Some heuristics
on the gaps between consecutive primes

Marek Wolf

e-mail: primes7@o2.pl

Abstract

We propose the formula for the number of pairs of consecutive primes $p_n, p_{n+1} < x$ separated by gap $d = p_{n+1} - p_n$ expressed directly by the number of all primes $< x$, i.e. by $\pi(x)$. As the application of this formula we formulate 7 conjectures, among others for the maximal gap between two consecutive primes smaller than x, for the generalized Brun's constants and the first occurrence of a given gap d. Also the leading term $\log \log(x)$ in the prime harmonic sum is reproduced from our guesses correctly. These conjectures are supported by the computer data.

Key words: *Prime numbers, gaps between primes, Hardy and Littlewood conjecture*
PACS numbers: 11A99, 11A41

> *"A subject that has attracted attention*
> *but concerning which the known results*
> *leave much to be desired,*
> *is that of the bahaviour of $p_{n+1} - p_n$,*
> *where p_n denotes the n-th prime."*
> H. Davenport, in [13, p.173]

1   Introduction.

In 1922 G. H. Hardy and J.E. Littlewood in the famous paper [18] have proposed 15 conjectures. The conjecture B of their paper states:

There are infinitely many primes pairs (p, p'), where $p' = p + d$, for every even d. If $\pi_d(x)$ denotes the number of prime pairs differing by d and less than x, then

$$\pi_d(x) \sim C_2 \prod_{p|d} \frac{p-1}{p-2} \frac{x}{\log^2(x)}. \tag{1}$$

Here the product is over odd divisors $p \geq 3$ of d and we use the notation $f(x) \sim g(x)$ in the usual sense: $\lim_{x \to \infty} f(x)/g(x) = 1$. The twin constant $C_2 \equiv 2c_2$ is defined by the infinite product:

$$C_2 \equiv 2c_2 \equiv 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2} \right) = 1.32032363169\ldots \tag{2}$$

Computer results of the search for pairs of primes separated by a distance $d \leq 512$ and smaller than x for $x = 2^{32}, 2^{34}, \ldots, 2^{44} \approx 1.76 \times 10^{13}$ are shown in Fig.1 and they provide a firm support in favor of (1). Characteristic oscillating pattern of points is caused by the product

$$\mathfrak{S}(d) = \prod_{p|d, p>2} \frac{p-1}{p-2} \tag{3}$$

appearing in (1). The main period of oscillations is $6 = 2 \times 3$ with overimposed higher harmonics $30 = 2 \times 3 \times 5$ and $210 = 2 \times 3 \times 5 \times 7$, i.e. when $\mathfrak{S}(d)$ has local maxima (local minima are 1 and they correspond to $d = 2^m$). The red lines present $\pi_d(x)/\mathfrak{S}(d)$ and they are perfect straight lines $C_2 x / \log^2(x)$.

There is large evidence both analytical and experimental in favor of (1). Besides the original circle method used by Hardy and Littlewood [18] there appeared papers [36] and [38] where other heuristic arguments were presented. Even the particular case of $d = 2$ corresponding to the famous problem of existence of infinitely many twin primes is not solved. In May 2004, in a preprint publication [2] Arenstorf attempted to prove that there are infinitely many twins. However shortly after an error in the proof was pointed out by Tenenbaum [44]. For recent progress in the direction of the proof of the infinity of twins see [23].

The above notation $\pi_d(x)$ denotes prime pairs not necessarily successive. Not much is known about gaps between *consecutive* primes, which seems to be more interesting and difficult than the case of pairs of arbitrary (not consecutive) primes treated by the Hardy–Littlewood
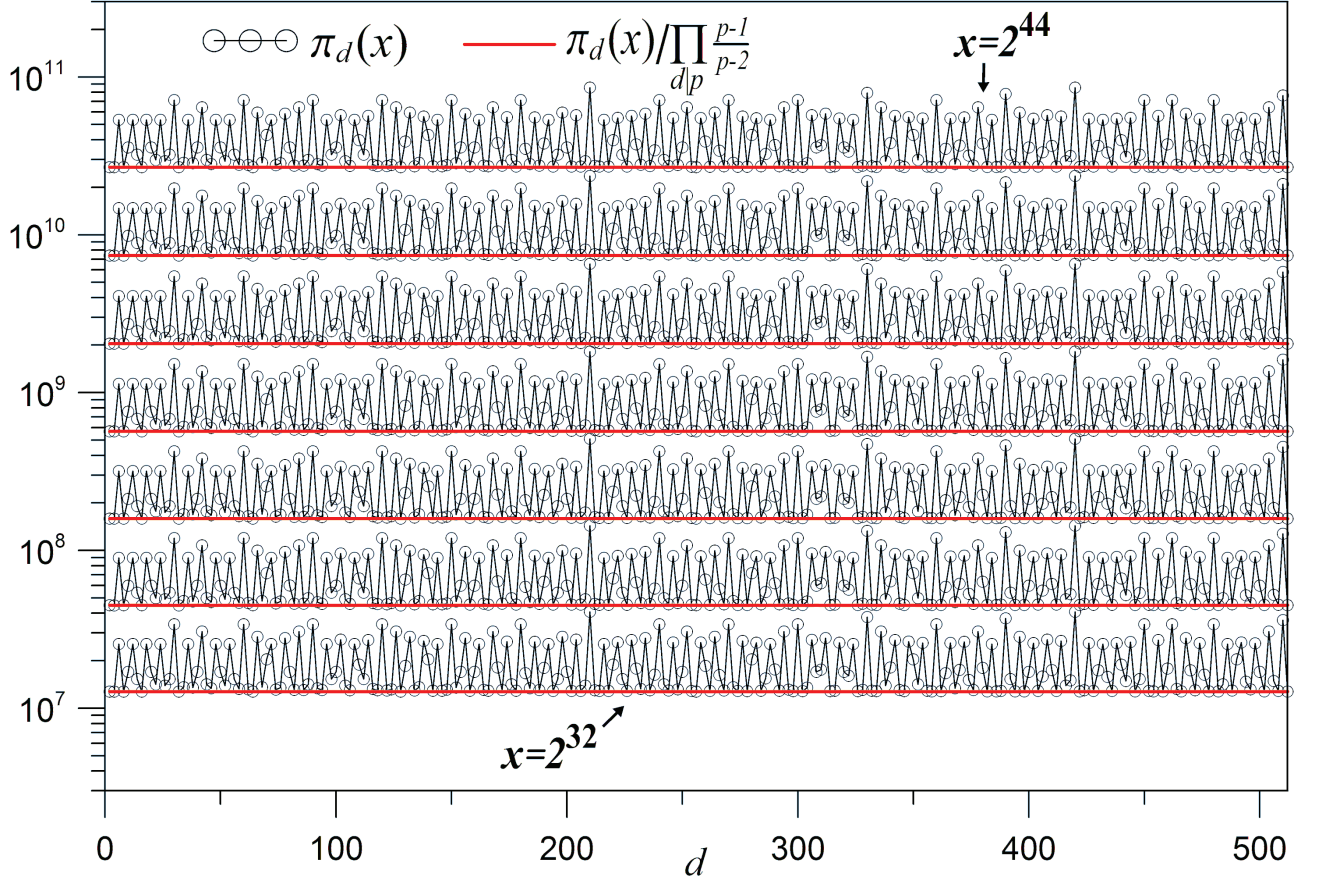
Figure 1: The plot of $\pi_d(x)$ (eq. (1)) obtained from the computer search for $d = 2, 4, \ldots, 512$ and for $x = 2^{32}, 2^{34}, \ldots, 2^{44}$. In red are the ratios $\pi_d(x)/\mathfrak{S}(d)$ plotted showing explicitly that a characteristic oscillating pattern with peaks at $6k, 30k, 210k$ is caused by the product $\mathfrak{S}(d)$.

conjecture B. Let $\tau_d(x)$ denote a number of pairs of *consecutive* primes smaller than a given bound x and separated by d:

$$\tau_d(x) = \{\text{number of pairs } p_n, p_{n+1} < x, \quad \text{with } d = p_{n+1} - p_n\}. \tag{4}$$

For odd d we supplement this definition by putting $\tau_{2k+1}(x) = 0$. The pairs of primes separated by $d = 2$ and $d = 4$ are special as they always have to be consecutive primes (with the exception of the pair (3,7) containing 5 in the middle). In this paper we will present simple heuristic reasoning leading to the formula for $\tau_d(x)$ *expressed directly by* $\pi(x)$ — the total number of primes up to x.

A few main questions related to the problem of gaps $d_n = p_{n+1} - p_n$ between consecutive primes can be distinguished. There are gaps of arbitrary length between primes: namely the n numbers $(n+1)! + 2, (n+1)! + 3, (n+1)! + 4, \ldots, (n+1)! + n + 1$ are all composite. But it is not known whether for every even d there exists a pair of consecutive primes $p_{n+1} - p_n$ with $d = p_{n+1} - p_n$. The growth rate of the form $d_n = \mathcal{O}(p_n^\theta)$ with different θ was proved in the past. A few results with θ closest to $1/2$ are the results of: C. Mozzochi [29] $\theta = \frac{1051}{1920}$, S. Lou and Q.

Yao obtained $\theta = 6/11$ [25], R.C. Baker and G. Harman have improved it to $\theta = 0.535$ [3] and recently R.C. Baker G. Harman and J. Pintz [4] have improved it by 0.01 to $\theta = 21/40 = 0.525$ which currently remains the best unconditional result. For a review of results on θ see [35]. The Riemann Hypothesis implies $d_n = \mathcal{O}(\sqrt{p_n} \log(p_n))$ and $\theta = \frac{1}{2} + \epsilon$ for any $\epsilon > 0$. On the other hand the second question about d_n concerns the existence of very large gaps. Let $G(x)$ denotes the largest gap between consecutive primes below a given bound x:

$$G(x) = \max_{p_n < x} (p_n - p_{n-1}). \tag{5}$$

For this function lower bounds are searched for: $G(x) > f(x)$. The Prime Number Theorem $\pi(x) \sim x/\log(x)$ trivially gives $G(x) > \log(x)$. Better inequality

$$G(x) \geq \frac{(ce^\gamma + o(1)) \log(x) \log\log(x) \log\log\log\log(x)}{(\log\log\log(x))^2}, \tag{6}$$

where $\gamma = 0.577216\ldots$ is the Euler-Mascheroni constant, was proved by H. Maier and C. Pomerance in [27] with $c = 1.31256\ldots$ and improved by J. Pintz to $c = 2$ in [?].

In the last few years a team of D. A. Goldston, J. Pintz, and C. Y. Yildirim has published a series of papers marking the breakthrough in some problems concerned with the prime numbers, for a review see [43], [35]. Among the results obtained by them are the following related to the subject of this paper:

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0 \tag{7}$$

and under appropriate unproved conjectures they also showed that there are infinitely many primes p_n, p_{n+1} such that:

$$p_{n+1} - p_n < 16. \tag{8}$$

In 1946 there appeared a paper [11], where the problem of different patterns of pairs, triplets etc. of primes was treated by the probabilistic methods. In particular the formula for a number of primes$< x$ and separated by a gap d was deduced on p. 57 from probabilistic arguments.

In 1974 there appeared a paper by Brent [6], where statistical properties of the distribution of gaps between consecutive primes were studied both theoretically and numerically. Brent had applied the inclusion–exclusion principle and obtained from (1) a formula for the number of consecutive prime pairs less than x separated by d. But his result (formula (4) in [6]) does not have a closed form and he had to produce on the computer a table of constants appearing in his formula (4). The attempt to estimate these sums and to write a closed formula for them was undertaken in [32]. However in this paper we will present a completely different approach to the problem of prime gaps.

