Sprimes: A Sparse Prime-Like Sequence of Natural Numbers Generated by the Goldbach Conjecture

Ian R. Harris
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275
USA
iharris@smu.edu

Abstract

Using a conjecture closely related to the Goldbach Hypothesis, we can generate a set of natural numbers, which we term *sprimes* that like the primes, have both regular and chaotic properties, and are a good deal sparser than the primes. This paper explains how to generate these and explores some of their properties.

1 Introduction

Suppose we define the set of primes to be P, and let $Q = \{1\} \cup P \backslash \{2\} = \{1, 3, 5, 7 \ldots\}$. A conjecture that is very similar to the Goldbach hypothesis on the primes is as follows

<u>Conjecture GH_Q</u>:
$\forall n \in \mathbb{N}\ \exists p, q$ such that $p, q \in Q$and $(p + q)/2 = n$.
(The set Q is chosen instead of the more conventional P as it fits in more nicely with the discussion and definitions to follow, but clearly the statement is very similar to the conventional Goldbach hypothesis.)
It is interesting to reverse the direction of the hypothesis, and seek sets of natural numbers $G \subseteq \mathbb{N}$ which are known by construction to satisfy the hypothesis. We will define a set G a *Goldbach Set* if it satisfies the *Goldbach Condition* as stated below:

<u>GC:</u>
$\forall n \in \mathbb{N}, \exists g_1, g_2 \in G$ such that $(g_1 + g_2)/2 = n$.

(Note that this definition is very similar to the concept of a Goldbach sequence as defined by Torelli [5]. A key difference is that averages of the members of our set are used to generate the natural numbers, rather than sums, and 0 is not allowed in constructing the averages.) Clearly it is easy to choose sets which satisfy GC. Trivial examples include $\{2k-1|k \in \mathbb{N}\}$ and $\{1+10k, 3+10k, 7+10k, 9+10k|k \in \mathbb{Z}^+\}$. None of the aforementioned Goldbach Sets exhibit complexity of behavior like Q. Indeed these sets are asymptotically far more dense than Q or P. One interesting questions is "does Q (and also P) only just satisfy the Goldbach Hypothesis?" In other words, would any smaller set (either a subset of Q or a less dense set than Q) also satisfy the hypothesis? If the answer is negative, this would imply proving the Goldbach Hypothesis may be harder than if the answer is affirmative.

This paper describes a set S, which we term *sprimes*, to draw analogy with the primes and also to reflect the fact that they are sparser (more spread out) than P (or Q). The set appears to suggest that P is denser than needed to satisfy the Goldbach Hypothesis and hence raises the possibility that the hypothesis might be satisfied by some suitable subset of P.

The set $S = \{s_i|i \in \mathbb{N}\}$ is generated sequentially as follows:

1) $s_1 = 1$

2) Given $\{s_1, \ldots s_k\} = S_k$, form the set $A = \{n \in \mathbb{N}|\exists s_i, s_j \in S_k, (s_i + s_j)/2 = n\}$.

3) Let $m = \min(\mathbb{N}\backslash A)$. This number is the smallest natural number which cannot be formed by averages of the current finite list of Sprimes.

4) Create a candidate set $C_k = \{n \in \mathbb{N}|n > s_k, \exists s_i \in S_k \quad (s_i+n)/2 = m\}$ of possible choices for s_{k+1}.

5) Evaluate the worth w_i of each member $i \in C_k$ by $w_i = |(\{(i+s_j)/2|s_j \in S_k\} \cup \{i\}) \cap \{\mathbb{N}\backslash A\}|$. (This assigns to each candidate a worth which is equal to the number of new values that will be added to set A if the candidate is chosen for S).

6) Choose for s_{k+1} the largest candidate with the highest worth, that is $s_{k+1} = \max\{i \in C_k|w_i = \max_{j \in C_k} w_j\}$.

Note that one could after step 4, choose the smallest natural number in the set, but this does not generate a sparse set, and the resulting set $\{1, 3, 5, \ldots\}$ is regular and uninteresting. The first 500 sprimes are listed in two tables in the appendix. Sprimes that are also prime (actually in Q) are in bold. Although the process described above that generates the sprimes is not a conventional sieve, in which arithmetic progressions of numbers are eliminated from $\mathbb{N}$, the process has similarities with the sieve methods. Note it is also possible to start with 3 instead of 1, which generates $R = \{r_i|i \in \mathbb{N}\}$, such that $r_i = s_i + 2$. In addition, by slightly altering the procedure for generating the set A, it is possible to start the sequence with 2 and 3, to make it more closely resemble the primes. Neither of these two changes in starting condition are important, in the sense that the sets generated exhibit the same general behavior as the sprimes.

