Surveys in Mathematics and its Applications

ISSN 1842-6298 (electronic), 1843-7265 (print)
Volume 13 (2018), 171 – 181

ON THE NUMBERS THAT DETERMINE THE DISTRIBUTION OF TWIN PRIMES

Antonie Dinculescu

Abstract. This paper is about a class of numbers indirectly connected to the twin primes, which have not been investigated so far. With the help of these numbers, we look at the set of twin primes from a different perspective and bring the reader's attention to some characteristics potentially useful in tackling the Twin Prime Conjecture. To this purpose, we simplify the problem by dividing the positive integers into two complementary sets: one whose members lead to a pair of twin primes by a simple algebraic operation but cannot be directly calculated, and another one whose members do not lead directly to twin primes but are responsible for their distribution and can be directly calculated. By analyzing the facts from the perspective of the numbers in the second set, we reveal some interesting patterns and properties that suggest new approaches to the problem.

1 Introduction

Around 300 B. C. Euclid gave an elementary but elegant proof that there is no largest prime number [6] and, hence, that there are infinitely many primes. Today, after more than 2000 years, we still do not know with certitude if there also are infinitely many twin primes. Since so far almost all attempts to solve this problem were based on probabilistic estimates, we are going to use a different approach and look at the problem from the point of view of an outside observer who just records the facts, and try to find some characteristics potentially useful in answering this question.

First, we simplify the problem by dividing the positive integers into two complementary sets: one whose members lead to a pair of twin primes by a simple algebraic operation but cannot be directly calculated, and another one whose members do not lead directly to twin primes, but determine their distribution and can be directly calculated. Next, we will look at the way the members of the set directly related to the twin primes are interspersed among the more numerous members of the other set. Third, mindful of the random nature of the prime numbers, we will

2010 Mathematics Subject Classification: 11A41.
Keywords: prime numbers; Twin Prime Conjecture; distribution of prime numbers; sieves

try to find some order among the members of the second set hoping that we will be able to see some useful patterns. Finally, after gathering all the data, we will pause to see which characteristics (if any) can be used in tackling the Twin Prime Conjecture (TPC), according to which there are infinitely many twin primes [2].

2 Twin ranks, the surrogates of twin primes

In what follows we denote by $(P_j)_j$ the increasing sequence of prime numbers. Two prime numbers P_j and P_{j+1} are called *twin primes* if $P_{j+1} - P_j = 2$. Examples:

3, 5; 5, 7; 11, 13; 17, 19; 29, 31; ... .

The first thing one notices when looking at the twin primes pairs is the fact that they all can be characterized by the number K between them called *twin index* (here in bold):

$$3, \mathbf{4}, 5; \quad 5, \mathbf{6}, 7; \quad 11, \mathbf{12}, 13; \quad 17, \mathbf{18}, 19; \quad 29, \mathbf{30}, 31 \ldots .$$

One also notices that with the exception of 4, all twin indices are divisible by 6. Let us call *twin ranks* (denoted k^*) all integers that lead to a pair of twin primes by multiplication with 6 followed by the addition/subtraction of 1 (ex. 1, 2, 3, 5, 7, 10, 12, 17, 18, 23, 25), and *non-ranks* (denoted k) the numbers that do not have this property (ex. 4, 6, 8, 9, 11, 13, 14, 15, 16, 19, 20, 22, 22, 24).
Despite their randomness, the twin ranks obey certain rules. For example:
- all numbers that end in 1, 4, 6, or 9 cannot be twin ranks; (this means we can discard 40% of the numbers in a given interval as not being twin ranks, by simply looking at their last digit);
- all twin ranks of the form $k^* = n^2$ are divisible by 5. (Throughout this paper n denotes a positive integer);
- all twin ranks of the form $k^* = n^3$ are divisible by 7;

The first two of the above properties can be easily explained by the fact that all twin prime indices larger than 4 must be divisible by 6. For the third one we do not have an explanation yet, but will be interesting to see if there are other twin ranks of the form $k^* = (an)^b$ with a determined by b and both of them different from the pairs 5, 2 and 7, 3.

