Proof of Polignac's conjecture for gap equal to six
(Proof that exist infinitely many sexy primes)

Marko V. Jankovic

Institute of Electrical Engineering “Nikola Tesla”, Belgrade, Serbia,

Abstract In this paper proof of the Polignac's Conjecture for gap equal to six is going to be presented. Consecutive primes with gap six are known as sexy primes. The proof represents an extension of the proof of the twin prime conjecture. It will be shown that sexy primes could be obtained through two stage sieve process, and that will be used to prove that infinitely many sexy primes exist.

1 Introduction

In number theory, Polignac's conjecture states: For any positive even number *g*, there are infinitely many prime gaps of size *g*. In other words: there are infinitely many cases of two consecutive prime numbers with the difference *g* [1].

In [2] it has been shown that exists infinitely many consecutive prime numbers that have gaps that are not bigger than 246. Recently, the Polignac's conjecture was proved for gaps of the size 2 and 4 [3]. The problem was addressed in generative space, which means that prime numbers were not analyzed directly, but rather their representatives that can be used to produce them. This paper represents an extension of the previous work [3]. Here, the gap equal to six is going to be analyzed. It will be shown that exist an infinite number of primes with gaps of the size 6 (also known as sexy primes). It will be shown that sexy primes could be generated by two stage recursion type sieve process. This process will be compared to other two stage recursion sieve process that leaves infinitely many numbers. Fact that sieve process that generate sexy primes leaves more numbers than the other sieve process will be used to prove that infinitely many sexy primes exist. Using very similar line of reasoning, it can be shown that Polignac's conjecture for gaps equal to 8 holds, too.

Remark 1: *In this paper any infinite series in the form* $c_1 \cdot l \pm c_2$ *is going to be called a thread defined by number* c_1 *(in literature these forms are known as linear factors – however, it seems that the term thread is probably better choice in this context). Here* c_1 *and* c_2 *are numbers that belong to the set of natural numbers (*c_2 *can also be zero and usually is smaller than* c_1*) and* l *represents an infinite series of consecutive natural numbers in the form (1, 2, 3, ...).*

2 Proof of the sexy prime conjecture

It is well known that all prime numbers can be expressed in one of the following form

$$ps_k = 6k - 1$$

$$pl_k = 6k + 1,\ k \in N.$$

As it was already explained, we will call numbers ps_k - numbers in *mps* form and numbers pl_k - numbers in *mpl* form.

If two consecutive prime numbers have the gap of the size 6 it is clear that both of those numbers have to be in *mps* form or both have to be in *mpl* form. In this paper we will only analyze the case of sexy primes in *mps* form. The case of sexy primes in *mpl* form can be analyzed analogously.

Here, we are going to present a two stage process that can be used for generation of the sexy primes in *mps* form. In the first stage we are going to produce prime numbers by removing all composite numbers from the set of natural numbers. In the second stage, we are going to remove all twin primes (since the prime numbers in *mps* form cannot have sexy pair from obvious reason), all prime numbers in *mpl* form that are left, and all prime numbers in *mps* form that have an bigger sexy odd neighbor (odd number that has gap 6 with the prime of interest) that is a composite number. At the end, only the prime numbers in the *mps* form, that represent the smaller number of a sexy prime pair, are going to stay. Their number is approximately the half of the number of sexy primes in *mps* form, or one quarter of all sexy primes (although a little bit bigger due to existence of sexy triplets and quadruplets – in that case only the last number in the series is going to be removed from the set; here, it is considered that number of triplets and quadruplets sexy primes is very small comparing to the number of sexy primes). It is going to be shown that that number is infinite. Second half of the sexy primes can be generated with similar procedure that takes care about sexy primes in *mpl* form.

STAGE 1

Prime numbers can be obtained in the following way:

First, we remove all even numbers (except 2) from the set of natural numbers. Then, it is necessary to remove the composite odd numbers from the rest of the numbers. In order to do that, the formula for the composite odd numbers is going to be analyzed. It is well known that odd numbers bigger than 1, here denoted by a, can be represented by the following formula

$$a = 2n + 1,$$

where $n \in N$. It is not difficult to prove that all composite odd numbers a_c can be represented by the following formula

$$a_c = 2(2ij + i + j) + 1 = 2((2j+1)i + j) + 1. \quad (1)$$

where $i, j \in N$. It is simple to conclude that all composite numbers could be represented by product $(i + 1)(j + 1)$, where $i, j \in N$. If it is checked how that formula looks like for the odd numbers, after simple calculation, equation (1) is obtained. This calculation is presented here. The form $2m + 1$, $m \in N$ will represent odd numbers that are composite. Then the following equation holds

