Prime Sequences and Distributions

Serge Dolgikh

December 29, 2020

Prime Sequences and Distributions

Serge Dolgikh [0000-0001-5929-8954]

Dept. of Information Technology,
National Aviation University

Abstract. The method of prime modulo cycles or prime harmonics was applied in the analysis of several types of consecutive prime distributions such as single primes; quadruplet primes; and several cases of Polignac's conjecture, allowing to produce heuristic estimates for the number of prime sequences in an arbitrary range. As well, a complete classification of consecutive prime sequences is given.

Keywords: Number theory; prime number distribution; prime sequences; Polignac's conjecture

1   Introduction

Extensive work has been done in the area of prime number distribution, including Prime Number Theorem a number of strong distance and number bounds [1,2] and many others. In this work we investigate distributions of sequences of consecutive prime pairs with the approach based on prime modulo cycles, or prime harmonics analysis proposed in the earlier works to describe distribution of twin primes [3].

2   Prime Sequences

2.1   Definitions

We will use prime modulo cycles or harmonic functions of prime orders p, $h(p, x)$ and cumulative harmonic $C(p, x)$ on the set of positive odd integers as defined in [3].

Definition [Prime sequence]
A prime sequence P of integer type N with the key k is defined as:

$$P(k, N) = (p, \ p_{12}), k, (p_2, p_{22}), k_1, \dots (p_N, p_{N2}) \qquad (1)$$

where p_k: prime, $3 \mid k_i$, N: the length of the sequence.

Clearly, k and N uniquely define a prime sequence. The case of $N = 1$, i.e. twin prime sequence was described in [3] and $N = 2$ corresponds to quadruplet (quad) primes [4]. Prime sequences are a special case of prime k-tuples [5] and the scope of this work unless stated otherwise will be limited to sequences of consecutive primes.

2.2   Harmonic Condition

We will now establish the harmonic condition for the key of a prime sequence of length $N > 1$. Recalling that according to the definition of the prime harmonic function, $h(p, x) = p - 1 \equiv -1$, $h(p, x) = p - 2 \equiv -2$ it can be seen immediately that to allow a minimum of two consecutive pairs of primes, the condition on the key of the sequence is $| C(p, k) | > 2$.

Lemma 1. *The harmonic condition for the key k of a prime sequence with length N > 1 is | C(p, k) | > 2.*

Proof. For the position x of the key, $3 | x$, $x - 2$, $x - 1$, $x + 1$, $x + 2$ must be prime and the condition for $(x + 1, x + 2)$ being prime was established in [3]: $C(p, x) > 2$.

For $(x - 2, x - 1)$ to be prime, $C(p, x)$ cannot be $-2 \equiv p - 2$ or $-1 \equiv p - 2$ and the proof is complete.

An immediate corollary of Lemma 1 is that they key of a sequence with $N > 1$ has to satisfy $h_5(x) = 0$ or $5 | k$, being the only value of $h_5(x)$ that satisfies the harmonic condition of Lemma 1.

It follows then that the minimal distance between the positions of possible keys of a prime sequence of type $N > 1$ is 3×5.

2.3   Maximum Length

Another immediate consequence of the harmonic condition is that the length of a prime sequence cannot be greater than 2. Indeed, if existed a sequence with $N > 2$ with the key k, then the harmonic condition requires $h_5(k) = 0$. Then, $h_5(k_1 = k + 3) = 2$ and the harmonic condition for the following pair (x_3, x_{32}) is not satisfied so they cannot be both prime.

Lemma 2. *The maximum length of a consecutive prime sequence with the key k > 3 is 2.*

Hence, for sequences with $N > 1$ only $N = 2$, or quadruplet primes is possible. The only exception is the sequence *(5,7) 9 (11,13) 15 (17,19)* of length $N = 3$ and the key $k = 3 < 5$, for which the condition of h_5 does not apply.