3 Non-ranks, the pillars of twin ranks

If one looks carefully at the set of positive integers, one notices that there are two non-ranks associated to a prime P symmetrically distributed at equal distances $[\pm P/6]$ from all numbers that are multiples of P.
Note: Here $[x]$ means *the nearest integer* to x. This notation is usually avoided in

number theory, but we are using it in this paper because the remainder from the division of any prime by 6 is either 1/6 or 5/6 but never 1/2.
One has, therefore:

$$k = nP_j \pm [P_j/6]\,, \tag{3.1}$$

where P_j is the jth prime. By writing (3.1) as $6k \pm 1 = P(6n \pm 1)$ it is easy to see that $6k$ cannot be twin index because $6k \pm 1$ is composite. With the above equation, by using all primes $5 \le P_i \le P_j < \sqrt{6M_{j+1}+1}$ and all necessary integers, one can find all non-ranks smaller than a *basic number*

$$M_{j+1} = \frac{P_{j+1}^2 - 1}{6}. \tag{3.2}$$

There are no other numbers in the set of positive integers besides twin ranks and non-ranks. Therefore, by knowing the number of non-ranks in a given interval one can find how many twin ranks are in the interval by simply subtracting that number from the length of the interval.

Let us call *basic interval* the interval between two consecutive basic numbers, and *basic prime* the largest prime corresponding to the non-ranks in a basic interval.

One has

$$\Delta M_j = \frac{P_{j+1}^2 - P_j^2}{6} = \frac{(P_{j+1} - P_j)\,(P_{j+1} + P_j)}{6}. \tag{3.3}$$

It is easy to see that when P_j and P_{j+1} are twin primes, the basic interval, here called *twin interval*, has the length $\Delta M_j = (2/3)K_j = 4k_j$, where K_j and k_j are the corresponding twin index and twin rank. Since there aren't too many linear correlations linking directly the prime numbers to relevant quantities, we note in passing two such correlations concerning the twin ranks:

- The number of twin ranks in a twin interval ΔM_j is proportional to j.
- The number of twin ranks up to a basic number M_j is proportional to j^2.

Let us, now, apply equation (3.1) to the non-ranks 6, 9, 13, 64, 145, using the primes 5, 7, 11, 13 (here shown in bold). Taking into account that $[5/6] = [7/6] = 1$ and $[11/6] = [13/6] = 2$ one has:

$$\begin{aligned}
6 &= 1 \times \mathbf{5} + 1 = 1 \times \mathbf{7} - 1 \\
9 &= 2 \times \mathbf{5} - 1 = 1 \times \mathbf{11} - 2 \\
13 &= 2 \times \mathbf{7} - 1 = 1 \times \mathbf{11} + 2 \\
64 &= 13 \times \mathbf{5} - 1 = 9 \times \mathbf{7} + 1 \\
145 &= 13 \times \mathbf{11} + 2 = 11 \times \mathbf{13} + 2.
\end{aligned}$$

Clearly, one can arrive at the same non-rank by using different primes (in our case 5_7, 5_11, 7_11, 5_7 and 11_13). In order to address this problem, we define a *parent prime* pertaining to a non-rank as the smallest prime required by equation

(3.1) to find that non-rank. In the above example, the parent primes of the non-ranks 6, 9, 13, 64 and 145 are 5, 5, 7, 5 and 11. The parent prime concept allows one to organize the non-ranks in such a manner as to be able to see some useful patterns. Here are several characteristics of this interesting class of numbers that nobody has explored so far:

- The non-ranks of the same parent prime P_j form an infinite number of consecutive sets here called *groups*, of length $L_j = \prod_{i=3}^{j} P_i$, each of them containing exactly G_j members (see below), symmetrically distributed with respect to the central gap. (Although L_j, G_j, S_j, R_j, etc. represents the number of terms in a set, for expediency we will use these symbols to designate the corresponding sets and intervals). Because L_j is divisible by all primes $5 \le P_i \le P_j$, by applying equation (3.1) to the interval L_j and eliminating each time the non-ranks of parent primes smaller than P_j, one obtains the following expression for the number of non-ranks of the same parent prime P_j in L_j,

