

Prime Harmonics and Twin Prime Distribution

Serge Dolgikh

National Aviation University, Kyiv 02000 Ukraine,

Abstract: Distribution of twin primes is a long standing problem in the number theory. As of present, it is not known if the set of twin primes is finite, the problem known as twin primes conjecture. An analysis of prime modulo cycles, or prime harmonics in this work allowed to define approaches in estimation of twin prime distributions with good accuracy of approximation and establish constraints on gaps between consecutive twin prime pairs. With technical effort, the bounds on the distance between consecutive prime pairs obtained in this work can prove sufficient to establish that the next twin prime exists within the estimated distance, leading to the conclusion that the set of twin primes is unlimited and reducing the infinitely repeating distance between consecutive primes to two. The methods developed in the study can be instrumental in the future analysis of prime distributions.

1 Problem Statement

Whereas many results were obtained in distribution of prime numbers, culminating in Prime Number Theorem [1] and numerous bounds on the distance between primes [2, 3, 4], less is known about twin primes i.e., pairs of consecutive odd primes greater than 3. At this time it is not known whether the set of twin primes is finite or infinite, the latter being the statement of the twin prime conjecture or Polignac's conjecture for $n = 2$ [5]. A proof of the conjecture that can be restated as: for a twin prime pair $p = p_1, p_2$, there exists a greater pair $p_n > p$; would also reduce the minimum infinitely repeating distance between consecutive primes to 2 [6].

In this work we apply methods of modulo analysis to distribution of twin primes and develop approaches to estimation of the distance between consecutive prime pairs. As well, methods of numerical analysis were used throughout the work to illustrate and verify the results of the theoretical approach.

2 Prime Harmonics

2.1 Definitions

We will consider the set of odd positive integers \mathbb{N}_1 , with an integer step of 2 separating consecutive elements: $x_{n+1} = x_n + 2$. In the text that follows, the distances between numbers and the operations such as addition and subtraction will often be expressed in odd steps of 2. Usually such cases are clear from the context or indicated explicitly. For a pair $x, p \in \mathbb{N}_1$ we will use modulo notation $x \bmod(p)$

in the form: $0, p-1, \dots, 2, 1$ with the value of 0 the highest in the cycle of length p . Trivially, the positions with the same value of the modulo p are separated by a minimum of p odd steps.

For a prime p , the prime harmonic function $h_p(x)$ can be defined on \mathbb{N}_1 as the modulo of x by p in the above format:

Definition 1. A single prime harmonic $h_p(x)$ is defined for an odd integer $x \geq 1$ as:

$$\begin{aligned} h_p(1) &:= (p-1)/2 \\ h_p(n+1) &:= h_p(n) - 1, \quad h_p(n) > 0 \\ h_p(n+1) &:= p-1, \quad h_p(n) = 0 \end{aligned}$$

Clearly, $h_p(x) = 0 \equiv p \mid x$.

For an odd $n > 1$ a prime set of order n is the set of odd primes in n .

Definition 2. A prime set of order n is the ordered set of non-trivial odd primes less or equal to n :

$$G(n) := \{p : 1 < p \leq n\}$$

Evidently:

1. The size of a prime set $G(n)$ is given by the prime counting function: $\text{card}(G(n)) = \pi(n) - 2$ [1].
2. If $P(n)$ is the maximum prime of n then $G(n) = G(P(n))$.

Hence, if the order n is a prime $P(n) = n$ and $G(n)$ contains odd primes up to, and inclusive of n ; whereas for a composite order, $G(n)$ contains odd primes up to $P(n)$.

For $n \in \mathbb{N}_1$ the prime harmonics function H_n is defined as the ordered set of prime harmonics in $G(n)$:

Definition 3. For $n, x \in \mathbb{N}_1$ the prime harmonics function $H_n(x)$ is defined as the ordered set of prime harmonics p_k in $G(n)$: $H_n(x) = [h_{p_1}(x), h_{p_2}(x), \dots, h_{p_k}(x)]$

For example, for $n = 5$, $H_5(x)$ is defined for odd x as $[\bmod_3(x), \bmod_5(x)]$ with values $[2, 0], [1, 4]$ for $x = 5, 7$ and so on.

2.2 Harmonic Functions

With the prime harmonics function H_n at an order n it would be useful to define some cumulative indicators of the state of harmonics h_k in H_n at a given position x . This can be achieved with the cumulative harmonic function $C_n(x)$ that can be defined as follows:

Definition 4. Cumulative harmonic function of the order n , $C_n(x)$ is a cumulative value of prime harmonics in $H_n(x)$

defined as:

$$C_n(x) := \begin{cases} m+1, & \text{if } h_k(x) > m \forall p_k \in G(n), m \geq 1 \\ 1 & \text{otherwise, i.e. } \exists p_k \in G(n) : h_{p_k}(x) = 1 \end{cases}$$

where as defined earlier, for harmonics h_k we consider the value of $0 \equiv k$. Evidently, $C_n(x)$ indicates the proximity of x to the next composite number with respect to prime orders in $G(n)$: $C_n(x) = 1$ means that at least one prime harmonic in $G(n)$ will be at 0 at $x+1$ and thus it must be a composite. On the contrary, the value of $C_n(x)$ above 2 entails by the definition that neither of the following positions $x+1, x+2$ can be composite with respect to $G(n)$ as for any prime harmonic in H_n , $h_k(x) \geq C_n(x) - 2 > 0$ that is a necessary (but not necessarily sufficient) condition for constituting a twin prime pair.

Clearly, for a given position x the condition of the cumulative harmonic function above becomes sufficient for sufficiently large order of the harmonic function n , namely such that there is no divisors in x greater than n . Then, the conditions for a prime, and twin prime pair following a given position x can be written as:

$$\begin{aligned} C_n(x) > 1 &\implies p = x+1 : \text{prime} \\ C_n(x) > 2 &\implies p = (x+1, x+2) : \text{twin prime} \end{aligned} \quad (1)$$

where as noted, $n > x/3$ and $m > n \implies m \nmid x$.

The harmonic of order 3, $H_3(x)$ is the first non-trivial prime harmonic. Trivially, only the zeroes of $h_3(x)$, or the maximums of $C_3(x)$ can precede a twin prime pair, and only the values (0, 2) can precede a prime.

2.3 Prime Traversal

With the prime harmonic functions H_p and C_p one can point to a simple method for identifying primes and twin prime pairs in an arbitrary range. Starting at 3 as the first non-trivial prime, one can calculate $C_3(3) = 3$ meaning that the next two positions at 5 and 7 must be primes by the cumulative harmonic condition (1) at order 3 and the fact that they cannot have greater divisors.

Then, the identified primes are added to the set of known primes G_{p_m} and C_{p_m} calculated at the maximums of h_3 in the range up to $3p_m$ (the greatest prime at the current iteration), producing new primes and so on.

This observation leads to the lemma of the prime condition that follows directly from the cumulative harmonic condition (1).

Lemma 1 (Prime and twin prime condition).

$$\begin{aligned} C_n(x) > 1 \text{ and } x \leq 3n+2 &\implies p = x+1 : \text{prime} \\ C_n(x) > 2 \text{ and } x \leq 3n+2 &\implies t = (x+1, x+2) : \text{twin prime} \end{aligned}$$

Proof. First let us assume that the order of the prime set is a prime. Then $C_n(x) > 2$ indicates that no prime harmonics

up to n would reach 0 in the next two steps and $x+1, x+2$ cannot be divisible by n or any prime below it. Also, they cannot be divisible by any prime above n and the proof is complete for prime orders n .

