The Seven Sisters: Prime Generation based on the Form $2p + k$

ALLEN PROXMIRE*

Abstract

This computational note explores the frequency with which simple linear offsets, in the form $2p + k$, generate additional primes from a given prime p. The analysis, conducted on the first one million primes, utilizes seven specific offsets: $k \in \{+1, +3, -3, -5, +7, +9, -9\}$. We find that this set of "Seven Sisters" produces additional primes at a remarkable **80.15%** success rate, with at least one offset yielding a prime in **64.71%** of cases.

Another key finding is the significant increase in success rates for offsets $k = +1$, $+7$, and -5 when candidates are filtered for divisibility by 3 and 5. This effect is a direct consequence of predictable modular relationships between the primes and these specific offsets. After this filtering, all successful offsets converge to a similar success rate of approximately 22%, with the notable exception of the $2p + 7$ offset, which consistently achieves a higher success rate of around 26.25%. This offset's unique behavior, along with the unique ability of $2p - 5$ to produce a prime in all ending-digit categories, represents a compelling pattern in prime number distribution and an efficient methodology for primality production.

Full data, code, and analysis are available at:
https://github.com/allen-proxmire/seven-sisters

1. Introduction

The study of prime numbers has fascinated mathematicians for centuries, with many questions still unanswered. One area of interest is the relationship between prime numbers and formulas that generate new primes, as potential glimpses into the underlying structure of prime distribution. This paper investigates the formula $2p + k$, a simple linear transformation of a prime number, with seven specific offsets, to see how frequently it results in another prime. We hypothesize that the success rates for these seven offsets are not random and are influenced by the modular properties of both the primes and the offsets themselves. By analyzing the data and filtering for common composite factors, we aim to uncover underlying patterns that dictate the success or failure of this formula.

We call the offsets: **$2p + 1$, $2p + 3$, $2p - 3$, $2p - 5$, $2p + 7$, $2p + 9$, and $2p - 9$**, the Seven Sisters. One of whom is Sophie Germain, who notably studied the form $2p + 1$.

*Unaffiliated

2. Methodology

Python scripts were developed to conduct this analysis. The scripts first generated a large list of primes using the Sieve of Eratosthenes to serve as an efficient lookup table. We then iterated through the first one million primes, skipping 2 and 5 as they do not end in 1, 3, 7, or 9. For each prime, the seven offsets were applied.

The resulting candidate numbers were then checked for primality using the pre-generated lookup table. The results were categorized by the ending digit of p, the original, input-prime (either 1, 3, 7, or 9). In a subsequent step, a filter was applied to remove any candidates that were multiples of 3 or 5, and the analysis was re-run to determine the new success rates.

3. Results and Key Findings

Python code and data are available at the GitHub link provided.

The initial analysis showed that the Seven Sisters are highly effective, collectively producing at least one prime 65% of the time, and producing a unique prime 80% of the time. The distribution of success counts are detailed in Table 1.

Table 1. Exact and Cumulative Results

Total primes considered	1000000	
Total primes produced by offsets	946543	
Unique primes produced by offsets	801493	80.15%
Exact success count distribution		
Offsets Succeeded	Count	Percentage
exactly 0 times	352946	35.29%
exactly 1 times	407786	40.78%
exactly 2 times	186686	18.67%
exactly 3 times	45466	4.55%
exactly 4 times	6613	0.66%
exactly 5 times	483	0.05%
exactly 6 times	20	0.00%
exactly 7 times	0	0.00%
Cumulative success count distribution		
Offsets Succeeded	Count	Percentage

at least 7	0	0.00%
at least 6	20	0.00%
at least 5	503	0.05%
at least 4	7116	0.71%
at least 3	52582	5.26%
at least 2	239268	23.93%
at least 1	647054	64.71%

The initial analysis also examined the success rates of each individual offset, which are summarized in Table 2. Results were varied with four offsets showing a success rate of 16.4% and the others between 8-11%.

Table 2. Individual Offset Success

Individual offset success counts		
Total primes considered	1000000	
Offset	Successes	Percentage
2p - 9	164067	16.41%
2p - 5	109637	10.96%
2p - 3	163897	16.39%
2p + 1	82237	8.22%
2p + 3	164061	16.41%
2p + 7	98463	9.85%
2p + 9	164181	16.42%

Additionally, the success rates of each individual offset were categorized by the ending-digit of p, the initial input-prime (either 1, 3, 7, or 9) as shown in Table 3. As expected, the success of the offsets are dependent upon the ending-digit of p. Also, as expected, the offsets work better for some ending-digits than others, but averaged together equal the totals in Table 2.

A significant observation is that the $2p - 5$ offset is the only offset to succeed in all prime ending-digit categories.

