Brun's Constant

The sum of reciprocals over the twin primes converges to a finite limit, known as Brun's Constant. Ed Rosenstiel decided to attempt the calculation on a micro, and made some interesting discoveries in the process.

> **$25,000 Prize**
>
> *Worldwide Computer Services is offering a $25,000 prize until 31 March 1987 to prove or disprove that there are infinitely many twin primes (the twin prime conjecture).*
>
> *Ed Rosenstiel's article illustrates how far down the road you can get with a micro today; previously the calculations shown have been done with minis and mainframes.*

One learns at school that the so-called harmonic series '1 + 1/2 + 1/3 + 1/4 + 1/5 + . . . and so on' diverges to infinity, *but so does* 1/2 + 1/3 + 1/5 + 1/7 + 1/11 + 1/13 + 1/17 . . . , that is, summing similarly but only over the primes.

Schur demonstrated this in a lecture in 1932 in Germany as follows:
Assume the contrary: that is, that the sum of the prime reciprocals converges to some limit, say, K.

Then, by a formula due to Euler, we have 1 + 1/2 + 1/3 + 1/4 + 1/5 + . . . + 1/n<(1 + 1/p_1 + 1/p_1^2+1/p_1^3+. .) * (1 + 1/p_2 + 1/p_2^2 + 1/p_2^3 +..)*...* (1 +1/p_m + 1/p_m^2 + 1/p_m^3 +..)
where the p_i on the right-hand side are just the m prime factors of all numbers from 1 to n.

A little bit of simple calculus then shows that *for all n:*

$$1 + 1/2 + 1/3 + 1/4 + 1/5 +..+ 1/n < \prod_{i=1}^{m} 1/(1 - 1/p_i) < \prod_{i=1}^{m} e^{2/p_i} < \exp[2*(1/2 + 1/3 + 1/5 + 1/7 + 1/11 + 1/13 + 1/17 + . . . \text{to infinity})] = e^{2K}$$

by the assumption, so the RHS is finite. Thus the sum of the reciprocals of all positive integers is also finite, which is false. Hence, so was the assumption. Therefore the sum of the reciprocals of *all* the primes also diverges to infinity!

Then Schur tantalised his audience by mentioning some of the problems connected with the so-called twin primes (3,5), (5,7), (11,13), (17,19), namely:
(i) it was an unsolved problem (*and still is!*), as to whether the list of twin primes ever ends; and
(ii) in 1919 Viggo Brun (who died only recently at the age of 92) stunned the mathematical world with a proof that the sum of reciprocals not over *all* the primes, but only over the twin primes (even if their number could be shown to be infinite) converges to a finite limit which is now known as **Brun's Constant** = say, S.

This much I remembered when, as part of a computer course at Birkbeck College in Pascal, I embarked on a project to calculate Brun's limit.

Writing a program in Basic to list twin primes and to evaluate the sums of their reciprocals is not difficult. The problem is that to find all the twins there is no other way but to compute almost *all* the primes, and this is a slow business on any computer. On a Commodore PET (since the machines operating Pascal were too busy most of the time), I went up to the last pair under 3020001, (later extended to 5000001), then made a