$$G_j = 2\prod_{i=3}^{j-1} (P_i - 2) \,. \tag{3.4}$$

The non-ranks of parent primes $5 \le P_i \le P_j$ form an infinite number of consecutive sets called *super-groups* of length L_j each of them containing exactly S_j members, symmetrically distributed with respect to the central term L_c situated between the numbers $L_{c1} = (L_j - 3)/2$ and $L_{c2} = (L_j + 3)/2$. By summing up the number of all non-ranks in the groups that form a super-group one obtains

$$S_j = L_j - \prod_{i=3}^{j} (P_i - 2) \,. \tag{3.5}$$

- Because $P_3 = 5$ is the smallest parent prime, all first groups and super-groups begin at $5 - 1 = 4$ and end at $L_j + 3$. Incidentally, the numbers 1, 2, 3, which are not covered by equation (3.1), are all twin ranks.
- Besides the first groups and super-groups which begin at $5 - 1 = 4$ and end at $L_j + 3$, there are an infinity of groups and super-groups of the same parent primes, with the same length, in which the distribution of non-ranks mirrors exactly the distribution in the first group or super-group. Once one knows a term in the first group or super-group, one can find the equivalent term in all the other groups or super-groups by simply adding to that term L_j multiplied by an integer.
- Since $S_j < L_j$, there always are numbers in a super-group L_j that do not belong to the set of non-ranks of parent primes $5 \le P_i \le P_j$. This set of numbers, here called *remnants*, have exactly $R_j = L_j - S_j$ members and contain twin ranks and non-ranks of parent primes larger than P_j. From (3.5) one obtains

$$R_j = \prod_{i=3}^{j} (P_i - 2) \,. \tag{3.6}$$

- The uniform distribution of numbers in the interval L_j allows the symmetry to be preserved after the subtraction of the non-ranks S_j from L_j. Consequently, the remnants R_j in a super-group are symmetrically distributed with respect to the central gap.

4 The covering process

We define the *covering process* as the operation of revealing the twin-ranks by crossing out (*covering*) the numbers situated at equal distances $[\pm P/6]$ from all numbers that are multiple of a prime P. Essentially this is a sieving process but with a different emphasis. One begins with $P_3 - [P_3/6] = 4$ and $P_3 + [P_3/6] = 6$ and *covers* all numbers $4 + nP_3$ and $6 + nP_3$, where $n = 0, 1, 2, \ldots$. Then, one repeats the procedure with $P_4 - [P_4/6] = 6$ and $P_4 + [P_4/6] = 8$ and and covers all numbers $6 + nP_4$ and $8 + nP_4$, then with $P_5 - [P_5/6] = 9$ and $P_5 + [P_5/6] = 13$ covering all numbers $9 + nP_5$ and $13 + nP_5$ and so on. One can visualize the process as an infinite wall containing a row of bricks numbered from 1 to infinity, and someone walking along it with a roll painter and covering in paint all bricks situated at equal distances $[\pm P/6]$ from the numbers that are multiple of a prime P. The bricks with numbers smaller than $M_{j+1} = \left(P_{j+1}^2 - 1\right)/6$ remained uncovered after using all rolls of sizes $5 \le P_i \le P_j$ are twin ranks.
Clearly, while the covering process started at the points $[P \pm P/6]$ goes forward to infinity, it does not go backward. This property ensures that once a number was shown to be a twin rank by remaining uncovered after using all primes up to a certain prime P_j, it is not going to be covered (i.e. shown to be a non-rank) by primes larger than P_j.
As one goes higher up in the number series, more and more primes are involved in the process, more and more numbers are covered by previous primes and the density of non-ranks increases. Since the twin ranks are the complements of non-ranks in the set of positive integers, *this explains why the density of twin primes decreases as the numbers go up*. The obvious question is: "Does the covering process reaches a point of saturation after which there is no need for a larger parent prime because all numbers have already been covered by the previous primes?" As shown below, this is not the case. In other words there is no largest parent prime. Indeed, following Euclid's approach let us assume that P_z is the largest parent prime, and let us apply equation (3.1) to the number $k_z = \prod_{i=3}^{z} P_i$. One has:

$$n = \frac{k_z}{P} \pm \frac{P \pm 1}{6P}. \tag{4.1}$$

Since all primes $5 \le P \le P_z$ divide k_z, it is easy to see that this Diophantine equation does not have an integer solution for $P \le P_z$. Therefore, if k_z is a non-rank, it must be of a parent prime larger than P_z; if k_z is a twin rank, clearly none of the primes $5 \le P \le P_z$ were able to cover it. The only way for the set of twin ranks to end at $M_{z+1} = \left(P_{z+1}^2 - 1\right)/6$ is for the primes larger than P_z to cover *all* subsequent numbers.