Table 3. Individual offset success by ending-digit of *p*

Prime Ending	Offset	Total Primes	Successful Primes	Other Composites	Success Rate (%)
ending_1	2p -9	249934	54815	195119	21.9318

ending_1	2p -5	249934	27293	222641	10.9201
ending_1	2p -3	249934	54740	195194	21.9018
ending_1	2p + 1	249934	27412	222522	10.9677
ending_1	2p + 3	249934	0	249934	0
ending_1	2p + 7	249934	32838	217096	13.1387
ending_1	2p + 9	249934	54844	195090	21.9434
Prime Ending	Offset	Total Primes	Successful Primes	Other Composites	Success Rate (%)
ending_3	2p -9	250110	54679	195431	21.862
ending_3	2p -5	250110	27276	222834	10.9056
ending_3	2p -3	250110	54860	195250	21.9343
ending_3	2p + 1	250110	27550	222560	11.0152
ending_3	2p + 3	250110	54650	195460	21.8504
ending_3	2p + 7	250110	32854	217256	13.1358
ending_3	2p + 9	250110	0	250110	0
Prime Ending	Offset	Total Primes	Successful Primes	Other Composites	Success Rate (%)
ending_7	2p -9	250014	1	250013	0.0004
ending_7	2p -5	250014	27405	222609	10.9614
ending_7	2p -3	250014	54296	195718	21.7172
ending_7	2p + 1	250014	0	250014	0
ending_7	2p + 3	250014	54649	195365	21.8584
ending_7	2p + 7	250014	32769	217245	13.1069
ending_7	2p + 9	250014	54800	195214	21.9188
Prime Ending	Offset	Total Primes	Successful Primes	Other Composites	Success Rate (%)
ending_9	2p -9	249940	54572	195368	21.834
ending_9	2p -5	249940	27662	222278	11.0675
ending_9	2p -3	249940	0	249940	0
ending_9	2p + 1	249940	27273	222667	10.9118
ending_9	2p + 3	249940	54760	195180	21.9093
ending_9	2p + 7	249940	0	249940	0
ending_9	2p + 9	249940	54535	195405	21.8192

Subsequently filtering the results in Table 3 for multiples of 3 and 5, results in the following observations, and are shown in Table 4. The seven offsets can be divided into two groups based on the effect of the filter:

- Group 1: No Effect on Success Rate: For the offsets $2p + 9$, $2p - 9$, $2p + 3$, and $2p - 3$, the success rate remained unchanged after filtering. This indicates that candidates from these offsets were rarely, if ever, multiples of 3 or 5, making the filter negligible.

- Group 2: Doubling of Success Rate: For the offsets $2p + 1$, $2p + 7$, and $2p - 5$, the success rates jumped dramatically, almost doubling in each case in which the offset worked. For instance, the initial success rate for the $2p - 5$ offset with primes ending in 1 was approximately 10.92%. After filtering out candidates that were multiples of 3 or 5, the success rate for this same offset and prime group jumped to approximately 21.82%.

Again, the $2p - 5$ offset is the only one to succeed in all prime ending categories. After filtering, all offsets that yield any successful primes achieve a similar success rate of approximately 22%, with the notable exception of $2p + 7$. This offset consistently achieved a higher success rate of approximately 26.25% in all ending-digit categories except when ending in 9, for which $2p + 7$ is not prime.

Table 4. Results filtered for multiples of 3 and 5

Prime Ending	Offset	Total Primes	Successful Primes	Not a Multiple of 3 and/or 5	Success Rate (%)
ending_1	2p -9	249934	54815	249934	21.9318
ending_1	2p -5	249934	27293	125094	21.818
ending_1	2p -3	249934	54740	249934	21.9018
ending_1	2p + 1	249934	27412	125094	21.9131
ending_1	2p + 3	249934	0	0	0
ending_1	2p + 7	249934	32838	125094	26.2507
ending_1	2p + 9	249934	54844	249934	21.9434
Prime Ending	Offset	Total Primes	Successful Primes	Not a Multiple of 3 and/or 5	Success Rate (%)
ending_3	2p -9	250110	54679	250109	21.8621
ending_3	2p -5	250110	27276	125066	21.8093
ending_3	2p -3	250110	54860	250110	21.9343
ending_3	2p + 1	250110	27550	125066	22.0284
ending_3	2p + 3	250110	54650	250109	21.8505

ending_3	2p + 7	250110	32854	125066	26.2693
ending_3	2p + 9	250110	0	0	0
Prime Ending	Offset	Total Primes	Successful Primes	Not a Multiple of 3 and/or 5	Success Rate (%)
ending_7	2p -9	250014	1	1	100
ending_7	2p -5	250014	27405	124959	21.9312
ending_7	2p -3	250014	54296	250014	21.7172
ending_7	2p + 1	250014	0	0	0
ending_7	2p + 3	250014	54649	250014	21.8584
ending_7	2p + 7	250014	32769	124959	26.2238
ending_7	2p + 9	250014	54800	250014	21.9188
Prime Ending	Offset	Total Primes	Successful Primes	Not a Multiple of 3 and/or 5	Success Rate (%)
ending_9	2p -9	249940	54572	249940	21.834
ending_9	2p -5	249940	27662	125050	22.1208
ending_9	2p -3	249940	0	0	0
ending_9	2p + 1	249940	27273	125050	21.8097
ending_9	2p + 3	249940	54760	249940	21.9093
ending_9	2p + 7	249940	0	0	0
ending_9	2p + 9	249940	54535	249940	21.8192

4. Discussion: The Role of Modular Arithmetic

The distinct behaviors observed in our two groups of offsets are not coincidental; they are a direct consequence of modular arithmetic, specifically how the offsets interact with the properties of the primes themselves. We can explain these behaviors by examining the value of $2p + k$ modulo 3.

