

On a Smarandache problem concerning the prime gaps

Felice Russo

Via A. Infante 7

67051 Avezzano (Aq) Italy

felice.russo@katamail.com

Abstract

In this paper, a problem posed in [1] by Smarandache concerning the prime gaps is analysed.

Let's p_n be the n -th prime number and d_n the following ratio:

$$d_n = \frac{p_{n+1} - p_n}{2} \quad \text{where } n \geq 1$$

If we indicate with $g_n = p_{n+1} - p_n$ the gap between two consecutive primes, the previous equation becomes:

$$d_n = \frac{g_n}{2}$$

In [1], Smarandache posed the following questions:

1. Does the sequence d_n contain infinite primes?
2. Analyse the distribution of d_n

First of all let's observe that d_n is a rational number only for $n=1$, being $p_1=2$, $p_2=3$. For $n>1$, instead, the ratio is always a natural number since the gap of prime numbers g_n is an even number ≥ 2 [2].

Moreover let's observe that the gap g_n can be as large as we want. In fact let's n be any integer greater than one and let's consider the following sequence of consecutive integers:

$$n!+2, n!+3, n!+4, \dots, n!+n$$

Notice that 2 divides the first, 3 divides the second, ..., n divides the n -1st, showing all of these numbers are composite. So if p is the largest prime smaller than $n!+2$ we have $g_n > n$. This proves our assertion.

Now let's check the first terms of sequence d_n :

n	1	2	3	4	5	6	7	8	9	10
d	0.5	1	1	2	1	2	1	2	3	1
p_n	2	3	5	7	11	13	17	19	23	29

Here p_n is the smallest prime relative to the gap d_n . As we can see, for the first 10 terms of sequence d_n we have 4 primes regardless if those are repeated or not. On the contrary, if we consider only how many distinct primes we have then this number is 2. So, the Smarandache question can be split in two sub-questions:

1. How many times the sequence d_n takes a prime value?
2. How many distinct primes the sequence d_n contains?

Proving both the questions is a very difficult task. Anyway, we can try to understand the behaviour of sequence d_n by using a computer search and then get a heuristic argument on the number of primes within it.

Thanks to an Ubasic code, the counting functions $p_1(N)$ and $p_2(N)$ have been calculated for N up to 10^9 .

$p_1(N)$ denotes how many times d_n takes a prime value for $n \leq N$ while $p_2(N)$ denotes the number of distinct primes in d_n , always for $n \leq N$. In table 1, the results of the computer search can be found.

In the third column, the number of distinct primes are reported whereas in the second one the number of all primes regardless of the repetitions are shown.

N	# primes	# distinct primes
10	0	0
100	14	2
1000	107	4
10000	695	7
100000	4927	11
1000000	37484	14
10000000	241286	19
100000000	2413153	24
1000000000	66593597	33

Table 1. Number of primes in d_n for different N values

Let's analyse the data of column 2. It is very easy to verify that those data grow linearly with N , that is:

$$p_1(N) \approx c(N) \cdot N \quad (1)$$

An estimation of $c(N)$ can be obtained using the following asymptotic relationship given in [3]:

$$h_N(d_n) \approx \frac{c_2 \cdot N}{\ln^2(N)} \cdot \prod_{p|2d_n, p>2} \frac{p-1}{p-2} \cdot e^{-\frac{2d_n}{\ln(N)}}$$

where $h_N(d_n)/N$ is the frequency of d_n for $n \leq N$ and p any prime number.

The constant c_2 is the twin prime constant defined in the following way:

$$c_2 \equiv 2 \cdot \prod_{p>2} \left(1 - \frac{1}{(p-1)^2} \right) = 1.320032 \dots$$