$$2m + 1 = (i_1 + 1)(j_1 + 1) \quad ,$$

where $i_1, j_1 \in N$. Now, it is easy to see that the following equation holds

$$m = \frac{i_1 j_1 + i_1 + j_1}{2}.$$

In order to have $m \in N$, it is easy to check that i_1 and j_1 have to be in the forms

$$i_1 = 2i \text{ and } j_1 = 2j,$$

where $i, j \in N$. From that, it follows that m must be in the form

$$m = 2ij + i + j = (2i + 1)\,j + i. \quad (2)$$

When all numbers represented by m are removed from the set of odd natural numbers bigger than 1, only the numbers that represent odd prime numbers are going to stay. In other words, only odd numbers that cannot be represented by (1) will stay. This process is equivalent to the sieve of Sundaram [4].

Let us denote the numbers used for the generation of odd prime numbers with $m2$ (here we ignore number 2). Those are the numbers that are left after the implementation of Sundaram sieve. The number of those numbers that are smaller than some natural number n, is equivalent to the number of prime numbers smaller than n. We denote with $\pi(n)$ number of primes smaller than n, than the following equation holds

$$\pi(n) \approx \frac{n}{\ln(n)}.$$

From [6] we know that following holds

$$\pi(n) > \frac{n}{\ln(n)}, \quad n \geqslant 17. \quad (3)$$

STAGE 2

What was left after the first stage are prime numbers. With the exception of number 2, all other prime numbers are odd numbers. All odd primes can be expressed in the form $2n + 1$, $n \in N$. It is simple to understand that their bigger sexy odd neighbor must be in the form $2n + 7$, $n \in N$. Now, we should implement a second stage in which we are going to **remove**:

A. Number 2 (since 2 cannot make a sexy pair);

B. All twin primes – so number of numbers that is going to be left is number of primes minus the number of twin primes (number 2 is ignored, and that has no impact on the analysis that follows). It is not difficult to prove that number of numbers left, is infinite. That will be done later in the text.

C. The rest of *mpl* primes – it is trivial to see that it can be done by one thread that is defined by 3 – so in this step it is going to be removed, approximately (having in mind that the number of *mps* primes is a bit bigger than the number of *mpl* primes), one half of the numbers that are left after step *B;*

D. All odd primes in the form $2m + 1$ such that $2m + 7$, $m \in N$ represents a composite number (all primes whose bigger "sexy" neighbor is composite number). If we make the same analysis, like in the Stage 1, it is simple to understand that m must be in the form

$$m = 2ij + i + j - 1 = (2i + 1)\, j + i \text{ -}3. \tag{4}$$

All numbers (in observational space) that are going to stay must be numbers in *mps* form and they represent a smaller primes of the sexy prime pairs in *mps* form. What has to be noticed is that thread in (4) that is defined by prime number 3 (for $i = 1$) is not going to remove any number from the numbers left, since it would remove same numbers as the thread defined by 3 used in Stage 1.

Let us mark the number of sexy primes with π_{G6}. Also, let us define the number of numbers that is left after two consecutive implementations of Sundaram sieve as *pd6*. The numbers obtained after recursive implementation of two Sundaram sieves (where the second Sundaram sieve is implemented on prime numbers from which number 2 and all twin primes are removed) are going to be called sexy double primes. The second stage sieve that is identical to the first stage sieve can be obtained if the prime numbers left after the first stage and removal off 2 and all twin primes, are lined up next to each other and then the numbers are removed from the exactly same positions like in the first stage. In that case it is easy to understand that the following equation would holds ($n \in N$)

$$pd6\,(n) \approx \frac{\pi(n) - \pi_{G2}(n)}{\ln\left(\pi(n) - \pi_{G2}(n)\right)} \quad , \tag{5}$$

where $pd6(n)$ represents the number of sexy double primes smaller than some natural number n. Since the *mps* sexy primes are obtained by implementation of the Sundaram sieve in the first stage and sieve that is similar to Sundaram sieve in the second stage, it can be intuitively concluded that numbers $\pi_{G6}/4$ (which is half of the sexy primes in *mps* form) and *pd6* should be comparable. However, in the case of generation of *mps* sexy primes the second stage sieve defined by (4) is not equivalent to the the first stage sieve since the second stage "Sundaram" sieve (defined by (4)) is applied on an incomplete set, that is depleted by previously implemented Sundaram sieve. Here, it will be shown that number of sexy double primes *pd6* is smaller than the number $\pi_{G6}/4$. In order to understand why it is so, we are going to analyze (2) and (4) in more detail.

