Introduction to twin primes and Brun's constant computation

Pascal Sebah and Xavier Gourdon
numbers.computation.free.fr/Constants/constants.html

July 30, 2002

Abstract

An introduction to the set of twin primes, to a sieving method to count them and to the latest estimation of Brun's constant.

1 Introduction

It's a very old fact (Euclid 325-265 B.C., in Book IX of the *Elements*) that the set of primes is infinite and a much more recent and famous result (by Jacques Hadamard (1865-1963) and Charles-Jean de la Vallee Poussin (1866-1962)) that the density of primes is ruled by the law

$$\pi(n) \sim \frac{n}{\log(n)}$$

where the *prime counting function* $\pi(n)$ is the number of prime numbers less than a given integer n. This result proved in 1896 is the celebrated *prime numbers theorem* and was conjectured earlier, in 1792, by young Carl Friedrich Gauss (1777-1855) and in 1798 by Adrien-Marie Legendre (1752-1833) who studied the repartition of those numbers in published tables of primes.

This approximation may be usefully replaced by the more accurate *logarithmic integral* $Li(n)$:

$$\pi(n) \sim Li(n) = \int_2^n \frac{dt}{\log(t)}.$$

However among the deeply studied set of primes there is a famous and fascinating subset for which very little is known and has generated some famous conjectures: the *twin primes* (the term *prime pairs* was used before [5]).

Definition 1 *A couple of primes* (p, q) *are said to be twins if* $q = p + 2$*. Except for the couple* $(2, 3)$*, this is clearly the smallest possible distance between two primes.*

Example 2 $(3, 5), (5, 7), (11, 13), (17, 19), (29, 31), ..., (419, 421), ...$ *are twin primes.*

2 Counting twin primes

As for the set of primes the most natural question is wether the set of twin primes is finite or not. But unlike prime numbers for which numerous and elementary proofs exist [10], the answer to this natural question is still unknown for twin primes ! Today, this problem remains one of the greatest challenge in mathematics and has occupied numbers of mathematicians. Of course, as we will see, there are some empirical and numerical results suggesting an answer and most mathematicians believe that there are infinitely many twin primes.

In 1849, Alphonse de Polignac (1817-1890) made the general conjecture that there are infinitely many primes distant from $2k$. The case for which $k = 1$ is the twin primes case.

It's now natural to introduce the *twin prime counting function* $\pi_2(n)$ which is the number of twin primes smaller than a given n.

2.1 Numerical results

Using huge table of primes (Glaisher 1878 [5], before computer age, enumerated $\pi_2(10^5)$) and with intensive computations during modern period (Shanks and Wrench 1974 [12], Brent 1976 [1], Nicely 1996-2002 [9], Sebah 2001-2002 [11], see also [10]) it's possible to compute the exact values of $\pi_2(n)$ for large n and its conjectured approximation $2C_2Li_2(n)$ (see next section for the definition).

The following array includes the relative error ε (in %) between the approximation and the real value.

n	$\pi_2(n)$	$2C_2Li_2(n)$	ε
10	2	5	150.00
10^2	8	14	75.00
10^3	35	46	31.43
10^4	205	214	4.39
10^5	1224	1249	2.04
10^6	8169	8248	0.97
10^7	58980	58754	-0.38
10^8	440312	440368	0.013
10^9	3424506	3425308	0.023
10^{10}	27412679	27411417	-0.0046
10^{11}	224376048	224368865	-0.0032
10^{12}	1870585220	1870559867	-0.0013
10^{13}	15834664872	15834598305	-0.00042
10^{14}	135780321665	135780264894	-0.000042
10^{15}	1177209242304	1177208491861	-0.000064
10^{16}	10304195697298	10304192554496	-0.000031

At present time (2002), Pascal Sebah has reached $\pi_2(10^{16})$ and his values are confirmed by Thomas Nicely up to $\pi_2(4.10^{15})$ who used an independent approach and implementation.

2.1.1 Sieving twin primes

Because the most convenient (in fact the only available) way to compute $\pi_2(n)$ is to find all twin primes and just count them, it's of great importance to improve as much as possible such an algorithm. All known methods use variations on the historical *Eratosthenes sieve*.

