Unexpected Regularities in the Distribution of Prime Numbers

Marek Wolf

Institute of Theoretical Physics, University of Wrocław
Pl.Maxa Borna 9, PL-50-204 Wrocław, Poland, e-mail: mwolf@ift.uni.wroc.pl

Abstract

The computer results of the investigation of the number $h_N(d)$ of gaps of the length d between two consecutive primes smaller than N are presented. The computer search was done up to $N = 2^{44} \approx 1.76 \times 10^{13}$. A few attempts to fit the obtained data by an analytical formula are given. As the applications two formulae are obtained: for the largest gap between consecutive primes below a given bound and the formula for "champions" — the most often occuring pairs of primes. Obtained data supports the conjecture that the number of Twins and primes separated by a gap of the length 4 ("Cousins") is almost the same and it determines a fractal structure on the set of primes.

1 Introduction

Recently there appeared a few papers where some mathematical facts about primes were applied to the study of quantum chaos [1]. For example, the Hardy–Littlewood conjecture [2] about the gaps between (not necessary consecutive) primes inspired some works on the correlations between periodic orbits, see [3]. In this paper we are going to look for the statistical properties of the distribution of gaps between *consecutive* primes.

The problem of the appearance of gaps between consecutive primes has a long history, see e.g. [5], [6], [7]. The main quantity studied here is defined as follows:

$$h_N(d) = \text{number of pairs } \; p_n, p_{n+1} < N \text{ with } d = p_{n+1} - p_n. \qquad (1)$$

I have made the computer search for prime numbers up to $N = 2^{44} \approx 1.76 \times 10^{13}$ and counted the number of gaps between consecutive primes [4]. The computer program was implemented on the DEC Alpha 300X/175MHz workstation in DECFortran which allows some assembly level instructions and it took 132 days of CPU time to reach 2^{44}. The obtained data allow to make conjectures on the form of the dependence of $h_N(d)$. The formulae presented in this paper are heuristic, but nevertheless are in perfect agreement with conjectures of other authors [2], [5], [6] and/or are very well confirmed by numerical calculations. There are no theorems nor proofs — only the computer results and the interpretation of them given by a physicist.

2 Exponential decreasing of $h_N(d)$ and oscillations

During the computer search the data representing the function $h_N(d)$ were stored at values of N forming the geometrical progression with the ratio 4, i.e. at $N = 2^{20}, 2^{22}, \ldots, 2^{42}, 2^{44}$. The resulting curves are plotted in the Fig.1.