It is not difficult to be seen that *m* in (2) and (4) is represented by the threads that are defined by odd prime numbers. For details see Appendix A. Now we are going to compare stages 1 and 2 step by step, for a few initial steps (analysis can be easily extended to any number of steps). Starting point for the second stage is after removal of number 2 and all twin primes

Table 1 Comparison of the stages 1 and 2 – threads defined by a few smallest primes

Step	Stage 1	Step	Stage 2
1	Remove even numbers (except 2) amount of numbers left is 1/2	1	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left is 1/2
2	Remove numbers defined by thread defined by 3 (obtained for $i = 1$) amount of numbers left is 2/3 of the numbers that are left after previous step	2	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left is 3/4 of the numbers that are left after previous step
3	Remove numbers defined by thread defined by 5 (obtained for $i = 2$) amount of numbers left is 4/5 of the numbers that are left after previous step	3	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left is 5/6 of the numbers that are left after previous step
4	Remove numbers defined by thread defined by 7 (obtained for $i = 3$) amount of numbers left is 6/7 of the numbers that are left after previous step	4	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left is 9/10 of the numbers that are left after previous step
5	Remove numbers defined by thread defined by 11 (obtained for $i = 5$) amount of numbers left is 10/11 of the numbers that are left after previous step	5	Remove numbers defined by thread defined by 13 (obtained for $i = 6$) amount of numbers left is 11/12 of the numbers that are left after previous step

From the table, it can be noticed that threads defined by the same number in first and second stage will not remove the same percentage of numbers. The reason is obvious – consider for instance the thread defined by 3: in the first stage it will remove 1/3 of the numbers left, but in the second stage it will remove ½ of the numbers left, since the thread defined by 3 in stage 1 has already removed one third of the numbers (odd numbers divisible by 3 in observation space). So, only odd numbers (in observational space) that give residual 1 and -1 when they are divided by 3 are left, and there is approximately same number of numbers that give residual -1 and numbers that give residual 1, when the number is divided by 3 (see Appendix A). Same way of reasoning can be applied for all other threads defined by the same prime in different stages.

From Table 1 can be seen that in every step, except step 1, threads in the second stage will leave bigger percentage of numbers than the corresponding threads in the first stage. This could be easily understood from the analysis that follows:

- suppose that we have two natural numbers j, k such that $j - 1 \geq k$ ($j, k \in N$), then the following set of equations is trivially true

$$j+k-1 \geqslant 2k$$

$$-j-k+1 \leqslant -2k$$

$$jk-j-k+1 \leqslant jk-2k$$

$$(j-1)(k-1) \leqslant (j-2)k$$

$$\frac{k-1}{k} \leqslant \frac{j-2}{j-1}$$

The equality sign holds only in the case $j = k + 1$. In the set of prime numbers there is only one case when $j = k + 1$ and that is in the case of primes of 2 and 3. In all other cases $p(i) - p(i - 1) > 1$, ($i > 1$, $i \in N$, $p(i)$ is i-th prime number). So, in all cases $i > 2$

$$\frac{p(i-1)-1}{p(i-1)} < \frac{p(i)-2}{p(i)-1} \quad .$$

From Table 1 (or last equation) we can see that bigger number of numbers is left in every step of stage 2 then in the stage 1 (except 1st step). From that, we can conclude that after every step bigger than 1, part of the numbers that is left in stage 2 is bigger than number of numbers left in the stage 1 (that is also noticeable if we consider amount of numbers left after removal of all numbers generated by threads that are defined by all prime numbers smaller than some natural number). From previous analysis we can safely conclude that the following equation holds

$$\frac{\pi_{G6}}{4} > pd6 = \lim_{n \to \infty} pd6(n) \quad .$$

From previous inequality it can be concluded that the following equation must hold

$$\pi_{G6} > pd6 = \lim_{n \to \infty} pd6(n) \quad .$$

Having in mind (3) and (5), we can say that for some n big enough the following inequality holds

$$pd6(n) > \frac{\pi(n) - \pi_{G2}(n)}{\ln\left(\pi(n) - \pi_{G2}(n)\right)} . \tag{6}$$