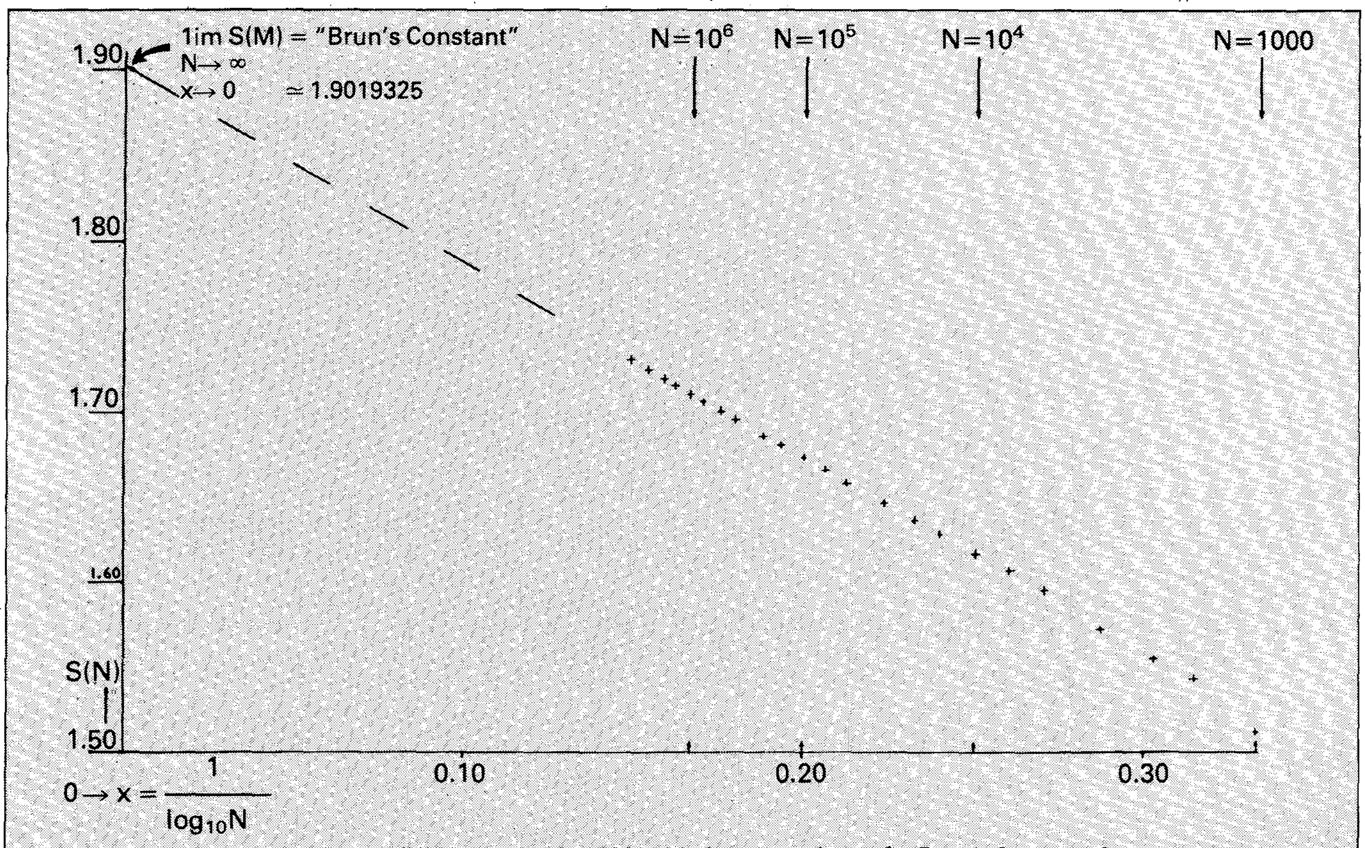

Fig 1 Graph of x = 1/log$_{10}$N, y = S(N) for N = 1000 to 5000001 gives an estimate for Brun's Constant S ≃ 1.9019325.

graph of necessity in logarithmic scale: that is, in powers of 10. It looked irregular at the lower end, but the gentle curve for the higher values looked promising and I also remembered that, according to Brun's Theorem, this curve would approach some horizontal line for very high values.

It seemed a good idea to eliminate the logarithmic scale, so I plotted 1/log N instead of N on the X-axis, and also left out the lower values under 10000 (Fig 1) and a *straight line* appeared.

It is remarkable in the wilderness of prime numbers, to come across an *apparently* straight line. Ignoring a professional mathematician's remark: '. . . if you take any kind of data and keep taking logs often enough, you will end up with a straight line. . .', my instinct told me this might be something original.