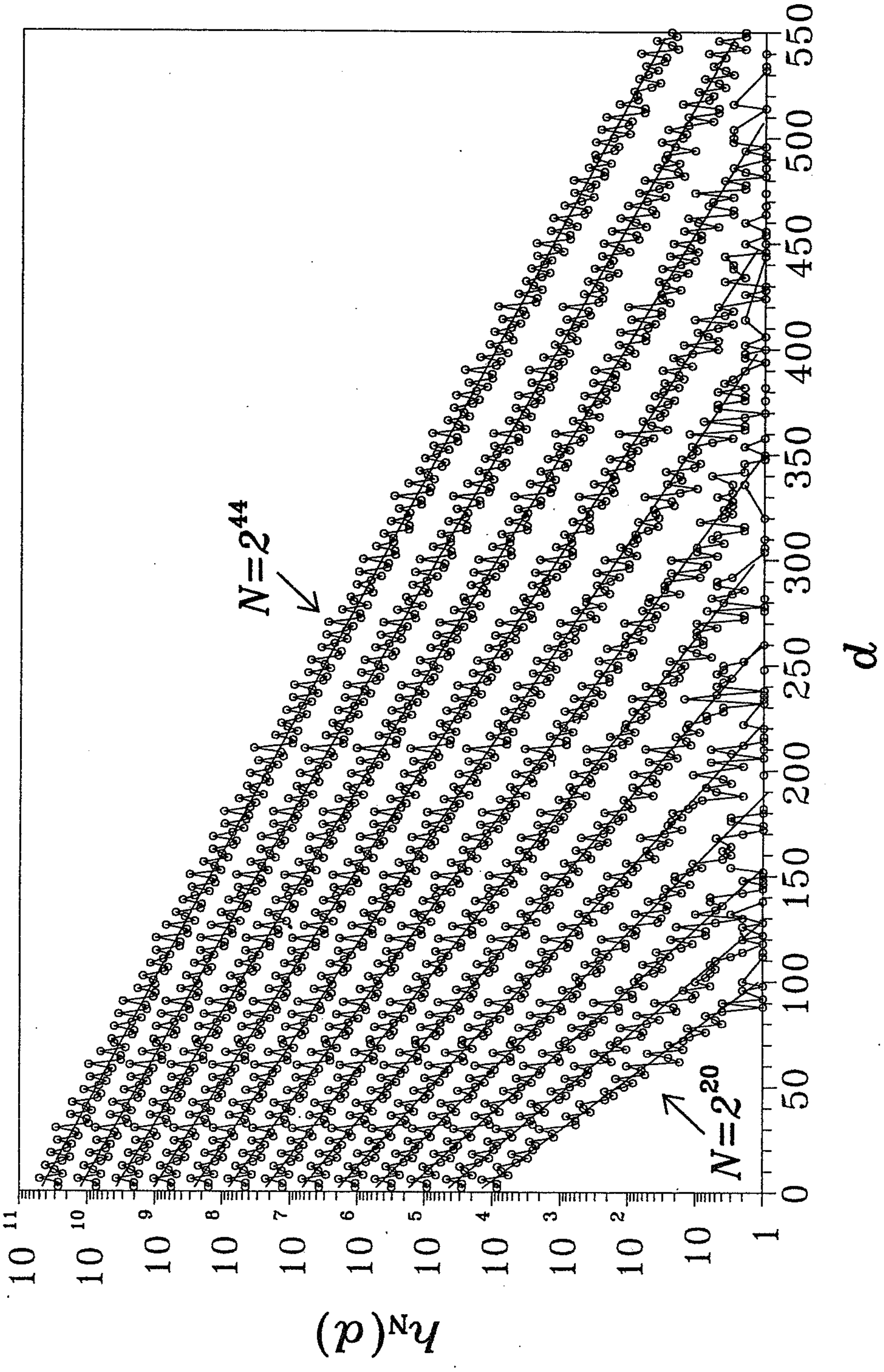

Figure 1: The plot showing the dependence of the histogram $h_N(d)$ on d at $N = 2^{20}, 2^{22}, \ldots, 2^{44}$. There is a logarithmical scale on the y-axis , while on the x-axis there is a linear scale. The values of $h_N(d)$ obtained from the computer search are represented by small circles. The straight lines are the best fits obtained by means of the least-square method. The points oscillate around the straight lines with period 6. Let us mention "local" spikes at $d = 30(= 2 \cdot 3 \cdot 5), 60, \ldots$. Especially well profound are spikes at $d = 210 = 2 \cdot 3 \cdot 5 \cdot 7$, and for $N = 2^{40}$, $N = 2^{42}$ and $N = 2^{44}$ also at its multiplicity $d = 420$ (second harmonic). More insights into the structure of the distribuion of circles can be gained when looking at sliding angles.

Because the points lie around the straight lines on the semi-logarithmic scale, we can infer from the Fig.1 the conclusion that $h_N(d)$ decreases exponentially with d:

$$h_N(d) \sim B_1(N)e^{-A_1(N)d}, \tag{2}$$

To determine the slope $A_1(N)$ and the intercept $B_1(N)$ of this let me remark, that there are two selfconsistency conditions that the functions $A_1(N)$ and $B_1(N)$ have to obey. First of all, the number of all gaps is by 1 smaller than the number of primes: $\sum_d h_N(d) = \pi(N) - 1$, where $\pi(N)$ denotes the number of primes smaller than N. The second selfonsistency condition comes from the observation, that the sum of distances between consecutive primes $p_n \leq N$ is equal to the largest prime $\leq N$. So for large N we can write: $\sum_d h_N(d)d \approx N$. For the Gauss estimation $\pi(N) \sim N/\ln(N)$ from these identities it follows, that the functions $A_1(N)$ and $B_1(N)$ have the form:

$$A_1(N) \sim \frac{1}{\ln(N)}, \qquad B_1(N) \sim \frac{2N}{\ln^2(N)}. \tag{3}$$