The paper is organized as follows: In Sect.2 we will present a formula for $\tau_d(x)$. As applications of this expression in Sections 3 and 4 we will give a formula for $G(x)$ and $\sum_{p_n \leq x} (p_n - p_{n-1})^2$ expressed directly by $\pi(x)$ and compare it with available computer data. In Section 5 we will consider the sums of reciprocals of all consecutive primes separated by a gap d and propose a compact expression giving the values of these sums for $d \geq 6$. In Sect.6 we will derive from formulas obtained in Sect. 5 the Euler–Mertens dependence of the harmonic prime sum $\sum_{p<x} 1/p \sim \log\log(x)$. Next, the heuristic formula for the first occurrence of a given gap between consecutive primes is proposed in Sect.7. In the last Sect. 8 a behaviour of the sequence $\sqrt{p_{n+1}} - \sqrt{p_n}$ is considered in connection with the Andrica conjecture [1].

2 The basic conjecture

We have collected during over a seven months long run of the computer program the values of $\tau_d(x)$ up to $x = 2^{48} \approx 2.8147 \times 10^{14}$. During the computer search the data representing the function $\tau_d(x)$ were stored at values of x forming the geometrical progression with the ratio 2, i.e. at $x = 2^{15}, 2^{16}, \ldots, 2^{47}, 2^{48}$. Such a choice of the intermediate thresholds as powers of 2 was determined by the employed computer program in which the primes were coded as bits. The data is available for downloading from `http://www.ift.uni.wroc.pl/~mwolf/gaps.zip`. The resulting curves are plotted in Fig.2.

In the plots of $\tau_d(x)$ in Fig.2 a lot of regularities can be observed. The pattern of points in Fig.2 *does not depend on* x: for each x the arrangements of circles is the same, only the intercept increases and the slope decreases. Like in the case of $\pi_d(x)$ the oscillations are described by the product $\mathfrak{S}(d)$, see the inset in Fig. 2. There is also a possibility of plotting $\tau_d(x)$ for a couple of values of d as a function of x, but such a graph does not reveal regularities seen in Fig.2. The fact that the points in Fig.2 lie around the straight lines on the semi-logarithmic scale suggest for $\tau_d(x)$ the following

Ansatz 1 :

$$\tau_d(x) = \mathfrak{S}(d)B(x)F^d(x), \tag{9}$$

where $F(x) < 1$ (because $\tau_d(x)$ decreases with d).

The essential point of the presented below considerations consists in a possibility of determining the unknown functions $F(x)$ and $B(x)$ by *assuming only the above exponential decrease of $\tau_d(x)$ with d and employing two identities fulfilled by $\tau_d(x)$ just by definition.* First of all, the number of all gaps is smaller by 2 than the number of all primes smaller than N:

$$\sum_{(p_n - p_{n-1}),\ p_n < x} 1 \equiv \sum_{d=2}^{G(x)} \tau_d(x) = \pi(x) - 2, \tag{10}$$

where $\pi(x)$ denotes the number of primes smaller than x and $G(x)$ is the largest gap below x and which was defined in (5). The second selfconsistency condition comes from an observation that the sum of differences between consecutive primes $p_n \leq x$ is equal to the largest prime $\leq x$ (minus 3 coming from the distance to $p_2 = 3$) and for large x we can write:

$$\sum_{p_n < x} (p_n - p_{n-1}) \equiv \sum_{d=2}^{G(x)} d\ \tau_d(x) \approx x. \tag{11}$$

Writing here x instead of the largest prime $\leq x$ leads in the worst case to the error of the order $G(x) \approx \log^2(x)$, see Sect.3. The erratic behavior of the product $\mathfrak{S}(d)$ is an obstacle in calculation of the above sums (10) and (11). We will replace the product $\mathfrak{S}(d)$ in the sums by its average value. In [5] E. Bombieri and H. Davenport have proved that:

$$\sum_{k=1}^{n} \prod_{p|k,p>2} \frac{p-1}{p-2} = \frac{n}{\prod_{p>2}(1 - \frac{1}{(p-1)^2})} + \mathcal{O}(\log^2(n)); \tag{12}$$

i.e. in the limit $n \to \infty$ the number $1/\prod_{p>2}(1 - \frac{1}{(p-1)^2})$ is the arithmetical average of the product $\prod_{p|k} \frac{p-1}{p-2}$. Thus we will assume that for functions $f(k)$ going to zero like $const^{-k}$ the
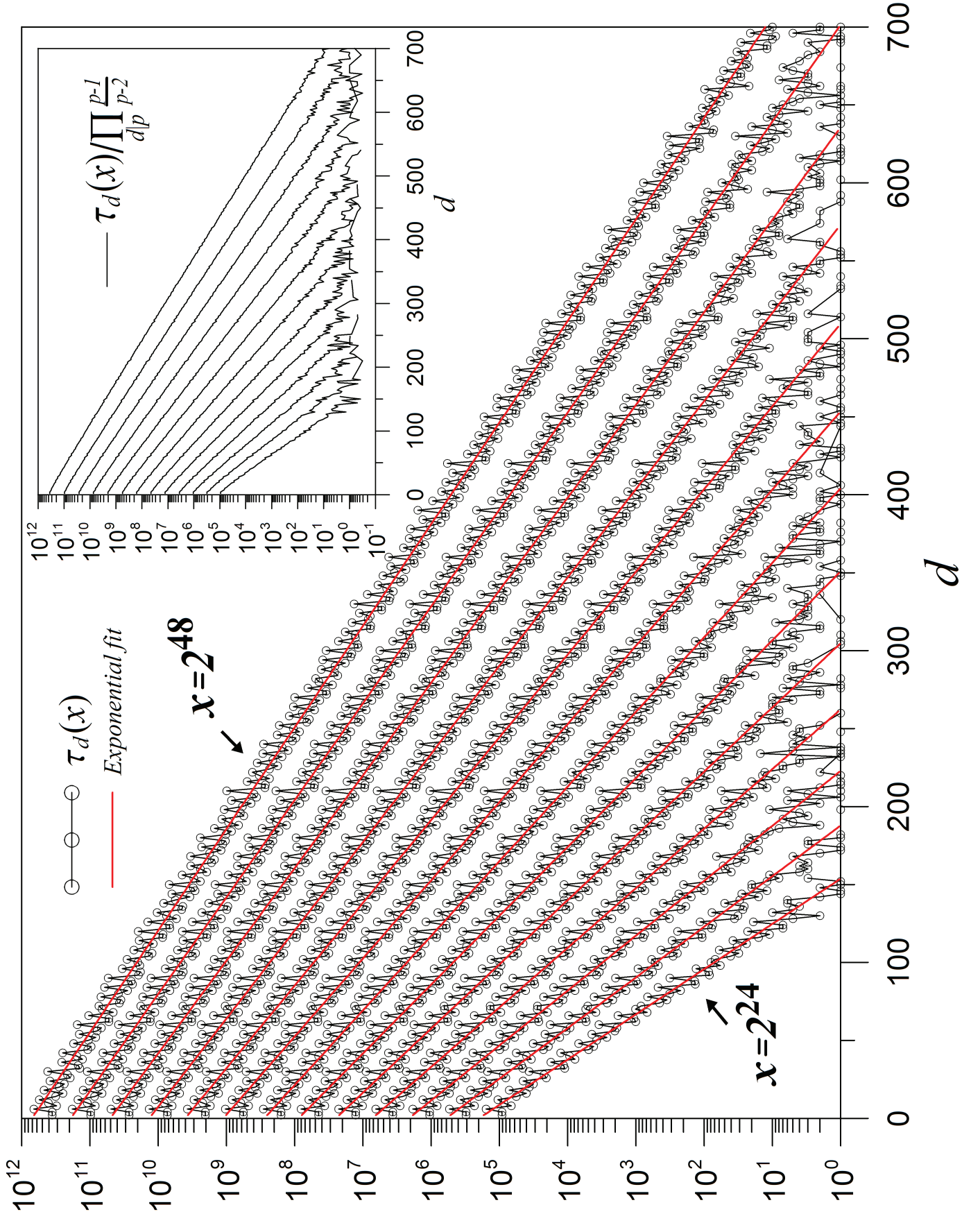
Figure 2: Plots of $\tau_d(x)$ for $x = 2^{24}, 2^{26}, \ldots, 2^{46}, 2^{48}$. In the inset plots of $\tau_d(x)/\mathfrak{S}(d)$ are shown for the same set of x. In red are exponential fits $a(x)e^{-db(x)}$ plotted.

following identity holds:

$$\sum_{k=1}^{\infty} \prod_{p|k,p>2} \frac{p-1}{p-2}\, f(k) = \frac{1}{\prod_{p>2}(1 - \frac{1}{(p-1)^2})} \sum_{k=1}^{\infty} f(k). \tag{13}$$

We extend in (10) and (11) the summations to infinity and using the Ansatz (9) and (13) we get the geometrical and differentiated geometrical series. For odd d we have defined $\tau_{2k+1}(x) = 0$. Then, writing $d = 2k$ we obtain:

$$\sum_{d=2}^{\infty} \tau_d(x) = \frac{B(x)}{c_2} \sum_{k=1}^{\infty} F^{2k}(x) = \frac{2}{c_2} \frac{B(x)F^2(x)}{1 - F^2(x)} \tag{14}$$

$$\sum_{d=2}^{\infty} d\tau_d(x) = \frac{2B(x)}{c_2} \sum_{k=1}^{\infty} kF^{2k}(x) = \frac{1}{c_2} \frac{B(x)F^2(x)}{(1 - F^2(x))^2}. \tag{15}$$

By extending summations in (10) and (11) to infinity $G(x) \to \infty$ we made an error of the order $\mathcal{O}(F(x)^{G(x)+2})$ in the first case and an error $\mathcal{O}(G(x)F(x)^{G(x)+2})$ in the second equation, both going to zero for $x \to \infty$, because for $x \to \infty$ we have $G(x) \to \infty$. That indeed $\mathcal{O}(G(x)F(x)^{G(x)+2})$ goes to zero for $x \to \infty$ can be checked *a posteriori* from the formulas for $F(x)$ and for $G(x) \sim \log^2(x)$, see Sect. 3. Thus we obtain two equations:

$$\frac{1}{c_2} \frac{B(x)F^2(x)}{1 - F^2(x)} = \pi(x), \qquad \frac{1}{c_2} \frac{2B(x)F^2(x)}{(1 - F^2(x))^2} = x \tag{16}$$

of which solutions are

$$B(x) = \frac{2c_2\pi^2(x)}{x} \frac{1}{(1 - \frac{2\pi(x)}{x})}, \qquad F^2(x) = 1 - \frac{2\pi(x)}{x} \tag{17}$$

and a posteriori the inequality $F(x) < 1$ holds evidently. Finally, we state the main:

Conjecture 1
The function $\tau_d(x)$ is expressed directly by $\pi(x)$:

$$\tau_d(x) = C_2 \prod_{p|d,p>2} \frac{p-1}{p-2} \frac{\pi^2(x)}{x} \left(1 - \frac{2\pi(x)}{x}\right)^{\frac{d}{2}-1} + \; error\; term(x,d) \; \text{for}\; d \geq 6 \tag{18}$$

For Twins ($d = 2$) and Cousins ($d = 4$) the identities $\tau_{2,4}(x) = \pi_{2,4}(x)$ hold. Because d is even the power of $(1 - 2\pi(x)/x)$ has a finite number of terms. The formula (18) consists of three terms. The first one depends only on d, the second only on x, but the third term depends both on d and x. In the usual probabilistic approach one should obtain $(1 - \frac{\pi(x)}{x})^{d-1}$, see e.g. [21], [43, p. 3]: to have a pair of adjacent primes separated by d there have to be $d - 1$ consecutive composite numbers in between and probability of such an event is $(1 - \pi(x)/x)^{d-1}$; then the term in front of it comes from the normalization condition.

