Journal of Integer Sequences, Vol. 8 (2005),
Article 05.4.2

Twin Prime Statistics

Harvey Dubner
85 Long Hill Road
Oakland, NJ 07436
USA
harvey@dubner.com

Abstract

Hardy and Littlewood conjectured that the the number of twin primes less than x is asymptotic to $2C_2 \int_2^x \frac{dt}{(\log t)^2}$ where C_2 is the twin prime constant. This has been shown to give excellent results for x up to about 10^{16}. This article presents statistics supporting the accuracy of the conjecture up to 10^{600}.

1 Introduction

The Twin Prime Conjecture asserts that there are an infinite number of twin primes. Proving this conjecture remains one of the great unsolved problems in mathematics. In 1923, Hardy and Littlewood [3] conjectured that the number of twin primes less than x is

$$L_2(x) \sim 2C_2 \int_2^x \frac{dt}{(\log t)^2} \quad , \tag{1.1}$$

where

$$C_2 = \prod_{p \text{ prime}>2} \left(1 - \frac{1}{(p-1)^2}\right) = 0.6601618158\cdots \tag{1.2}$$

is the "twin-prime constant." The reasoning that leads to C_2 has been explained by many authors and is described by Ribenboim in [5]. Although this conjecture (1.1) has never been proved, it gives remarkably good results. For example, Crandall and Pomerance [2, p13] report that by actual count

$$\pi_2(2.75 \cdot 10^{15}) = 3049989272575,$$

while from eq.(1.1)

$$L_2(2.75 \cdot 10^{15}) \approx 3049988592860.$$

In this article we present statistics supporting the accuracy of conjecture (1.1) up to 10^{600}.

2 Method

When dealing with large numbers it is obvious that using exact counts of all twin primes eventually becomes impractical, so that incremental counts have to be substituted for total counts. Equation (1.1) becomes

$$L_2(x_0, x_0 + x) \sim 2C_2 \int_{x_0}^{x_0+x} \frac{dt}{(\log t)^2} \approx 2C_2 \frac{x}{(\log x_0)^2} \quad , \tag{2.1}$$

which assumes that x_0 is large enough so that $\log x$ is virtually constant over x.

To find the number of twin primes in x we actually count all the prime gaps in x. This is accomplished by processing segments of 256,000,000 numbers, first by sieving to eliminate all composite numbers with a prime factor less than 10^9, then by subjecting the remaining numbers to a Fermat test for primality. Those that pass the Fermat test are assumed to be prime even though there is a slight chance of a small number being composite. However, any such error would not have a significant effect on the twin prime count. See below for a discussion of this topic.

We started collecting gap counts at $x_0 = 10^{20}$. Finding 10^9 gaps, of which 28,667,923 were twins, took about 2 hours on a 2.0 GHz computer. This required the determination of the primality of about $23 \cdot 10^9$ odd numbers. The same gap count problem starting at $x_0 = 10^{50}$ took about 82 hours.

2.1 Pseudoprimes

A number which passes a base b Fermat test might not be prime but could be psp, a base b pseudoprime. There have been various studies estimating psp frequencies but the results are asymptotic in nature and usually do not give useful information for relatively small numbers. We investigated psp's near 10^{19} because we could use 64-bit arithmetic to sieve such numbers so as to accurately and efficiently determine the composite numbers. These were all subjected to a base 3 Fermat test to determine which were psp's.

$2000 \cdot 10^9$ consecutive odd numbers near 10^{19} were tested and 17 psp's were found. Thus the probability of an odd number near 10^{19} being a psp is about

$$\text{Prob(psp)} = 17/(2000 \cdot 10^9) = 8.5 \cdot 10^{-12} \quad .$$

Applying the psp probability for 10^{19} to the $23 \cdot 10^9$ odd numbers above 10^{20} gives an expected error of

$$\text{expected error} = (23 \cdot 10^9) \cdot (8.5 \cdot 10^{-12}) = 0.20$$

and so we would expect to have at most one extra twin prime gap appearance in error when counting 10^9 gaps. Actually the error is considerably less than that shown since the sieving process would have eliminated most of the psp's. For larger values of 10^n the number of expected psp's continues to decrease, therefore it is clear that using a Fermat test to determine twin prime statistics is well justified.