If the order n of the cumulative harmonic function n is composite, then $C_n(x)$ condition still means that the pair following x cannot be divisible by any prime less than n . And because it cannot be divisible by any prime greater than n , the proof is complete for all cases.

The condition for the primes can be proven similarly. \square

The condition on the range of x in the lemma will be referred to as "the completeness condition", as it is sufficient for the value of $C_n(x)$ to indicate a true prime or twin prime pair. As can be seen immediately, it can actually be extended inclusive to $3 p_{next} - 1$, the next prime after n , that however, may not be known precisely though a number of strong bounds exist [2]-[4].

Based on this result we will refer to positions for which the conditions of harmonic function C_n are satisfied but not necessarily the completeness condition as "the candidates", in contrast to the true primes and twin prime pairs, for which both harmonic and completeness conditions are satisfied.

The diagram in Fig.1 provides an illustration of the regions of distribution of the candidates vs genuine twin primes in Lemma 1.

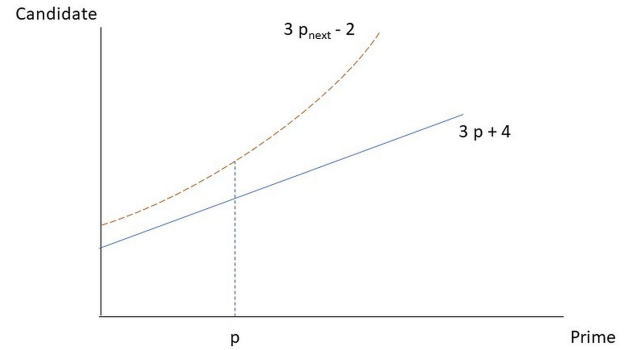


Figure 1: Prime and twin prime conditions

Though elementary, the lemma shows that the prime nature of an integer x is controlled entirely by its first prime harmonics at $x/3$ i.e. in the prime set $G(x/3)$. The values of C_n thus mark prime and twin-prime candidates that become primes and twin primes if and when the completeness condition is satisfied. Further in the text, the positions with values of C_n that are the necessary condition for prime and twin prime pair will be referred to as p , tp -candidates of order p respectively:

$$C_p(x) > 2 : \text{tp-candidate at order } p$$

$$C_p(x) > 1 : \text{p-candidate at order } p$$

In the rest of this work the focus will be on twin prime distributions and refer to tp -candidates as candidates. A trivial observation was already mentioned that a tp -candidate

x of any order n must be a maximum of h_3 harmonic as only $h_3(x) = 0$ satisfies the cumulative harmonic condition of Lemma 1.

2.4 Prime Set Cycle

When a new prime p is added to a lower order prime set, the change in H_p and C_p is controlled by two factors: the values of the lower prime set harmonics H_{p-1}, C_{p-1} and the prime harmonic of order p , h_p . A complete combination of these values can be described by a cycle K_p with the length of $L(p) = p \times L(p-1) = p\#$ (the primorial of p) after which the harmonics of H_{p-1} and h_p will reach the same state and the cycle repeats.

Evidently, prime set cycles must be closely related to cyclic groups of prime and composite orders; this relation will be explored in another work.

An essential observation that follows immediately is that for an order p that is sufficiently large, the position of p in its prime set cycle will be very near its center, in the sense that the ratio of the distance from the center of K_p to p to the length of K_p tends to zero with p .

Indeed, as seen above, the length of K_p , L_p being $p\#$ whereas its end marked by the maximum conditions of K_{p-1} and h_p achieved at position $Z_p = p L(p-1) = p\# \gg p$ and:

$$p/Z_p = \frac{1}{(p-1)\#} \rightarrow 0 \quad (2)$$

For example, for $p = 11$, the the length of the cycle L_{11} is 1155, and p is less than 1% of the "radius" of C_{11} from the center.

It is straightforward to estimate the number of prime and twin prime candidates $R_p, R_t(p)$ in a p -set cycle. Indeed, given that any combination of H_{p-1} values and p -harmonic will happen exactly once in the p -cycle, one immediately derives:

$$\begin{aligned} R_p(p) &= (p-1) R_p(p-1) = \prod_{\text{prime } m=3}^p (m-1) \\ R_t(p) &= (p-2) R_t(p-1) = \prod_{\text{prime } m=3}^p (m-2) \end{aligned} \quad (3)$$

where m are the members of G_p . The analysis that follows will focus on twin prime pairs and $R(p)$ would signify the number of twin prime candidates in the prime set cycle of G_p .

From (2) and (3) one can easily obtain the density of tp -candidates in a p -cycle as the number of candidates by its length:

$$d(p) = R(p)/L_p = \prod_{\text{prime } m=3}^p \left(1 - \frac{2}{m}\right) \quad (4)$$

but of greater interest would be the number of candidates within a range of length p that is given by $d_r(p) = p d(p) = R(p) / L_{p-1}$. As is easy to see, $d_r(19) \sim 1.5$ and increases monotonously with p .

It can be noted that the reciprocal of d_r has a clear meaning as the average gap $g(p)$ in the cycle of order p , relative to p .

Then, from above and Mertens's second theorem [7] the average density of candidates can be estimated as $\propto p/\log^2 p$, and the average gap size in a prime set cycle of order p , as:

$$g(p)/p \propto \frac{\log^2 p}{p} \quad (5)$$

i.e. that the average gap between twin prime candidates in C_p relative to p must tend to 0 as p increases.

One may note that Lemma 1 and the estimate in (5) offer a statistical hint for the validity of the twin prime conjecture. Indeed, if the gaps were distributed uniformly, at certain order p the average gap between the candidates would decrease to a value that is small enough value to satisfy the completeness condition of Lemma 1 that is sufficient to make them a true twin prime pair.

Of course, there is no reason to expect that the uniformity of gap distributions would hold at higher prime orders and more detailed analysis of gap distributions at arbitrary order is needed.

3 Gap Distributions

In this section we will look at the distribution of gaps between given values of $C_p(x)$, such as those that define tp -candidates, Section 2.3 in a general prime set cycle.

Some preliminary notes:

We will consider the upper, i.e. the positive half of the prime set cycle of an arbitrary order p with the center position, in odd steps, at -1 , so that the integer value of 1 is one odd step from the center of the cycle and so on. The positions of gaps and intervals in this section will be indicated in odd steps relative to the center position unless stated otherwise.

Definition 5. A gap distribution of order p is defined as an ordered set of gaps between consecutive tp -candidates of the order: $D_p = \{ g_k(p) \} = \{ tp_{k+1}(p) - tp_k(p) \}$

Clearly, with a gap distribution D_p the positions of all tp -candidates of the given order can be identified by sequential summing of gaps.

As can be seen immediately, in the first prime set of order 3 the distribution is uniform, with the maximum gap same as minimum, and the length of the cycle, $D_3 = [3]$. As outlined above, the first maximum of C_3 occurs at the odd position 2 from the center (that is equal to integer value 3), and every $L_3 = 3$ steps thereupon. Clearly, only the maxima of C_3 can precede tp -candidates, with the obvious conclusion that all gaps in any order must be divisible by 3.

For higher prime orders p , the distribution of gaps will be controlled by the interaction of the higher p -harmonic with the cumulative harmonic function of the previous order C_{p-1} that defines the gap distribution D_{p-1} . If at the

position of a certain C_{p-1} candidate c_k $h_p(c) \leq 2$, then the following pair $(x+1, x+2)$ cannot be twin prime and the candidate at order $p-1$ is “erased” by a collision with the higher harmonic.