3   Prime Distributions

The results of [1-3] and Section 2 allow to provide a complete classification of consecutive prime sequences:

1. *Regular single primes (RS)*: $p = k + 1$, $3 | k$. This case corresponds to $C(k / 3, k) = 2$ [d]
2. *Irregular single primes (IS)*: $p = k + 2$, $3 | k$. The harmonic condition for this case is $C(k / 3, k) = 1$, and $h_p(k) = 1$ or $h_p(k) > 2$. A special value can be added for this case in the definition of the cumulative harmonic function (Def.4, [3]) to associate this case with a unique value of $C(p, x)$.

3. *Twin primes (T)*: the harmonic condition is $C(k / 3, k) > 2$. This distribution was analyzed in [3].
4. *Quad primes (Q)*: as per Lemma 1, the harmonic condition for sequences of quad primes is $| C(k / 3, k) | > 2$.
5. *No-Prime (NP)*: pairs identified by a key $3 | k$ are not required to contain at least one prime so this case should also be considered for the classification to be complete. The harmonic condition for this case is $C(k / 3, k) = 1$ and $h_p(k) = 2, 4, 5, \ldots$ according to the lengths of the NP-sequence, as discussed in Section 4.

Immediately, it follows that $Q \subset T$.

We will now note that the cumulative harmonic approach defined in [3] for twin prime distributions can be readily applied to regular single prime (RS, IS) and quad prime (Q) distributions.

3.1 Regular Single Primes

Following the analysis in [3], the initial distribution of key positions is still D_3: $3 | x$ and the harmonic condition for the following position $x + 1$ being a prime candidate at a prime order p is defined by $C(p, x) > 1$. Hence, higher harmonics in this case have a single position collision range in each p-modulo cycle defined by $h_p(x) = 1$.

As in the analysis of twin prime distributions, a candidate position can be eliminated by a "collision" with a higher harmonic that erases it and creates a non-trivial prime gap. Then repeating derivations (15)-(18) [3] one can obtain an estimate for the number of prime candidate positions up to distance d as:

$$N_{can}(p, d) = \frac{d}{3} \left(1 - S(p)\right) \tag{2}$$

$$S(p) = S(p_{n-1}) + \frac{1}{p_{n-1}} \left(1 - S(p_{n-1})\right); \; 1 - S(p) = \prod_5^p (1 - \frac{1}{p_l}) \sim \frac{A}{\log p}$$

$$N_{can}(p, d) = \frac{1}{3} \frac{A \, d}{\log p} + O(\frac{1}{\log^2 p})$$

with Mertens theorem [6].

Finally, to obtain an estimate of the number of true primes in the range of n the order p has to be greater than $p_{\_m}$, the maximum prime that can produce collisions at n, $p_m \lesssim n \mathbin{/\mkern-5mu/} 3$:

$$N_{RS}(d) = \frac{1}{3} \frac{A \, n}{\log n/3} + O \tag{3}$$

The estimate above includes all pairs with the first member, $x + 1$, a prime i.e. regular single primes (RS) and twin primes (T).

Interestingly, the same result can be obtained for irregular single primes (IS). In this case the collision range still has a single position $h_p(k) = 2$; $h_p = 1$ does not result in a collision in this case as $k + 1$ is not required (but allowed) to be a prime. Then the

derivation for regular primes can be repeated. This result confirms the intuitive expectation that the number of RS and IS primes should be asymptotically equal:

$$\lim_{n \to \infty} N_{RS} = \lim_{n \to \infty} N_{IS} = \frac{1}{3} \frac{A\,n}{\log n/3} + O \tag{4}$$

Indeed, the number of RS vs IS primes is 5 / 4 for $n = 100$, 24 / 23, $n = 500$ and so on.

The total number of primes in the range n can then be obtained straightforwardly as the double of (3):

$$N_{PR}(n) = \frac{2}{3} \frac{A\,n}{\log n/3} + O(\frac{n}{\log^2 n}) \tag{5}$$

in agreement with **PNT** [1].

Note that as commented in [3], the estimate assumes the existence of rigorous bounds that account for the granularity effects at incomplete cycles of multiple intersections of prime harmonics.

3.2   Quadruplet Prime Distribution

In the case of quad primes, the initial distribution has the period of 3×5 (Corollary, Section 2.2) and collision ranges of higher harmonics, four positions: $h_p(x) = $ -2; -1; 1; 2.