Using a TI-59 program which works out the least squares fit of a polynomial, I soon confirmed that I had found a much more accurate straight line than had it merely been deduced from a graph (Table 1a). And some extrapolations to values higher than those used for the least squares approximation were later found to agree with their computer counterparts to four significant digits!

Looking seriously at what was behind these findings, I decided to retrace the steps which had led me to such an extraordinary result: the 'gentle curve' prompted me to look for some closed mathematical expression to graph it and I had noticed that:

a) it was convex; and

b) it was asymptotic to a line parallel to the x-axis by Brun's Theorem, so I had thought of curves which might fit. By chance I had hit on the right answer straight away, namely on $y = S - 1/x$, the 'upside down hyperbola', although I had meant to consider also $y = S - 1/\exp(x)$ if $y = S - 1/x$ would not work.

The next step was to make a thorough literature search. Brun's Constant had indeed been calculated by several workers (3,4), and the most recent *probable* value given (4) was: $1.9021604 \pm 5 * 10^{-7}$

However, all the calculations had assumed that the famous conjecture made in 1923 by Hardy and Littlewood (6) is true. This says that the number of twin prime pairs up to some number X is closely approximated by:

$L_2(X) = 2c_2 *_2 \int^X dt/(1n\ t)^2 \sim 2c_2X/(1n\ X)^2$

that is, neglecting terms of order $X/(1nX)^3$, where $c_2 = 0.66016181\ ..$ is the 'twin prime' constant as given by Brent (4).

Furthermore, Brent estimates, making the assumption that twin primes are randomly distributed with density $2c_2/(1nx)^2$, (which implies that Brun's series is an infinite series):

that $1\text{im}\ S(n) - S(X) \sim 4c_2 * x\int^{\infty} dt/ n \to \infty\ t*(\text{I}nt)^2 = 4c_2/1nX$

which is the 'Straight Line Conjecture' that I had come up with on the PET (Table 1b), with $c_2 \simeq 0.25*k*1n10 = 0.6596417\ldots$

Does this show that, 60 years after two brilliant mathematicians had deduced a (so far unproven, but, in practice, very accurate) formula for the number of twin primes, by taking the opposite route, from the Straight Line Conjecture to the Hardy-Littlewood approximation, a mere tyro *could* have discovered this celebrated formula on a micro?

Computations

All computations were done on a Commodore PET with a simple program. These were cross-checked on a faster 'sieve' program which leaves out division by multiples of the first primes 2,3,5,7, and 11, and other checks were made against printouts of primes from a TI-59 calculator.

Most results were just copied from the VDU, but a complete printout of all twin primes less than 100000 allowed a manual count of 1224 in agreement with figures previously published by Brent (4). It was interesting to compare the calculation speed of the sieve with that of the simple program: it took the latter 25.3 days to reach the twins up to N = 1700000, while the sieve program needed only 12.2 days, a saving of ≃52%! (The sieve program took 54 days for a complete run up to N + 5000001.)

From the least squares fit (Table 1a) it will be seen that the value derived for S, on the assumption that the Straight Line Conjecture is true:

that Brun's Constant

$S = 1\text{im}\ S(N) = S(N) + k/\log N + \text{error}\ (N),\ N \to \infty$

is S = 1.90074 . .

which agrees with Brent (4) for three significant digits,

while from $k = 1.1396\ .. \simeq 4c_2 * 1n10$

we have $c_2 \simeq 0.6560\ldots$

However, there is something rather unsatisfactory in the above approach, where values below some arbitrarily chosen N are ignored for the extrapolation to S, and it is then observed that all higher values appear to lie on a straight line — not exactly, but to a high degree of 'accuracy'. (This mimics the quite different situation in physical experiments, where data is inevitably tainted due to to observational errors.) I was thus led to consider the question whether 'better' estimates for Brun's Constant might be obtainable by using a statistical approach to curve fitting.