For example, with $p=5$, the position of the first maximum of h_5 is at 3, and the collision positions of the harmonic $h_5(x) = 1, 2$ are (6, 7); (11, 12) and so on, colliding with the C_3 candidates at positions: 11, 17 and so on, and creating gaps at the positions of collision.

A direct but important in the analysis that follows observation is that for a harmonic of order p there are exactly two such collision positions at $h_p(x) = 1, 2$ per each modulo cycle of h_p of length p , so it is worthwhile to introduce the notion of order-related intervals in the analysis of gap distributions.

3.1 Intervals

In the analysis of candidate and gap distributions in the prime set cycle of order p it is convenient to define the intervals of length p , starting from the central position of the cycle.

Definition 6. The k^{th} interval of a prime harmonic h_p is a range of length p from the k^{th} maximum of h_p :

$$J_k(p) := \left[\frac{p+1}{2} + (k-1)p, \frac{p+1}{2} + kp \right], k \geq 1$$

$$J_0(p) := \left[1, \frac{p+1}{2} \right]$$

The k^{th} collision range of a prime harmonic h_p , $I_k(p)$ is defined as the last two positions in the interval $J_k(p)$:

$$I_k(p) = (J_k[p-1], J_k[p])$$

From the definition, $I_k = (J_k[1] + p - 2, p - 1) = (J_{k+1}[1] - 2, 1)$. Also straightforwardly, the distance between the corresponding positions of the consecutive collision ranges is p , whereas the minimum and maximum distance between the positions of the consecutive collision ranges is $p-1, p+1$ respectively.

Let us consider a harmonic of order p . An intersection of a collision range with a gap boundary in the gap distribution of the previous order D_{p-1} would signify a collision with a p -candidate resulting in the elimination of the candidate and a merger of adjacent gaps in the distribution D_p .

One can note that while mergers change the size of gaps, they do not change the positions of gap boundaries in the initial C_3 cycle. Hence, in the analysis of distributions of any order p one can conclude that if collision ranges of h_p did not intersect with the gap boundaries in the initial prime cycle C_3 , a collision of the p -harmonic with gap distributions of any lower order is not possible.

Consider the gap distribution D_3 , starting from the center and extending upwards indefinitely. Then, for a given order p , consider the positions of collision ranges I_k that immediately precede $(k+1)^{\text{th}}$ maximum of h_p as defined

previously.

Now incrementing the order of the cycle: $p_n = p + 2$ from Def.6, the positions of the maxima and the collision ranges change as:

$$I_k(p_n) = I_k(p) + (2k+1) \quad (6)$$

i.e., as the order p increases, the intervals and collision ranges effectively move outwards with different “speeds”: 1, 3, 5, ... for intervals 0, 1, ..., respectively. Thus, superimposing the movement of collision intervals over the initial gap distribution one can identify the positions of possible collisions and the resulting gap distributions in the higher orders.

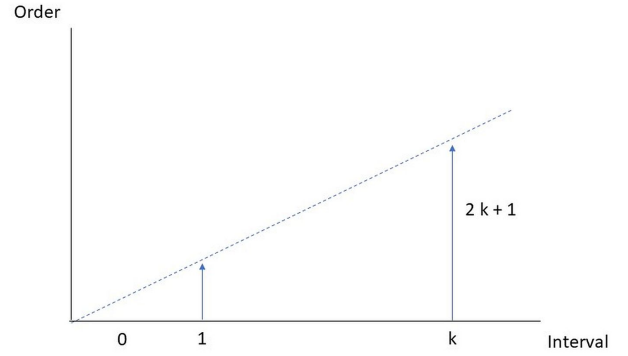


Figure 2: Collision ranges in gap distributions

In conclusion, a note on the zero-th interval $J_0(p)$ and collision range $I_0(p)$. Its upper boundary is positioned at $(p+1)/2$ and moves upwards by 1 with the order p and therefore is bound to reach any position in the cycle at certain order p_0 that can be easily calculated. Because for any position $x \in J_0 < p$, collisions with a p -harmonics cannot occur in this interval, whereas for a composite p such a collision, if possible, would have already happened with a harmonic of a prime factor of p . Consequently, no new collisions can occur in this interval and all candidates of the lower distribution D_{p-1} are immune at least up to the first collision range $I_1(p)$.

3.2 Gap Distribution Properties

From the definitions and observations in the previous sections some properties of gap distributions D_p can be pointed out immediately.

Corollary 1 (Central symmetry). For $p > 3$ gap distribution D_p is near-symmetrical with respect to the center of the prime set cycle K_p .

The proof follows from the observation that a prime set cycle must be invariant to the reflection with respect to the center with simultaneous reversal of the direction of the modulo iteration, so that the collision points and gaps at the same distance from the edges of the cycle must have same values. Note that the exact symmetry is achieved at the position $L_p - 3$, so that the last gap in any p -cycle has

a length of 3.

Next, it can be readily derived from Def.2 of the composite harmonic function C_n that for the orders n that are composite, the gap distribution will coincide with the distribution of the prime set of n , therefore:

Corollary 2 (Composite orders). *For a cycle K_n of a composite order n with p_m the highest prime in n , $p_m < n$, there will be no new collisions between cumulative harmonics C_{p_m} and the n -harmonic, and the gap distribution D_n is identical to D_{p_m} .*

The proof follows from the fact that the maximums and therefore, collision ranges of a composite harmonic will always coincide with those of its prime factors. An obvious consequence is that new gap creation in consecutive orders is possible only for twin prime orders.

In the analysis of gap distributions in the center of the cycle, i.e. lower values of prime harmonics, and at its edges i.e. higher values in the prime set cycle of a given order, it will be useful to define a collision function of an (odd number) position x and a harmonic of the order p as an indicator of intersection of a collision range of h_p with x :

$$\begin{aligned} \phi(x, p) &:= 1 \text{ if } \exists k : x \in I_k(p) \\ \phi(x, p) &:= 0 \text{ else} \end{aligned} \quad (7)$$

Similarly, a cumulative collision function for a set of orders $S = \{ p_k \}$ can be defined as the sum of collision functions for $p_k \in S$:

$$\phi_S(x) = \sum_{i \in S} \phi_i(x) \quad (8)$$

The meaning of the cumulative collision function at a position x is evidently, the number of times x is intersected by prime harmonics in the range S . A special case of the cumulative collision function is when S is a prime set, $S = G(p)$. In this case, $\phi_S(x) = 0$ indicates that the position at x is "immune" up to the order p and if $x \in D_3$ then x is also a candidate of order p :

$$\phi_{G(p)}(x) = 0 \text{ and } x \in D_3 \implies x \in D_p$$

3.3 Central Gap Distributions

In this section we shall examine gap distributions near the center of a prime set cycle i.e. at positions starting from 1 and upwards. The central area is remarkable because for any order p it is at the same time, the middle of the middle cycle of the prime set cycle; and the middle of the p -harmonic modulo cycle.

As was shown in Section 3.1, collisions with the p harmonic cannot happen at the interval J_0 . It can be seen easily that collisions cannot happen in the next interval J_1 either: because the collision range in this interval increments by multiples of 3, if it did not intersect with C_3 at a certain order, it could not for any higher one. This indeed is the case for $p = 5$ with $I_1(5) = (6, 7)$ and does not intersect with C_3 gap boundaries at 5, 8. An alternative way to

prove it is to notice that the second maximum of $h_p = 3p$ in J_2 has to be in D_3 and therefore, I_1 cannot intersect with D_3 .