Torelli [5] suggested a similiar construction using the concept of a "basis", in which a set A is a basis for the natural numbers if any natural number can be written as a sum of members of the set A. A Goldbach sequence as defined by Torelli is then such a basis written in increasing order, without the element 0 (see also Gunturk and Nathanson [1]). In this paper we only insist that sums of members of a set A cover the even positive integers (that is averages would cover all the natural numbers), and 0 is not allowed in any sum. These conditions are essentially the same as the conditions classically stated for the Goldbach Hypothesis. "Lucky" numbers (Hawkins and Briggs [3]) and "random primes" (Hawkins[2]) also appear to be Goldbach sets in the sense described in this paper, but there is no known proof of this assertion, only empirical evidence. Heyde [4] shows that numbers generated by the random Hawkins sieve almost surely satisfy the Riemann hypothesis.

2 Some properties and conjectures for the sprimes motivated by the primes

1) By construction, $|S| = \infty$, which is true also of P and Q, as proven by Euclid.

2) The members of Q can be generated iteratively by the sieve of Eratosthenes. The sprimes S are generated by a sieve-like process.

3) There is strong empirical evidence that Q is a Goldbach set (that is GH is true). By construction S is a Goldbach set.

4) There is empirical evidence that there are an infinite number of twin primes, ie $\{n, m \in Q | m = n + 2\}| = \infty$. It is not known if S has this property, but we note there are 8 pairs of twin sprimes in the first 500, namely {(1,3),(11,13),(77,79),(115,117),(3791,3793), (6855,6857),(14551,14553), (31231,31233)}.

5) The prime number theorem states that $\pi(n) \sim n/\ln(n)$. More accurately if one defines $Li(n) = \int_2^n \frac{1}{\ln t} dt$,

then $\pi(n) \sim Li(n)$. Empirical evidence (see section 3) suggests that the sprime counting function $\bar{\sigma}(n) \sim n^{0.5709}$. However, this is very misleading as a power function in n also provides a very accurate approximation to $\pi(n)$ for "small" n. As shown in section 3, there is an apparent linear relationship between $\log(s_k)$ and $\log(p_k)$, which if true for larger n, would imply $\bar{\sigma}(n) \sim \frac{n^{0.763}}{0.763 \log n}$, or some such similar relationship.

6) (Bertrand's postulate, proved for P by Chebyshev). $\forall n \geq 2, \exists p \in P$ st $n \leq p \leq 2n$. This property is true for any Goldbach set, hence is true for S.
7) $\sqrt{p_{n+1}} - \sqrt{p_n} < 1$ (Andrica's conjecture). Numerical evidence is strongly in favor of this for primes. For sprimes, the conjecture is false, although it is possible that the result is true with a higher bound. Note that $\sqrt{s_{316}} - \sqrt{s_{315}} = \sqrt{23973} - \sqrt{23145} = 2.7$.
8) (Sophie Germain conjecture). It is conjectured that there are infinite n such that $(n, 2n+1) \in P$. In the first 500 sprimes, there are 8 such pairs: {(1,3),(3,7),(13,27),(117,235),(5421,10843), (6159,12319), (12565,25131),(25523,51047)}.
9) It seems reasonable that $|Q \cap S| = \infty$, although like other results of this type, the proof is probably hard. A related question is whether the density of the primes amongst the sprimes is what one would expect if they were "distributed at random". Probablistic investigations in section 3 suggest that sprimes are slightly more likely to be primes than non-sprime odd integers.
10) Does $\sum_{k=1}^{\infty} \frac{1}{s_k}$ converge? We know that the corresponding sum for primes diverges, but the sprimes are more spread out. Direct examination of a sum of a finite number of terms is of no use as if the sum diverges, it will do so very slowly. If indeed the sprimes counting function behaves as $\bar{\sigma}(n) \sim \frac{n^{0.763}}{0.763 \log n}$, then we would expect the sum to converge.

3 Probabilistic analysis:

Here we perform various statistical analyses of the sprimes, under assumptions that they can be modelled as random variables. This method has a long tradition in number theory.
A plot of the $\log p_k$ versus $\log s_k$ for the first 500 terms of each sequence reveals a near straight line relationship (correlation coefficient of 0.999), which implies that

$$\log s_k \approx \alpha + \beta \log p_k.$$

In fact a relationship with $\alpha = 0$ is possible. For simplicity, we will take this to be the case, and assume $p_k = s_k^{\gamma}$ for some power γ, and use the empirical relationship to derive the form of the counting function $\bar{\sigma}(n)$ for the sprimes.