Insisting on absolute accuracy in determining primes would have increased the computer time by a factor of at least 1000, making this project virtually impossible.

2.2 Poisson distribution

Although the exact distribution of counts of twin primes is not known, it seems reasonable to assume that such counts might be approximated by a Poisson probability distribution since this is true for almost all distributions of rare phenomena. We could then present the error between the actual and estimated twin prime counts as the number of standard deviations which effectively normalizes the error. For the Poisson distribution the standard deviation is the square root of the twin prime count. Using a common probability distribution to describe twin prime counts is not new. For example, R. Brent used that approach in a 1975 paper [1], as did T.Nicely in 1999 in evaluating Brun's constant [4]. We are only using probability concepts to normalize the error in predicting twin prime counts.

3 results

The twin prime counts shown here were obtained as part of a larger project to obtain the frequencies of all prime gaps near 10^n for a wide range of n. For a given n, gap data was collected until the total of all gaps, g_t, equalled a selected limit such as 10^9. Because of the Prime Number Theorem this means that about $g_t \cdot \log(10^n)$ numbers have to be checked for primality. Substituting this into eq.(2.1),

$$L_2(10^n, 10^n + g_t \cdot \log(10^n)) \approx 2C_2 \frac{g_t \cdot \log(10^n)}{(\log(10^n))^2} = 2C_2 \frac{g_t}{\log(10^n)} \tag{3.1}$$

which is the predicted number of twin primes, shown in column four of Tables 1 and 2. The error measured in Standard Deviations,

$$SD = \frac{\text{error}}{\sqrt{\text{actual}}}$$

is shown in the last column.

It is clear from examining the data in Tables 1 and 2 that the Hardy Littlewood conjecture for the count of twin primes, eq. (1.1), is "accurate" up to primes of 600 digits. The error is consistantly of the order of the square root of the actual twin prime count. Unfortunately we do not yet have any theory for improving the error prediction.

4 Acknowledgements

We wish to thank Jens Kruse Andersen for the use of his excellent program for counting pseudoprimes near 10^{19}.

References

[1] R. P. Brent, Irregularities in the distribution of primes and twin primes, *Math. Comp.* **29**, (1975), 43–56.

[2] R. Crandall and C. Pomerance, *Prime Numbers: a Computational Perspective*, Springer-Verlag, New York, 2001.

[3] G. H. Hardy and J. E. Littlewood, Some problems on partitio numerorum, III: On the expression of a number as a sum of primes. *Acta Math.* **44**, (1923), 1–70.

[4] T. Nicely, Enumeration to $1.6 * 10^{15}$ of the twin primes and Brun's constant, (1999). **http://www.trnicely.net/twins/twins2.html**

[5] P. Ribenboim, *The Little Book of Bigger Primes*, second edition, Springer-Verlag, New York, 2004.

2000 *Mathematics Subject Classification*: Primary 11B05; Secondary 11A41.
Keywords: prime gaps, twin primes, twin prime constant.

(Concerned with sequences A001097, A001359, and A006512.)

Received May 3 2005; revised version received August 25 2005. Published in *Journal of Integer Sequences*, August 29 2005.

Return to Journal of Integer Sequences home page.