It can be realized that n that is big enough is $n \geq 211$, since 211 is the 17th prime left, when number 2 and all twin prime numbers are eliminated from the prime numbers set. It is not difficult to understand that the number of primes that are left when 2 and all twin primes are eliminated, is infinite. That follows in an elementary way from the fact that the sum of the reciprocal of prime numbers is infinite and Brun's theorem [7]. From the fact that sum of reciprocals of all primes is infinite and fact that the sum of reciprocal of twin primes is finite [7] (also reciprocal of 2 is 1/2), follows that the sum of reciprocals of primes that are not twin primes or 2, is infinite. Since all reciprocal of primes are smaller than 1, it is easy to conclude that the number of primes that are not twin primes and 2, has to be infinite. Having that in mind it it easy to show that following holds

$$\lim_{n \to \infty} \frac{\pi(n) - \pi_{G2}(n)}{\ln\left(\pi(n) - \pi_{G2}(n)\right)} = \infty .$$

Then, the following equation holds

$$pd6 = \lim_{n \to \infty} pd6(n) = \infty .$$

Now, we can safely conclude that the number of sexy primes is infinite. That concludes the proof.

Here we will state the following conjecture: (for *n* big enough, and under assumption that number of sexy *k*-plets (triplets and quadruplets; there is only one quintuplet – quintuplet that starts with 5; other quintuplets are not possible since one of five consecutive numbers that are six apart must be divisible by 5, so it is not a prime number – number 5 is the only exception; from the same reason, *k*-plets, where $k > 5$, do not exist) is much smaller than the number of sexy primes (this can be proved by using a procedure similar to one used in this paper)), number of sexy primes is given by the following equation

$$\pi_{G6}(n) \sim 4C_2 \cdot \left(\frac{(\pi(n) - \pi_{G2}(n))}{\ln(\pi(n) - \pi_{G2}(n))} + \frac{(\pi(n) - \pi_{G4}(n))}{\ln(\pi(n) - \pi_{G4}(n))} \right),$$

where C_2 represents the twin prime constant [5], and $\pi_{G4}(n)$ is number of cousin primes smaller than *n* and comes from the analysis of sexy primes in *mpl* form. Why it is reasonable to make such conjecture is explained in Appendix B. Since $\pi_{G2}(n) \approx \pi_{G4}(n)$ we have

$$\pi_{G6}(n) \sim 8C_2 \cdot \left(\frac{(\pi(n) - \pi_{G2}(n))}{\ln(\pi(n) - \pi_{G2}(n))} \right),$$

If we mark the number of primes smaller than some natural number *n* with $\pi(n) = f(n)$, where function $f(n)$ gives good estimation of the number of primes smaller than *n,* than $\pi_{G6}(n)$, for *n* big enough, is given by the following equation (under assumption $\pi_{G2}(n) \approx \pi_{G4}(n)$)

$$\pi_{G6}(n) \sim 8C_2 \cdot \left(f\left(f(n) - 4C_2 f(f(n)) \right) \right).$$

References

[1] A. de Polignac. (1849). Recherches nouvelles sur les nombres premiers. Comptes Rendus des Séances de l'Académie des Sciences.

[2] D.H.J. Polymath. (2014) Variants of the Selberg sieve and bounded intervals containing many primes. Research in the Mathematical Sciences. 1(12), arxiv:1407.4897.

[3] M. V. Jankovic. (2021). A Proof of the Twin Prime Conjecture (together with the Proof of Polignac's Conjecture for Cousin Primes), hal-02549967, version 12.

[4] V. Ramaswami Aiyar. (1934) Sundaram's Sieve for prime numbers, The Mathematics Student, 2(2) , 73.

[5] G.H. Hardy, J.E. Littlewood. (1923) Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes, Acta Math. 44, pp.1-70.

[6] J.B. Rosser, L. Schoenfeld. (1962) Approximate formulas for some functions of prime numbers. Illinois Journal of Mathematics, 6(1), pp. 64-94.

[7] V. Brun. (1919) La serie 1/5+1/7+1/11+1/13+1/17+/19+1/29+1/31+1/41+1/43+1/59+1/61+ … , où les dénominateurs sont numberes premiers jumeaux est convergente ou finie. Bulletin des Sciences Mathématiques, 43, pp 100-104, 124-128.

APPENDIX A.

Here it is going to be proved that m in (2) is represented by threads defined by odd prime numbers. Now, the form of (2) for some values of i will be checked.