In order to accelerate the sieve, a possible idea is to represent integers modulo a base m, so that any integer has the form $mk + r$ with $0 \leq r < m$. Primes numbers are such as m and r are relatively primes and for any value of m, there are $\phi(m)$ numbers r which are prime with m (this is the definition of Euler's ϕ *totient* function [6]).

Modulo 6 For example modulo 6 all integers have one of the form

$$6k, 6k+1, 6k+2, 6k+3, 6k+4, 6k+5$$

but $6k$ may not be prime, $6k+2$ and $6k+4$ are divisible by 2, $6k+3$ is divisible by 3, therefore primes (except 2 and 3) must be of the form

$$6k+1, 6k+5$$

or which is more convenient to sieve twin primes

$$(6k+5, 6k+7)\,.$$

This allows to sieve a proportion of only $\phi(m)/m = 2/6$ or 33.3% of all the numbers.

Modulo 30 The same kind of approach modulo 30 gives for candidates

$$30k+1, 30k+7, 30k+11, 30k+13, 30k+17, 30k+19, 30k+23, 30k+29$$

and here we need to sieve only $\phi(m)/m = 8/30$ or 26.66% of the integers to find all primes (except 2,3 and 5).

But remember that we are only trying to sieve twin primes hence are left only the candidate couples

$$(30k+11, 30k+13)\,, (30k+17, 30k+19)\,, (30k+29, 30k+31)$$

and the proportion drops to 6/30 or 20% of all integers.

This suggest to introduce the function $\phi_2(m)$ which is the number of pairs of integer $0 \leq r < m$ such as r and $r+2$ are relatively prime with m. We observe from the last two examples that

$$\phi_2(6) = 1, \phi_2(30) = 3.$$

Example Let's illustrate this on a numerical example. The enumeration of twin primes modulo 30 up to 10^{10} gives, respectively, for each of the 3 previous couples:

$$9137078, 9138015, 9137584$$

twin primes. So that (don't forget to count the two couples $(3, 5)$ and $(5, 7)$!)

$$\pi_2(10^{10}) = 9137078 + 9138015 + 9137584 + 2 = 27412679.$$

And, for the same couples, up to 10^{12}:

$$623517830, 623557143, 623510245$$

which produces

$$\pi_2(10^{12}) = 623517830 + 623557143 + 623510245 + 2 = 1870585220.$$

It's interesting to observe that the contribution to the enumeration of the twin primes of each couple is almost equivalent. This was also observed during all numerical estimations for other modulo like 210, 2310, 30030, ... [11].

This result is well known when enumerating just prime numbers but may be *conjectured for twin primes.*

Other modulo In this table we show the proportion $2\phi_2(m)/m$ of integer to sieve in order to count twin primes as a function of the modulo m:

m	$\phi_2(m)$	%
2	.	50.0
6	1	33.3
30	3	20.0
210	15	14.3
2310	135	11.7
30030	1485	9.9
510510	22275	8.7

The smallest ratios are obtained for values of m which are the product of the first primes $(2\# = 2, 3\# = 2 \times 3, 5\# = 2 \times 3 \times 5, 7\# = 2 \times 3 \times 5 \times 7, ...)$, that is the first values of the *primorial* # function. For example in [11], the sieves were made modulo 30030 and 510510, therefore less than 10% of the set of integers were considered by the algorithm. In some others implementations sieves modulo 6 or 30 are used.

2.2 Twin prime conjecture

Based on heuristic considerations, a law (the *twin prime conjecture*) was developed, in 1922, by Godfrey Harold Hardy (1877-1947) and John Edensor Littlewood (1885-1977) to estimate the density of twin primes.

According to the *prime number theorem* the probability that a number n is prime is about $1/\log(n)$, therefore, if the probability that $n+2$ is also prime was independent of the probability for n, we should have the approximation

$$\pi_2(n) \sim \frac{n}{\log^2(n)},$$

but a more careful analysis shows that this model is too simplified (an argument is given in [6]). In fact we have the following and more accurate conjecture (called *conjecture B* in[7]).