5 Summary and suggestions

Now, that with the help of twin ranks and their complements, non-ranks, we have a clearer picture about the twin primes and the way they are generated, we can pause to evaluate what we learned so far and see which characteristics (if any) can be used as a possible approach to the TPC? But first, let us briefly recapitulate our main findings.

Here is what we found:

(i) The twin ranks and non-ranks complement each other in the set of positive integers; once one knows the non-ranks in a number interval, one can find all twin ranks in that interval by simply subtracting those non-ranks from the set;

(ii) There are two non-ranks associated to a prime P symmetrically distributed at equal distances $[\pm P/6]$ from all numbers that are multiples of P; they can be found with the help of equation (3.1);

(iii) The covering process goes in steps from one basic interval to the next one, each transition requiring the use of a larger prime;

(iv) There is no largest parent prime. Sooner or later, as one goes higher up in the number series, one will need a new parent prime in order to be able to cover the larger numbers.

(v) The density of non-ranks increases as the numbers go up;

(vi) By ordering the non-ranks according to their parent primes, one can exactly calculate their number in certain intervals using equation (3.5).

And here is what we suggest:

A. Based on the first two properties, solve the following problem:

A set of sinusoidal curves of wavelength $\lambda_a = 6n-1$ *and* $\lambda_b = 6n+1$ (where n takes consecutive integer values), *originate at different points on the abscissa. Knowing that the curves of wavelength* λ_a *begin at the points* $(6-1)n-1$ *and* $(6+1)n-1$, *while the curves of wavelength* λ_b *begin at the points* $(6-1)n+1$ *and* $(6+1)n+1$, *find a formula for the number of distinct points at which the above curves intersect the abscissa on a given interval.*

Since the above solutions are all non-ranks, knowing their number in a given interval is equivalent with knowing the number of twin ranks in that interval. Note that in formulating the problem we did not require the use of prime numbers.

B. Based on the next two properties, consider the following approach: *Take the interval $M_z...M_{z+1}$ of length $\Delta M_z = M_{z+1} - M_z$.* As shown, one can use only the primes $5 \leq P \leq P_z$ to cover this interval, with the uncovered numbers being twin ranks. Since there are an infinity of intervals like this and no largest parent prime, chances are some of these intervals still contain uncovered terms after one has used all but the basic prime P_z. Because the covering process doesn't go backwards, the only chance to have all terms in the interval covered by P_z is for this prime to cover the rest. But this prime can cover maximum

$$N_c = 2\Delta M_z / P_z \tag{5.1}$$

terms (because there are only two non-ranks associated to P_z at the distances $[\pm P_z/6]$ from the numbers that are multiples of P_z), and many of these terms have already been covered by smaller primes. Prove that this is impossible for *all* intervals (an infinity of them) and, hence, the set of twin ranks cannot end at M_z.

C. Based on the last two properties (v and vi), approximate the number of twin ranks up to a given number as following:

Consider an interval of length $L_j = \prod_{i=3}^{j} P_i$. As shown, there are $S_j = L_j - \prod_{i=3}^{j}(P_i - 2)$ non-ranks of parent primes $5 \leq P_i \leq P_j$ in this interval. (We stress that this is an exact value and not an approximation). Their average density is:

$$\sigma_j = \frac{S_j}{L_j} = 1 - \prod_{i=3}^{j} \frac{P_i - 2}{P_i}. \tag{5.2}$$

