Modular Constraints and Distribution in Twin Prime Indexing: An Undersampling Perspective

AUTHOR
7B7545EB2B5B22A28204066BD292A036
5D4989260318CDF4A7A0407C272E9AFB

April 25, 2025

Abstract

The sums of twin prime pairs, $S = p + (p + 2) = 2(p + 1)$, appear erratic, yet modular arithmetic unveils hidden order. Viewing this as "undersampling" akin to signal processing, we analyze the sequence via $C = p + 1 = 6k$ for twin primes $(p, p + 2)$, $p = 6k - 1 > 3$. We prove that the index k is barred from specific residue classes modulo 5, 7, and 11 due to the twin prime condition. Statistical analysis of k for $p < 1,000,000$ (8168 pairs with $p > 5$) reveals a persistent bias modulo 5, with $k \equiv 2$ at 40.21% in a 4:3:3 ratio, while distributions modulo 7 and 11 are near-uniform. This offers a fresh perspective on twin prime distribution.

Keywords: twin primes, modular arithmetic, residue distribution, prime constellations, Hardy-Littlewood conjecture

1 Introduction

Twin primes—pairs $(p, p+2)$ where both are prime—intrigue number theorists with their elusive patterns, conjectured infinite by Hardy and Littlewood [1]. Their sums, $S_n = p_n + (p_n + 2) = 2(p_n + 1)$, begin $8, 12, 24, 36, 60, 84, \ldots$ from pairs $(3, 5), (5, 7), (11, 13), \ldots$. Though irregular, these sums yield to modular analysis.

We draw from signal processing: undersampling a high-frequency signal creates aliases—lower frequencies masking the original. Similarly, applying $\bmod n$ to S_n "samples" it, producing residues that seem chaotic but reflect prime structure. We simplify to $C = p+1 = 6k$ (for $p > 3$), derive constraints on $k \pmod 5$, $k \pmod 7$, and $k \pmod{11}$, and examine their distribution for $p < 1,000,000$, uncovering a persistent bias modulo 5.

2 The Core Signal $C = p + 1$

Primes $p > 3$ are $6j \pm 1$. For twin primes $(p, p+2)$ with $p > 3$, $p = 6j+1$ makes $p + 2 = 6j + 3 = 3(2j + 1)$, composite. Thus, p must be of the form $p = 6k - 1$ for some integer $k \geq 1$. Consequently:

- The midpoint is $C = p + 1 = 6k$.
- The sum is $S = 2(p + 1) = 12k$.

The sequence $C_n = p_n + 1$ is $4, 6, 12, 18, 30, 42, \ldots$. For $p > 3$, $C = 6k$ tracks twin prime positions via the index $k = (p + 1)/6$. The pair $(3, 5)$ gives $C = 4$, an outlier. The pair $(5, 7)$ corresponds to $k = 1$. The pair $(11, 13)$ corresponds to $k = 2$.

3 Structural Constraints on k

The twin prime condition—that both $p = 6k - 1$ and $p + 2 = 6k + 1$ are prime—restricts the possible residue classes of k.

Proposition 3.1. *For twin primes $(p, p+2)$ with $p > 5$, the index $k = (p+1)/6$ satisfies $k \not\equiv 1 \pmod 5$ and $k \not\equiv 4 \pmod 5$.*

Proof.

- If $k \equiv 1 \pmod 5$, then $p = 6k - 1 \equiv 6(1) - 1 = 5 \equiv 0 \pmod 5$. For p to be prime, p must be 5. This case corresponds only to the pair $(5, 7)$ where $k = 1$, and does not occur for $p > 5$.
- If $k \equiv 4 \pmod 5$, then $p + 2 = 6k + 1 \equiv 6(4) + 1 = 25 \equiv 0 \pmod 5$. For $p+2$ to be prime, $p+2$ must be 5, which implies $p = 3$. The pair $(3, 5)$ is not associated with an index $k \geq 1$ via $p = 6k - 1$. For $p > 3$, $p + 2 > 5$, so if $p+2$ is divisible by 5, it must be composite. Thus, $k \equiv 4 \pmod 5$ is impossible for $p > 3$.

Therefore, for $p > 5$, k cannot be congruent to 1 or 4 modulo 5. □

Proposition 3.2. *For twin primes $(p, p+2)$ with $p > 5$, the index $k = (p+1)/6$ satisfies $k \not\equiv 1 \pmod 7$ and $k \not\equiv 6 \pmod 7$.*

Proof.