Conjecture 3 (Twin prime conjecture) *For large values of* n*, the two following equivalent approximations are conjectured*

$$\pi_2(n) \sim 2C_2 Li_2(n) = 2C_2 \int_2^n \frac{dt}{\log^2(t)} \tag{1}$$

or

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

Note that C_2 is the *twin prime constant* and is defined by

$$C_2 = \prod_{p \geq 3} \left(\frac{p(p-2)}{(p-1)^2} \right) = 0.6601618158468695739278121100145...$$

This new constant occurs in some asymptotic estimations involving primes and it's interesting to observe that it may be estimated using properties of the *Riemann Zeta function* to thousand of digits (Sebah computed it to more than 5000 digits).

Remark 4 *The function* $Li_2(n)$ *occuring in (1) may be related to the logarithmic integral* $Li(n)$ *by the trivial relation*

$$Li_2(n) = Li(n) + \frac{2}{\log(2)} - \frac{n}{\log(n)}.$$

2.2.1 Generalizations

In fact, Hardy and Littlewood made a more general conjecture on the primes separated by a gap of d. A natural generalization of the twin primes is to search for primes distant of $d = 2k$ (which should be infinite for any d according to Polignac's conjecture). The case $d = 2$ is the twin primes set, $d = 4$ forms the cousin primes set, $d = 6$ is the sexy primes set, ...

If we denote $\pi_d(n)$ the number of primes $p \leq n$ such as $p + d$ is also prime (observe that here p and $p + d$ may not be consecutive), Hardy-Littlewood's conjecture states (in [7]) that for $d \geq 2$:

$$\pi_d(n) \sim 2C_2 R_d \int_2^n \frac{dt}{\log^2(t)},$$

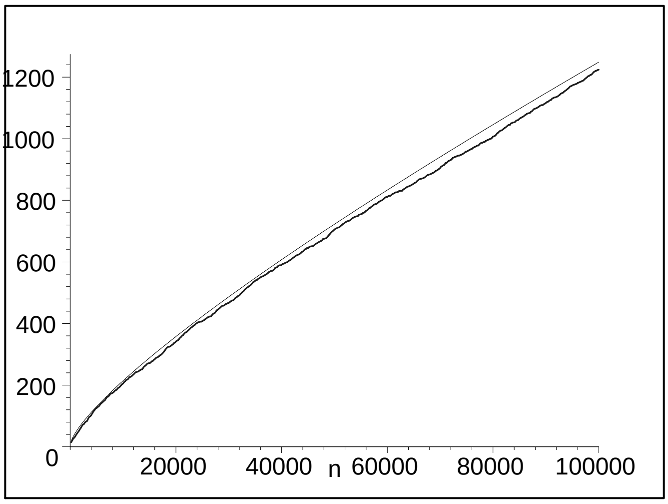

Figure 1: $\pi_2(n)$ and $2C_2Li_2(n)$

with $R_d \geq 1$ being the rational number

$$R_d = \prod_{p|d,p>2} \frac{p-1}{p-2}, \qquad p \text{ is prime.}$$

The first values of the function R_d are

d	2	4	6	8	10	12	14	16	18	20
R_d	1	1	2	1	4/3	2	6/5	1	2	4/3

According to this conjecture the density of twin primes is equivalent to the density of cousin primes. For example, the exact computed values up to 10^{12} are:

$$\pi_2(10^{12}) = 1870585220$$
$$\pi_4(10^{12}) = 1870585458,$$

which can be compared to the predicted value 1870559867 by the conjecture. Marek Wolf has studied the function

$$W(n) = \pi_2(n) - \pi_4(n)$$

and its fractal properties and approximate dimension of 1.48 ([13]).

3 Brun's constant

3.1 From Euler's constant to Brun's constant

Euler's constant It's very natural to understand the nature of the *harmonic numbers*

$$H(n) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... + \frac{1}{n}$$

when n becomes large and the sum takes all integers in account. We know since Euler, for instance, that

$$\left| \begin{array}{l} H(n) \sim \log(n) \\ H(n) - \log(n) \sim \gamma \end{array} \right.$$

so that the harmonic numbers tends to infinite like $\log(n)$. Note that γ is *Euler's constant* and may be evaluated to million of digits.