However the points in the Fig.1 very clearly oscillate around the lines being the best least-square fits. In fact, the period of this oscillation is exactly 6. I claim that these oscillations are described by the following modification of the formula (2):

$$h_N(d) \sim B_2(N) \prod_{p|d,p>2} \frac{p-1}{p-2} e^{-A_2(N)d}, \tag{4}$$

where $p \mid d$ denotes such primes p which divide d. The product appearing here in front of the exponent I have borrowed from the n-tuple conjecture of Hardy and Littlewood [2]. All regularities seen in the Fig.1 and oscillations of different periods are caused by the product:

$$\prod_{p|d,p>2} \frac{p-1}{p-2}. \tag{5}$$

The points obtained from formula (4) reproduce quite well data obtained from the computer search, see [4]. In the physical language the formula (2) can be regarded as a zero order (in the periods of oscillations) approximation to the formula (4). The "first order" approximation, taking into consideration only the period 6, is discussed in [4]. The problem is that for (4) I am not able to write down in the closed form the selfonsistency conditions. But the product (5) increases very slowly; e.g. for the $N < 2^{44}$ it takes values between 1 and 3.2 (3.2 appears only twice: for $d = 210$ and $d = 420$), so I suggest that the functions $A_2(N)$ and $B_2(N)$ are also given by eqs.(3).

3 Two Applications

As the application of the formula (2) we can obtain the length $G(N)$ of the largest gap between consecutive primes below a given bound N. Simply, the largest gap G appears only once, so it is equal to the value at which $h_N(d)$ crosses the d-axis on the Fig.1: $h_N(G(N)) = 1$; it gives:

$$G(N) \sim \ln(N)(\ln(N) - 2\ln\ln(N) + c), \qquad c = 1/2. \tag{6}$$

The above formula agrees well with the actual computer data, see [4] and for large N passes into the Cramer [5] conjecture (see also [6]):

$$G(N) \sim \ln^2(N). \tag{7}$$

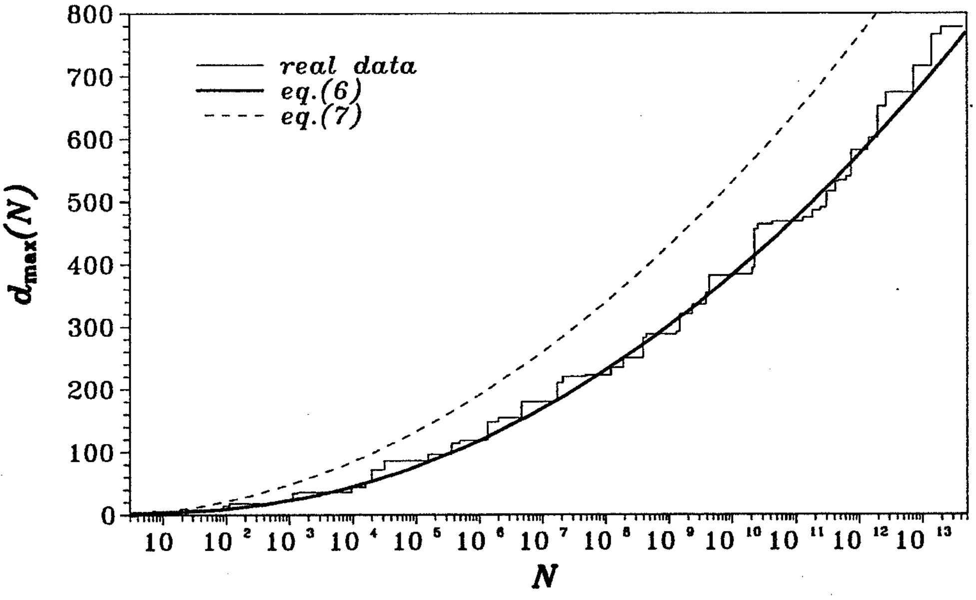

Figure 2: Plot of $G(N)$ for N up to 5×10^{13}. The results from the computer search are drawn by thin solid line, eq.(6) is bold line and the conjecture of Cramer (7) is shown by dashed line.