With the help of the Applied Statistics Module for the TI-59 (7), I re-evaluated the results obtained, and also computed the correlation coefficient 'r'. Next I tried to *improve* 'r' by excluding in turn one value, arguing that because of the locally irregular distribution of primes one particular value might perhaps unduly influence the final result. As was not altogether surprising, the coefficient was improved by omitting either of the two *lowest* values for N, so I felt justified to omit both and to start calculating from N = 100001 upward, using higher values for S(N), which had come to hand. From Table 1b N = 734001 was omitted when calculating the final figures. These were: S = 1.901932526, k = 1.14591496., therefore $c_2 \simeq 0.6596417.$, where c_2 differs by 0.079%, S by 0.012% from the published results already mentioned. (The correlation coefficient was: r = 0.9999908.).

Conclusion

What I called the Straight Line Conjecture is not new, but during simple micro computations it suggested itself in a most obvious way; yet there was no hint about how to estimate *independently* the errors with these methods. If one uses the most recently published estimates for S and c_2 to calculate error terms for each N of Table 1b; that is, $\text{error}\ (N) = S - S(N) - 4c_2/_{1nN}$, then by a simple calculator exercise we have: $|\text{error}(N)| < 2/_N{}^{0.66}$, so k/1nN dominates the approximation.

An essential difference between Brun's and other converging series is seen when comparing it with Gregory's well-known series (which was also discovered independently by Leibniz): $\pi = 4[1 - 1/3 + 1/5 - 1/7 +-\ ..\ - 1/(2n-1)] + 1/n + \text{error}(N)$, where the error consists of terms of the form $\text{constant}/_n(2k+1)$ with $k>0$.

Now the square-bracket expression converges to $\pi/4$ with any desired number of decimals, (although much too slowly without the correction 1/n to be of any practical use), provided that a sufficient number of terms is computed (8). To show that the same is true for Brun's series still requires proofs of conjectures of one kind or another, even if better estimates were obtained for Brun's Constant by the use of more powerful computers. It will be remembered that to determine S to only three significant figures by computing its partial sums, requires a program to 'look' at *all* prime numbers up to 10^{1000}.

Until new theories are discovered, one can still only make 'plausible' estimates, — however well these might seem to fit with computation carried out so far.

Thus the mysteries of Brun's series still beckon: only one of the many unsolved problems of The Theory of Numbers.

It is not known whether Brun's converging series S = 1/3 + 1/5 + 1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + . . . + . . has an infinite number of terms, but if so then it probably converges *very slowly* indeed with the largest error term ≃ 2.64/1nN. This has been compared with Gregory's infinite series for π which has as largest error term 1/N, thus converging too slowly for practical computation, but still much faster than Brun's series. A more well-behaved series (although a rather trivial example) is the geometric series $G = 2 = G(N) + 1/_{2^n}$ with $G(N) = (1 + 1/2 + 1/4 + 1/8 +\ ..\ + 1/_{2^n})$ where the error term is exactly $1/_{2^n}$ and convergence is correspondingly fast.

Table 1a
Plotting S(N) against $\log_{10}N$

N	$\log_{10}N$	S(N)
51	1.708	1.2700
71	1.851	1.3032
101	2.004	1.3310
151	2.179	1.3969
201	2.303	1.4286
301	2.479	1.4602
501	2.700	1.4861
701	2.846	1.5061
1001	3.000	1.5180
1501	3.176	1.5426
2001	3.300	1.5549
3001	3.477	1.5722
5001	3.699	1.5947
7001	3.845	1.6067
10001	4.000	1.6169
15001	4.176	1.6279
20001	4.301	1.6359
30001	4.477	1.6462
50001	4.699	1.6585*
70001	4.845	1.6652*
100001	5.000	1.6728*
150001	5.176	1.6806*
200001	5.301	1.6858*
350001	5.544	1.6953*
500001	5.699	1.7007*
734001	5.866	1.7064
1020001	6.009	1.7111
1142001	6.058	1.7127
1420001	6.152	1.7156
1500001	6.176	1.7164
1700001	6.230	1.7180
1800001	6.255	1.7188
2000001	6.301	1 7201
3020001	6.480	1.7251

......