Although (18) is postulated for $d \geq 6$, we get from it for $d = 2$:

Conjecture 2

$$\tau_2(x) \equiv \pi_2(x) = C_2 \, \frac{\pi^2(x)}{x} \; + \; error \; term(x) \tag{19}$$

instead of the usual conjectures

$$\pi_2(x) \sim C_2 \frac{x}{\log^2(x)} \tag{20}$$

or

$$\pi_2(x) \sim C_2 \int_2^x \frac{du}{\log^2(u)} \equiv C_2 \mathrm{Li}_2(x). \tag{21}$$

Remark: The equation (19) expresses the intuitively obvious fact that the number of twins should be proportional to the square of $\pi(x)$. Of course (19) for $\pi(x) \sim x/\log(x)$ goes into (20).

We have checked with the available computer data that (19) is better than (20) but worse than (21). Because $\mathrm{Li}_2(x)$ in (21) monotonically increases while there are local fluctuations in the density of primes and twins, the above formula (19) incorporates all irregularities in the distribution of primes into the formula for the number of twins. Since both $d = 2$ and $d = 4$ gaps are necessarily consecutive, we propose the identical expression (19) for $\tau_4(x) \equiv \pi_4(x) \approx \pi_2(x)$, see [46].

It is possible to obtain another form of the formula for $\tau_d(x)$, more convenient for later applications. Namely, let us represent the function $F(x)$ in the form: $F(x) = e^{-A(x)}$, i.e. now the Ansatz 1 has the form:

Ansatz 1′

$$\tau_d(x) \sim B(x)\mathfrak{S}(d)e^{-A(x)d}, \tag{22}$$

where $A(x)$ is the slope of the lines plotted in red in Fig. 2 and as we can see $A(x)$ goes to zero for $x \to \infty$. In the equations (16) we use in the nominators the approximation $e^{-2A(x)} \approx 1 - 2A(x)$ and in the denominators $1 - e^{-2A(x)} \approx 2A(x)$ for small $A(x)$ obtaining finally

Conjecture 1′

$$\tau_d(x) = C_2 \frac{\pi^2(x)}{x - 2\pi(x)} \prod_{p|d,p>2} \frac{p-1}{p-2} e^{-d\pi(x)/x} + error \; term(x, d) \; \text{ for } d \geq 6. \tag{23}$$

For large x we can skip $2\pi(x)$ in comparison with x in the denominator and obtain finally the following pleasant formula:

Conjecture 1″

$$\tau_d(x) = C_2 \frac{\pi^2(x)}{x} \prod_{p|d,p>2} \frac{p-1}{p-2} e^{-d\pi(x)/x} + error \; term(x, d) \; \text{ for } d \geq 6. \tag{24}$$

In equations (23) and (24) the term in the exponent has a simple interpretation: difference d is divided by the mean gap $x/\pi(x)$ between consecutive primes. Because for small u an

approximation $\log(1-u) \approx -u$ holds, we can turn for large x the conjecture (18) to the form of conjecture 1′:

$$\left(1 - \frac{2\pi(x)}{x}\right)^{\frac{d}{2}} = e^{\frac{d}{2}\log\left(1 - \frac{2\pi(x)}{x}\right)} \approx e^{-\frac{d\pi(x)}{x}} \tag{25}$$

Next we see that for large x both (18) and (23) go into the conjecture (24).

Putting in (24) $\pi(x) \sim x/\log(x)$ and comparing with the original Hardy–Littlewood conjecture we obtain that the number $\tau_d(x)$ of *successive* primes $(p_{n+1},\ p_n)$ smaller than x and of the difference $d\ (= p_{n+1} - p_n)$ is diminished by the factor $\exp(-d/\log(x))$ in comparison with the number of *all* pairs of primes (p, p') apart in the distance $d = p' - p$:

$$\tau_d(x) \sim \pi_d(x) e^{-d/\log(x)} \quad \text{for } d \geq 6. \tag{26}$$

Heuristically, this relation encodes in the series for $e^{-d/\log(x)}$ the inclusion-exclusion principle for obtaining $\tau_d(x)$ from $\pi_d(x)$. The above relation is confirmed by comparing the Figures 1 and 2. R.P. Brent in [6] using the inclusion-exclusion principle has obtained from the B conjecture of Hardy and Littlewood the formula for $\tau_d(x)$, which agrees very well with computer results. However the formula of Brent (eq.(4) in the paper [6]) is not of a closed form: it contains a double sequence of constants $A_{r,k}$, which can be calculated only by a direct use of the computer, what is very time consuming, see discussion of S. Herzog at the web site http://mac6.ma.psu.edu/primes. R. P. Brent in [6] in Table 2 compares the number of actual gaps $d = 2, \ldots, 80$ in the interval $(10^6, 10^9)$ with the numbers predicted from his formula finding perfect agreement. Analogous method to determine the values of $\tau_d(x)$ was employed in [32, see eq.(2-8) and the preceding formula]. The formula (2-8) from [32] adapted to our notation has the form:

$$\tau_d(x) \sim C_2 \mathfrak{S}(d) \int_2^x \frac{\exp(-d/\log(u))}{\log^2(u)} du. \tag{27}$$

Integrating the above integral once by parts gives a term $x e^{-d/\log(x)}/\log^2(x)$ corresponding to (24) with $\pi(x) \sim x/\log(x)$.

It is not possible to guess an analytical form of error terms in formulas (18), (23) and (24) at present (let us remark that the error term in the twins conjectures (20) or (21) is not known even heuristically). The only way to obtain some information about the behaviour of *error term*(x, d) is to compare these conjectures with actual computer counts of $\tau_d(x)$. Of course, the best accuracy has the formula (18). We have compared it with generated by the computer actual values of $\tau_d(x)$ — i.e. we have looked at values of

$$\Delta(x, d) \equiv \tau_d(x) - C_2 \mathfrak{S}(d) \frac{\pi^2(x)}{x}\left(1 - \frac{2\pi(x)}{x}\right)^{\frac{d}{2}-1}. \tag{28}$$

The values of $\Delta(x, d)$ were stored for 105 values of $d = 2, 4, \ldots, 210(= 2 \cdot 3 \cdot 5 \cdot 7)$ at the arguments x forming the geometrical progression $x_k = 1000 \times (1.03)^k$. Additionally the values of $|\Delta(x, d)| < 9$ were stored to catch sign changes of $\Delta(x, d)$. It is difficult to present these data for all values of d. We have found that for some gaps d there was monotonic increase of $\Delta(x, d)$, for other gaps there were sign changes of the difference $\Delta(x, d)$, see Fig.3. For 30
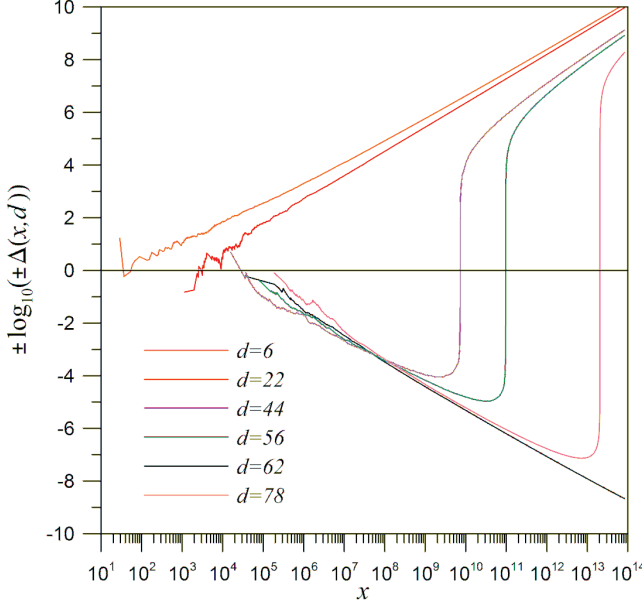
Figure 3: Plots of $\Delta(x,d)$ on the double logarithmic scale for $d = 6, 22, 44, 56, 62, 78$. On the y axis we have plotted $\log_{10}(\Delta(x,d))$ if $\Delta(x,d) > 0$ and $-\log_{10}(-\Delta(x,d))$ if $\Delta(x,d) < 0$.
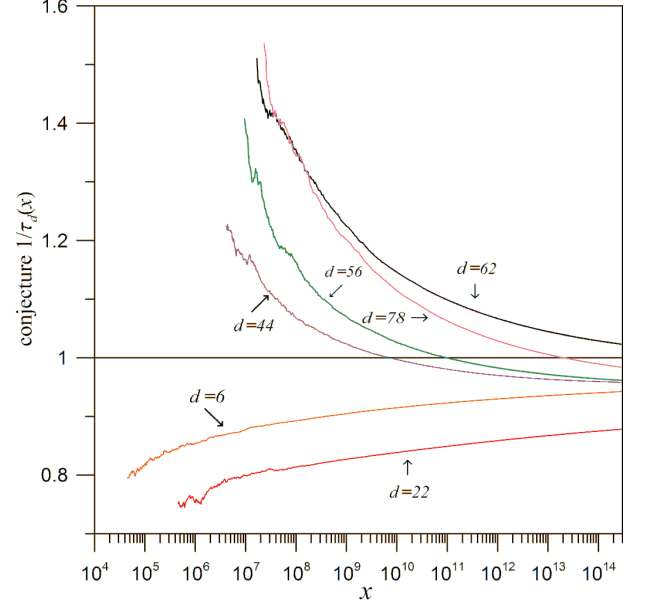
Figure 4: Plots of ratios of the values predicted from the Conjecture 1 to the real values of $\tau_d(x)$ for $d = 6, 22, 44, 56, 62, 78$. The plots begin at such x that $\tau_d(x) > 1000$ to avoid large initial fluctuations of these ratios (see initial parts of curves in the previous Figure).

values of d of all 105 looked for we have found sign changes for $x < 8 \times 10^{13}$. Surprising is the steep growth of $\Delta(x,d)$ for $d = 44, 56, 78$ (the same behaviour we have seen for other values of d) in the region of crossing the $y = 0$ line. In fact, there were 76 sign changes of $\Delta(x,54)$, 109 sign changes of $\Delta(x,56)$ and 207 sign changes of $\Delta(x,78)$. The general rule is that the ratio $\tau_d(x)/C_2\mathfrak{S}(d)\frac{\pi^2(x)}{x}(1-\frac{2\pi(x)}{x})^{\frac{d}{2}-1}$ tends to 1, see Fig. 4. Thus we formulate the

Conjecture 3

For every d there are infinitely many sign changes of the functions $\Delta(x,d)$. For fixed d we guess

$$\lim_{x\to\infty} \frac{\text{Conjecture}_{1,1',1''}(d,x)}{\tau_d(x)} = 1. \tag{29}$$

We can test the conjecture (24) with available computer data plotting on one graph the scaled quantities:

$$T_d(x) = \frac{x\tau_d(x)}{C_2\mathfrak{S}(d)\pi^2(x)}, \qquad D(x,d) = \frac{d\pi(x)}{x}. \tag{30}$$

From the conjecture (24) we expect that the points $(D(x,d), T_d(x))$, $d = 2, 4, \ldots, G(x)$ should coincide for each x — the function $\tau_d(x)$ displays scaling in the physical terminology. In Fig. 5 we have plotted the points $(D(x,d), T_d(x))$ for $x = 2^{28}, 2^{38}, 2^{48}$. If we denote $u = D(x,d)$ then all these scaled functions should lie on the pure exponential decrease e^{-u} (Poisson distribution, see [42, p.60]), shown in red in Fig. 5. We have determined by the least square method slopes