Again, following the interval analysis [3], (15)-(18) one obtains:

$$N_{can}(p, d) = \frac{d}{15} (1 - 4\,S(p)) \tag{6}$$

$$S(p) = S(p_{n-1}) + \frac{1}{p_{n-1}} \left(1 - 4\,S(p_{n-1})\right); \quad 1 - 4\,S(p) = \prod_{7}^{p} (1 - \frac{4}{p_l})$$

$$N_Q(n) = \frac{1}{15} \frac{B\,n}{\log^4 n/3} + O(\frac{1}{\log^5 n})$$

where B can be calculated from (5), the *quad prime constant* (as discussed in the next section).

Quad χ Function Asymptotics
Quad χ function is defined from (5) as:

$$\chi(p_n) = S(p_{n+1})$$
$$S(p = 7) = \frac{1}{7}$$
$$S(p) = S(p_{n-1}) + \frac{1}{p_{n-1}} \left(1 - 4\,S(p_{n-1})\right)$$
$$1 - 4\,S(p) = \prod_{p_l = 7}^{p_i < p} (1 - \frac{4}{p_l})$$

The product can be calculated based on a similar derivation of the product of $(1 - \frac{2}{p})$ [7]. Indeed, consider the equality:

$$\left(1 - \frac{2}{p}\right)^2 \times \left(1 - \frac{4}{(p-2)^2}\right) = 1 - \frac{4}{p}$$

Then,

$$\prod_7^n \left(1 - \frac{2}{p}\right)^2 \times \prod_7^n \left(1 - \frac{4}{(p-2)^2}\right) = \prod_7^n \left(1 - \frac{4}{p}\right)$$
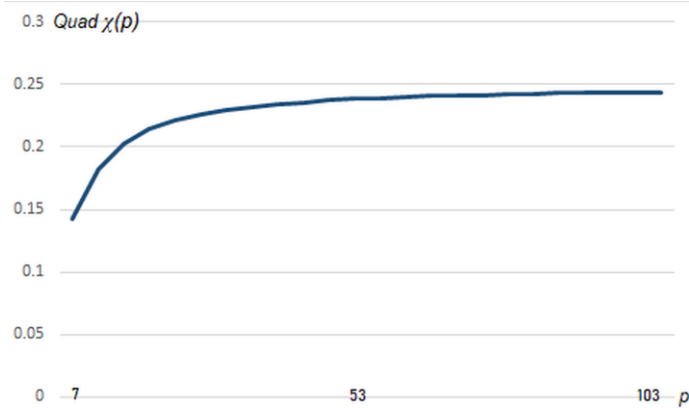
The first term on the left was calculated [7] as: $\frac{C\,\Pi_2}{\log^2 n} + O(\frac{1}{\log^3 n})$ and one obtains for the sought product:

$$S(n) = \prod_7^n \left(1 - \frac{4}{p}\right) = \frac{C\,\Pi_2^2\,\Pi_4}{\log^4 n} + O(\frac{1}{\log^5 n})$$

where *quad prime constant:*

$$\Pi_4 = \lim_{n \to \infty} \prod_7^n \left(1 - \frac{4}{(p-2)^2}\right)$$

The plot of the Quad χ function in the first prime orders is given in Figure 1.

Fig. 1. Quad χ in the first prime orders

Again, a robust accounting of the granularity effects in incomplete cycles of multiple intersections is needed to complete the formal proof of this result. It is hoped that once such bounds are obtained it would be possible to prove infiniteness of quad primes based only on the number bound (5), by showing that for a given distance n always exists n_{next} such that $N_Q(n) < N_Q(n_{next})$. That would allow to bypass a more complex analysis of gap distributions as for twin distribution in [3].

4 No-Prime (NP) Sequences

In accordance with the classification of prime sequences, Section 3, a no-prime (NP) sequence can be defined as:

$$P_{NP,L} = (\dots) \, k \, (x,x) \, k_1(x,x) \, (\dots)$$

In analogy with consecutive prime sequences (Section 2.1), no-prime sequences can be defined for any length L and well-known examples show that unlike prime case, there is no limit for possible length of NP-sequence.

