Sieving primes from composite pairs within sets of $6n \pm 1$ numbers

Adrian M. Stokes

11th August 2024

Abstract

The prime numbers ≥ 5 within a finite sequence of natural numbers can be found by calculating all of the values given by $6n \pm 1$ that fall within the sequence and subtracting the composites given by $(6n_1 \pm 1)(6n_2 \pm 1)$, where n is a natural number. A test model for finding primes based on this method uses three reference sub-set multiplication tables to calculate composites and then matches these to the corresponding values in the sets $\{6n - 1\}$ and $\{6n + 1\}$. The unmatched numbers are primes. Although this model provides a useful proof of concept, it is impractical at scale. A new method that replaces the sub-set tables with an equation to calculate $6n \pm 1$ composite pairs forms the basis of an improved model using sieve methodology.

1. Introduction

All prime numbers ≥ 5 belong to either $\{6n - 1\}$ or $\{6n + 1\}$ and can be found by subtracting composite sub-sets from these sets as follows:

$$\{6n-1\} - \{(6n_1-1)(6n_2+1)\} = \{6n-1\}_p$$

And;

$$\{6n+1\} - \{(6n_1-1)(6n_2-1)\} - \{(6n_1+1)(6n_2+1)\} = \{6n+1\}_p$$

where $_p$ is prime and n, n_1 and n_2 are natural numbers [1].

This offers the potential to develop a model for finding all of the primes ≥5 in a finite sequence of natural numbers utilising only a third of the numbers in the sequence. This is because the $6n \pm 1$ sets exclude all of the composites in the sequence that are divisible by 2 and/or 3. The remaining composites can only be divided by other $6n \pm 1$ numbers.

A test model based on these equations found all of the primes in the number sequence 5 to 50,000 using three reference tables, one for each of the sub-sets, to find the composites in the sets $\{6n - 1\}$ and $\{6n + 1\}$. The unmatched values were all primes in accordance with the equations above. Some improvements to this model were gained by removing the majority of duplicated and excessive values, which enabled a larger number sequence (5 to 100,000) to be tested and the distribution of primes analysed within the context of sets of $6n \pm 1$ numbers [2].

Notwithstanding the improvements, however, the biggest problem with this method is the number of results held in the three composite sub-set tables, which become ever larger with longer number sequences. This is clearly an undesirable overhead when the only data required are the prime numbers of the number sequence being searched.

2. The $6n \pm 1$ composite pairs

Developing a more practical model relies on being able to discard the composite sub-set tables altogether by finding a more direct method for identifying composites within the sets $\{6n - 1\}$ and $\{6n + 1\}$.

An answer to the problem emerges when considering a trial division method for finding primes in a finite number sequence. This method combines the sets $\{6n - 1\}$ and $\{6n + 1\}$ that fall within the number sequence into a single set $\{6n \pm 1\}$ and arranges it in a numerically ordered list. Dividing these numbers by the first number in the list, 5, and finding whole number results >1 allows composites to be identified and deleted. A new iteration starts with the next number on the list, 7, and the process is repeated. Given the lowest composite in the starting list is 25 and each iteration finds and deletes the lowest composite remaining after the previous iteration, the number selected as the divisor is always guaranteed to be prime. The problem with this model is the number of trial division calculations required but an analysis of the relationship between composites by prime divisor reveals another possibility.

Table 1 shows a limited number of trial division results for a set of $6n \pm 1$ numbers from 5 to 133 for each of the first six prime divisors. The whole number results of a second factor indicating a composite are highlighted in yellow. It is clear that in each column there are pairs of composites (bounded by a bold outline) that are separated by a fixed gap dependent on the divisor. The gap between each number in the pair is $2P$, where P is the prime divisor. Adding P to the lower number of the pair or conversely subtracting it from the higher number therefore gives the same central value for a given pair, which is equivalent to $6Pn$, where n is a natural number. The equation for calculating all $6n \pm 1$ composite pairs is therefore:

$$6Pn \pm P = C$$

where C is composite.