$s(x)$ of the fits $a(x)e^{-s(x)u}$ to the linear parts of $(D(x,d), \log(T_d(x)))$. The results are presented in Fig. 6. The slopes very slowly tend to 1: for over 6 orders of x they change from 1.187 to 1.136.
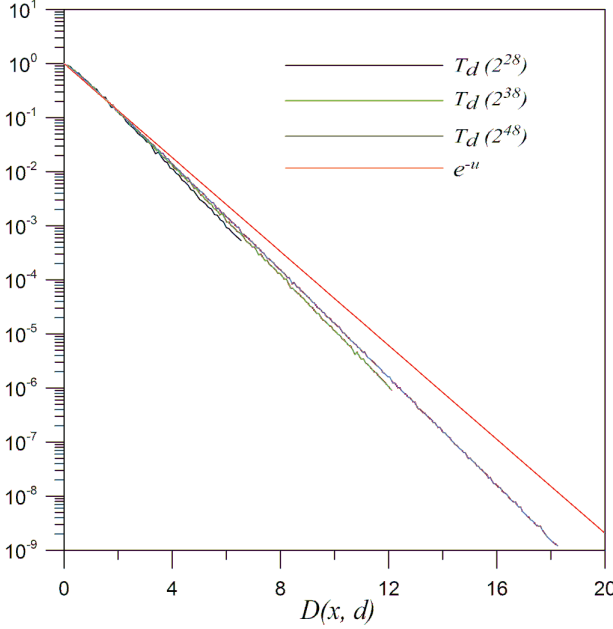

Figure 5: Plots of $(D(x,d), T_d(x))$ for $x = 2^{28}, 2^{38}, 2^{48}$ and in red the plot of e^{-u}. Only the points with $\tau_d(x) > 1000$ were plotted to avoid fluctuations at large $D(x,d)$.
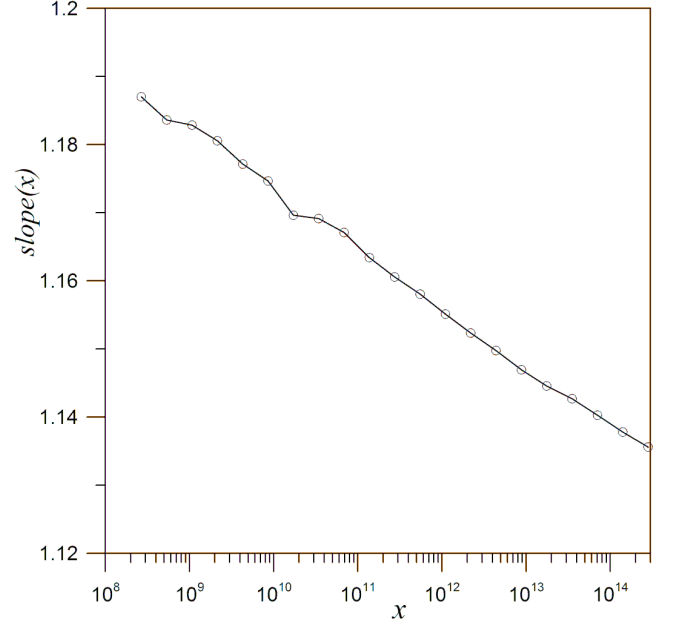
Figure 6: Plot of slopes obtained from fitting straight lines to $(D(d,x), \log(T_d(x)))$ for $x = 2^{28}, 2^{29}, \ldots, 2^{48}$.

3   Maximal gap between consecutive primes

From (18) or (24) we can obtain approximate formula for $G(x)$ assuming that maximal difference $G(x)$ appears only once, so $\tau_{G(x)}(x) = 1$: simply the largest gap is equal to the value of d at which $\tau_d(x)$ touches the d-axis on Fig.2. Skipping the oscillating term $\mathfrak{S}(d)$, which is very often close to 1, we get for $G(x)$ the following estimation expressed directly by $\pi(x)$:

Conjecture 4

$$G(x) \sim g(x) \equiv \frac{x}{\pi(x)} \big(2\log(\pi(x)) - \log(x) + c\big), \tag{31}$$

where $c = \log(C_2) = 0.2778769\ldots$.

Remark: The above formula explicitly reveals the fact that the value of $G(x)$ is connected with the number of primes $\pi(x)$: more primes means smaller $G(x)$. It is intuitively obvious: if we draw randomly from a set of natural numbers $\{1, 2, \ldots N\}$ some subset of different numbers $r_1, r_2, \ldots r_k$ and calculate differences $\delta_i = r_{i+1} - r_i$, then for larger k we will expect smaller δ_k — more elements in the subset smaller gaps between them.

For the Gauss approximation $\pi(x) \sim x/\log(x)$ the following dependence follows:

$$G(x) \sim \log(x)(\log(x) - 2\log\log(x) + c) \tag{32}$$

and for large x it passes into the well known Cramer's [12] conjecture:

$$G(x) \sim \log^2(x). \tag{33}$$

The examination of the formula (31) and the formula (33) with the available results of the computer search is given in Fig.7. The lists of known maximal gaps between consecutive primes we have taken from our own computer search up to 2^{48} and larger from web sites www.trnicely.net and www.ieeta.pt/~tos/gaps.html. The largest known gap 1476 between consecutive primes follows the prime $1425172824437699411 = 1.42\ldots \times 10^{18}$. On these web sites tabulated values of $\pi(x)$ can be also found and we have used them to plot the formula (31). Let $\nu_G(T)$ denotes the number of sign changes of the difference $G(x) - g(x)$ for $2 < x < T$. There are 33 sign changes of the difference $G(x) - g(x)$ in the Fig.7 and $\nu_G(T)$ is presented in the inset in Fig. 7. The least square method gives for $\nu_G(T)$ the equation $0.786 \log(T) + 0.569$.

There appeared in literature a few other formulas for $G(x)$, see e.g. [40], [10]; in particular D.R. Heath-Brown in [21, p. 74] gives the following formula:

$$G(x) \sim \log(x)(\log(x) + \log\log\log(x)). \tag{34}$$

A. Granville argued [16] that the actual $G(x)$ can be larger than that given by (33), namely he claims that there are infinitely many pairs of primes p_n, p_{n+1} for which:

$$p_{n+1} - p_n = G(p_n) > 2e^{-\gamma}\log^2(p_n) = 1.12292\ldots\log^2(p_n). \tag{35}$$

where $\gamma = 0.577216\ldots$ is the Euler–Mascheroni constant. The estimation (35) follows from the inequalities proved by H.Maier in the paper [26], which put into doubts Cramer's ideas. For other contradiction between Cramer's model and the reality, see [34].

4 The Heath-Brown conjecture on the $\sum_{p_n \leq x}(p_n - p_{n-1})^2$

As the application of the formula (18) we consider the conjecture made by D.R. Heath-Brown in [20]. Assuming the validity of the Riemann Hypothesis and the special form of the Montgomery conjecture on the pair correlation function of zeros of the $\zeta(s)$ function, Heath-Brown has conjectured in this paper that

$$\sum_{p_n \leq x}(p_n - p_{n-1})^2 \sim 2x\log(x). \tag{36}$$

Erdös conjectured that the r.h.s. should be *const* $x\log^2(x)$, see [17, bottom of p.20]. From the guessed formula (18) we obtain the above sum expressed directly by $\pi(x)$ (we have extended the summation over d up to infinity and used (13); then the dependence on c_2 drops out):

Conjecture 5

$$\sum_{p_n < x}(p_n - p_{n-1})^2 = \sum_{d=2,4,6,\ldots} d^2\tau_d(x) \sim \frac{2x^3}{\pi(x)(x - 2\pi(x))}\left(1 - \frac{3\pi(x)}{x} + \frac{2\pi^2(x)}{x^2}\right) \tag{37}$$
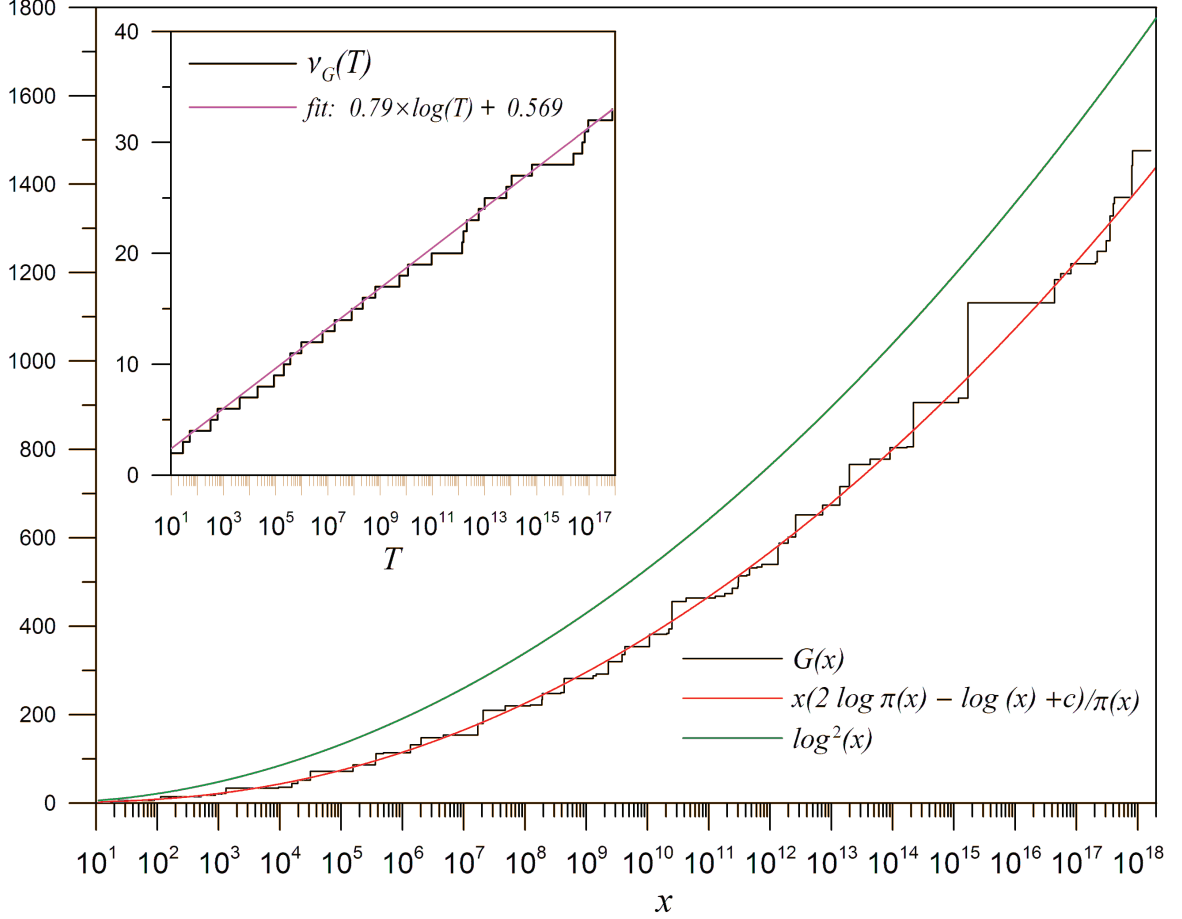
Figure 7: The comparison of $G(x)$ and $g(x)$ as well as of the Cramer conjecture. In the inset there is the plot of a number of crossings of curves representing $G(x)$ and $g(x)$. This figure should be compared with the figure on page 12 in [48].

For large x we can reduce the above formula to a simple form:

$$\sum_{p_n < x} (p_n - p_{n-1})^2 \sim \frac{2x^2}{\pi(x)} \tag{38}$$

which for $\pi(x) \sim x/\log(x)$ gives exactly (36). The above equation is intuitively obvious: the sum of squares of $(p_n - p_{n-1})$ is proportional to x^2 and inversely proportional to $\pi(x)$, because more primes means smaller differences $p_n - p_{n-1}$ on average. The same formula $2x^2/\pi(x)$ is obtained from the conjecture (24) in the limit of large x. In the Table I the comparison between the predictions (36) and (37) and real computer data is shown. As it is seen from the column 3 the convergence towards 1 of the ratio of the Heath-Brown to the real data is very slow, while the expression (37) predicts the actual numbers $\sum_{p_n < x} (p_n - p_{n-1})^2$ better.