RESULTS:

From the starred values by the TI-59 pakette(2) program:

$S \simeq 1.90074..$
$k = 1.139594148$
$c_2 \simeq 0.65600..$

Table 1b
Plotting S(N) against $1/\log_{10}N$
where $S(N) = \sum [1/p + 1/(p+2)]$
$p<N$, (p and p+2 prime)

N	$1/\log_{10}N$	S(N)	least squares fit to S(N)
100001	0.1999998263	1.67279958	1.672750.
150001	0.1931958674	1.68055034	1.680546.
200001	0.1886425074	1.68584216	1.685764.
350001	0.1803729262	1.69527377	1.695240.
500001	0.1754702774	1.70071693	1.7008585
734001*	0.1704827337	1.70642789	1.706574.
1020001	0.1664281031	1.71108006	1.711220.
1142001	0.1650800688	1.71268937	1.712765.
1420001	0.1625411382	1.71564571	1.715674.
1500001	0.1619146983	1.71635648	1.716342.
1700001	0.1605020716	1.71802810	1.718011.
1800001	0.1598651315	1.71877363	1.718741.
2000001	0.1587042065	1.72013171	1.720071.
3020001	0.1543208189	1.72513665	1.725094.
5000001	0.1492766778	1.73097675	1.730874.
....			
$2*10^{10}$	0.0970776709	----------	1.7906898.
$1*10^{99}$	0.0101010101	----------	1.8903576.
....			

RESULTS:

$S \simeq 1.9019325..$ [cf. Brent (4) who gives a probable value for S as: $1.9021604 \pm 5*10^{-7}$]

$k = 1.14591496$, hence

$c_2 \simeq 0.6596417...$ and $r = 0.9999908...$

where r is the correlation coefficient computed by the TI-59 Bivariate Data Transform Program ST-12 (6). (The starred value 734001 was not used for calculating these results,cf.p.7)

IN BOTH TABLES: $S = \lim_{N \to \infty} S(N)$

$S - S(N) \sim k/\log_{10}N$

$k \simeq 4c_2/\ln 10$ and $c_2 = 0.660161181$

[$\simeq$ is used for 'approximately equal to', ~ means 'asymptotically equal to' in the strict mathematical sense (cf.LeVeque (5)) and c_2 is the 'twin prime constant' as given by Brent (4).]

References

1 **C L Baker and F J Gruenberger,** *Primes in the thousandth million,* P-1245, The Rand Corporation, Santa Monica, 1958

2 **J Walston,** *Least-mean-square fit of a polynomial,* Texas Instruments Inc, 1978

3 **D Shanks and J W Wrench, Jr** *Brun's Constant,* Math Comp, v28, 1974 pp293-299

4 **R P Brent,** *Irregularities in the distribution of primes and twin primes,* Math Comp, v29, 1975, pp43-56

5 **W J Leveque,** *Fundamentals of Number Theory,* p134, Addison-Wesley Publishing Company, London, 1977.

6 **G H Hardy and J E Littlewood,** *Some problems of partitio numerorum,* Acta Math, v44, 1923, pp1-70

7 **Applied Statistics,** *Solid State Software Module,* Texas Instruments Inc, 1977

8 **C W Kilmister,** *Summing infinite series,* The Gresham Lectures 1983

With thanks to the staff of Birkbeck College, Department of Mathematics.

Ed Rosenstiel came to England from Germany in the 1930s. He was a practising dentist until his retirement in 1978, when he became a maths undergraduate at Birkbeck College, London.

END