A detailed examination of the second collision range shows that collisions are not possible there either, leading to following statement:

Lemma 2 (Collision suppression at lower intervals). *Collisions with the lower distribution D_{p-1} are not possible in the first and second intervals of prime harmonic $p > 7$.*

Proof. The lemma was proven for I_1 above, and we will outline the proof for I_2 with the reference order $p_0 = 7$. As explained in the note to (6) it needs to be proven that at any order that is greater than p_0 , the collisions at this interval cannot create new gaps compared with the distribution D_5 . The position of the third maximum of h_p at $p_0 = 7$ is 18, the second collision range $I_2(p_0) = (16, 17)$ and given that I_2 advances with the rate of 5 with each increment of the order, the positions of I_2 at the next two orders after p_0 are: (21, 22) and (26, 27). As can be seen immediately, only the positions 17, 26 intersect with D_3 .

Secondly, considering the collision ranges at order 5 $I_k(5)$, it follows that both positions above intersect with the collision ranges of h_5 harmonic and are eliminated in the D_5 distribution: $17 = (5 + 1)/2 + 3 \times 5 - 1$; $26 = (5 + 1)/2 + 5 \times 5 - 2$. For that reason, higher harmonics in the considered range of orders cannot intersect with new tp-candidates and create new gaps.

Finally, one can recall that the length of the prime set cycle C_5 is 15 and the gap pattern of D_5 repeats with the same period. Then, as was shown above, the only collisions of I_2 with D_3 within a range of the length 15 are also those of h_5 , no new collisions can happen in that range, and given the period of C_5 , in any other range of the same length, completing the proof. \square

A similar but more tedious analysis of D_7 distribution shows that collisions are not possible in the third interval J_3 as well at orders higher than 7. As in the case of J_2 , the proof involves sequential comparison of the D_7 gap positions versus D_3 . As the rate of expansion of I_3 collision range is 7, being a divisor of L_7 , the gap position analysis needs to be done for the first fifteen of h_7 collision ranges with the pattern of gaps and collision intervals repeating thereon.

It is straightforward to prove that collisions cannot occur in intervals $1 + 3k$ at any order p because collision ranges in these intervals do not intersect with the initial candidate distribution D_3 : $I_1 + 3kp = I_{3k+1} = d_3 - 1, 2 + 3kp$:

Corollary 3 (Third intervals). *For $p > 3$ collisions are not possible in the h_p intervals $1 + 3k$.*

The next observation allows to estimate an order-dependent bound on the position of the nearest possible collision in the central distribution of any order.

Let us consider the first $(p + 1)/2 \sim p/2$ intervals and

collision ranges at an order p . As was noted previously, for any harmonic only the intersections of collision ranges with the initial, C_3 distribution can result in collisions. As can be seen immediately, every third collision range in any order does not overlap with C_3 and can be eliminated. For the rest it is possible to identify the conditions of overlapping of between the collision ranges of p with those of the lower harmonics p_k . If such an overlapping happens, a collision would not create a new gap as the candidate at the position has already been eliminated by the lower harmonic. These arguments lead to the following lemma:

Lemma 3 (Minimum collision distance). *The minimum distance to the first possible collision of h_p and C_{p-1} $t_m \sim 1/2 p^2 \geq (p^2 - 3)/2$ for $p > 3$.*

Proof. Taking any interval $k < (p - 1)/2$ that intersects with D_3 at one of the values in the collision range I_k , the condition of an overlap for h_p with a lower harmonic p_l can be written as:

$$\begin{aligned} (p+1)/2 - [1,2] + k p &= \\ (p_l+1)/2 - [1,2] + k_l p_l & \\ (2k+1) p = (2k_l+1) p_l - 2 \times [1,2] + 2, \text{ or} & \\ (2k+1) p = (2k_l+1) p_l - 2 \times [1,2] + 4 & \end{aligned} \quad (9)$$

where the left side has to intersect with D_3 and for the given k, p the solution sought is p_l, k_l . Clearly, the condition above can always be satisfied by choosing $p_l = 2k + 1 < p$, $k_l = (p - 1)/2$ and the appropriate choice of the collision range position of p_l , 1 or 2. This means that for any interval range J_k , $k < (p - 1)/2$ where a collision is possible, there will be a lower harmonic with an overlapping collision range and correspondingly, no new gaps can be created.

If p_l above is a composite then an overlap would still happen with one of its prime factor harmonics. The only exception is if $p_l = 2k + 1$ is a power of 3, but in that case as is easy to see from Cor. 3, collisions with D_3 would not be possible in these intervals.

Finally, for the collision range $I_{(p-1)/2}$ the proof will not work as p_l cannot be greater than p and a collision is possible. In this case the minimum position of the collision would be at the offset of -2 of the next interval maximum i.e.:

$$P_{min} = (p^2 + 1)/2 - 2 \sim p^2/2 \quad (10)$$

□

As can be shown, the condition for the collision at the last range is equivalent to $p^2 - 2$ being a composite. This condition in turn depends on factorization of the expression $2x^2 - 1$ that has monotonously growing number of prime factors at higher p , x providing an heuristic argument for the conjecture that gap creation at the last collision range of the "immune zone" has to be suppressed at higher p .

As seen in this section, gap creation and growth is suppressed in the central distributions by several factors:

1. Interval-specific constraints, J_1 to J_3 and J_{1+3k} .
2. The minimal distance to the first collision constraint, Lemma 3.
3. General suppression of new gap creation due to increasing number of composite orders where no new gaps are created (Cor.2).

It follows from these results that a characteristic feature of the central distributions in higher orders has to be suppression of collisions and gap creation in the area near the center of the cycle including the range protected from collisions that expands from the center quadratically with the order p . For these reasons it can be expected that gap distributions at higher orders will be more uniform in the ranges close to the center.

This conclusion is confirmed by numerical modeling of the central distributions as illustrated below.

D_7 : [2, 3, 3, 6, 6, 9, 6, 15, 3, 15, 6, 9, 6, 6, 3, 6, 6, 3, 6, 6, 9, 6, 15 ...]

D_{19} : [[D_7], 18, 36, 6, 15, 30, 24, 15, 9, 12, 9, 75, 6, 3, 6, 6, 3, 12 ...]

D_{383} : [[D_{19}], 69, 6, 9, 54, 12, 15, 3, 60, 6, 24, 15, 12, 33, 42, 3, 27, 9, 24, 15 ...]

The examples show a more uniform pattern of gap distribution near center in the lower intervals. The results of numerical modeling for the mean and maximum gap sizes in the immune from collisions range $p^2/2$ relative to the order p decrease from, respectively, (0.51, 2.58) at $p = 29$ to (0.11, 0.82) at $p = 383$, in agreement with the conclusions of the analysis in this section.

3.4 Edge Gap Distributions

As follows from Cor.1, edge distributions are symmetrical with respect to the center, so we can consider the distributions at the beginning of the p -cycle without loss of generality. Edge distributions begin at a maxima of both C_{p-1} and p -harmonic with first collision interval at $(p - 2, p - 1)$. There are no known constraints for collisions and gap creation at the edges other than the general one of the composite order Cor.(2). In fact, a straightforward analysis of the collision function near the beginning of the cycle shows that in these distributions the first gap will grow indefinitely as the order increases.

It follows immediately from the observation that for any gap of size g_{ed} , the harmonic of the order $g_{ed} + 2$ or if composite, one of its prime factors will have a collision with the candidate at the boundary of the gap creating a larger gap that is thus bound to grow indefinitely.