In the past in the literature there were studied sums over large differences between consecutive primes, see e.g. [22], [14]. For example D. Goldston has proved assuming the Riemann

Hypothesis that

$$\sum_{\substack{p_n < x \\ p_n - p_{n-1} \geq H}} (p_n - p_{n-1}) = \mathcal{O}\left(\frac{x \log(x)}{H}\right) \tag{39}$$

uniformly for $H \geq 1$, while from (18) we get

$$\sum_{\substack{p_n < x \\ p_n - p_{n-1} \geq H}} (p_n - p_{n-1}) = \sum_{d \geq H} d \; \tau_d(x) \sim \frac{x^2}{(x - 2\pi(x))}\left(1 - \frac{2\pi(x)}{x}\right)^{H/2}\left(1 + \frac{(H-2)\pi(x)}{x}\right). \tag{40}$$

For $H = 2$ it gives the correct value x on the r.h.s. of the above equation. Putting in r.h.s of (40) $\pi(x) = x/\log(x)$ and expanding with respect to $1/\log(x)$ for large x we obtain:

$$\sum_{\substack{p_n < x \\ p_n - p_{n-1} \geq H}} (p_n - p_{n-1}) \sim x + \frac{H(H-2)}{2}\frac{x}{\log^2(x)} + \mathcal{O}\left(\frac{1}{\log^3(x)}\right), \tag{41}$$

which for x so large that $\log(x) > H$ is indeed smaller than upper bound (39) of Goldston. In general, expressions for sums of the form $\sum_{H \leq d \leq K} f(d)$ can be obtained in closed form if the sums are differentiated geometrical series in d.

TABLE I

The sum of squares of gaps between consecutive primes. In the second column the numbers obtained by a computer are given, while in the third one values obtained from eq.(36) and in the fifth from eq.(37) are presented. The fourth and sixth columns contain the appropriate ratios.

x	$\sum_{p_n < x}(p_n - p_{n-1})^2$	eq.(36)	$\frac{\sum_n d_n^2}{eq.(36)}$	eq.(37)	$\frac{\sum_n d_n^2}{eq.(37)}$
2^{24}	444929861	558195733	0.7971	488725881	0.9104
2^{26}	1959715561	2418848443	0.8102	2141587523	0.9151
2^{28}	8565851937	10419653325	0.8221	9313220996	0.9198
2^{30}	37168128501	44655665552	0.8323	40239313423	0.9237
2^{32}	160316134721	190530845965	0.8414	172900857995	0.9272
2^{34}	687851546609	809756094320	0.8495	739353131559	0.9303
2^{36}	2938092559089	3429555231277	0.8567	3148372990028	0.9332
2^{38}	12499933597193	14480344308470	0.8632	13357112013493	0.9358
2^{40}	52993288896469	60969870777867	0.8692	56482296752813	0.9382
2^{42}	223959886541173	256073457287370	0.8746	238142313949083	0.9404
2^{44}	943825347126665	1073069725777350	0.8796	1001414251864841	0.9425
2^{46}	3967383251021137	4487382489617471	0.8841	4201009869963194	0.9444
2^{48}	16638404184530149	18729944304492034	0.8883	17585360374792679	0.9462

5 Generalized Brun's constants

In 1919 Brun [9] has shown that the sum of the reciprocals of all twin primes is finite:

$$\mathcal{B}_2 = \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{1}{5} + \frac{1}{7}\right) + \left(\frac{1}{11} + \frac{1}{13}\right) + \ldots < \infty. \tag{42}$$

Sometimes 5 is included only once, but here we will adopt the above convention. The analytical formula for $\mathcal{B}_2$ is unknown and the sum (42) is called the Brun's constant [41]. The numerical estimations give [30] $\mathcal{B}_2 = 1.90216058\ldots$. Here we are going to generalize the above $\mathcal{B}_2$ to the sums of reciprocals of all consecutive primes separated by gap d and to propose a compact expression giving the values of these sums for $d \geq 6$.

Let $\mathcal{T}_d$ denote the set of consecutive primes separated by distance d:

$$\mathcal{T}_d = \{(p_{n+1}, p_n) : p_{n+1} - p_n = d\}. \tag{43}$$

We define the generalized Brun's constants by the formula:

$$\mathcal{B}_d = \sum_{p \in \mathcal{T}_d} 1/p. \tag{44}$$

We adopt the rule, that if a given gap d appears two times in a row: $p_n - p_{n-1} = p_{n+1} - p_n$, the corresponding middle prime p_n is counted two times (in the case of $\mathcal{B}_2$ only 5 appears two times); e.g. for $d = 6$ we have the terms $\ldots + 1/47 + 1/53 + 1/53 + 1/59 + \ldots$ and next $\ldots + 1/151 + 1/157 + 1/157 + 1/163 + \ldots$.

B.Segal has proved [39] that the sum in (44) is convergent for every d, thus generalized Brun's constants are finite. Because of that the sums (44) can be called Brun–Segal constants for $d > 2$.

Let us define partial (finite) sums:

$$\mathcal{B}_d(x) = \sum_{p \in \mathcal{T}_d, p < x} 1/p. \tag{45}$$

We have computed on the computer quantities $\mathcal{B}_d(x)$ for x up to $x = 2^{46} \approx 7.037 \times 10^{13}$. The partial generalized Brun's constants $\mathcal{B}_d(x)$ were stored at $x = 2^{15}, 2^{16}, \ldots, 2^{46}$ and data is available for download from `http://www.ift.uni.wroc.pl/~mwolf/Brun.zip`. In Fig. 8 we present a part of the obtained data.

The dependence of $\mathcal{B}_2(x)$ on x is usually (see [41], [7]) obtained by appealing to the conjecture (20) (i.e. Hardy–Littlewood conjecture (1) for $d = 2$). It gives that the probability to find a pair of twins in the vicinity of x is $2c_2/\log^2(x)$, so the expected value of the finite approximation to the Brun constant can be estimated as follows:

$$\mathcal{B}_2(x) = \mathcal{B}_2(\infty) - \sum_{p \in \mathcal{T}_2, p > x} \frac{1}{p} \approx \mathcal{B}_2 - 4c_2 \int_x^{\infty} \frac{du}{u \log^2(u)} = \mathcal{B}_2 - \frac{4c_2}{\log(x)}. \tag{46}$$

It means that the plot of finite approximations $\mathcal{B}_2(x)$ to the original Brun constant is a linear function of $1/\log(x)$. The same reasoning applies *mutatis mutandis* to the gap $d = 4$.

To repeat the above reasoning for $d = 2$, 4 for larger d an analog of the Hardy–Littlewood conjecture for the pairs of *consecutive* primes separated by distance d is needed and we will use the form (24) for $\tau_d(x)$ (the integrals occurring below can be calculated analytically also for (18)). Putting in the equation (24) $\pi(x) = x/\log(x)$ we obtain for $d \geq 6$:

$$\mathcal{B}_d(x) = \mathcal{B}_d(\infty) - \sum_{p \in \mathcal{T}_d, p > x} \frac{1}{p} \approx \mathcal{B}_d - 4c_2 \prod_{p \mid d} \frac{p-1}{p-2} \int_x^{\infty} \frac{e^{-d/\log(u)}}{u \log^2(u)} du. \tag{47}$$
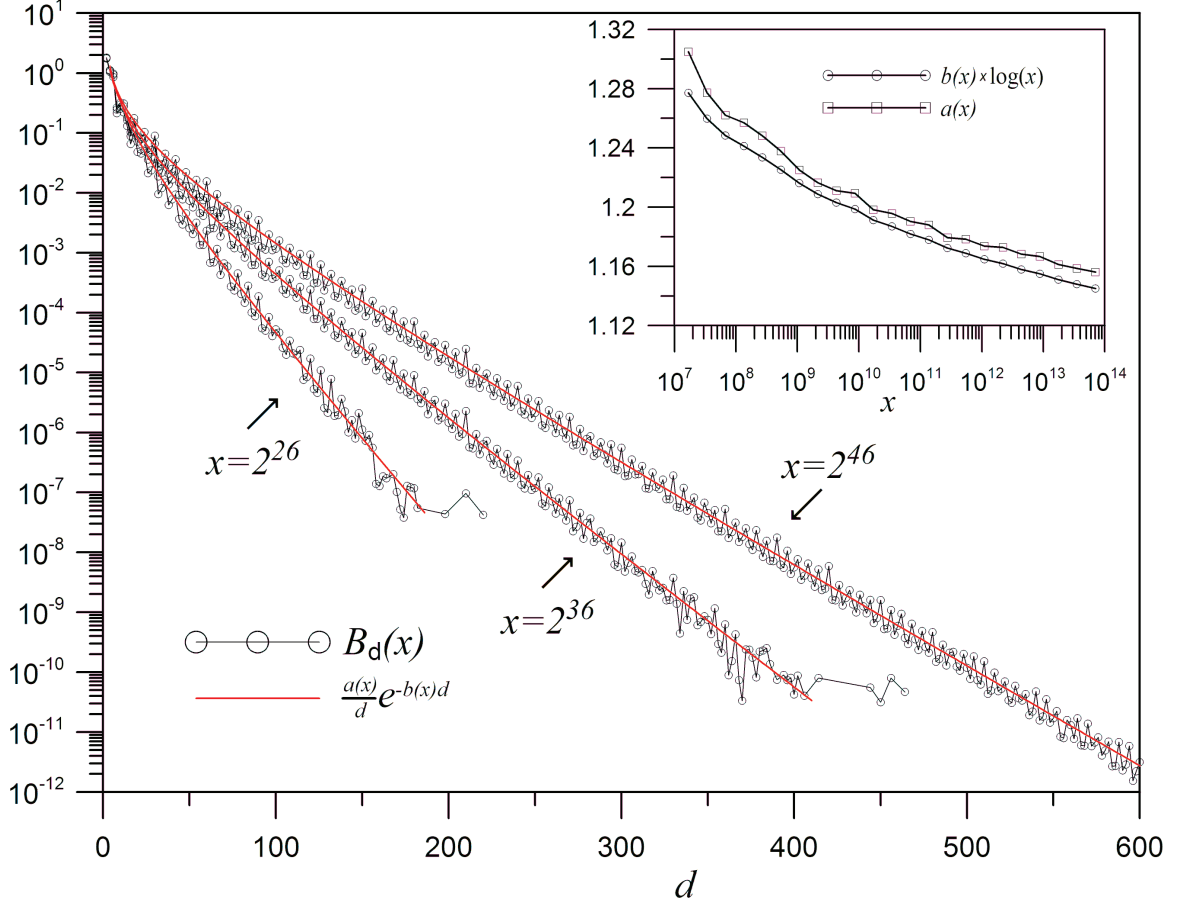
Figure 8: The plot of $\mathcal{B}_d(x)$ for $x = 2^{26} = 6.71... \times 10^7, 2^{36} = 6.87... \times 10^{10}, 2^{46} = 7.04... \times 10^{13}$. The fit $a(x)e^{-b(x)d}/d$ to $\mathcal{B}_d(x)/\mathfrak{S}(d)$ obtained by the least square method is plotted in red. In the inset the values of $a(x)$ as well as differences between conjectured slope $-1/\log(x)$ and actual fit $b(x)$ are shown for $x = 2^{24}, 2^{25}, \ldots, 2^{46}$.

and the integral can be calculated explicitly:

$$\mathcal{B}_d(x) = \mathcal{B}_d(\infty) + \frac{2C_2}{d} \prod_{p|d} \frac{p-1}{p-2} \left(e^{-d/\log(x)} - 1\right). \tag{48}$$

From this, it follows that the partial sums $\mathcal{B}_d(x)$ for $d \geq 6$ should depend linearly on $e^{-d/\log(x)}$ instead of linear dependence on $1/\log(x)$ for $\mathcal{B}_2(x)$ and $\mathcal{B}_4(x)$.