Mertens' constant The next step is to take in account only the primes numbers in the sum that is

$$P(p) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + ... + \frac{1}{p}$$

and we have the beautiful results that

$$\left| \begin{array}{l} P(p) \sim \log\left(\log(p)\right) \\ P(p) - \log\left(\log(p)\right) \sim M \end{array} \right.$$

Therefore the sum diverges (and this was also observed by Euler) but at the very low rate $\log\left(\log(p)\right)$ and M is the interesting *Mertens' constant* which may be evaluated to much less digits than γ, say a few thousands.

Brun's Constant In the last step we only take in account the twin primes less than p in the sum

$$B_2(p) = \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{1}{5} + \frac{1}{7}\right) + \left(\frac{1}{11} + \frac{1}{13}\right) + ...$$

and here comes the remarkable result due to Norwegian Mathematician Viggo Brun (1885-1978) in 1919 [2].

Theorem 5 *The sum of the inverse of the twin primes converges to a finite constant B_2.*

We write this result as

$$\lim_{p \to \infty} B_2(p) = B_2.$$

Note that this theorem doesn't answer to the question of the infinitude of twin primes, it just says that the limit exists (and may or may not contains a finite number of terms !). The proof is rather complex and based on a majoration of the density of twin primes ; a more modern one may also be found in [8].

Unlike Euler's constant or Mertens' constant, *Brun's constant* is one of the hardest to evaluate and we are not even sure to know 9 digits of it. By mean of very intensive computations, we only have guaranteed minorations !

3.2 Estimation of Brun's constant

3.2.1 Direct estimation

In the following table we have try to estimate this constant by computing the partial sums $B_2(p)$ up to different values of p.

p	$B_2(p)$
10^2	1.330990365719...
10^4	1.616893557432...
10^6	1.710776930804...
10^8	1.758815621067...
10^{10}	1.787478502719...
10^{12}	1.806592419175...
10^{14}	1.820244968130...
10^{15}	1.825706013240...
10^{16}	1.830484424658...

From this, we observe that the convergence is extremely slow and irregular. If we expect to find even just a few digits, we have to make some assumptions.

3.2.2 Extrapolation

An easy consequence of the twin prime conjecture is that we may write the numbers $B_2(p)$ as (see [4] and [9])

$$B_2(p) = B_2 - \frac{4C_2}{\log(p)} + O\left(\frac{1}{\sqrt{p}\log(p)}\right)$$

and thanks to this relation, the extrapolated value

$$B_2^*(p) = B_2(p) + \frac{4C_2}{\log(p)}$$

converges much faster to Brun's constant B_2.

Let's take a look to numerical values:

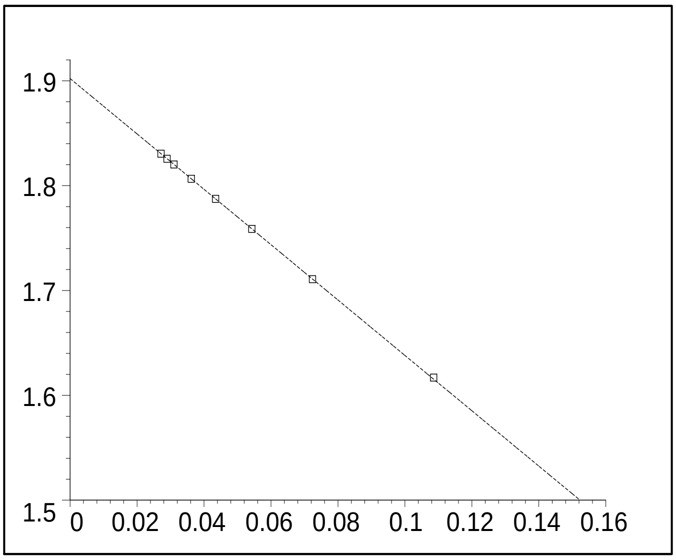

Figure 2: $B_2(p) = f\left(\frac{1}{\log(p)}\right)$

p	$B_2^*(p)$
10^2	1.904399633290...
10^4	1.903598191217...
10^6	1.901913353327...
10^8	1.902167937960...
10^{10}	1.902160356233...
10^{12}	1.902160630437...
10^{14}	1.902160577783...
10^{15}	1.902160582249...
10^{16}	1.902160583104...

which suggest that the value of B_2 should be around 1.902160583... (a similar value was first proposed by Nicely after intensive computations and checked later by Sebah, see [9] and [11]).