This observation is confirmed by the examples of edge distributions obtained in numerical modeling:

D_{19} : [30, 15, 12, 3, 12, 9, 24, 21, 21, 9, 6, 3, 6, 6, 15, 18, 15, 6, 24, 6 ...]

D_{43} : [105, 21, 30, 9, 12, 15, 18, 15, 30, 6, 6, 9, 6, 9, 21, 3, 156 ...]

D_{317} : [660, 12, 45, 15, 24, 81, 24, 15, 9, 96, 15, 21, 75, 60, 33, 15, 12, 24, 45, 24, 12, 3, 45 ...],

A conclusion that can be drawn from these observations is that the gap distribution at a given order p , with the average gap size tending to zero relative to the order (5) should have larger gaps tending toward the edge of the cycle (i.e. higher positions in the cycle) with the distributions near the centre, on the contrary, more uniform and dominated by smaller gaps.

3.5 TP-candidates in Central Distributions

In the analysis that follows it would be helpful to estimate the number of tp-candidates in the region near the center of the cycle defined by the distance d from the center. To achieve this one would need to estimate the number of collisions at an arbitrary order p with the original candidates in D_3 .

First, we will observe that in the distribution D_5 there can be no collisions with the lower harmonics and the distribution of candidates and gaps is defined entirely by collisions of h_5 with D_3 , resulting in a near uniform periodical gap distribution with the pattern 6 – 6 – 3, period of 15 and the average gap size of 5.

At an arbitrary order p , the number of candidates in a region within the distance d to the center, N_{can} will be determined by these factors:

1. The number of D_3 candidates in the interval $[1, d]$: $N_3(d)$.
2. The number of D_3 candidates in the interval that collided with h_p and one or more of the lower harmonics in G_{p-1} : $N_{3,p-1}(d)$.
3. The number of collision ranges of h_p in the interval that intersect with D_3 : $N_{col,3}(p, d)$.
4. The number of collision ranges of h_p that intersect with D_3 and overlap with a collision range of one or more of the lower orders $l < p$: $N_{col,3,l,p}(p, d)$.

Then, the number of new collisions at the order p with the candidates of D_3 can be estimated as:

$$\begin{aligned} N_{can}(p, d) &= N_3(d) - \sum_{k \in G_p} N_{col}(k, d) \\ N_{col}(k, d) &= N_{col,3}(k, d) - \\ &\quad - \sum_{l < k} N_{col,3,l,k}(l, k, d) \end{aligned} \quad (11)$$

where the sum is over the prime set of p , G_p .

To estimate the number of collisions between a higher and a lower harmonics p , l at D_3 candidates $N_{c3,l,p}(p, d)$ one can use the condition of overlap for h_p , h_l , (9).

Immediately it can be observed that for each collision range of order p that intersects with D_3 only one of the offsets from the maximum of h_p , (1, 2) can result in a collision and thus the offset value in (9) should be considered

fixed leading to the following conditions for the overlap with the lower harmonics:

$$\begin{aligned} (2k+1)p &= (2k_l+1)l \\ (2k+1)p &= (2k_l+1)l+2 \\ (2k+1)p &= (2k_l+1)l-2 \end{aligned} \quad (12)$$

Further, it is easy to see that in each of the cases above 1) the solutions, if exist, repeats every l of p -intervals; and 2) every third solution in each case intersects with a D_3 candidate. The complete solution of the overlap condition (12) for each prime harmonic $l < p$ then consists of three sequences: $(2k+1) \bmod l = 0; \pm 1$ for the k^{th} collision range of h_p and for a given order p and a lower prime order $l < p$ we have exactly 4 overlap positions in the interval of length $3p$: two with the collision range offsets of the same value for p and l ; and one for each possible combination of different values.

This observation allows to estimate the number of collisions in a central interval of sufficient length d for an arbitrary order p via **PIE**-style summation, where we want to count each position of overlapping D_3 collisions of p and lower order l exactly once:

$$N_{col}(p, d) = N_{col,3}(p, d) - \sum_{l < p} N_{ov}(p, l) \quad (13)$$

where l , $l_1 < p$, are the lower harmonics, N_{ov} , overlaps of collision ranges of h_p with the lower harmonics and N_{int} , the number collision ranges of h_p in the central range d . Then, in the above, from Cor.3, (12) and the earlier observations in this section,

$$\begin{aligned} N_{int}(p, d) &\approx 2d/3p \\ N_{ov}(p, l) &\approx 4d/3p \left(\sum_{l < p} 1/l - N_{ov}(p, l, \dots) \right) \\ N_{ov}(p, l, l_1) &\approx \frac{2}{l l_1} \end{aligned} \quad (14)$$

where the last term in the above represents multiple overlaps of collision ranges p with several lower harmonics l, l_1, \dots for which every position is counted only once and the approximation is caused by the possibility of incomplete intervals in the range d (fringe or granularity effects). It is straightforward to see that in higher-order overlaps for each position of an overlapping condition of D_3 , h_p , and multiple lower harmonics h_l there will be exactly two possible positions of an overlap with a collision range of another lower harmonic l_1 , for each of the two possible values of collision offset of l_1 . Consequently, in each next iteration of overlaps the range of the full cycle of possible overlaps is multiplied by l_1 and the number of overlaps in the range, by 2.

Then combining all terms in the above:

$$\begin{aligned} N_{col}(p, d) &\approx 2d/3p \left(1 - 2 \left(\sum_{l < p} \frac{1}{l} - \right. \right. \\ &\quad \left. \left. - 2 \sum_{l, l_1 < p} \frac{1}{l l_1} + 4 \sum_{l_k < p} \frac{1}{l l_1 l_k} \pm 2^k \sum_{l_k < p} \frac{1}{l l_1 \dots l_k} \right) \right) \\ &= 2d/3p (1 - 2\chi(p)) \end{aligned} \quad (15)$$

As follows from the above, the number of collisions at a central interval of a sufficient length at an arbitrary order p is controlled by the twin prime collision function $\chi(p)$. Rearranging the terms in (15) from the lower to higher, one obtains:

$$\chi(p) = \frac{1}{l_1} + \frac{1}{l_2} \left(1 - \frac{2}{l_1}\right) + \frac{1}{l_3} \left(1 - \frac{2}{l_1} - \frac{2}{l_2} + \frac{4}{l_1 l_2}\right) + \sum_{l < p} \frac{1}{l} (1 - 2f_k) = S_{lm-1} + \frac{1}{l_m} (1 - 2S_{lm-1}) \quad (16)$$

Thus, as follows immediately from (15) and the above, the twin collision function $\chi(p)$ and the total number of collisions $N_{col}(p, d)$ at order p are defined by the recursive sequence:

$$\begin{aligned} S_n &= S_{n-1} + \frac{1}{p_n} (1 - 2S_{n-1}) = \frac{1}{p_n} + \left(1 - \frac{2}{p_n}\right) S_{n-1} \\ S_1 &= 1/5 \\ 1 - 2S_n &= \left(1 - \frac{2}{p_{n-1}}\right) (1 - 2S_{n-1}) = \\ &= \prod_{3 < p < p_n} \left(1 - \frac{2}{p}\right) \end{aligned} \quad (17)$$

with p being prime orders starting with 5, as:

$$\begin{aligned} \chi(p_n) &= S_{n-1} = \frac{1}{p_{n-1}} + \left(1 - \frac{2}{p_{n-1}}\right) \chi(p_{n-1}) \\ N_{col}(p_n, d) &= \frac{2d}{3} S_n \\ N_{can}(p_n, d) &= \frac{d}{3} - \frac{2d}{3} S_n = \frac{d}{3} (1 - 2S_n) \end{aligned} \quad (18)$$

Some properties of the twin χ function can be pointed:

1. First values $\chi(5) = 0$; $\chi(7) = 1/5$; $\chi(11) = 2/7$ and so on according to (18)
2. Monotonously increasing: $\chi(p_{k+1}) > \chi(p_k)$ (follows from (17), 3)
3. Upper limit at $p \rightarrow \infty$: 0.5, asymptotics discussed further below

The equations (15)-(18) allow to conclude that the number of collisions with D_3 and the resulting central gap distribution at an arbitrary order p and sufficiently large distance d is controlled by the behavior of $\chi(p)$ and is approximately proportional to the distance, as long as it is sufficiently large to make the granularity effects negligible. Figure 3 plots the behavior of twin χ at lower orders, obtained with numerical modeling of (18).