Because $\mathcal{B}_d(x)$ is 0 for $x = 1$ (in fact each $\mathcal{B}_d(x)$ will be zero up to the first occurrence of the gap d, see Sect. 7), we take in (48) the limit $x \to 1^+$ and obtain

$$\mathcal{B}_d(\infty) \equiv \mathcal{B}_d = \frac{4c_2}{d} \prod_{p|d} \frac{p-1}{p-2} \quad \text{for } d \geq 6. \tag{49}$$

Thus the formula expressing the x dependence of $\mathcal{B}_d(x)$ has the form:

$$\mathcal{B}_d(x) = \frac{2C_2}{d} \prod_{p|d} \frac{p-1}{p-2} e^{-d/\log(x)} + error\ term(d, x). \tag{50}$$

The characteristic shape of the dependence of $\mathcal{B}_d(x)/\mathfrak{S}(d)$ on d is described by the relation $\log(\mathcal{B}_d(x)/\mathfrak{S}(d)) \sim -\log(d) - d/\log(x)$: if $d/\log(x) > \log(d)$ the linear dependence on d preponderates. In these linear parts we have fitted by least square method the dependence $\log(a(x)) - db(x)$ to the actual values of $\log(d\mathcal{B}_d(x)/2C_2\mathfrak{S}(d))$. We have obtained, that indeed $b(x)$ tends to $1/\log(x)$ and $a(x)$ tends to 1 with increasing x, see the inset in Fig. 8.

The comparison of the formula (49) with the values extrapolated from the partial approximations $\mathcal{B}_d(2^{46})$ by the formula

$$\mathcal{B}_d(\infty) = \mathcal{B}_d(2^{46}) + \frac{2C_2}{d} \prod_{p|d} \frac{p-1}{p-2} \left(1 - e^{-d/46\log(2)}\right) \tag{51}$$

obtained from the equation (48), is shown in Fig. 9 for $d \geq 6$ — predicted by (49) values for $d = 2$ and $d = 4$ are skipped. Because on average the product $\mathfrak{S}(d)$ is equal to $1/c_2$, we can write $\mathcal{B}_d \approx 4/d$. Let us mention that $4/d$ provides remarkably good approximations to $\mathcal{B}_2 = 1.90216058\ldots$ and $\mathcal{B}_4 = 1.19705\ldots$.

The outcome of the above analysis allow us to make the

Conjecture 6

$$\mathcal{B}_d(\infty) \equiv \mathcal{B}_d = \frac{4c_2}{d} \prod_{p|d} \frac{p-1}{p-2} + error\ term(d), \quad \text{for } d \geq 6. \tag{52}$$

The data shown in the inset in Fig.9 suggest that the error term should decrease with d.

6   The Merten's Theorem on the prime harmonic sum.

It is well known, that the sum of reciprocals of all primes smaller than x is given by [28], [19, Theorems 427 and 428], [45]:

$$\sum_{p<x} \frac{1}{p} = \log(\log(x)) + M + o(1); \tag{53}$$

here $M = 0.2614972\ldots$ is the Mertens constant and it has a few representations:

$$M = \sum_p (\log(1 - 1/p) + 1/p) = \gamma + \sum_{k=2}^{\infty} \mu(k) \log(\zeta(k))/k, \tag{54}$$

where $\mu(n)$ is the Moebius function and $\zeta(s)$ is the Riemann zeta function. On the other hand, the sum $\sum_{p<x} 1/p$ can be expressed by finite approximations to the generalized Brun's
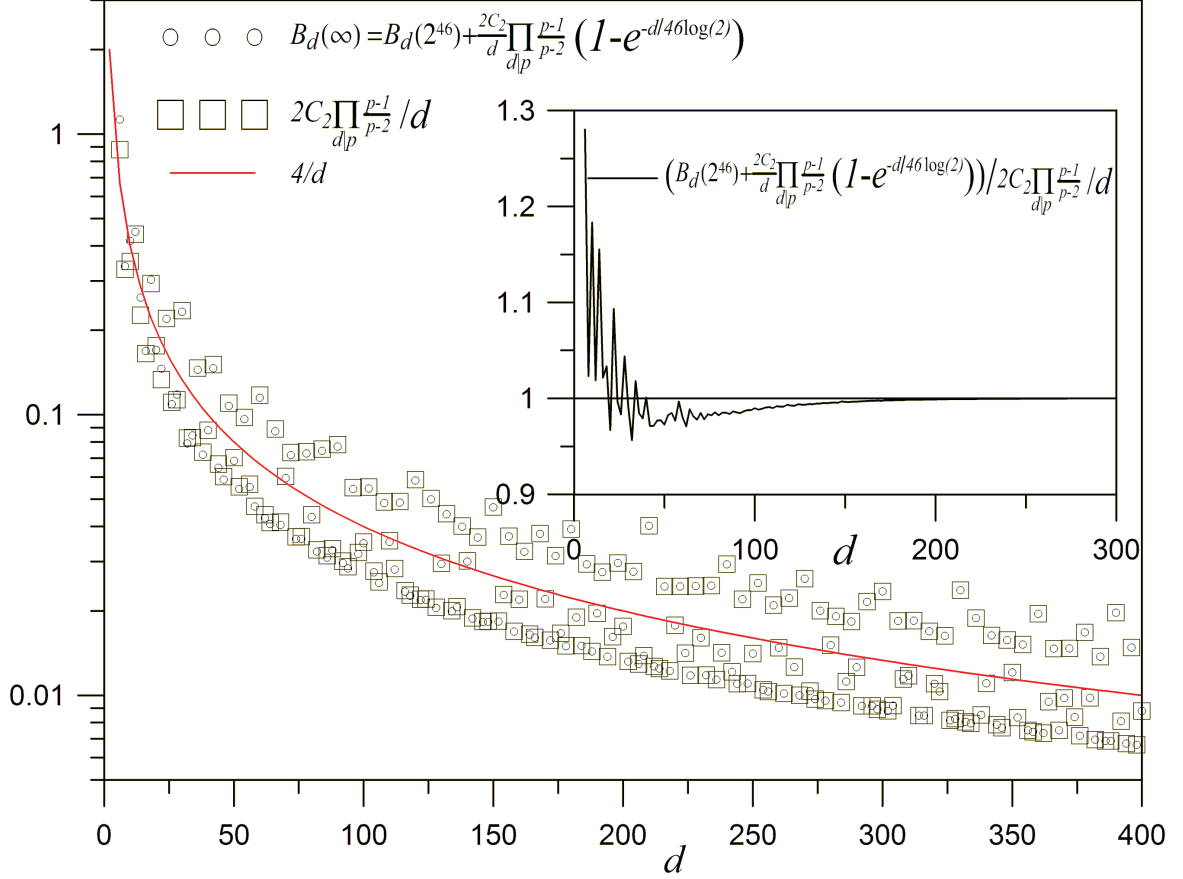
Figure 9: The plot of the generalized Brun's constants $\mathcal{B}_d$ extrapolated from (51) marked by circles and predicted by (49) marked by squares. In the inset the ratio of the values obtained from these two equations is plotted.

constants:

$$\sum_{p<x}\frac{1}{p}=\frac{1}{2}+\frac{1}{6}+\frac{1}{2}\sum_d\mathcal{B}_d(x)=M'+\frac{2}{3}+C_2\sum_{d=2}^{G(x)}\frac{1}{d}\prod_{p|d}\frac{p-1}{p-2}e^{-d/\log(x)} \tag{55}$$

Because each prime except 2 and 3 (hence the terms $1/2$ and $1/6$) appears as the right and left end of the adjacent pairs, we have to divide the sum by $1/2$ (we remind that we have adopted in Sect.5 the convention that if a given gap d appears two times in a row: $p_n-p_{n-1}=p_{n+1}-p_n$ — the corresponding middle prime p_n is counted two times). We have introduced here the constant M' which accounts the sum of the unknown errors terms in (50) as well as incorporates the fact that the dependence of $B_2(x)$ and $B_4(x)$ on x is not described by the formula (50) but by (46). The sum in (55) runs over even d and extends up to the greatest gap $G(x)$ between two consecutive primes smaller than x. For $G(x)$ we will use the Cramer's formula (33). To get rid

of the product $\mathfrak{S}(d)$, we will make use of the (13) and we obtain:

$$\sum_{p<x}\frac{1}{p} = M' + \frac{2}{3} + 2\sum_{d=2}^{G(x)}\frac{1}{d}e^{-d/\log(x)} = M' + \frac{2}{3} + \sum_{k=1}^{\frac{1}{2}G(x)}\frac{1}{k}q^k, \quad q = e^{-2/\log(x)}. \tag{56}$$

Expanding $\log(1-q)$, where $0 < q < 1$, into the series we obtain

$$\sum_{k=1}^{n}\frac{1}{k}q^k = -\log(1-q) + \int_0^q \frac{u^n}{u-1}du. \tag{57}$$

For large x the term with logarithm goes into:

$$\log(1 - e^{-2/\log(x)}) = -\log(\log(x)) + \log(2) + \mathcal{O}(1/\log(x)). \tag{58}$$

Now, by the weighted mean value theorem we calculate the integral:

$$\mathcal{I} = \int_0^q \frac{u^n}{u-1}du = \frac{1}{(\theta q - 1)}\frac{q^{n+1}}{(n+1)}, \quad 0 < \theta < 1. \tag{59}$$

But $q = \exp(-2/\log(x)) < 1$ and:

$$|\frac{1}{\theta q - 1}| < \frac{1}{1-q} = \frac{e^{2/\log(x)}}{e^{2/\log(x)}-1} < \frac{\log(x)}{2}e^{2/\log(x)} = \mathcal{O}(\log(x)). \tag{60}$$

For $x \gg 1$ we have on the virtue of the Cramer conjecture that in our case $n \sim \frac{1}{2}\log^2(x)$, thus:

$$|\mathcal{I}| = \mathcal{O}(1/x\log(x)). \tag{61}$$

Finally we obtain from (50) and (55):

$$\sum_{p<x}\frac{1}{p} = \log(\log(x)) + M' + \frac{2}{3} - \log(2) + \mathcal{O}(1/\log(x)) \tag{62}$$

Because 2/3 is practically equal to $\log(2)$ to require consistency with the Merten's theorem, we have to postulate that $M' \approx M$. The comparison of the Mertens estimation for $\sum_{p<x}1/p$ with data obtained by a computer is shown in Fig.10. By the separate run of the computer program we have checked that up to 1.4×10^{14} there are almost 550000 sign changes of the difference $\sum_{p<x}\frac{1}{p} - \log(\log(x)) - M$; the first sign change appears at $5,788,344,558,967$.