Case i = 1: $m = 3j + 1$,

Case i = 2: $m = 5j + 2$,

Case i = 3: $m = 7j + 3$,

<u>Case i = 4:</u> $m = 9j + 4 = 3(3j + 1) + 1$,

<u>Case i = 5:</u> $m = 11j + 5$,

<u>Case i = 6:</u> $m = 13j + 6$,

<u>Case i = 7:</u> $m = 15y + 7 = 5(3j + 1) + 2$,

<u>Case i = 8:</u> $m = 17j + 8$,

It can be seen that m is represented by the threads that are defined by odd prime numbers. From examples (cases $i = 4$, $i = 7$), it can be seen that if $(2i + 1)$ represent a composite number, m that is represented by thread defined by that number also has a representation by the the thread defined by one of the prime factors of that composite number. That can be proved easily in the general case, by direct calculation, using representations similar to (2). Here, that is going to be analyzed. Assume that $2i + 1$ is a composite number, the following holds

$$2i + 1 = (2l + 1)(2s + 1)$$

where ($l, s \in N$). That leads to

$$i = 2ls + l + s.$$

The simple calculation leads to

$$m = (2l + 1)(2s + 1)j + 2ls + l + s = (2l + 1)(2s+1)j + s(2l + 1) + l$$

or

$$m = (2l+1)((2s+1)j + s) + l$$

which means

$$m = (2l + 1)f + l,$$

and that represents the already exiting form of the representation of m for the factor $(2l + 1)$, where

$$f = (2s + 1)j + s.$$

In the same way this can be proved for (4), (5) and (7).

<u>Note</u>: It is not difficult to understand that after implementation of stage 1, the number of numbers in residual classes of some specific prime number are equal. In other words, after implementation of stage 1, for example, all numbers divisible by 3 (except 3, but it does not affect the analysis) are removed. However, the number of numbers in the forms 3k + 1 and 3k + 2 (alternatively, 3k – 1) are equal. The reason is that the thread defined by any other prime number (bigger than 2) will remove the same number of numbers from the numbers in the form 3k + 1 and from the numbers in the form 3k + 2. It is simple to understand that, for instance, thread defined by number 5, is going to remove 1/5 of the numbers in form 3k + 1 and 1/5 of the numbers in form 3k + 2. This can be proved by elementary calculation. That will hold for all other primes and for all other residual classes.

APPENDIX B.

Here asymptotic density of numbers left, after implementation of the first and second Sundaram sieve is calculated.

After first k steps of the first Sundaram sieve, after removal of all composite even numbers, density of numbers left is given by the following equation

$$c_k = \frac{1}{2} \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j)}\right),$$

where $p(j)$ is j-th prime number.

In the case of second "Sundaram" sieve the density of numbers left after the first k-steps is given by the following equation

$$c2_k = \prod_{j=2}^{k+1} \left(1 - \frac{1}{p(j) - 1}\right) = \prod_{j=2}^{k+1} \left(\frac{p(j) - 2}{p(j) - 1}\right).$$

So, if implementation of first sieve will result in the number of prime numbers smaller than n which we denote as $\pi(n)$, than implementation of the second sieve on some set of size $\pi(n)-\pi_{G2}(n)$ should result in the number of numbers $gp(n)$ that are defined by the following equation (for some big enough n)

$$gp(n) = r_{S2S1}(n) \cdot \frac{\pi(n) - \pi_{G2}(n)}{\ln\left(\pi(n) - \pi_{G2}(n)\right)} \quad ,$$

where $r_{S2S1}(n)$ is defined by the following equation (k is the number of primes smaller or equal to $n1$ = $sqrt(n)$, where $sqrt$ marks square root function)

$$r_{S2S1}(n) = \frac{c2_k}{c_k} = \frac{\prod_{p>2,\, p\leq n1}\left(\frac{p-2}{p-1}\right)}{\prod_{p\leq n1}\left(\frac{p-1}{p}\right)} = 2\prod_{p>2,\, p\leq n1}\left(\frac{p-2}{p-1}\right)\left(\frac{p}{p-1}\right) \approx 2C_2 .$$

For n that is not big, $gp(n)$ should be defined as

$$gp(n) = f_{COR}(n) \cdot 2C_2 \cdot \frac{\pi(n) - \pi_{G2}(n)}{\ln\left(\pi(n) - \pi_{G2}(n)\right)},$$

where $f_{COR}(n)$ represents correction factor that asymptotically tends toward 1 when n tends to infinity.