An estimate for the number of NP-sequences of lengths $N = 1$ (i.e., a single pair of non-primes) in a range n can be obtained from the single prime distributions, Section 3.1. One can observe that a key preceding such configuration cannot be either of regular single prime (RS) or irregular one (IS), Then:

$$N_0(n) = \frac{n}{3} - N_{RS}(n) - N_{IS}(n) + N_T(n) \sim \frac{n}{3} \left(1 - \frac{2A}{\log \frac{n}{3}} + \frac{2D}{\log^2 \frac{n}{3}} \right) + O \qquad (7)$$

and the number of NP pairs in a range n grows as $\sim n \left(1 - \frac{A}{\log n} \right)$.

4.1 NP Distributions

NP-sequences and distributions with $N \geq 1$ will be discussed in more detail in another work. However, general approach to understanding distributions of NP-sequences based on prime modulo cycles is quite straightforward. The necessary and sufficient condition for a sequence of K consecutive non-prime pairs (x, y) and the key k is $h(p, k) \in \{(1, 2), (4, 5), (7, 8), \dots\}$ for at least $2K$ prime harmonics p. Each combination of harmonic values satisfying this condition repeats with a period of $3 \, \Pi \, p_k$ leading to the estimate of the number of NP-sequences of length K in a given interval d:

$$N_{NPK}(d) = \frac{d \, 2K!}{3} \; (C_{2K}(p_m) - N_{over}(p)) \qquad (8)$$

where C_{2K} is the *cyclic prime sum* of order $2K$, p_m the maximum prime that can be a factor in the range d, $p_m \sim d / 3$ and N_{over}, the count of overlaps between combinations of prime orders at certain positions.

A cyclic prime sum of order N is defined as:

$$C_N(P) = \sum_{P_l}^{p_l < P, p_i \neq p_j} \frac{1}{p_1 p_2 .. p_K} \qquad (9)$$

so, $C_1(P) = \sum^{p_l < P} \frac{1}{p_l}$, and will be discussed in more detail elsewhere.

5 Cases in Polignac's Conjecture

Whereas the general case of the conjecture [8] with an arbitrary even N may not be immediately straightforward to approach with harmonic analysis due to large number of possible configurations, certain advances can be made in some specific cases where the number of configurations is limited.

The case of $N = 2$ constitutes the twin prime conjecture that was discussed in [3].

5.1 N = 4 Conjecture

As before, we consider two pairs of odd positions separated by a key $k \mid 3$. As can be concluded immediately, for $N = 4$ there is only one possible configuration of positions with a gap of 4: an IS prime in the first pair and an RS one in the following one:

$$P_4 = (x, p_1)\, k\, (p_2, x)$$

The harmonic conditions for the initial distribution can then be easily established as $\mid h_5(k) \mid\, \neq 1$, i.e. $h_5(k) = (0; 2; -2 \equiv 3)$. This produces initial distribution D_{P4} with 3 strands corresponding to allowed values of h_5 and a period of 3×5.

The collision of positions in the initial distribution with a higher harmonic occurs if the higher harmonic cancels one of the prime candidate positions at k i.e. $h_p(k) = -1$ or 1, resulting in a two-position collision range, as in the case of twin prime distribution. Then the collision estimation as in [3] and previous sections can be repeated, producing the estimate on the number of P_4-candidate positions in a range d as:

$$N_{P4\_can}(p, d) = \frac{3}{15} \frac{c\, d}{\log^2 p} + O \tag{10}$$

$$N_{P4}(n) = \frac{1}{5} \frac{c\, n}{\log^2 n/3} + O$$

again, allowing to obtain an estimate of the number of prime positions in an arbitrary range n.