7 First occurrence of a given gap between consecutive primes

In this section we will present the heuristical reasoning leading to the formula for the first appearance of a given gap of length d, see e.g. [24], [8], [47], [31].
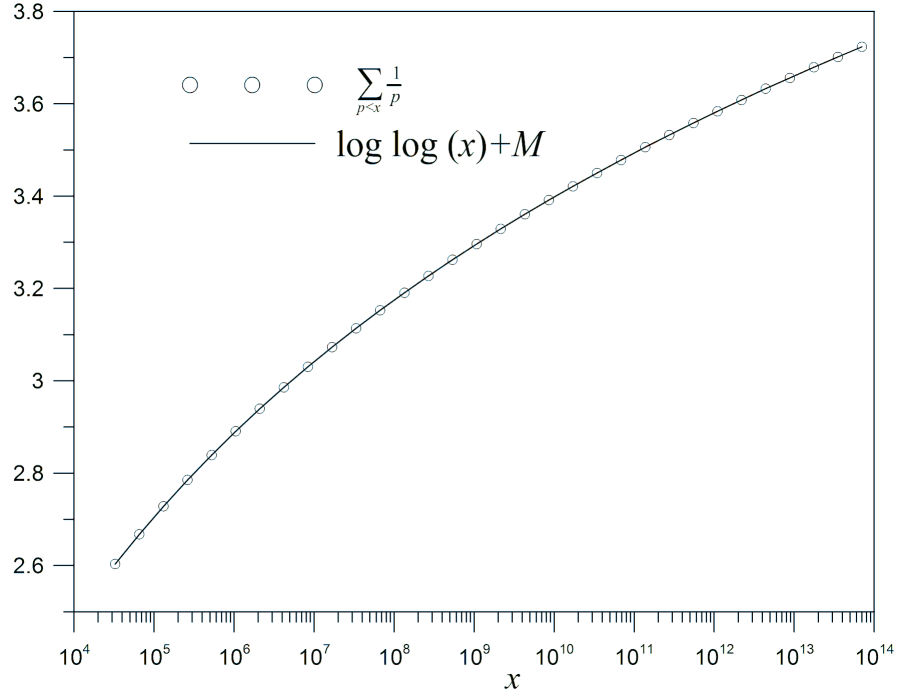
Figure 10: The plot of the prime harmonic sum up to $x = 2^{15}, 2^{16}, \ldots, 2^{46}$ and the Merten's approximation to it. The original of this figure has y axis of the length 8 cm and spans the interval (2.5, 3.8), so if the x axis would be plotted in the linear scale instead of logarithmic, then it should be $5.33(3) \times 10^9$ km long — that is the size of the Solar System.

We will use the conjecture (50) to estimate the position of the first appearance of a pair of primes separated by a gap of the length d. More specifically, let:

$$p_f(d) = \begin{cases} \text{minimal prime, such that the next prime } p' = p_f(d) + d \\ \infty \text{ if there is no pair of primes } p_{n+1} - p_n = d. \end{cases} \tag{63}$$

It is not known whether gaps of arbitrary (even) length exist or not, in other words the answer to the question: Is it true that for every d there is $p_f(d) < \infty$? is unknown [8].

We can obtain the heuristic formula for $p_f(d)$ by remarking that the finite approximations to the generalized Brun's constants are for the first time different from zero at $p_f(d)$ and then they are equal to $2/p_f(d)$:

$$\frac{4c_2}{d} \prod_{p|d} \frac{p-1}{p-2} e^{-d/\log(p_f(d))} = \frac{2}{p_f(d)}. \tag{64}$$

Referring to the argument that on average $\mathfrak{S}(d)$ is equal to $1/c_2$, we skip $\mathfrak{S}(d)$ and c_2. Neglecting the $\log(2) = 0.69314\ldots$, we end up with the quadratic equation for $t = \log(p_f(d))$:

$$t^2 - t\log(d) - d = 0$$

The positive solution of this equation gives:

Conjecture 7

$$p_f(d) = \sqrt{d} \; e^{\frac{1}{2}\sqrt{\log^2(d)+4d}}. \tag{65}$$

The comparison of this formula with the actual available data from the computer search is shown in Fig. 11. Most of the points plotted on this figure come from our own search up to $2^{48} = 2.815\ldots \times 10^{14}$. First occurrences $p_f(d) > 2^{48}$ we have taken from http://www.trnicely.net and http://www.ieeta.pt/~tos/gaps.html. In the Fig.11 there is also a plot of the conjecture made by Shanks [40]:

$$p_f(d) \sim e^{\sqrt{d}}, \tag{66}$$

while from (65) for large d it follows that

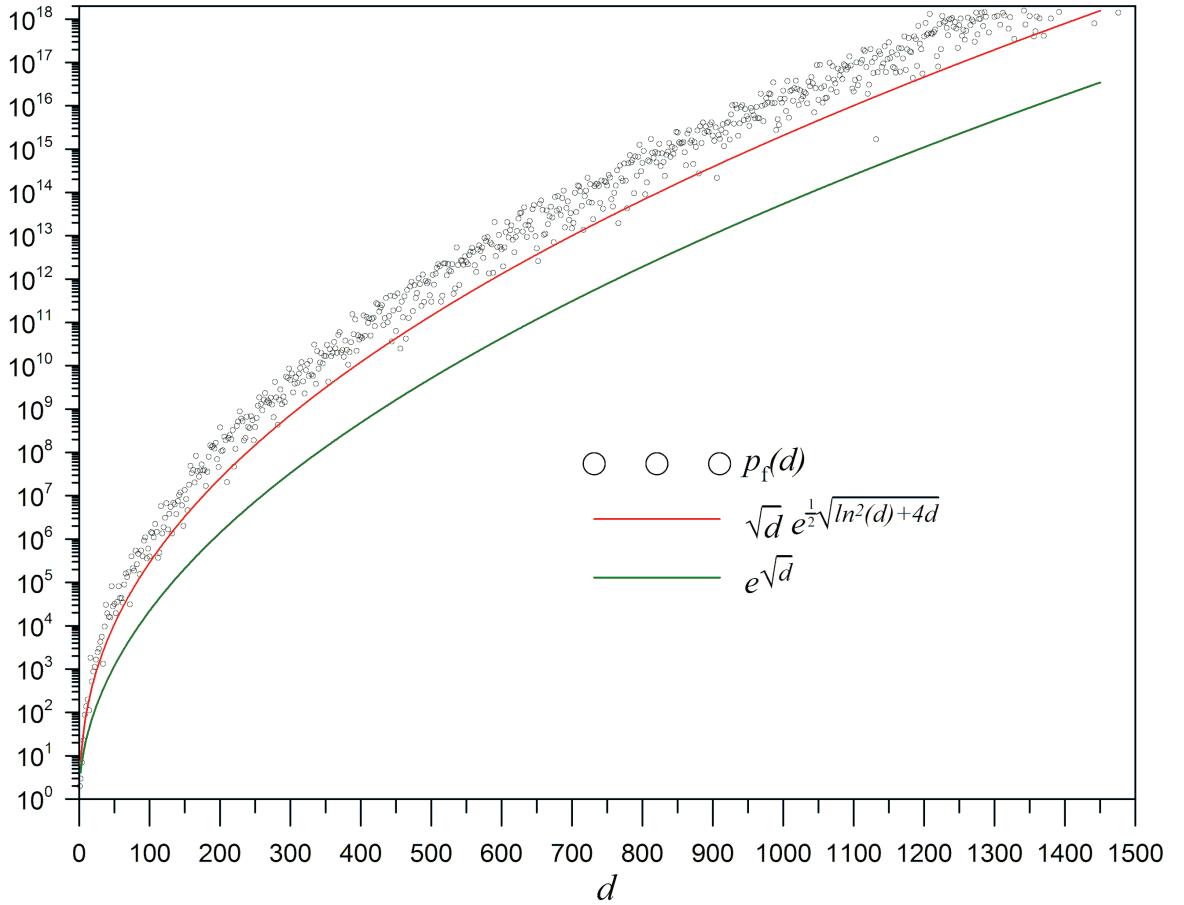$$p_f(d) \sim \sqrt{d} \; e^{\sqrt{d}}. \tag{67}$$

Figure 11: The plot of $p_f(d)$ and approximation to it given by (65) and (66).

8 The Andrica Conjecture

In the last section we will make use of most of the conjectures formulated so far. The Andrica conjecture [1] (see also [17, p. 21] and [37, p. 191]) states that the inequality:

$$A_n \equiv \sqrt{p_{n+1}} - \sqrt{p_n} < 1, \tag{68}$$

where p_n is the n-th prime number, holds for all n. Despite its simplicity it remains unproved. In Table II the values of A_n are sorted in descending order (it is believed this order will persist forever).

We have

$$\sqrt{p_{n+1}} - \sqrt{p_n} = \frac{p_{n+1} - p_n}{\sqrt{p_{n+1}} + \sqrt{p_n}} < \frac{d_n}{2\sqrt{p_n}}. \tag{69}$$

From this we see that the growth rate of the form $d_n = \mathcal{O}(p_n^\theta)$ with $\theta < 1/2$ will suffice for the proof of (68), but as we have mentioned in the Introduction, currently the best unconditional result is $\theta = 21/40$ [4].

For twins primes $p_{n+1} = p_n + 2$ there is no problem with (68) and in general for short gaps $d_n = p_{n+1} - p_n$ between consecutive primes the inequality (68) will be satisfied. The Andrica conjecture can be violated only by extremely large gaps between consecutive primes. Let us denote the pair of primes $< x$ comprising the largest gap $G(x)$ by $p_{L+1}(x)$ and $p_L(x)$, hence we have

$$G(x) = p_{L+1}(x) - p_L(x). \tag{70}$$

Thus we will concentrate on the values of the difference appearing in (68) corresponding to the largest gaps and so let us introduce the function:

$$R(x) = \sqrt{p_{L+1}(x)} - \sqrt{p_L(x)}. \tag{71}$$

Then we have:

$$A_n \leq R(p_n). \tag{72}$$

TABLE **II**

n	p_n	p_{n+1}	d_n	$\sqrt{p_{n+1}} - \sqrt{p_n}$
4	7	11	4	0.6708735
30	113	127	14	0.6392819
9	23	29	6	0.5893333
6	13	17	4	0.5175544
11	31	37	6	0.5149982
2	3	5	2	0.5040172
8	19	23	4	0.4369326
15	47	53	6	0.4244553
46	199	211	12	0.4191031
34	139	149	10	0.4167295

For a given gap d the largest value of the difference $\sqrt{p + d} - \sqrt{p}$ will appear at the first appearance of this gap: each next pair $(p', p' + d)$ of consecutive primes separated by d will produce smaller difference (see (69)):

$$\sqrt{p' + d} - \sqrt{p'} < \sqrt{p + d} - \sqrt{p}. \tag{73}$$

Hence, we have to focus our attention on the first occurrences $p_f(d)$ of the gaps. Using the conjecture (67) we calculate

$$\sqrt{p_f(d)+d} - \sqrt{p_f(d)} = \sqrt{\sqrt{d}e^{\sqrt{d}}+d} - \sqrt{\sqrt{d}e^{\sqrt{d}}} =$$
$$\sqrt{\sqrt{d}e^{\sqrt{d}}}\left(\sqrt{1+\frac{d}{\sqrt{d}e^{\sqrt{d}}}} - 1\right) = \frac{1}{2}d^{\frac{3}{4}}e^{-\frac{1}{2}\sqrt{d}} + \dots. \tag{74}$$

Substituting here for d the maximal gap $g(x)$ given by the Coniecture 4 (31) we obtain the approximate formula for $R(x)$:

Conjecture 8
$$R(x) = \frac{1}{2}g(x)^{3/4}e^{-\frac{1}{2}\sqrt{g(x)}} + error\ term. \tag{75}$$

The comparison with real data is given in Figure 12.

The maximum of the function $\frac{1}{2}x^{\frac{3}{4}}e^{-\frac{1}{2}\sqrt{x}}$ is reached at $x = 9$ and has the value $0.57971\dots$. The maximal value of A_n is $0.6708735\dots$ for $d = 4$ and second value is $0.6392819\dots$ for $d = 14$. Let us remark that $d = 9$ is exactly in the middle between 4 and 14.

Because in (75) $R(x)$ contains exponential of $\sqrt{g(x)}$, it is very sensitive to the form of $g(x)$. The substitution $g(x) = \log^2(x)$ leads to the form:

$$R(x) = \frac{\log^{3/2}(x)}{2\sqrt{x}}. \tag{76}$$

This form of $R(x)$ is plotted in Fig.12 in green. If we will use the guess $p_f(d) \sim e^{\sqrt{d}}$ (66) made by D. Shanks then we will get the expression:

$$\sqrt{p_f(d)+d} - \sqrt{p_f(d)} = \frac{1}{2}de^{-\frac{1}{2}\sqrt{d}} \tag{77}$$

instead of (74). Substitution here for d the form (32) leads to the curve plotted in Fig.12 in blue.