Table 1: Twin Prime data for 10^{20} to 10^{55}

	total	twin primes			percent	error
size	gaps	actual	predicted	error	error	in SD
10^{20}	10^9	28667923	28670463	-2540	-0.0089	-0.47
10^{21}	10^9	27297642	27305203	-7561	-0.0277	-1.45
10^{22}	10^9	26069405	26064057	5348	0.0205	1.05
10^{23}	10^9	24933403	24930837	2566	0.0103	0.51
10^{24}	10^9	23885814	23892052	-6238	-0.0261	-1.28
10^{25}	10^9	22942786	22936370	6416	0.0280	1.34
10^{26}	10^9	22060593	22054202	6391	0.0290	1.36
10^{27}	10^9	21243727	21237380	6347	0.0299	1.38
10^{28}	10^9	20482302	20478902	3400	0.0166	0.75
10^{29}	10^9	19773012	19772733	279	0.0014	0.06
10^{30}	10^9	19113536	19113642	-106	-0.0006	-0.02
10^{31}	10^9	18503641	18497073	6568	0.0355	1.53
10^{32}	10^9	17913954	17919039	-5085	-0.0284	-1.20
10^{33}	10^9	17372782	17376038	-3256	-0.0187	-0.78
10^{34}	10^9	16868189	16864978	3211	0.0190	0.78
10^{35}	10^9	16387161	16383121	4040	0.0247	1.00
10^{36}	10^9	15931851	15928035	3816	0.0240	0.96
10^{37}	10^9	15496561	15497547	-986	-0.0064	-0.25
10^{38}	10^9	15090215	15089717	498	0.0033	0.13
10^{39}	10^9	14707109	14702801	4308	0.0293	1.12
10^{40}	10^9	14334329	14335231	-902	-0.0063	-0.24
10^{41}	10^9	13982239	13985591	-3352	-0.0240	-0.90
10^{42}	10^9	13650736	13652601	-1865	-0.0137	-0.50
10^{43}	10^9	13337659	13335099	2560	0.0192	0.70
10^{44}	10^9	13029542	13032028	-2486	-0.0191	-0.69
10^{45}	10^9	12740605	12742428	-1823	-0.0143	-0.51
10^{46}	10^9	12471625	12465418	6207	0.0498	1.76
10^{47}	10^9	12201271	12200197	1074	0.0088	0.31
10^{48}	10^9	11948039	11946026	2013	0.0168	0.58
10^{49}	10^9	11703340	11702229	1111	0.0095	0.32
10^{50}	10^9	11464673	11468185	-3512	-0.0306	-1.04
10^{55}	10^9	10426568	10425623	945	0.0091	0.29

Table 2: Twin Prime data for 10^{60} to 10^{600}

	total	twin primes			percent	error
size	gaps	actual	predicted	error	error	in SD
10^{60}	10^8	956273	955682	591	0.0618	0.60
10^{70}	10^8	819450	819156	294	0.0359	0.32
10^{80}	10^8	716327	716761	-434	-0.0606	-0.51
10^{90}	10^8	636623	637121	-498	-0.0782	-0.62
10^{100}	10^8	572885	573409	-524	-0.0915	-0.69
10^{110}	10^8	520799	521281	-482	-0.0926	-0.67
10^{120}	10^8	477439	477841	-402	-0.0842	-0.58
10^{130}	10^8	440294	441084	-790	-0.1794	-1.19
10^{140}	10^8	409384	409578	-194	-0.0474	-0.30
10^{150}	10^8	382170	382272	-102	-0.0267	-0.16
10^{160}	10^7	36006	35838	168	0.4666	0.89
10^{180}	10^7	32107	31856	251	0.7818	1.40
10^{200}	10^7	28652	28670	-18	-0.0628	-0.11
10^{220}	10^7	26213	26064	149	0.5684	0.92
10^{240}	10^7	23638	23892	-254	-1.0745	-1.65
10^{260}	10^7	22281	22054	227	1.0188	1.52
10^{280}	10^7	20458	20478	-20	-0.0978	-0.14
10^{300}	10^7	19181	19113	68	0.3545	0.49
10^{320}	$5 \cdot 10^6$	9041	8959	82	0.9070	0.86
10^{340}	$5 \cdot 10^6$	8584	8432	152	1.7707	1.64
10^{360}	$5 \cdot 10^6$	8090	7964	126	1.5575	1.40
10^{380}	$5 \cdot 10^6$	7410	7544	-134	-1.8084	-1.56
10^{400}	$5 \cdot 10^6$	6997	7167	-170	-2.4296	-2.03
10^{450}	$11 \cdot 10^6$	14035	14016	19	0.1354	0.16
10^{500}	$7 \cdot 10^6$	8140	8027	113	1.3882	1.25
10^{550}	$5 \cdot 10^6$	5084	5212	-128	-2.5177	-1.80
10^{600}	$5 \cdot 10^6$	4734	4778	-44	-0.9294	-0.64