5.2 N = 6

In the case of $N = 6$ there are two possible configurations with a gap of 6:

$$P_{61} = (p_1, x)\, k\, (p_2, x),\ P_{62} = (x, p_1)\, k(x, p_2)$$

Similar to the previous case of $N = 4$ the conditions for the initial distribution and collisions can be written for these cases as:

P_{61}: $h_5(k) = (0, 2, 4)$; $h_p(k)$ collision range: (1, 3)

P_{62}: $h_5(k) = (0, 1, 3)$; $h_p(k)$ collision range: (2, 4)

Each of these configurations produces three strand initial distributions D_{P61}, D_{P62} with a period of 3×5. The estimate of P_4, Section 4.1 can be used then, with the following result for the number of prime configurations with the gap of 6 in a range n:

$$N_{P6}(n) = 2\, N_{P4}(n) = \frac{2}{5} \frac{c\, n}{\log^2 n/3} + O \tag{11}$$

5.3 N = 8

For $N = 8$, there is a single possible configuration P_8 with a gap of 8 that is symmetric to P_4 in the case $N = 4$:

$$P_8 = (p_1, x)\, k\, (x, p_2)$$

Similar to the case of N = 4 the conditions for the initial distribution and collisions can be written as:

$$P_8: h_5(k) = (0, \pm 1); \; h_p(k) \text{ collision range: } (\pm 2)$$

As in P_4, this configuration produces a three-strand initial distribution D_8, with a period of 3×5 and the estimate of P_4 for the number of P_8 key positions within a range n should apply:

$$N_{P8}(n) = N_{P4}(n) = \frac{1}{5} \frac{C\,n}{\log^2 n/3} + O \tag{12}$$

5.4   Polignac's Conjecture with N > 8

$N = 8$ is the last gap configuration that is possible between two consecutive pairs. As can be seen immediately, any configuration with a gap that is greater than 8 must contain an NP-sequence:

$$P_{N>8} = (P)\, k\, (NP)\, k_n\, (P)$$

where P is one of the configurations with primes. The configurations with a gap $N > 8$ can then be described, similarly to the cases $N = 4, 6, 8$ discussed earlier:

$$P_{6k+4} = (x, p_1)\, k\, (NP)\, (p_2, x) \tag{13}$$

$$P_{6k+6} = (p_1, x)\, k\, (NP)\, (p_2, x); \; (x, p_1)\, k\, (NP)\, (x, p_2)$$

$$P_{6k+8} = (p_1, x)\, k\, (NP)\, k_n(x, p_2)$$

These configurations can then be described as combinations of one of the four boundary configurations in (13) with an NP-sequence of an appropriate length. The harmonic condition on $C(k/3, k)$ for $N > 8$ configurations would be a combination of the boundary pair conditions discussed in Sections 5.1 – 5.3 with more complex constraints on specific harmonics of NP-sequence. These cases will be discussed in another study.

6   Conclusion

It was shown that prime harmonics analysis can be instrumental in understanding and estimating the distributions of consecutive prime sequences. With the classification of prime sequences, Section 3, the distribution of prime sequences can be described as the combination of distributions of single primes (RS, IS); twin primes (T); quadruplet primes (Q) and no-prime (NP) sequences described in Sections 3, 4 and previous works [1-7].

Harmonic analysis was as well instrumental in approaching a number of specific cases of Polignac's conjecture (Section 5).

It can be concluded that prime harmonics analysis offers a simple and consistent framework to approach distributions of prime sequences and related problems in this area of research and with attention and development, could produce further interesting results.

References

1. Ingham, A.E. The distribution of prime numbers. Cambridge University Press (1990).
2. Schoenfeld, L. Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. Mathematics of Computation, 30 (134) 337 – 360 (1976).
3. Dolgikh, S. Prime Harmonics and Twin Prime Distribution. Cambridge Open Engage (preprint), doi:10.33774/coe-2020-c9x75-v2 (2020).
4. The top twenty: Quadruplet. PrimePages, UTM, https://primes.utm.edu/top20/page.php?id=55
5. Forbes, T. Prime clusters and Cunningham chains. Math. Comput. 68, 1739-1748, (1999).
6. Mertens, F. Ein Beitrag zur analytischen Zahlentheorie. J. reine angew. Math. 78 (1874).
7. 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).
8. Tattersall, J.J. Elementary number theory in nine chapters, Cambridge University Press, 112 (2005).