This density increases as one goes up on the number series. Now, if one tries to approximate the number of all non-ranks in an interval $I > M_{j+1}$ by the product $N_{kI} = \sigma_j I$ one obtains a value smaller than the actual value. This is because beyond M_{j+1} there are non-ranks of parent primes larger than P_j which were not taken into consideration by equation (5.2). On the other hand, if one applies the same method to a much smaller interval (as for example $1...P_j$), and approximate the total number of non-ranks inside by $N_{kp} = \sigma_j P_j$, one obtains a value larger than the actual value. This is because by using a larger average density based on equation (5.2), one takes into account all primes $5 \leq P_i \leq P_j$, despite the fact that one needs only the primes up to $P_a \approx \sqrt{6P_j + 1} \ll P_j$ to cover the interval. All non-ranks of parent primes larger than P_a fall outside the interval $1...P_j$. Example:

Let $j = 9$. One has

$P_j = 23,\ P_{j+1} = 29,\ M_{j+1} = (P_{j+1}^2 - 1)/6 = 140,\ L_j = \prod_{i=3}^{9} P_i = 37182145,$
$S_j = L_j - \prod_{i=3}^{9}(P_i - 2) = 29229970$ and $\sigma_j = S_j/L_j \approx 0.78613$.

The approximate number of non-ranks in an interval $I = 1...1000 > M_{j+1}$, for example, is $N_{kI} = \sigma_j I \approx 786$. The actual number is $858 > N_{kI}$. On the other hand, the approximate number of non-ranks in the interval $1...P_j$ is $N_{kp} = \sigma_j P_j \approx 18$. The actual number is $13 < N_{kp}$. They are: 4, 6, 8, 9, 11, 13, 14, 15, 16, 19, 20, 21, 22. None of these non-ranks is of a parent prime larger than $P_a = 13 \approx \sqrt{6P_j + 1}$. Since the interval $1...P_j$ is well inside the interval $1...M_{j+1}$, all numbers remained after subtracting the above non-ranks are twin ranks. They are: 1, 2, 3, 5, 7, 10, 12, 17, 18, 23.
Therefore, because the density of non-ranks in an interval $1...P_j$ is much smaller than the density of the non-ranks of parent primes $5 \le P_i \le P_j$ in the corresponding super group L_j, one can conservatively approximate the number of twin ranks in the interval by $T_P = P_j - N_{kp}$. One has

$$T_P = P_j(1 - \sigma_j) = P_j \prod_{i=3}^{j} \frac{P_i - 2}{P_i}. \tag{5.3}$$

Recalling that $R_j = \prod_{i=3}^{j}(P_i - 2)$ represents the number of remnants in L_j (see equation 6) it is easy to see that the product

$$\chi_j = \prod_{i=3}^{j} \frac{P_i - 2}{P_i} \tag{5.4}$$

represents the average density of remnants of parent primes $5 \le P_i \le P_j$ in the super-group L_j. From (5.4) one concludes that:

Given a prime P_j, the number of twin ranks up to P_j is larger than P_j multiplied by the average density of remnants of parent primes $5 \le P_i \le P_j$ in the interval $4...L_j + 3$.

Since there is a one-to-one correspondence between the twin ranks in an interval $a...b$ and the twin indices in the interval $6a...6b$, it follows that:

Given a prime P_j, the number of twin prime pairs up to $6P_j$ is larger than P_j multiplied by the average density of remnants of parent primes $5 \le P_i \le P_j$ in the corresponding super-group.

From Mertens' 3^{rd} theorem [5],

$$\prod_{i=1}^{j} \frac{P_i - 1}{P_i} \simeq \frac{e^{-\gamma}}{\log P_j}, \tag{5.5}$$

(where $\gamma = 0.57721...$ is the Euler 's constant [4]), and the twin prime constant [1],

$$C_2 = \prod_{i=2}^{j} \frac{P_i(P_i-2)}{(P_i-1)^2} = 0.66016..., \tag{5.6}$$

one obtains

$$T_P \simeq 12C_2P_j\left(\frac{e^{-\gamma}}{\log P_j}\right)^2. \tag{5.7}$$

Here the factor of 12 comes from the fact that the products in equations (5.4), (5.5) and (5.6) start from different primes (P_3 =5, P_1 =2, and P_2 =3, respectively).