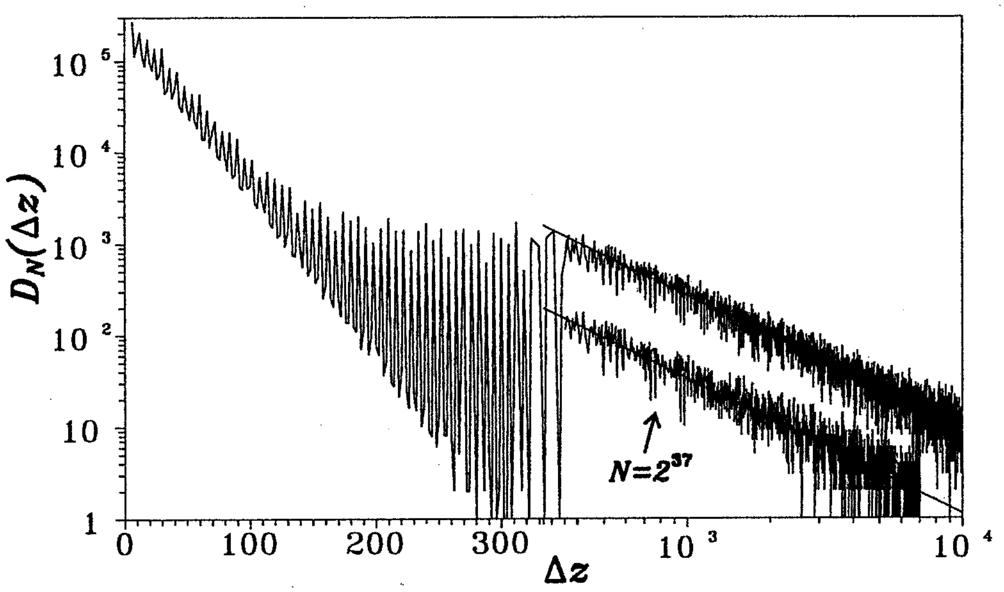

Figure 3: The plot showing the dependence of the distribution $\mathcal{D}_N(\Delta z)$ of spacings Δz between consecutive $p^{(z)}$. There is a logarithmic scale on the y-axis , while on the x-axis there is a linear scale up to $\Delta z = 360$ while for larger Δz the scale is logarithmic. There is also the power-like part of $\mathcal{D}_N(\Delta z)$ for $N = 2^{37}$ plotted and for this case the slope is $\gamma \approx 1.47$

The comparison of the eqs.(6) and (7) with real data is shown in the Fig.2.

As an application of the formula (4) let us consider the most often occurring distances between primes, called "high-jumpers" or "champions". The straightforward checking shows, that up to 389 the largest number of gaps is $d = 2$, while the most often one becomes the gap $d = 6$ — the circle corresponding to $d = 6$ is the highest one on the plots for each N in the Fig.1. Odlyzko [8] has shown by some heuristic arguments as well as by numerical search that around $N \sim 10^{35}$ the gaps of the length $d = 6$ are overtaken by pairs of primes separated by $d = 30$. He also claims that at approximately $N \sim 10^{450}$ the most often occuring gap becomes $d = 210$. In the Fig.1 the local spikes appear at multiplicities of $30 = 2 \cdot 3 \cdot 5$. For larger values of N even higher spikes at $d = 210 = 2 \cdot 3 \cdot 5 \cdot 7$ can be seen. As N increases the slopes $A(N)$ decrease and at some value $N^{(2)}$ the peak at $d = 30$ will be greater than that at $d = 6$. At much larger $N^{(3)}$ the spike at $d = 210$ will take over $d = 30$. From this observation and eq.(4), where I assume that $A_2(N) = 1/\ln(N)$ (the dependence on $B_2(N)$ drops out), I can obtain the general estimation for the values of $N^{(n)}$ at which the consecutive products $D^{(n)} = 2 \cdot 3 \cdot \ldots \cdot p_n$ become the champions:

$$\ln(N^{(n)}) = \frac{2 \cdot 3 \cdot \ldots \cdot p_{n-1}(p_n - 1)}{\ln((p_n - 1)/(p_n - 2))}. \tag{8}$$

The values of $N^{(n)}$ obtained from (8) are shown in the Table 1. Obtained from (8) for $D^{(1)} = 6$ the value of $N^{(1)} = 321$ quite well agrees with the actual value 389.

$D^{(n)}$	$N^{(n)}$
6	3.21×10^{2}
30	1.70×10^{36}
210	5.81×10^{428}
2310	1.48×10^{8656}
30030	1.30×10^{138357}
510510	8.02×10^{3233259}
9699690	8.50×10^{69820169}
223092870	$5.14\times10^{1992163572}$
6469693230	$3.56\times10^{74595540317}$
200560490130	$2.10\times10^{2486392448589}$
$\vdots$	$\vdots$

Table 1.