By definition of $p_1(N)$ function we have:

$$p_1(N) = \sum_{d_n=2}^{d_{\max}} \frac{c_2 \cdot N}{\ln^2(N)} \cdot \prod_{p|2d_n, p>2} \frac{p-1}{p-2} \cdot e^{-\frac{2 \cdot d_n}{\ln(N)}} \quad (2)$$

where the above summation is extended on all prime values of d_n up to d_{\max} . But the largest gap d_{\max} for a given N can be approximated by [2],[3]:

$$d_{\max} \approx \frac{1}{2} \cdot \ln^2(N)$$

and then (2) can be rewritten as:

$$p_1(N) \approx \frac{2 \cdot N}{\ln^2(N)} \sum_{d_n=2}^{\frac{1}{2} \ln^2(N)} e^{-\frac{2 \cdot d_n}{\ln(N)}} \quad (3)$$

where the function :

$$J(d_n) = \prod_{p|2d_n, p>2} \frac{p-1}{p-2}$$

has small values of order 1 and then has been replaced by its mean value $\frac{2}{c_2}$ [3].

Since, as N goes to infinity, the summation:

$$\sum_{d_n=2}^{\frac{1}{2} \ln^2(N)} e^{-\frac{2 \cdot d_n}{\ln(N)}}$$

is the number of primes in the range 1 to $\frac{1}{2} \cdot \ln^2(N)$, we can write:

$$\sum_{d_n=2}^{\frac{1}{2} \ln^2(N)} e^{-\frac{2 \cdot d_n}{\ln(N)}} \leq \pi\left(\frac{1}{2} \cdot \ln^2(N)\right)$$

where $\pi(N)$ is the counting function of prime numbers [2]. Using the Gauss approximation [2] for it, we have:

$$\sum_{d_n=2}^{\frac{1}{2} \ln^2(N)} e^{-\frac{2 \cdot d_n}{\ln(N)}} \leq \frac{\frac{1}{2} \cdot \ln^2(N)}{\ln\left(\frac{1}{2} \cdot \ln^2(N)\right)}$$

and then:

$$p_1(N) \approx c(N) \cdot N \leq \frac{N}{\ln\left(\frac{1}{2} \cdot \ln^2(N)\right)}$$

by using (1) and (3), that implies:

$$c(N) \leq \frac{1}{\ln\left(\frac{1}{2} \cdot \ln^2(N)\right)}$$

According to those experimental data the following conjecture can be posed:

Conjecture A: The sequence d_n takes infinite times a prime value.

Let's now analyse the data reported in table 1, column 3. By using the least square method, we can clearly see that the best fit is obtained using a logarithmic function like:

$$p_2(N) \approx c(N) \cdot \ln(N) \quad (4)$$

where $c(N)$ can be estimated using the following approximation:

$$p_2(N) \approx \pi(0.5 \cdot \ln^2(N))$$

being $p_2(N)$ the number of primes in the range 1 to d_{\max} .
Therefore:

$$\frac{\ln^2(N)}{2 \cdot \ln(0.5 \cdot \ln^2(N))} \approx c(N) \cdot \ln(N)$$

$$\Rightarrow c(N) \approx \frac{\ln(N)}{2 \cdot \ln(0.5 \cdot \ln^2(N))}$$

In table 2, the comparison of (4) with calculated values $p_2(N)$ shown in table 1 (column 3) is reported. Notice the good agreement between $p_2(N)$ and its estimation as N increase.

According to those data, also $p_2(N)$ like $p_1(N)$ goes to the infinity as N increase, although $p_2(N)$ more slowly than $p_1(N)$. Then this second conjecture can be posed:

Conjecture B: The sequence d_n contains an infinite number of distinct primes

$$\frac{\ln^2(N)}{2 \cdot \ln(0.5 \cdot \ln^2(N))} \approx c(N) \cdot \ln(N)$$

$$\Rightarrow c(N) \approx \frac{\ln(N)}{2 \cdot \ln(0.5 \cdot \ln^2(N))}$$

In table 2, the comparison of (4) with calculated values $p_2(N)$ shown in table 1 (column 3) is reported. Notice the good agreement between $p_2(N)$ and its estimation as N increase. According to those data, also $p_2(N)$ like $p_1(N)$ goes to the infinity as N increase, although $p_2(N)$ more slowly then $p_1(N)$. Then this second conjecture can be posed:

Conjecture B: The sequence d_n contains an infinite number of distinct primes

N	$P_2(n)$	$c(N) \cdot \ln(N)$	ratio
10	0	2.719152	0
100	2	4.490828	0.445352
1000	4	7.521271	0.531825
10000	7	11.31824	0.618471
100000	11	15.80281	0.696079
1000000	14	20.93572	0.668713
10000000	19	26.69067	0.711859
100000000	24	33.04778	0.726221
1000000000	33	39.9911	0.825184

Table2. Comparison of $p_2(N)$ with the approximated formula $c(N) \cdot \ln(N)$. In the third column the ratio $p_2(N) / c(N) \cdot \ln(N)$

Let's analyse now the distribution of d_n , as always requested by Smarandache. Thanks to a Ubasic code the frequency of prime gaps up to $N=3601806621$ have been calculated. The plot of those frequencies versus d_n for $n > 1$ is reported in Fig1. It shows a clear jigsaw pattern superimposed onto an exponential decay. The jigsaw pattern is due to a double population that is clearly visible in the two plots of fig 2. The frequency of d_n for n being a multiple of 3 (or equivalently for n multiple of 6 for g_n) is always larger than adjacent differences.

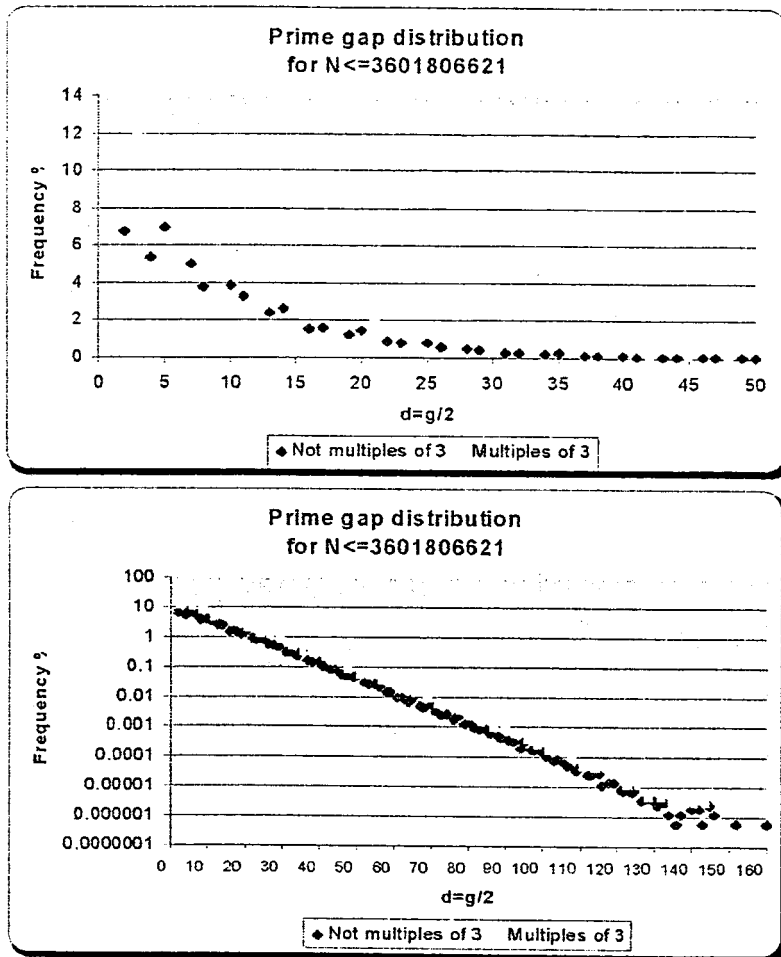


Fig 2. Prime gap distribution. The second plot uses a logarithmic scale for the Y-axis.

According to the conjecture 1 reported in [3] and already mentioned above , the number of pairs $p_n, p_{n+1} < N$ with $d_n = \frac{p_{n+1} - p_n}{2}$ is given by:

$$h_N(d_n) \approx \frac{c_2 \cdot N}{\ln^2(N)} \cdot \prod_{p|2d_n, p>2} \frac{p-1}{p-2} \cdot e^{-\frac{2d_n}{\ln(N)}}$$

Let's $f(p) = \frac{p-1}{p-2}$ where p is any prime number greater than 2. As it can be seen in fig 3. this function approaches 1 quickly, with the maximum value at $p=3$.