Asymptotic Behavior Of interest here will be the asymptotic behavior of $1 - 2\chi(p)$ because from (15) it is directly related to the density of new collisions at order p , $N_{col}(p, d)/3d$.

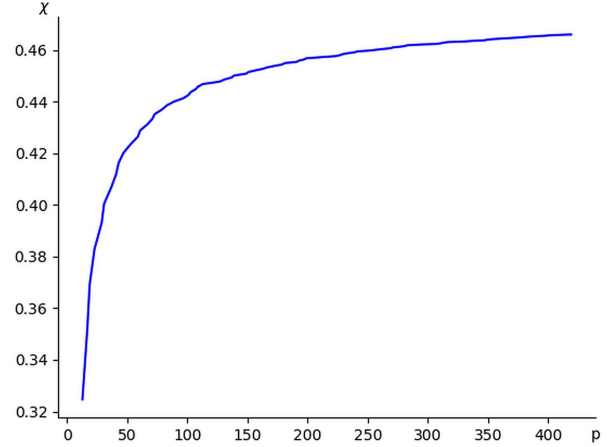


Figure 3: Twin prime χ at first prime orders

From (17), (18) and Mertens theorem [7] it can be obtained [8] that $\chi(p)$ in a central range with distance d behaves asymptotically as:

$$1 - 2\chi(p) \propto \frac{1}{\log^2 p} = \frac{A}{\log^2 p} + O\left(\frac{e^{-c\sqrt{\log p}}}{\log^2 p}\right) \quad (19)$$

where $A/3 = 4e^{-2\gamma}\Pi_2 = 0.83243\dots$, the twin prime density constant, without consideration of the granularity / fringe effects at shorter ranges, as commented further.

As can be concluded from (15), the full range at which all overlaps with all lower harmonics take place at an order p has to be in the order of $3 \times p \#$ (primorial) and at shorter ranges overlaps with higher harmonics can be suppressed due to incomplete range. A direct application of the results in (17) and (18) to these ranges would make an assumption of proportional distribution of collision and overlap positions in incomplete ranges that may not be substantiated. For this reason, a more detailed analysis of the effect of incomplete ranges is needed to obtain robust bounds on the number of tp-candidates at arbitrary distances.

Collision and Candidate Estimate As was shown in (18), an estimate on bounds on $\chi(p)$ at higher orders can lead directly to the estimates on the number of collisions and candidates at an arbitrary order p , as the density of both is directly related to $S_n = \chi(p_{n+1})$ and obtaining rigorous bounds on the number of collisions and overlaps in (15), (16) at an arbitrary central distance will be essential in the understanding of distribution of tp-candidates at higher prime orders. This investigation will be the subject of another work, but for now we will use the estimate (19). Then, straightforwardly from (18),

$$\begin{aligned} N_{can}(p_n, d) &= \frac{d}{3} (1 - 2\chi(p_{n+1})) \approx \frac{A d}{\log^2 p_{n+1}} + \\ &+ O\left(\frac{e^{-c\sqrt{\log p_{n+1}}}}{\log^2 p_{n+1}}\right) \end{aligned} \quad (20)$$

and there exist a number of strong bounds on the distance between p and p_{n+1} , the next prime order [3], [4]. This estimate can be applied to the no-collision range to evaluate the number of candidates immune to collisions at higher orders. Immediately, with $d = p^2/2$ one obtains:

$$N_{can}(p, d \sim p^2/2) = \frac{A p^2}{2 \log^2 p} \quad (21)$$

Numerical modeling yields excellent agreement of (18) with the actual distribution of twin primes in the verified range of up to 200 first pairs (Figure 4).

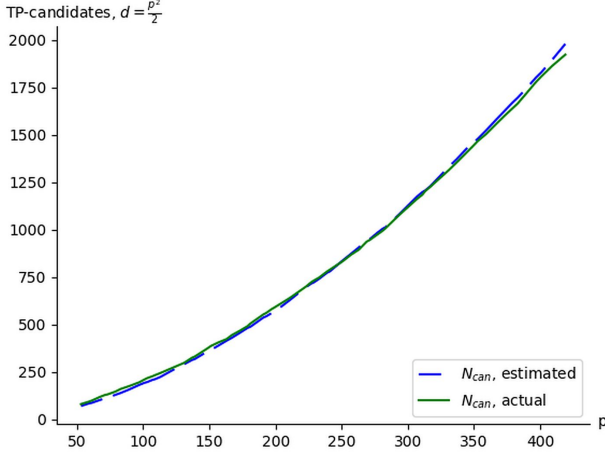


Figure 4: TP-candidates, estimated vs actual at $d = \frac{p^2}{2}$

The bounds (19) - (20) also yield good agreement with the actual distribution of twin primes at higher orders.

In conclusion of this section, we will note a result that follows immediately from the estimates (19)-(21) and will be used further in establishing the bounds on gaps between the candidates in the lower intervals. A rigorous proof of this result will need to be given to include ranges with the length $d \sim p^2 \ll 3 \times p \#$ and therefore account for the granularity effects of incomplete ranges at these distances.

Lemma 4 (Candidate Distribution). *For sufficiently large orders $p > P$ there is at least one tp-candidate c_p of the gap distribution D_p in the interval between collision-immune ranges of orders p and p_n : $(p^2 + 1)/2 < c_p < (p_n^2 - 3)/2$.*

where p_n is the next prime after p .

Proof. Assuming that granularity / incomplete interval effects were accounted for, from (21) for the number of tp-candidates in the ranges p, p_n one gets:

$$\begin{aligned} N_{can}(p, d(p)) &= \frac{A p^2}{2 \log^2 p} + \text{higher orders} \\ N_{can}(p, d(p_n)) &= \frac{A p_n^2}{2 \log^2 p_n} + \text{higher orders} \\ N_{can}(p, d(p, p_n)) &= N_{can}(p, d(p_n)) - N_{can}(p, d(p)) \sim \\ &\frac{2 A p}{\log^2 p} + \text{higher orders} \end{aligned}$$

where $d(p, p_n)$, the range between no-collision ranges of orders p, p_n .

The main term in the above is unbounded and there will be P such that $N_{can}(p, d(p, p_n)) > 1$. \square

Numerical modeling confirms that Lemma 4 holds for $p > 17$ in all verified prime orders. The number of tp-candidates in the range of length $2p < d(p, p_n)$ in the interval between collision-free ranges of consecutive prime orders is shown in Table 1.

Table 1: Tp-candidates in the boundary range

Ordinal range	5-20	40-70	90-120	150-180
Tp-candidates	2 - 5	4 - 18	10 - 26	21 - 38

3.6 Gap Estimate in Central Distributions

The analysis in the previous sections led to understanding of the gap creation process in the central and edge areas of prime set cycles, with the estimate on the number of collisions and candidates at an arbitrary order and range in the central distributions. Gap creation at any candidate position x in the initial distribution D_3 in an arbitrary order is then controlled by the harmonics h_p of the current order and cumulative harmonic function C_{p-1} of the lower orders at the candidate position.