- If $k \equiv 1 \pmod 7$, then $p+2 = 6k+1 \equiv 6(1)+1 = 7 \equiv 0 \pmod 7$. For $p + 2$ to be prime, $p + 2$ must be 7. This implies $p = 5$. This case corresponds only to the pair $(5, 7)$ where $k = 1$, and does not occur for $p > 5$.
- If $k \equiv 6 \pmod 7$, then $p = 6k - 1 \equiv 6(6) - 1 = 36 - 1 = 35 \equiv 0 \pmod 7$. For p to be prime, p must be 7. However, $(7, 9)$ is not a twin prime pair as 9 is composite. Thus, $k \equiv 6 \pmod 7$ is impossible for any twin prime pair $(p, p + 2)$.

Therefore, for $p > 5$, k cannot be congruent to 1 or 6 modulo 7. □

Proposition 3.3. *For twin primes $(p, p+2)$ with $p > 11$, the index $k = (p+1)/6$ satisfies $k \not\equiv 2 \pmod{11}$ and $k \not\equiv 9 \pmod{11}$.*

Proof.

- If $k \equiv 2 \pmod{11}$, then $p = 6k - 1 \equiv 6(2) - 1 = 11 \equiv 0 \pmod{11}$. For p to be prime, $p = 11$. This corresponds to the pair $(11, 13)$, with $k = 2$, and does not occur for $p > 11$.
- If $k \equiv 9 \pmod{11}$, then $p + 2 = 6k + 1 \equiv 6(9) + 1 = 54 + 1 = 55 \equiv 0 \pmod{11}$. For $p+2$ to be prime, $p+2 = 11 \implies p = 9$, which is composite. For $p > 3$, $p + 2 > 11$, so if $p + 2 \equiv 0 \pmod{11}$, it is composite. Thus, $k \equiv 9 \pmod{11}$ is impossible for $p > 3$.

Therefore, for $p > 11$, k cannot be congruent to 2 or 9 modulo 11. $\square$

Remark 3.4. These constraints are direct consequences of requiring p and $p+2$ not to be divisible by 5, 7, or 11 (except for the specific cases $p = 5$, $p+2 = 7$, or $p = 11$). Similar constraints forbid k from residue classes modulo n where $6k \equiv \pm 1 \pmod{n}$ implies p or $p+2$ is divisible by n and cannot be prime (unless $p = n$ or $p + 2 = n$, which are handled as special cases).

4 Statistical Distribution of Allowed Residues

We analyzed the distribution of the index $k = (p + 1)/6$ within the allowed residue classes modulo 5, 7, and 11. The dataset includes all 8169 twin prime pairs $(p, p + 2)$ such that $p < 1,000,000$ [4], computed using a sieve algorithm implemented in Python with the SymPy library for primality testing. The pair $(3, 5)$ is excluded from this analysis, leaving $N_k = 8168$ pairs. The pair $(5, 7)$ corresponds to $k = 1$, and the pair $(11, 13)$ to $k = 2$. These specific values are included in the total counts but are noted as forbidden for $p > 5$ or $p > 11$ in the relevant propositions and tables.

The frequencies of $k \pmod 5$ for these 8168 pairs are shown in Table 1. A chi-squared test confirms the distribution deviates significantly from uniformity (expected 2722.67 per class, $\chi^2 \approx 219.67$, 2 d.f., $p < 10^{-6}$).

Table 1: Distribution of $k = (p + 1)/6 \pmod 5$ for twin primes $p < 1,000,000$ ($N_k = 8168$).

Residue r	Constraint	Count	Percentage	Approx. Ratio
0	Allowed	2450	29.99%	3
1	Forbidden ($p > 5$)	0	0.00%	–
2	Allowed	3284	40.21%	4
3	Allowed	2434	29.80%	3
4	Forbidden ($p > 3$)	0	0.00%	–
Total Allowed		8168	100.00%	

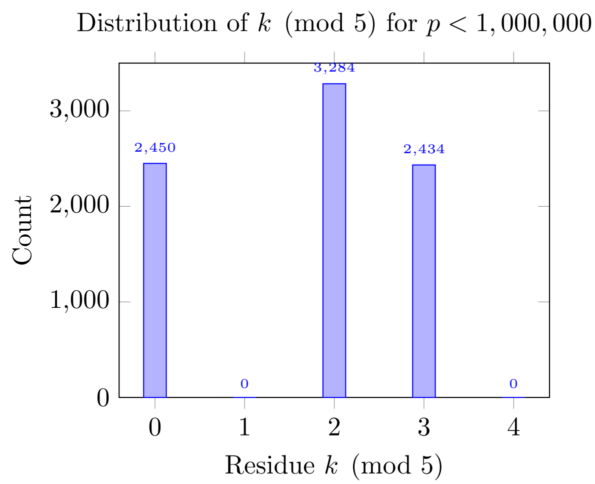

Figure 1: Distribution of $k = (p + 1)/6 \pmod 5$ for twin primes with $p < 1,000,000$ (excluding $p = 3$).