Group 1: Unaffected Offsets

For the offsets $k \in \{+3, -3, +9, -9\}$, the resulting candidate number $2p + k$ is never a multiple of 3. We can prove this by examining the congruence classes of primes. Every prime $p > 3$ is either congruent to 1 or 2 modulo 3.

- If $p \equiv 1 \pmod{3}$**:** The candidate is $2p + k \equiv 2\,(1) + k \pmod{3}$. For $k \in \{+3, -3, +9, -9\}$, we have $k \equiv 0 \pmod{3}$. Therefore, $2\,(1) + k \equiv 2\,(1) + 0 \equiv 2 \pmod{3}$.
- If $p \equiv 2 \pmod{3}$**:** The candidate is $2p + k \equiv 2\,(2) + k \pmod{3}$. For $k \in \{+3, -3, +9, -9\}$, we have $k \equiv 0 \pmod{3}$. Therefore, $2\,(2) + k \equiv 4 + 0 \equiv 1 \pmod{3}$.

In both cases, the resulting candidate is never divisible by 3. This explains why filtering out multiples of 3 had no effect on the success rates for this group of offsets.

Group 2: Filter-Affected Offsets

For the offsets $k \in \{+1, +7, -5\}$, the resulting candidate number $2p + k$ is guaranteed to be a multiple of 3 for one of the two prime congruence classes (either $p \equiv 1 \pmod{3}$ or $p \equiv 2 \pmod{3}$).

- Offset $2p + 1$**:**
  - If $p \equiv 1 \pmod{3}$, then $2p + 1 \equiv 2\,(1) + 1 \equiv 3 \equiv 0 \pmod{3}$.
  - If $p \equiv 2 \pmod{3}$, then $2p + 1 \equiv 2\,(2) + 1 \equiv 5 \equiv 2 \pmod{3}$.
- Offset $2p - 5$**:**
  - If $p \equiv 1 \pmod{3}$, then $2p - 5 \equiv 2\,(1) - 5 \equiv -3 \equiv 0 \pmod{3}$.
  - If $p \equiv 2 \pmod{3}$, then $2p - 5 \equiv 2\,(2) - 5 \equiv -1 \equiv 2 \pmod{3}$.
- Offset $2p + 7$**:**
  - If $p \equiv 1 \pmod{3}$, then $2p + 7 \equiv 2\,(1) + 7 \equiv 9 \equiv 0 \pmod{3}$.
  - If $p \equiv 2 \pmod{3}$, then $2p + 7 \equiv 2\,(2) + 7 \equiv 11 \equiv 2 \pmod{3}$.

Since the population of primes is split almost equally between the two congruence classes, approximately half of all candidates generated by these offsets are guaranteed to be multiples of 3. Our filter effectively removes this large pool of guaranteed composites, revealing the primes generated by the other half, which is why the success rate for this group appears to double.

Unique Behaviors

The data reveals two especially compelling patterns:

- Offset $2p - 5$: This offset is the only one to produce primes across all four ending-digit categories of the input prime. While it is obvious that $2p - 5$ cannot equal a multiple of 5, it is not obvious how it avoids other prime factors. This suggests a unique and stable relationship that warrants further investigation.
- Offset $2p + 7$: After filtering, this offset consistently outperformed all others, achieving a success rate of 26.25% in three of the four categories. This superior performance is a

crucial finding that challenges the notion of a uniform prime distribution even after modular filtering and deserves dedicated future study.

This illustrates how a simple filter based on a number's prime factors can reveal deeper, non-obvious patterns in the distribution of primes.

5. Conclusion

This study demonstrates that the **Seven Sisters** — a simple linear formula of the form $2p + k$ applied to a given prime — can serve as an efficient engine for prime number generation. The high success rates observed, with a unique prime generated in over 80% of cases and at least one prime in over 64% of cases, underscore the remarkable generative power of this method. The application of a filter for the multiples of 3 and 5 creates a more accurate picture of each offset's true effectiveness by removing predictably composite candidates.

Specifically, the unique behaviors of the $2p - 5$ and $2p + 7$ offsets highlight that prime distribution is not uniform even within the confines of this formula. The ability of $2p - 5$ to produce a prime in every prime ending-digit category and the consistently superior performance of $2p + 7$ after filtering warrant further research.

This methodology presents a practical and effective approach to primality production and offers a rich area for continued computational and theoretical exploration into the patterns underlying prime number sequences.