The harmonic collision function $\phi_p(x)$ was defined earlier (8) as the number of prime harmonics $p_l \leq p$ that collided with the initial distribution D_3 at position x . A collision at x indicated by a positive value of ϕ creates a gap at that position and simultaneously, as discussed in the previous section, prevents all further collisions at this position.

Numerical analysis of the behavior of ϕ at the lower positions near the center shows that there exist positions that are "immune" to collisions, for example, $x = 53, 95, 119, \dots$. At these positions, the collision function at the maximum order where gap creation is still possible $p_{max} \sim \sqrt{2x}$ (Lemma 3), $\phi_{p_{max}}(x) = 0$ and collisions are avoided, whereas at the higher orders they are no longer possible due to the distance constraint of Lemma 3. Such immune positions satisfying the condition $\phi_{\sqrt{2x}}(x) = 0$ form the permanent distribution of twin primes in the central zone of x and the process continues to the higher orders indefinitely.

In the general case where $\phi_{p_{max}}(x) > 0$, a collision will happen at the position x and some order $p_k \leq p$, creating a gap with two boundaries, (x_l, x_h) and the possibility of further collisions will be controlled by ϕ at these positions till both boundary positions enter the immune zone and further gap creation stops. Investigating this process would be an interesting problem in prime combinatorics. Based on the results of Section 3.5 some important observations can be made about central gap distributions

in the collision-immune range $p^2/2$:

1. Runaway gap creation, similar to the edge distributions where the outward gap boundary expands faster than the order are not possible in central distributions at least up to certain distance from the center.
2. A constraint on the maximum gap size in the no-collision range $p^2/2$.

The validity of p.1 follows immediately from Lemma 4 as the last gap g_l in the interval that begins at position x_l below or at $p^2/2$ is bound to end at $x_m < (p_n^2 - 2)/2$ that as is the first possible collision position of the next prime harmonic $p_n \geq p + 2$ and therefore is protected from collisions with p_n and higher harmonics.

An estimate of the maximum gap size at $p^2/2$ can be obtained based on the results of the previous section.

At a prime order p with consecutive orders $p_p < p < p_n$, consider a central interval $d(p)$ with the upper boundary of the no-collision range at $(p^2 + 1)/2$, the same interval in the next prime order $d(p_n)$ and the maximum tp-candidates of orders p, p_n in $d(p), d(p_n)$: $x_m(p) \leq d(p), x_m(p_n) \leq d(p_n)$.

Lemma 4 provides that at least one candidate of D_p exists in the range $]d, d_n[$ and so, $x_m(p_n) > x_m(p), x_m(p_n) > d(p)$. Then, $d(p_n) - x_m(p_n) < \Delta(d, d_n)$, where $\Delta = d(p_n) - d(p)$.

This argument is valid for all orders, then:

$$\begin{aligned} d(p) - x_m(p) &< \Delta(d_p, d) \\ x_m(p_n) - d(p) &\leq \Delta(d, d_n) \\ x_m(p_n) - x_m(p) &\leq \Delta(d, d_n) + \Delta(d_p, d) = \Delta(p) \end{aligned} \quad (22)$$

Clearly, the interval at the boundary of the collision-immune range $[x_m(p), x_m(p_n)]$ contains all gaps of D_p in this range from which it can be concluded that all gaps in that interval must be limited by the condition (22).

Then, as discussed earlier in Section 3 one can recall that as the order p increases, the boundaries of the intervals $I_k(p)$ move upwards (6) eventually reaching the boundary of the non-collision region of some lower order. This observation allows to establish the constraint on the maximum gap size in the first central intervals that is of interest for the analysis of the twin prime distribution.

First we will prove a straightforward statement that a gap in one of the first intervals J_k corresponds to a gap in the boundary region of some lower order p_g , that is defined by the position of the gap boundary, with the constraint (22) on the size.

Corollary 4. For a gap in an interval $g \in J_k, \exists p_g : g \leq \Delta(p_g)$.

Proof. Let us consider a gap g in $J_k(p)$, with the lower boundary position $((2k + 1)p + 1)/2 \leq x_l(g) < ((2k + 3)p + 1)/2$.

For prime orders $p_l < p$ consider the sequence $d(p_l)$ of

boundaries of no-collision ranges $(p_l^2 + 1)/2$ and the maximum candidates of D_p in them, $x_m(p_l)$.

Then for the lower intervals it can be shown that there exist consecutive prime orders p_g, p_{g+1} such that:

$$\begin{aligned} d(p_g) &\leq x_l(g), \\ x_l(g) &< d(p_{g+1}) \end{aligned}$$

Then g satisfies the conditions of (22) at order p_g and must be limited by it. Furthermore, g is immune to collisions in at orders higher than p_g the size of g cannot change at higher orders and the condition of (22) must hold:

$$g \leq \Delta(p_g)$$

□

Now, based on the noted previously relation between the intervals at different prime orders we can now attempt to establish the constraint on the size of gaps in the first central intervals.

Lemma 5 (Maximum Gap Estimate). The maximum gap between adjacent tp-candidates in a central distribution at an order $p > P$ with the lower boundary $x_l(g)$ in $I_2(p)$, i.e. $(3p + 1)/2 \leq x_l(g) < (5p + 1)/2$ is less than or equal to, $p + 1$.

Proof. As just proven, for a gap g in $I_2(p)$ at an order p there exists an order p_g such that the condition of (22) holds, i.e., $g \leq \Delta(p_g)$. Then if $x_l(g), x_r(g)$ are the positions of the lower and upper boundary of g , under the conditions of the lemma the following must be true:

$$\begin{aligned} x_l &< (5p + 1)/2 \\ x_l &\geq (3p + 1)/2 \end{aligned}$$

Then, choosing p_g as in Cor.4 and denoting the boundary of no collision interval $b = (p^2 + 1)/2$:

$$\begin{aligned} x_l &\geq x_{m,g} > b_{g-1} \\ x_l &< x_{m,g+1} \leq b_{g+1} \end{aligned}$$

and combining the two,

$$\begin{aligned} p_{g-1}^2 &\leq 5p \\ p_{g+1}^2 &\geq 3p \end{aligned}$$

imposing upper and lower bounds on p_g, p_{g-1} relative to p .

Now one can use known bounds on the distance between consecutive primes, such as [2, 3] to express $\Delta(p_g)$ in terms of p_{g-1} .

$$\begin{aligned} p_g &\leq (1 + \alpha) p_{g-1} \\ p_{g+1} &\leq (1 + \alpha) p_g = (1 + \alpha)^2 p_{g-1} \\ \Delta(p_g) &= 1/2 (p_{g+1}^2 - p_{g-1}^2) \leq 2\alpha (1 + \beta) p_{g-1}^2 \end{aligned}$$

where $\alpha, \beta \ll 1$. Finally, from the upper estimate on p_{g-1}^2 earlier one can obtain the estimate of g in terms of the current order p :

$$\begin{aligned}\Delta(p_g) &\leq 2\alpha (1 + \beta) p_{g-1}^2 \\ p_{g-1}^2 &\leq 5p \\ g &\leq \Delta(p_g) \leq 10\alpha (1 + \beta) p\end{aligned}\quad (23)$$

and given that for sufficiently large P the bounds on α can be substantially stronger than $1/10$ [3, 4], the proof is complete. \square

4 Twin Prime Distribution

4.1 Distance Bound

In the analysis of twin prime distributions in this section we will use Lemmas 1, 2 and 5. For a twin pair $p = (p_1, p_2)$ we will denote the integer key $k(p) = (p_1 - 2)/3$ and consider the prime set G_k and twin prime candidate distribution D_k of order $k(p)$.