Table 1

	Prime Divisor -->					
6*n* +/-1	**5**	**7**	**11**	**13**	**17**	**19**
5	1	0.714286	0.454545	0.384615	0.294118	0.263158
7	1.4	1	0.636364	0.538462	0.411765	0.368421
11	2.2	1.571429	1	0.846154	0.647059	0.578947
13	2.6	1.857143	1.181818	1	0.764706	0.684211
17	3.4	2.428571	1.545455	1.307692	1	0.894737
19	3.8	2.714286	1.727273	1.461538	1.117647	1
23	4.6	3.285714	2.090909	1.769231	1.352941	1.210526
25	5	3.571429	2.272727	1.923077	1.470588	1.315789
29	5.8	4.142857	2.636364	2.230769	1.705882	1.526316
31	6.2	4.428571	2.818182	2.384615	1.823529	1.631579
35	7	5	3.181818	2.692308	2.058824	1.842105
37	7.4	5.285714	3.363636	2.846154	2.176471	1.947368
41	8.2	5.857143	3.727273	3.153846	2.411765	2.157895
43	8.6	6.142857	3.909091	3.307692	2.529412	2.263158
47	9.4	6.714286	4.272727	3.615385	2.764706	2.473684
49	9.8	7	4.454545	3.769231	2.882353	2.578947
53	10.6	7.571429	4.818182	4.076923	3.117647	2.789474
55	11	7.857143	5	4.230769	3.235294	2.894737
59	11.8	8.428571	5.363636	4.538462	3.470588	3.105263
61	12.2	8.714286	5.545455	4.692308	3.588235	3.210526
65	13	9.285714	5.909091	5	3.823529	3.421053
67	13.4	9.571429	6.090909	5.153846	3.941176	3.526316
71	14.2	10.14286	6.454545	5.461538	4.176471	3.736842
73	14.6	10.42857	6.636364	5.615385	4.294118	3.842105
77	15.4	11	7	5.923077	4.529412	4.052632
79	15.8	11.28571	7.181818	6.076923	4.647059	4.157895
83	16.6	11.85714	7.545455	6.384615	4.882353	4.368421
85	17	12.14286	7.727273	6.538462	5	4.473684
89	17.8	12.71429	8.090909	6.846154	5.235294	4.684211
91	18.2	13	8.272727	7	5.352941	4.789474
95	19	13.57143	8.636364	7.307692	5.588235	5
97	19.4	13.85714	8.818182	7.461538	5.705882	5.105263
101	20.2	14.42857	9.181818	7.769231	5.941176	5.315789
103	20.6	14.71429	9.363636	7.923077	6.058824	5.421053
107	21.4	15.28571	9.727273	8.230769	6.294118	5.631579
109	21.8	15.57143	9.909091	8.384615	6.411765	5.736842
113	22.6	16.14286	10.27273	8.692308	6.647059	5.947368
115	23	16.42857	10.45455	8.846154	6.764706	6.052632
119	23.8	17	10.81818	9.153846	7	6.263158
121	24.2	17.28571	11	9.307692	7.117647	6.368421
125	25	17.85714	11.36364	9.615385	7.352941	6.578947
127	25.4	18.14286	11.54545	9.769231	7.470588	6.684211
131	26.2	18.71429	11.90909	10.07692	7.705882	6.894737
133	26.6	19	12.09091	10.23077	7.823529	7

3. A $6n \pm 1$ prime sieve

The new equation is able to completely replace the sub-set multiplication tables of the original model and permits the development of a sieve model instead. In contrast to the trial division method described above, the composites are calculated directly by the equation thereby significantly reducing the number of calculations required. The composite values are then stored in temporary reference tables that are initially created to calculate and store the pairs of composites given by the lowest prime factor, 5. The results in the tables are matched against the same numbers in the $6n \pm 1$ list and those numbers are then deleted from it. The test model moves on to the next number in the list, 7, guaranteed to be prime as described for the trial division model, and repeats the process replacing the values in the temporary reference tables as it proceeds. The iterations continue until the $6n \pm 1$ list only contains primes.

Note that as the prime factors become larger the gap between each composite in a pair and the gap between successive pairs also becomes larger (see Table 1), thus reducing the number of composite calculations required per prime factor iteration. The matching process also quickens as the $6n \pm 1$ list reduces in size as a result of the composites that are deleted with each iteration.

A further efficiency in the model is created by excluding calculations lower than the square of each prime factor. This avoids a significant amount of replicated calculations. Using 11 as a prime factor example, it can be seen that including the calculations 11 x 5 and 11 x 7 would be repeating the 5 x 11 and 7 x 11 calculations of the previous iterations. Starting the calculations with 11^2 sidesteps these unnecessary duplications.

Nevertheless there are still composites that are calculated more than once. This happens when a composite has more than two prime factors. When this occurs, the model is required to 'skip' a composite that has already been identified. The method adopted here is to do that at the matching phase. In other words the calculation proceeds but a repeated composite cannot be matched as it has already been found and deleted and is thus ignored. Obviously, the larger a composite the greater the probability of it being calculated more than once but this is mitigated by the gaps between composites becoming larger with each model iteration and by the model excluding composites less than the square of primes. For a number sequence with an upper limit of *X*, this combined effect reduces the opportunities for repeated composite results. An analysis of the composite calculations for the number range 5 – 1024, for example, reveals only 33 repeated composite results of which 9 arise for the composites generated by prime factor 7 reducing to 1 for prime factor 29 after taking the mitigating effects into account.

A key benefit of this sieve model is that it resolves the main problem of the composite sub-set reference tables by storing very little data since it deletes composites and replaces the temporary composite reference tables with each prime factor iteration. Only the primes in the selected number sequence are retained.

4. Model results and performance

The test model is written as a Visual Basic for Applications (VBA) script to run in Microsoft Excel 2021 (64 bit) on a Microsoft Windows 10 Home edition operating system using a HP OMEN Laptop x64 with an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 2592 Mhz with 6 cores and 12 logical processors.

A run covering the number sequence 5 – 300,001 was verified to have correctly identified the 25,995 primes belonging to this sequence without error or omission. There were no incidences of numbers being incorrectly identified as primes either. The run time was 6 mins 48secs on the computer described above. For this test the CPU averaged 4.5Ghz with a maximum of 20% usage across the 12 logical processors. RAM was minimally utilised.

5. Potential enhancements

Undoubtedly the model would benefit considerably from running on more suitable software and on a more powerful computer. The VBA script writes the results to worksheets with many in-out operations, which creates some unwanted processing overhead. This could be significantly improved with the use of in-memory arrays and writing only the final results to a worksheet for the output. There may be some model design improvements too to further minimise the number of calculations. In principle the economic use of data should enable the model to be scalable, possibly enhanced by segmentation and running on multiple devices, for example.

References

[1] A. M. Stokes. Isolating the Prime Numbers. *http://viXra.org/abs/2402.0163*. 2024.

[2] A. M. Stokes. The Distribution of Prime Numbers ≥ 5 Within Sequences of Natural Numbers. *http://viXra.org/abs/2404.088*. 2024.

Acknowledgement

With deep gratitude to my brother, Duncan, for coding the model and making it come to life. It would have been way harder without your willingness, expertise and support. Thank you.

Contact

adestokes71@live.com