Finally, let us remark that from the above analysis it follows that

$$\lim_{n\to\infty}\left(\sqrt{p_{n+1}} - \sqrt{p_n}\right) = 0 \tag{78}$$

The above limit was mentioned on p. 61 in [15] as a difficult problem (yet unsolved).

9   Conclusions

We have formulated eight conjectures on the gaps between consecutive primes, in particular we have expressed maximal gap $G(x)$ directly by $\pi(x)$. The guessed formulas are well confirmed by existing computer data. The proofs of them seem to be far away and in conclusion we quote here the following remarks of R.Penrose from [33], p.422:
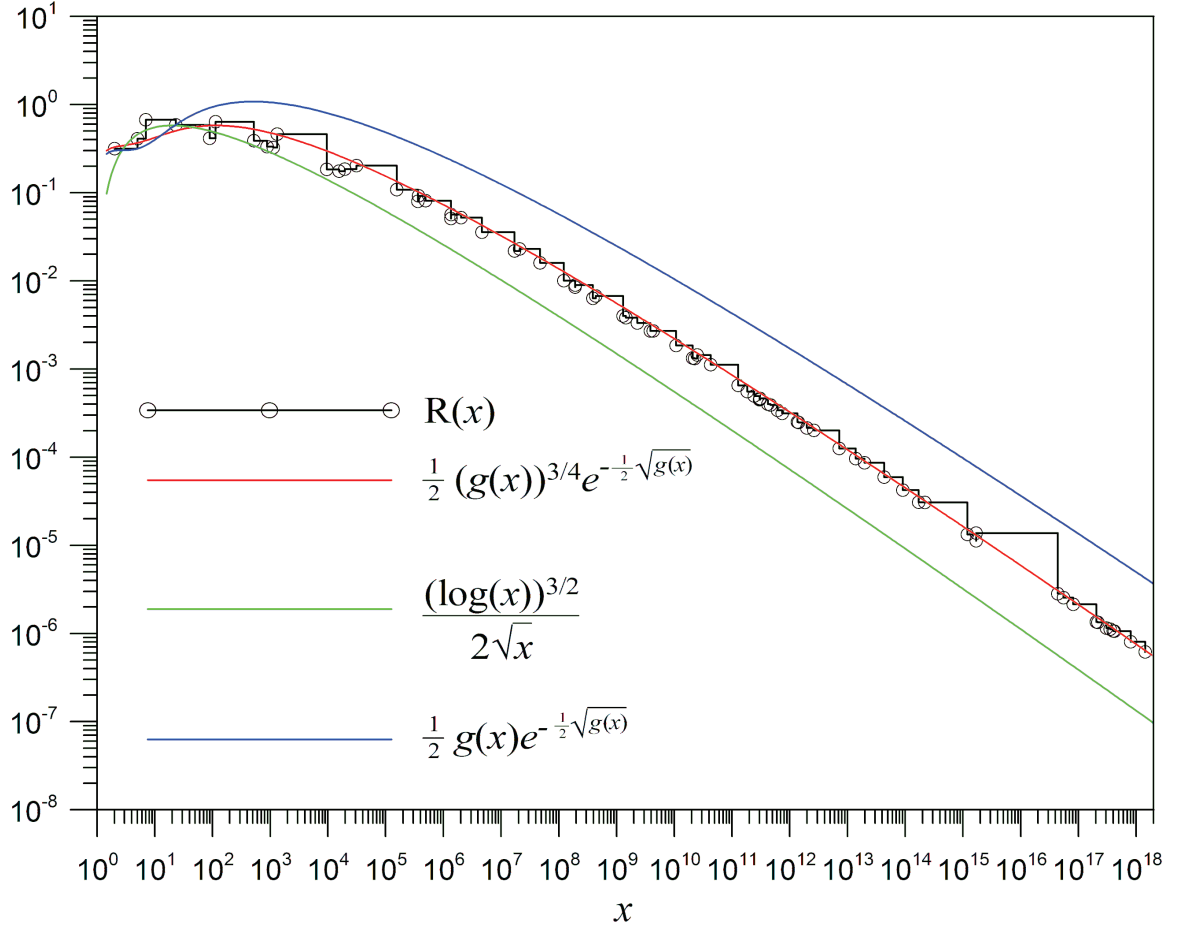
Figure 12: The plot of $R(x)$ and approximations to it given by (75), (76) and (77). The are 75 maximal gaps available currently and hence there are 75 circles in the plot of $R(x)$. To calculate $g(x)$ given by (31) we have used tabulated values of $\pi(x)$ available at the web sites www.trnicely.net and www.ieeta.pt/~tos/primes.html. There are over 50 crossings of our formula (75) with $R(x)$.

Rigorous argument is usually the last step! Before that, one has to make many guesses, and for these, aesthetic convictions are enormously important — always constrained by logical arguments and known facts.

References

[1] D. Andrica. Note on a conjecture in prime number theory. *Studia Univ. Babes-Bolyai Math.*, 31:44–48, 1986.

[2] R. Arenstorf. There are infinitely many prime twins, 26–th May 2004. `http://arxiv.org/abs/math/0405509v1`.

[3] R. Baker and G. Harman. The difference between consecutive primes. *Proc. Lond. Math.Soc. III*, Ser. 72, No.2:261–280, 1996.

[4] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes, II. *Proc. London Math. Soc.*, 83(3):532–562, 2001.

[5] E. Bombieri and H. Davenport. Small differences between prime numbers. *Proc. Royal Soc.*, A293:1–18, 1966.

[6] R. Brent. The distribution of small gaps between successive primes. *Mathematics of Computation*, 28:315–324, 1974.

[7] R. Brent. Irregularities in the distribution of primes and twin primes. *Mathematics of Computation*, 29:43–56, 1975. available from `http://wwwmaths.anu.edu.au/~brent/pd/rpb024.pdf`.

[8] R. Brent. The first occurrence of certain large prime gaps. *Mathematics of Computation*, 35:1435–1436, 1980.

[9] V. Brun. La serie $1/5 + 1/7 + ...$ est convergente ou finie. *Bull. Sci.Math.*, 43:124–128, 1919.

[10] J. H. Cadwell. Large intervals between consecutive primes. *Mathematics of Computation*, 25(116):909–913, 1971.

[11] L. Cherwell. Note on the Distribution of the intervals between prime numbers. *Quarterly Journal of Mathematics (Oxford)*, 17:46–62, 1946.

[12] H. Cramer. On the order of magnitude of difference between consecutive prime numbers. *Acta Arith.*, II:23–46, 1937.

[13] H. Davenport. *Multiplicative Number Theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1980. Revised by Hugh L. Montgomery.

[14] D. A. Goldston. *Large Differences Between Consecutive Prime Numbers*. PhD thesis, University of California, Berkeley, 1982. Available from www.math.sjsu.edu/∼goldston/thesis81.pdf.

[15] S. W. Golomb. Problem E2506: Limits of differences of square roots. *Amer. Math. Monthly*, 83:60–61, 1976.

[16] A. Granville. Harald Cramer and the distribution of prime numbers. *Scandanavian Actuarial J.*, 1:12–28, 1995.

[17] R. K. Guy. *Unsolved Problems in Number Theory*. Springer-Verlag, 2nd ed. New York, 1994.

[18] G. H. Hardy and J. E. Littlewood. Some problems of 'Partitio Numerorum' III: On the expression of a number as a sum of primes. *Acta Mathematica*, 44:1–70, 1922.

[19] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford Science Publications, 1980.

[20] D. R. Heath-Brown. Gaps between primes, and the pair correlation of zeros of the zeta-function. *Acta Arithmetica*, XLI:85–99, 1982.

[21] D. R. Heath-Brown. Differences between consecutive primes. *Jahresber. Deutsch. Math.-Verein*, 90:71–89, 1988.

[22] A. Ivìc. On Sums of Large Differences between Consecutive Primes. *Mathematische Annalen*, 241:1–9, 1979.

[23] J. Koreevar. Distributional Wiener–Ikehara theorem and twin primes. *Indag. Mathem., N.S.*, 16:3749, 2005. Available from `http://staff.science.uva.nl/~korevaar/DisWieIke.pdf`.

[24] L. J. Lander and T. Parkin. On First Apperance of Prime Differences. *Mathematics of Computations*, 21:483–488, 1967.

[25] S. Lou and Q. Yao. A Chebyshev's type of prime number theorem in a short interval. II. *Hardy-Ramanujan J.*, 15:1–33, 1992.

[26] H. Maier. Primes in short intervals. *Michigan Math. J.*, 42:221–225, 1985.

[27] H. Maier and C. Pomerance. Unusually large gaps between consecutive primes. *Trans. Amer. Math. Soc.*, 322:201–237, 1990.

[28] F. Mertens. Ein Beitrag zur analytischen Zahlentheorie. *Journ. Reine Ang. Math.*, 78:46–62, 1874.

[29] C. Mozzochi. On the difference between consecutive primes. *Journal Number Theory*, 24:181–187, 1986.

[30] T. Nicely. Enumeration to 1.6×10^{15} of the twin primes and Brun's constant. `http://www.trnicely.net/twins/twins2.html`.

[31] T. Nicely. New maximal prime gaps and first occurrences. *Mathematics of Computation*, 68:1311–1315, 1999.

[32] A. Odlyzko, M. Rubinstein, and M. Wolf. Jumping champions. *Experimental Mathematics*, 8(2):107–118, 1999.

[33] R.Penrose. *The Emperor's New Mind*. Oxford University Press, 1989.

[34] J. Pintz. Cramer vs. Cramer. On Cramers probabilistic model for primes. *Functiones et Approximatio*, XXXVII:361–376, 2007.

[35] J. Pintz. Landaus problems on primes. *Journal de thorie des nombres de Bordeaux*, 21:357–404, 2009.

[36] G. Polya. Heuristic Reasoning in the Theory of Numbers. *American Math. Monthly*, 66:375–384, 1959.

[37] P. Ribenboim. *The Little Book of Big Primes.* 2ed., Springer, 2004.

[38] M. Rubinstein. A simple heuristic proof of Hardy and Littlewood conjecture B. *Amer. Math. Monthly*, 100:456–460, 1993.

[39] B. Segal. Generalisation du theoreme de Brun. *Dokl. Akad. Nauk SSSR*, 1930:501–507, 1930.

[40] D. Shanks. On maximal gaps between successive primes. *Mathematics of Computation*, 18:646–651, 1964.

[41] D. Shanks and J. W. Wrench Jr. Brun's constant. *Mathematics of Computation*, 28(125):293–299, 1974.

[42] K. Soundararajan. The Distribution of Prime Numbers. In A. Granville and Z. Rudnick, editors, *Equidistribution in Number Theory*, volume 237 of *NATO Advanced Study Institute*, pages 59–84. Springer, 2007.

[43] K. Soundararajan. Small gaps between prime numbers: The work of Goldston-Pintz-Yildirims. *Bull. Amer. Math. Soc.*, 44:1–18, 2007.

[44] G. Tenenbaum. Re: Arenstorf's paper on the twin prime conjecture. NMBRTHRY@listserv.nodak.edu mailing list. 8 Jun 2004. `http://listserv.nodak.edu/cgi-bin/wa.exe?A2=ind0406&L=nmbrthry&F=&S=&P=1119`.

[45] M. B. Villarino. Mertens' proof of Mertens' theorem. *arXiv:math/0504289*, Apr 2005.

[46] M. Wolf. Random walk on the prime numbers. *Physica*, **A**250:335–344, 1998.

[47] J. Young and A. Potler. First occurrence prime gaps. *Mathematics of Computation*, 52:221–224, 1989.

[48] D. Zagier. The first 50 million prime numbers. *Mathematical Intelligencer*, 0:7–19, 1977.