Assume first that $k(p)$ itself is a prime. Lemma 1 then ascertains that at least up to $3p_n \geq 3k + 2$, p_n being the next prime after k , tp-candidates identified by the harmonic conditions of C_k in that range will be true twin primes. The obtained bounds on gap distributions in the first central intervals then allow to estimate the maximum gap to the next twin prime pair:

Lemma 6 (Prime key). *For a twin prime pair with a prime key $k(p) > P$ there will be at least one tp-candidate c_n of order k satisfying the condition: $c_n > p$; $c_n - 3k \leq k + 1$.*

Proof. Given that $p_1 - 1 = 3k$ belongs to the second interval $J_2(k)$ of the prime order k : $p_1 - 1 = (k + 1)/2 + k$, Lemma 5 provides that the distance to the next tp-candidate $c_n(k) > p$ in D_k cannot be greater than $k + 1$: $c_n \leq (3k + 1)/2 + k + 1 = (5k + 3)/2$.

Unless c_n collides with one of the harmonics of order greater than k , it will produce the next twin prime pair after p : $p_n = c_n + 1$.

Now consider the next prime order after k , $k_n \geq k + 2$. The first collision of the prime harmonic h_{k_n} with D_k cannot happen at the second and third intervals of k_n by Lemma 2 and the next collision range of k_n will be in $2k_n - 2$ steps from $3k_n$: $t_n \geq (3k_n + 1)/2 + 2k_n - 2$. Then:

$$\begin{aligned}c_n &\leq (k + 1)/2 + k + k + 1 = (5k + 3)/2 \\ t_n &\geq (3k_n + 1)/2 + 2k_n - 2 \geq (7k + 7)/2 \\ &\Rightarrow c_n < t_n\end{aligned}\quad (24)$$

and it follows that the collision ranges at the next prime order cannot intersect with c_n . Clearly for the higher orders than k_n the first possible position can only be greater than t_n and no harmonic $n > k$ can produce a collision with c_n . Then with $n > c_n/3$ the completeness condition of Lemma 1 is satisfied and c_n must precede the next twin prime pair. \square

With this result one can approach the main statement of the twin prime distance theorem.

Theorem 1 (Twin prime distance). *For sufficiently large twin prime pair p with a key $k(p)$ there exists the next twin prime pair within the distance of $k + 1$ from $3k = p - 1$.*

Proof. The case of a prime key $k(p)$ has been proven earlier in Lemma 6 so the remaining case is that of a composite key.

For a twin prime pair p we will denote as k_l, k_n the nearest primes to $k(p)$: $k_l < k < k_n$. Then given the known bounds on the distance between consecutive primes, the position of the key $k(p)$ relative to its nearest lower prime k_l falls within the second interval I_2 of order k_l . Indeed:

$$(3k + 1)/2 - (3k_l + 1)/2 \leq 3/2 \alpha k_l < k_l$$

for sufficiently large k and the position of $3k$ will be in the second interval of the order k_l : $(3k + 1)/2 \in J_2(k_l)$ and the bounds of Lemma 6 on the maximum distance to the next tp-candidate apply. Then, as in the proof of Lemma 6 earlier, with c_n the next tp-candidate of the order k_l ,

$$\begin{aligned}t_n &\geq (7k + 7)/2 \\ c_n &\leq (3k + 1)/2 + k_l + 1 < (5k + 3)/2 \Rightarrow \\ c_n &< t_n\end{aligned}\quad (25)$$

where as before, $k_n \geq k + 2$.

Again, as in Lemma 6 the next tp-candidate c_n is immune from collisions with the higher harmonics and by Lemma 1 must be the key of the next twin prime pair, completing the proof of the theorem. \square

Numeric modeling with the first 100,000 twin prime pairs [9] confirmed the bound of the Theorem 1 on the distance to the next twin prime pair.

In conclusion we will reformulate the statement of the theorem in terms of integer (versus odd step) values:

For sufficiently large twin prime pair p with the key $k(p) = (p - 2)/3$ there exists the next twin prime pair within the distance $2(k + 1)$ from p .

4.2 Twin Prime Number

The results in (19),(20) may point to a direction toward an estimate of the number of twin primes in an arbitrary range, an analog of PNT for prime numbers [1]. Indeed, as follows from Lemma 1 for an arbitrary n , all tp-candidates of the order $p_n = n/3$ in the range up to n are twin primes. Then, from the result on the number of candidates (20) one readily obtains an estimate for the number of twin primes in the range $(3, n)$, i.e. *twin prime number*:

$$\begin{aligned}N_{tp}(p_n, n) &= N_{can}(p_n, n) = \frac{A n}{\log^2 p_n} + O\left(n \frac{e^{-c\sqrt{\log p_n}}}{\log^2 p_n}\right) \\ &\sim \frac{A n}{\log^2(n/3)}\end{aligned}\quad (26)$$

A very good agreement of the estimate of χ -candidates based on twin χ with the actual distribution in Section 3.5 may point to weaker influence of granularity effects due to suppression of collisions and overlaps at higher harmonics, for which there is a number of heuristic arguments. Providing robust bounds on the number of candidates in (20) with consideration of incomplete ranges may thus allow to advance to a formal proof of the twin prime number estimate of (26).

5 Conclusion

The approach based on understanding of the interaction and cumulative behavior of prime harmonics allowed to advance toward a formal proof of the statements of the maximum distance to the next twin prime pair (Theorem 1, *twin prime distance*) and the number of twin primes at an arbitrary range (26), *twin prime number*. With certain further technical effort in establishing robust bounds on the number of candidates taking into account granularity effects in (18)-(21), the proof of these results outlined in this work will be complete.

Prime harmonic analysis can be applied in investigation of prime distributions as well. As is easy to see, the condition on candidate distribution is less restrictive in the case of single prime candidates: the condition of the cumulative harmonic function for prime candidates is $C_p(x) > 1$ versus $C_p(x) > 2$ for twin primes, and the collision range of harmonic h_p has only one position, $h_p = 1$, resulting in higher frequency of primes and smaller gaps in gap distributions.

References

- [1] Ingham, A.E.: The distribution of prime numbers. Cambridge University Press (1990) 2–5
- [2] Nagura, J.: On the interval containing at least one prime number. Proceedings of the Japan Academy, Series A **28** (4) (1952) 177–181
- [3] Schoenfeld, L.: Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. Mathematics of Computation, **30** (134) (1976) 337 – 360
- [4] Dusart, P.: Estimates of some functions over primes without R.H. arXiv:1002.0442 [math.NT] (2010)
- [5] Tattersall, J.J.: Elementary number theory in nine chapters. Cambridge University Press (2005) 112
- [6] Zhang, Yitang: Bounded gaps between primes. Annals of Mathematics **179** (3) 1121–1174
- [7] Mertens, F.: Ein Beitrag zur analytischen Zahlentheorie. J. reine angew. Math. **78** (1874) 46–62
- [8] Naslund, E: Computing the product of $p/(p - 2)$ over the odd primes. URL (version: 2018-01-15): <https://math.stackexchange.com/q/22435> (2018)
- [9] PrimePages, UTM: The First 100,000 Twin Primes <https://primes.utm.edu/lists/small/100ktwins.txt>