$$

So, positiveness of numerator depends on the following expression

$$\sqrt{x} \cdot (0.5 - 0.2149 \ln x). \tag{20}$$

Because of $\ln(x)$, this amount is positive at first and then it turns negative; in this way, the numerator will become negative too, considering that we can find a sum of all negative terms. As a result, the approximated conjecture is rejected.

Now we can assume $\pi(x) \sim \frac{1}{k}\frac{x}{\ln x} \sim \mathrm{Li}(x)$ for some upper orders depending on k choice where $k\ln(x)$ for $0 < k < 1$ is the average distance between twin primes.

Theorem 2. *Using Theorem (1) but choosing as average distance*

$$\frac{1}{k} \cdot \frac{x}{\ln(x)},$$

where $0 < k < 1$, the approximating function $\tilde{\Upsilon}(x)$ is positive for each $x > 141.83$.

Proof. On the basis of the previous demonstration we obtain

$$\frac{\sqrt{x}-x}{6} + \frac{1}{k}\left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln\sqrt{x}}\right) + \frac{1}{k}\cdot\frac{(k\ln x - 5.5)}{6}\left(\frac{x}{\ln x} - \frac{\sqrt{x}}{\ln\sqrt{x}}\right) \geq 0,$$

from which

$$\frac{0.5}{6}\cdot\frac{1}{k}\left(\frac{x-2\sqrt{x}}{\ln(x)}\right) - \frac{\sqrt{x}}{6} \geq 0$$

and so

$$\frac{\sqrt{x}(\frac{0.5}{k}\cdot\sqrt{x} - \frac{1}{k} - \ln x)}{\ln x} \geq 0,$$

verifying the hypothesis as the numerator is a logarithmic function like Theorem (1). $\qquad\square$

The conjecture has been demonstrated to all orders of $\mathrm{Li}(x)$ function as it is sufficient assume that

$$\frac{1}{k} = \sum_{t=0}^{\infty}\frac{t!}{(\ln x)^t}.$$

References

[1] Brun, V. (1920). *Le crible d'Eratosthène et le théorème de Goldbach*. Kristiania: En Commission chez J. Dybwad.

[2] De la Vallée Poussin, C. J. (1896). Recherches analytiques la théorie des nombres premiers. *Annales de la Société scientifique de Bruxelles*, 20, 183–256.

[3] Derbyshire, J. (2004). *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics*, New York: Penguin.

[4] Erdős, P. (1940). The difference of consecutive primes. *Duke Mathematical Journal*, 6(2), 438–441.

[5] Gauss, C. F. (1863). *Werke, Band 10*, Teil I. p. 10.

[6]   Goldston, D. A., Motohahi, Y., Pintz, J., & Yıldırım, C. Y. (2006). Small gaps between primes exist. *Proceedings of the Japan Academy, Series A, Mathematical Sciences*, 82(4), 61–65. DOI: 10.3792/pjaa.82.61.

[7]   Greaves, G. (2001). *Sieves in Number Theory*. Springer, Berlin.

[8]   Hadamard, J. (1896). Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France*, 24, 199–220.

[9]   Hardy, G. H., & Wright, E. M. (1979). *An Introduction to the Theory of Numbers*, 5th ed., Oxford Science Publications, Oxford, England: Clarendon Press, pp. 354–358.

[10] Havil, J., & Dyson, F. (2003). Problems with Primes. In: *Gamma: Exploring Euler's Constant*. Princeton; Oxford: Princeton University Press, pp. 163–188.

[11] Iverson, K. E. (1962). *A Programming Language*. New York: Wiley, p. 12.

[12] Littlewood, J. E. (1914). Sur les distribution des nombres premiers. *Comptes rendus de l'Académie des Sciences*, 158, 1869–1872.

[13] Naldi, G., & Pareschi, L. (2008). *Matlab, Concetti e Progetti*, Apogeo Education.

[14] Pintz, J. (1997). Very large gaps between consecutive primes. *Journal of Number Theory*, 63(2), 286–301.

[15] Riesel, H. (1994). *Prime Numbers and Computer Methods for Factorization*. Series Progress in Mathematics. Vol. 126 (second ed.). Boston, MA: Birkhäuser.