N	$h_N(2)$	$h_N(4)$	$h_N(2)/h_N(4)$
2^{20}	8535	8500	1.0041176
2^{22}	27995	27764	1.0083201
2^{24}	92246	91995	1.0027284
2^{26}	309561	309293	1.0008665
2^{28}	1056281	1057146	0.9991818
2^{30}	3650557	3650515	1.0000115
2^{32}	12739574	12740283	0.9999443
2^{34}	44849427	44842399	1.0001567
2^{36}	159082253	159089620	0.9999537
2^{38}	568237005	568225073	1.0000210
2^{40}	2042054332	2042077653	0.9999886
2^{42}	7378928530	7378989766	0.9999917
2^{44}	26795709320	26795628686	1.0000030

Table 2.

4 Twins and Cousins

There is a lot of regularities seen in the Fig.1, however the most striking one is the fact that on each plot (i.e. for every N) the first two circles, corresponding to $d = 2$ and $d = 4$, are lying on the same attitude, see also Table 2. It means, that numbers of pairs of primes separated by the distance $d = 2$ and $d = 4$ are almost the same:

$$h_N(2) \approx h_N(4). \tag{9}$$

By analogy with Twins a pair of primes separated by a gap $d = 4$ I will name "Cousins". As it is seen from the Table 2 the ratio $h_N(2)/h_N(4)$ is sometimes larger than 1, and for other N is smaller than 1. It means that there is a set of such N that the numbers of Twins and Cousins are equal. Because $h_N(d)$ can change value only for N being prime, it is reasonably to look for such primes $p^{(z)}$ that the difference $h_{p^{(z)}}(2) - h_{p^{(z)}}(4)$ is zero. I have checked [9] by direct computer search that up to $N = 2^{43}$ there are 2823290 primes $p^{(z)}$, at which the numbers of Twins and Cousins are exactly the same. The distribution $\mathcal{D}_N(\Delta z)$ of spacings Δz between consecutive $p^{(z)} < N$ is shown in the Fig.3. There are 2790362 spacings shown on this figure — only 32928 distances betwenn consecutive $p^{(z)}$ were larger than 10^4. In fact these larger distances have very scattered values (the largest gap between two clusters of the same number of Twins and Cousins was 314267840234) and they appeared only once — it resembles the intermittent behaviour in some dynamical systems. There is a cross-over from the exponential dependence of $\mathcal{D}_N(\Delta z)$ to the power-like decrease. It means that the set of $p^{(z)}$ is formed by "clusters" separated by distances obeying the power law, what in turn is characteristic for fractal sets. Because inside the clusters $h_{p^{(z)}}(2)$ and $h_{p^{(z)}}(4)$ do not change their values, spacings Δz between $p^{(z)}$ inside clusters follow the usual rule (4) and it is the explanation for the left part of the plot in the Fig.3. The exponent γ in the power–like part $\mathcal{D}_N(\Delta z) \sim (\Delta z)^{-\gamma}$ has the value of $\gamma \approx 1.46$ and it *does not* depend on N. The possible relevance of this intermittent and fractal behaviour of $p^{(z)}$ for quantum chaos is now under study.

Acknowledgments. I thank prof. A.M. Odlyzko for e-mail feedback and Prof. W.Narkiewicz for discussions. The author receives support from the KBN research grant No 2 P302 057 07.

References

[1] R.Aurich et al, "Universal Signatures of Quantum Chaos", *Phys.Rev.Lett.* **73** (1994), p.1356; "Quantum Chaos", edts. G.Casati et al (North Holland, Amsterdam, 1993)

[2] G.H.Hardy and J.E. Littlewood, "Some problems of 'Partitio Numerorum' III", *Acta Mathematica* **44** (1922), p.1-70

[3] E.Bogomolny and P.Leboeuf, "Statistical Properties of the Zeros of Zeta Functions — Beyond Riemann case", *Nonlinearity* **7** (1994), p.1155-67

[4] M.Wolf, preprint IFTUWr 894/95, available via anonymous ftp from orchid.ift.uni.wroc.pl (156.17.88.118), directory pub/primes.

[5] H.Cramer, "On the order of magnitude of difference between consecutive prime numbers", *Acta Arith.* **2** (1937), p.23-46

[6] D.Shanks, "On Maximal Gaps between Succesive Primes", *Math.Comp.* **18** (1964), p.464

[7] H.Maier, C.Pomerance, "Unusually large gaps between primes", *Trans. Amer. Math.Soc.* **322** (1990), p.201-237

[8] A.M. Odlyzko, unpublished — private communication

[9] M.Wolf, to be published; the file containing zeros of the difference $h_N(2) - h_N(4)$ for N being prime numbers up to 2^{43} is over 37MB long.