The relation

$$B_2(p) \sim B_2 - \frac{4C_2}{\log(p)}$$

invites us to draw the function $B_2(p) = f\left(1/\log(p)\right)$ (see Figure 2) which should be a line with a negative slop with a value near $-4C_2 \approx -2.64064726$.

The intersection of the line with the vertical axis (that is $p = \infty$) is Brun's constant if the twin prime conjecture is valid. And according to this line the direct estimation $B_2(p)$ should reach 1.9 not before the value $p \sim 10^{530}$ which is far beyond what could be reach with any computer!

4 Twin prime characterization

There is a result from Clement (1949, [3]) which permits to see if a couple $(p, p+2)$ is a twin primes pair. This theorem extends Wilson's famous theorem on prime numbers.

Theorem 6 *Let $p \geq 3$, the integers $(p, p+2)$ form a twin primes pair if and only if*

$$4\left((p-1)!+1\right) \equiv -p \ mod \ p(p+2)$$

Example 7 *For $p = 17$, $4\left((p-1)!+1\right) = 83691159552004 \equiv 306 \ mod \ 323$ and $-p \equiv 306 \ mod \ 323$, therefore $(17, 19)$ is a twin prime pair.*

The huge value of the factorial makes this theorem of no practical use to find large twin primes.

4.1 Large twin primes

Today, thanks to modern computers, a lot of huge twin primes are known. Many of those primes are of the form $k \times 2^n \pm 1$ because there are efficient primality testing algorithms for such numbers when k is not too large.

The following theorem due to the French farmer François Proth (1852-1879) may be used.

Theorem 8 (Proth's theorem - 1878) *Let $N = k.2^n + 1$ with $k < 2^n$, if there is an integer a such as*

$$a^{(N-1)/2} \equiv -1 \ mod \ N$$

then N is prime.

To help finding large pairs, an idea is to take a value for n and then to start a sieve in order to reduce the set of possible values for the k. It should take a few hours to find twin primes with a few thousands digits.

For example the following numbers are twin prime pairs (some are given from [10]):

459.2^{8529} Dubner, 1993

594501.2^{9999}

6797727.2^{15328} Forbes, 1995

697053813.2^{16352} Indlekofer & Janai, 1994

318032361.2^{107001} Underbakke & Carmody & ..., 2001

The last one is a twin primes pair of more than 32000 digits !

References

[1] R.P. Brent, *Tables Concerning Irregularities in the Distribution of Primes and Twin Primes Up to* 10^{11}, Math. Comput., (1976), vol. 30, p. 379

[2] V. Brun, *La série 1/5+1/7+1/11+1/13+1/17+1/19+1/29+1/31+1/41+1/43+1/59+1/61+..., où les dénominateurs sont nombres premiers jumeaux est convergente ou finie*, Bulletin des sciences mathématiques, (1919), vol. 43, p. 100-104 and p. 124-128

[3] P.A. Clement, *Congruences for sets of primes*, American Mathematical Monthly, (1949), vol. 56, p. 23-25

[4] C.E. Fröberg, *On the sum of inverses of primes and twin primes*, Nordisk Tidskr. Informationsbehandling (BIT), (1961), vol. 1, p. 15-20.

[5] J.W.L. Glaisher, *An enumeration of prime-pairs*, Messenger of Mathematics, (1878), vol. 8, p. 28-33

[6] G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford Science Publications, (1979)

[7] 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, (1922), vol. 44, p. 1-70

[8] W.J. LeVeque, *Fundamentals of Number Theory*, New York, Dover, (1996)

[9] T. Nicely, *Enumeration to* 10^{14} *of the Twin Primes and Brun's Constant*, Virginia J. Sci., (1996), vol. 46, p. 195-204

[10] P. Ribenboim, *The new Book of Prime Number Records*, Springer, (1996)

[11] P. Sebah, *Counting Twin Primes and estimation of Brun's Constant up to* 10^{16}, Computational project at *http://numbers.computation.free.fr/Constants/constants.html*, (2002)

[12] D. Shanks and J.W Wrench, *Brun's Constant*, (1974), Math. Comput., vol. 28, p. 293-299

[13] M. Wolf, *On the Twin and Cousin primes*, (1996), See *http://www.ift.uni.wroc.pl/~mwolf/*