Example:

$j = \{10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, 10^{12}\}$;

$P_j = \{7919, 104729, 1299709, 15485863, 179424673, 2038074743, 22801763489, 252097800623, 2760727302517, 29996224275833\}$;

$T_P = \{245, 1957, 16374, 141064, 1240213, 11074501, 100080604, 913209584, 8399277629, 77768995855\}$;

$\pi_2(6P_j) = \{541, 4653, 40874, 364352, 3285526, 29919745, 274659396, 2538564662, 23599816507, 220502074373\}$.

It is important to realize that in the above formula one can use any number instead of P_j. For example, in the case of a basic number $M_j = (P_j^2-1)/6$ the number of twin ranks up to M_j, is on the order of

$$N_{tr} \simeq 12C_2M_j\left(\frac{e^{-\gamma}}{\log M_j}\right)^2. \tag{5.8}$$

This generalization is possible because the logarithmic part of (5.8) (which represents the density of remnants in the super-group) corresponds to a super-group of maximum parent prime on the order of M_j, while for covering all numbers smaller than M_j one needs only the parent primes up to $\sqrt{6M_j+1}$. Since

$$\frac{M_j}{\log^2 M_j} \propto \frac{P_j^2}{\log^2 P_j}, \tag{5.9}$$

based on the Prime Number Theorem, it is easy to see that $N_{tr} \approx j^2$. This explains the remark at the beginning of the paper regarding the number of twin ranks up to a basic number.

According to the so called *Hardy-Littlewood conjecture B* (see [3] for details), the number of twin primes pairs up to a number $n = 6P_j$ (and, hence, of twin ranks up to P_j) is on the order of

$$\pi_2(n) \approx 2C_2 \frac{n}{\log^2 n}.$$

Interestingly, although we used a completely different approach and did not resort to any of the considerations that lead Hardy and Littlewood to the above conjecture, for large numbers the ratio of the two approximations is constant and on the order of $e^{-2\gamma} = 0.315...$.

If one can rigorously prove that the number of twin ranks up to a prime P approximated with equation (5.7) is always smaller than the real value, one has proved the TPC, since P tends to infinity.

6 Conclusions

With this paper we do not pretend to have solved the Twin Prime Conjecture. Our objective was only to reveal new paths and approaches based on a class of numbers that nobody has studied so far (the non-ranks). The fact that these numbers are the complements of twin ranks in the set of positive integers makes them very interesting and potentially useful in our search for a solution to this centuries-old problem.

References

[1] S. R. Finch, Hardy-Littlewood constants, §2.1 in *Mathematical Constants*. Encyclopedia of Mathematics and Its Applications **94**, Cambridge University Press, Cambridge, 2003, 84-94. MR2003519. Zbl 1054.00001.

[2] R. K. Guy, *Unsolved Problems in Number Theory*. 3rd ed. Problem Books in Mathematics, Springer-Verlag, New York, 2004, 32-35. MR2076335. Zbl 1058.11001.

[3] G. H. Hardy and J. E. Littlewood, *On some problems of 'Partitio numerorum'*; III: On the expression of a number as a sum of primes, Acta Matematica, **44** (1923), 1-70. MR1555183. JFM 48.0143.04.

[4] J. Havil, *Gamma. Exploring Euler's constant.* With a foreword by Freeman Dyson, Princeton University Press, Princeton, 2003. MR1968276. Zbl 1375.26002.

[5] F. Mertens, *Ein Beitrag zur analytischen Zahlentheorie (A contribution to analytic number theory)*, *Journal fur die Reine und Angewandte Mathematik* **78** (1874), 46-62. MR1579612. JFM 06.0116.01.

[6] H. Tietze, *Famous Problems of Mathematics: Solved and unsolved mathematics problems from antiquity to modern times*, Authorized translation from the second (1959) revised German edition. Edited by Beatrice Kevitt Hofstadter and Horace Komm. (English), Graylock Press, New York, 1965, 7-9. MR0181558. Zbl 0133.24102.

Antonie Dinculescu
4148, NW 34th Drive, Gainesville, Florida, 32605, USA.
email: adinculescu@cox.net

License