We know $k = \pi(p_k) \approx p_k / \log p_k$, so substitution of the approximate form for s_k gives

$$k = \frac{s_k^\gamma}{\gamma \log s_k}.$$

Thus we can suggest

$$\bar{\sigma}(n) \approx \frac{n^\gamma}{\gamma \log n}.$$

Empirically we have $\gamma \approx 0.763$. It is of course possible that $\bar{\sigma}(n)$ follows some totally different form, and the empirical results are merely artifacts of the "law of small numbers". A reasonable empirical alternative is that $\bar{\sigma}(n) \approx n^{0.571}$.

The table below gives actual values for $\bar{\sigma}(n)$, along with values for the approximations $n^{0.763}/0.763 \log n$ and $n^{0.571}$ for various n.

n	$\bar{\sigma}(n)$	$n^{0.571}$	$n^{0.763}/0.763 \log n$
50	9	9.3	6.6
100	13	13.9	9.5
500	34	34.7	24.2
1000	50	51.6	36.9
5000	127	129.3	102.2
10000	190	192.1	160.4
15000	246	242.2	209.4
20000	287	285.4	253.1
25000	324	324.2	293.5
30000	360	359.7	331.4
40000	416	424.0	401.5
50000	479	481.6	466.2

The n power law seems a better fit to $\bar{\sigma}(n)$ than the law derived from the linear relationship, but again, this could easily be an artifact of the aforementioned law. Indeed the power law $(2/3)n^{0.81}$ is a better fit to the prime counting function $\pi(n)$ than the known correct asymptotic form $n/\log n$ for $n \approx 3000$.

Roughly speaking about $2/\log n$ of the odd integers up to n are prime, so applying this to the 500 sprimes terminating in $53,003$ we would expect about $2(500)/\log 53003 \approx 90$ of the first 500 sprimes to be prime. The actual number of prime-sprimes in the first 500 sprimes is 110, suggesting perhaps sprimes are slightly more likely to be prime than randomly chosen odd integers, but again this is likely to merely be related to the fact that the $n/\log n$ formula undercounts primes for small n.

We can see how many sprimes are divisible by 3,5,7 etc. The table below gives the proportion of such in the first 500 sprimes.

	Expected proportion	Actual proportion
3	33.33%	32.6%
5	20%	20%
7	14.2%	15.6%
11	9.1%	8.6%
13	7.7%	7.6%
17	5.9%	7.2%
23	4.3%	4.4%

Note, the expected proportion is calculated assuming a random sample from the odd integers. The observed proportions of multiples of 3,5 etc are what one would expect if the sprimes were randomly chose from the odd integers.

4 Conclusion:

This paper introduces the sprimes (denoted by S), which are specifically constructed to satisfy a Goldbach type property (namely that any even positive integer can be represented as a sum of two sprimes). The construction is recursive, where each member of the set S is determined by a simple algorithm based on the current members. The set S is constructed in such a way that the set is as sparse as possible in a certain sense. Empirically the sprimes follow as similar distribution law as the primes, but are sparser. This suggests that the primes might be redundant in satisfying Goldbach's conjecture, that is they are denser than is needed to form a basis for even positive integers.

Appendix 1:
The 1st to 500th sprimes.