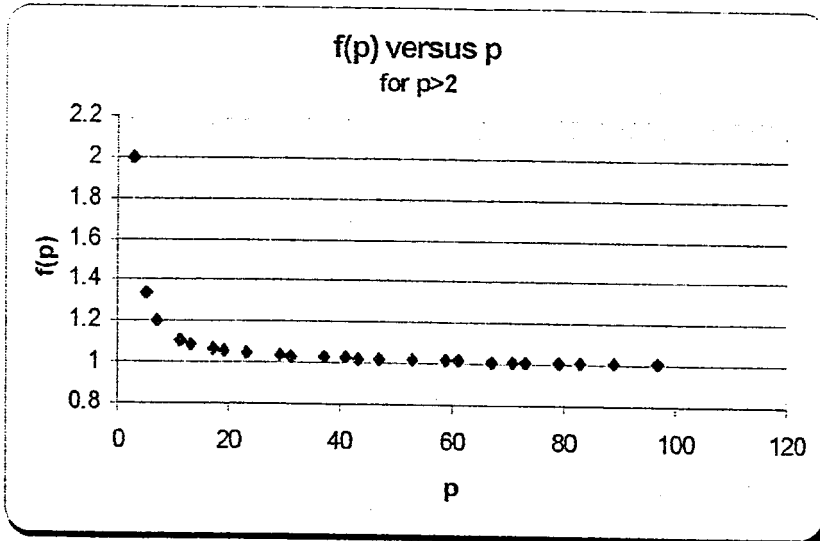


Fig.3: Plot of function $f(p)$ versus p

Being $f(p)$ maximum for $p=3$ means that $h_N(d_n)$ has a relative maximum every time $2d_n$ has 3 as prime factor, that is when $2d_n$ is a multiple of 3.

This explains the double population seen in the Fig 2 and then the jigsaw pattern of the fig 1.

In fig. 4, the distribution of d_n obtained by computer search and the one estimated with the use of $h_N(d_n)$ formula is reported. Notice the very good agreement between them.

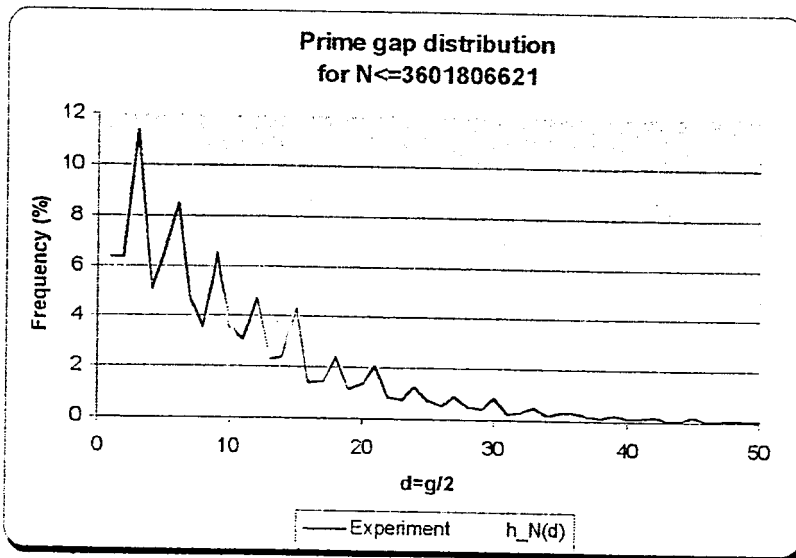


Fig 4: Prime gap distribution comparison. The good agreement between the experimental and the estimated data has to be noticed.

References.

- [1] F. Smarandache, *Only problems not solutions*, Xiquan publication, Phoenix, 1993.
- [2] E. Weisstein, *CRC Concise Encyclopedia of Mathematics*, CRC Press, 1999.
- [3] M. Wolf, *Some conjectures on the gaps between consecutive primes*, preprint at <http://www.ift.uni.wroc.pl/~mwolf/>