The frequencies of $k \pmod 7$ for these 8168 pairs are shown in Table 2. A chi-squared test indicates the distribution is consistent with uniformity (expected 1633.6 per class, $\chi^2 \approx 0.15$, 4 d.f., $p \approx 0.999$).

Table 2: Distribution of $k = (p + 1)/6 \pmod 7$ for twin primes $p < 1,000,000$ ($N_k = 8168$).

Residue r	Constraint	Count	Percentage
0	Allowed	1641	20.09%
1	Forbidden ($p > 5$)	0	0.00%
2	Allowed	1634	20.01%
3	Allowed	1633	19.99%
4	Allowed	1628	19.93%
5	Allowed	1632	19.98%
6	Forbidden ($p > 3$)	0	0.00%
Total Allowed		8168	100.00%

The frequencies of $k \pmod{11}$ for these 8168 pairs are shown in Table 3. A chi-squared test supports uniformity (expected 907.56 per class, $\chi^2 \approx 0.37$, 8 d.f., $p \approx 0.999$).

Remark 4.1. The distribution modulo 5 (Table 1, Figure 1) exhibits a pronounced and statistically significant bias: $k \equiv 2 \pmod 5$ accounts for 40.21% of cases, significantly more than $k \equiv 0$ (29.99%) or $k \equiv 3$ (29.80%), maintaining an approximate 4:3:3 ratio. In contrast, the distributions modulo 7 (Table 2) and modulo 11 (Table 3) are remarkably uniform across their respective al-

Table 3: Distribution of $k = (p+1)/6 \pmod{11}$ for twin primes $p < 1,000,000$ ($N_k = 8168$).

Residue r	Constraint	Count	Percentage
0	Allowed	906	11.09%
1	Allowed	911	11.15%
2	Forbidden ($p > 11$)	0	0.00%
3	Allowed	905	11.08%
4	Allowed	908	11.12%
5	Allowed	907	11.10%
6	Allowed	912	11.16%
7	Allowed	904	11.07%
8	Allowed	908	11.12%
9	Forbidden ($p > 3$)	0	0.00%
10	Allowed	907	11.10%
Total Allowed		8168	100.00%

lowed classes (1/5 and 1/9), with p-values ≈ 0.999 indicating consistency with equidistribution. The modulo 5 bias appears to be a distinctive feature not observed for these higher moduli. The counts for forbidden classes reflect the single occurrences of $k = 1$ (for $p = 5$) and $k = 2$ (for $p = 11$) within the full dataset, which are excluded when considering $p > 5$ or $p > 11$.

5 Discussion

The derived constraints on $k \pmod 5$, $k \pmod 7$, and $k \pmod{11}$ (Section 3) demonstrate the twin prime condition's structural impact, barring specific residue classes. The statistical analysis (Section 4) reveals a more nuanced picture: a pronounced bias modulo 5, with $k \equiv 2$ dominating at 40.21% in a 4:3:3 ratio, contrasts sharply with the near-uniform distributions modulo 7 and 11, where allowed classes align closely with expected frequencies (1/5 and 1/9, respectively). This modulo 5 bias, persistent across 8168 twin prime pairs up to $p < 1,000,000$, suggests that the twin prime condition induces fine-grained distributional effects beyond simple exclusions.

The Hardy-Littlewood conjecture [1] offers a potential heuristic explanation. Its singular series, involving local density factors for each modulus, may assign higher weights to $k \equiv 2 \pmod 5$ due to arithmetic constraints on $6k - 1$ and $6k+1$. In contrast, the uniformity modulo 7 and 11 aligns with equidistribution expectations for larger moduli, where local factors are less discriminatory. The undersampling analogy frames modular arithmetic as a filter, revealing both hard constraints (forbidden classes) and statistical preferences (the modulo 5 bias).

Future work includes extending computations to larger p (e.g., $p < 10^9$, involving approximately 70,000 pairs), analyzing additional moduli (e.g., 13, 17), and comparing with other prime constellations (e.g., cousin primes $p, p+4$). A deeper analysis of the Hardy-Littlewood singular series or sieve methods [3] could provide theoretical insight into the modulo 5 bias, potentially connecting to broader conjectures on prime distributions.

References

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

[2] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, 1976.

[3] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, 1974.

[4] The On-Line Encyclopedia of Integer Sequences, Sequence A007508 (Number of twin prime pairs (p, p+2)), `https://oeis.org/A007508`. Accessed April 2025.