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
1-10	**1**	**3**	**7**	**11**	**13**	27	**31**	35	49	**61**
11-21	77	**79**	93	**101**	115	117	133	**163**	183	187
21-30	**193**	235	245	**257**	**271**	279	323	335	343	381
31-40	399	**439**	481	497	507	535	549	**569**	**619**	669
41-50	681	693	713	**739**	815	833	**863**	905	**941**	973
51-60	**1033**	1053	1089	1119	1141	**1163**	**1181**	**1259**	1285	1317
61-70	1341	1401	1419	1431	1555	1565	1647	1685	1691	**1699**
71-80	**1747**	1841	1853	1875	1981	2001	2041	2051	2095	2149
81-90	2195	2259	**2281**	2299	**2339**	**2377**	2511	**2557**	2611	**2663**
91-100	2669	2781	2815	**2857**	2889	3027	**3041**	3087	3125	3129
101-110	3249	3309	3379	3501	**3559**	3577	3667	3675	3791	**3793**
111-120	**3853**	3933	4043	4055	4125	4181	4185	**4253**	4317	4389
121-130	4465	4539	**4583**	4689	4781	**4877**	4977	5069	5137	5215
131-140	5259	5307	**5387**	5421	**5443**	5497	5523	5661	5739	5747
141-150	5763	5837	5871	**5881**	5965	5977	6065	6153	6159	6235
151-160	6465	6535	6539	6855	**6857**	6889	**6907**	6957	6993	**7103**
161-170	7131	7295	7379	7389	7445	7735	7739	7939	7955	8133
171-180	8175	8207	8279	8403	8437	**8447**	8639	8687	**8747**	8775
181-190	8881	**8929**	**9049**	9095	9207	9393	9581	9673	9693	9745
191-200	10003	10073	10129	**10169**	10269	10299	10307	10385	10603	10717
201-210	10797	10813	10843	**10903**	10911	11079	11227	11269	11331	11461
211-220	**11807**	11877	12035	12085	12099	12175	12217	12319	12525	12565
221-230	12587	**12647**	12727	12811	12845	13005	**13163**	13371	13389	**13591**
231-240	13945	14117	**14153**	14157	14181	**14221**	14239	14305	14337	14409
241-250	14467	**14551**	14553	14799	14853	14981	15113	15385	15483	15533

	1	2	3	4	5	6	7	8	9	10
251-260	15743	15833	16167	**16223**	**16273**	16413	16557	**16889**	16915	17051
261-270	17197	17335	17407	17473	17481	17603	**17683**	17847	17875	**17971**
271-280	18101	**18379**	18403	**18661**	18695	18817	18827	**19073**	**19141**	19153
281-290	19223	19579	19731	19745	19757	19831	19883	20075	**20101**	20331
291-300	20355	20385	20403	20783	20927	20955	20991	21135	21339	21549
301-310	21665	21781	21935	22001	**22037**	22047	22165	**22397**	**22483**	22521
311-320	**22807**	22815	**22871**	23085	23145	23793	24067	24163	**24469**	24615
321-330	24627	24703	24769	**24841**	25131	25177	**25439**	**25523**	**25643**	25787
331-340	**25903**	26055	26063	26213	26291	**26431**	26491	26617	26659	**26693**
341-350	**26729**	26899	27035	27123	27377	27511	**27581**	27619	27693	28241
351-360	28569	28601	28805	29001	29045	**29221**	**29587**	29619	29957	29985
361-370	30179	30377	**30859**	31107	31157	**31231**	31233	31625	31945	31977
371-380	31987	32041	32109	32559	32667	32721	**32993**	33033	33299	33465
381-390	34125	34285	34379	34453	34477	34515	34569	34901	34915	35079
391-400	35629	35679	35763	35913	35929	35975	36021	**36209**	36225	**36599**
401-410	36665	37059	37125	37151	37179	37575	37631	37083	38063	**38609**
411-420	38703	**38833**	**39043**	39125	39291	39529	40041	40135	40233	40315
421-430	**40433**	40513	40573	40641	40811	**41161**	**41981**	**42023**	**42101**	42275
431-440	42319	42393	42445	42561	42865	43185	43269	43369	**43457**	43493
441-450	43719	43811	44077	44209	44329	**44357**	44767	44807	44933	45149
451-460	45311	**45329**	45479	45623	45695	45909	46085	46501	46545	**46633**
461-470	46969	47085	**47111**	47209	47685	47083	**48091**	48545	48795	48803
471-480	**48907**	49127	49259	**49417**	49513	49567	49593	49793	49929	50205
481-490	50295	50641	50669	50701	50985	51017	**51407**	51443	51495	**51631**
491-500	52199	**52379**	52389	52499	52513	52663	52749	52777	**52981**	**53003**

References

[1] Gunturk C. S. and Nathanson M. B. (2006) A new upper bound for additive bases. *Acta Arithmetica*, 124, 235-255.

[2] Hawkins D.(1958) The random sieve. *Mathematics Magazine*. 31, 1-3.

[3] Hawkins, D. and Briggs, W. E. (1958). The lucky number theorem. *Mathematics Magazine*. 31, 81-84.

[4] Heyde, C. C. (1978). A log log improvement to the Riemann hypothesis for the Hawkins random sieve. *Annals of Probability*, 6, 850-875.

[5] Torelli, M. (2006). Increasing integer sequences and Goldbach's conjecture. *RAIRO-Theoretical informatics and applications*, 40, 107-121.

2010 *Mathematics Subject Classification*: Primary 11A41.

Keywords: prime counting function, Goldbach conjecture, quasi-primes.

(Concerned with sequence A192358.)
