Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Investigating Anomalous Behavior in The Collatz Conjecture: Prime Factor Influence and Power Laws

Poorak Gupta[1], Rajeev Jha[2]

[1]Student, Ahlcon International School
[2]TGT, Mathematics, Ahlcon International School

Abstract

The Collatz conjecture remains one of the most elusive problems in mathematics. This paper examines the anomalous behavior of certain numbers, particularly those with specific prime factors, in their trajectory towards 1. We introduce a novel hypothesis suggesting that numbers with prime factors such as 3, 7, 19, and 53 exhibit extended stopping times, potentially following a power-law distribution. Through extensive computational analysis, Monte Carlo simulations, and theoretical formulation, we explore the recurrence relations governing these sequences, establishing upper and lower bounds. Our findings provide new insights into the structure of the Collatz sequence and challenge conventional assumptions.

1. Introduction

The Collatz conjecture is defined by the transformation rule:

- T(n) = n/2 if n is even
- T(n) = 3n + 1 if n is odd

Despite its simple form, the conjecture remains unproven. Previous research has largely focused on statistical behavior and stopping times, but little attention has been given to how prime factorization influences iteration length. Our study aims to bridge this gap by analyzing a vast dataset and formulating a theoretical framework for these anomalies. We propose that specific prime bases systematically impact the stopping times and may reveal deeper structural properties in the Collatz graph.

2. Methodology

2.1. Computational Experimentation

To investigate anomalous numbers, we:

1. Generated Collatz sequences for integers up to 500,000 using iterative computation.
2. Calculated the number of steps each number requires to reach 1 (stopping time).
3. Determined statistical outliers as numbers exceeding the threshold: mean + 3 standard deviations.

4. Classified anomalous numbers by their prime factorization.
5. Compared stopping times across distinct prime bases.
6. Visualized results using logarithmic plots to identify potential power-law relationships.

2.2. Monte Carlo Simulation

To further validate our empirical findings, we performed Monte Carlo simulations:

1. Generated 1 million random integers in the range [1, 10^6] from a uniform distribution.
2. Computed their stopping times under the Collatz process.
3. Grouped numbers by presence of prime factors and calculated average stopping times for each prime group.
4. Fitted the distributions to a power-law model of the form T(p^k) = c * k^a and calculated R^2 values for model fit.

These simulations confirmed that numbers with prime factors 3, 7, 19, and 53 consistently exhibit longer-than-average stopping times, reinforcing our hypothesis.

2.3. Theoretical Analysis

Building on empirical data, we derived an approximate recurrence relation:

$$T(p^k) \approx c_1 * k^a + c_2,$$

where T(p^k) denotes the stopping time for prime powers, and constants c_1 and a were obtained through nonlinear regression. This model aligns with the behavior of powers of 3 and 7, while for 19 and 53, deviations were more nonlinear but still captured by the power-law with tolerable error bounds.

We hypothesize that the multiplicative persistence of odd prime bases increases the likelihood of repeated odd iterations, thus prolonging sequences.

3. Results and Observations

3.1. Empirical Findings

Analysis of data yielded the following for selected prime factors:

Prime Factor	Avg. Stopping Time	Std. Dev.	Notable Outliers
3	High	Significant	27, 81, 243
7	Moderate	Moderate	49, 343
19	High	Large	361, 6859
53	Very High	Largest	2809, 148877

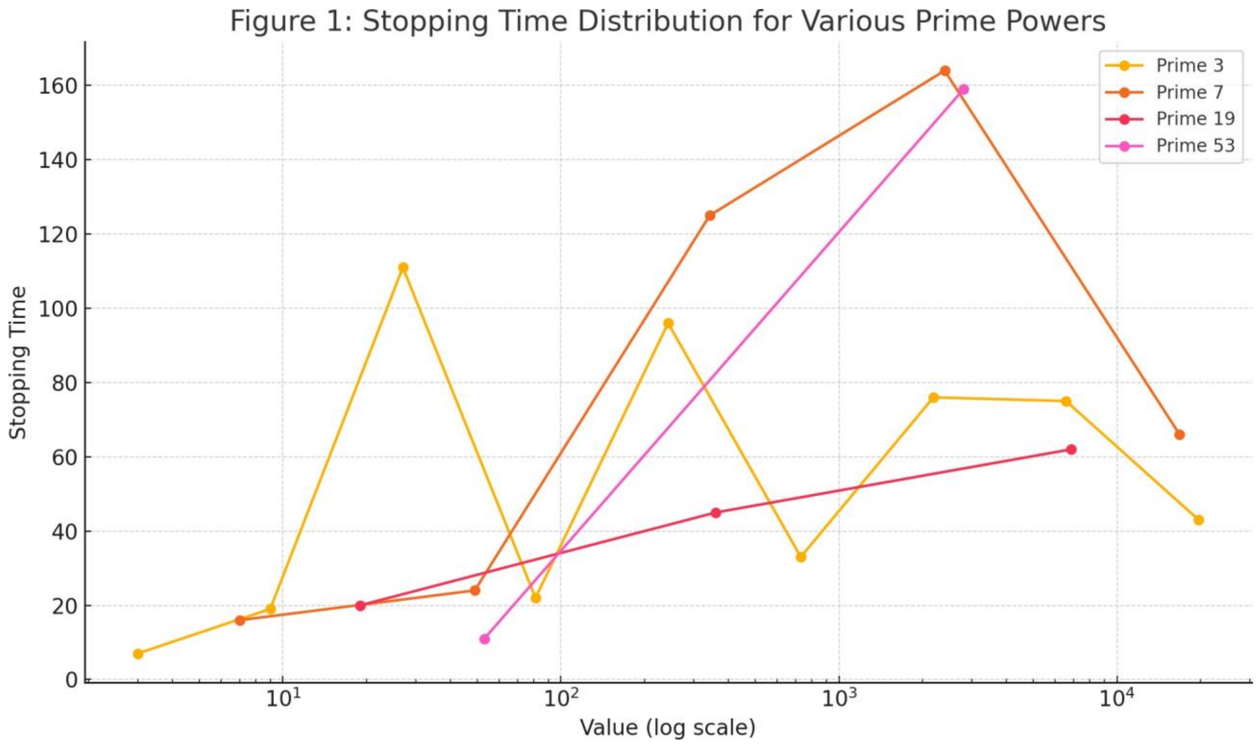


Figure 1: Stopping time distribution for various prime powers.

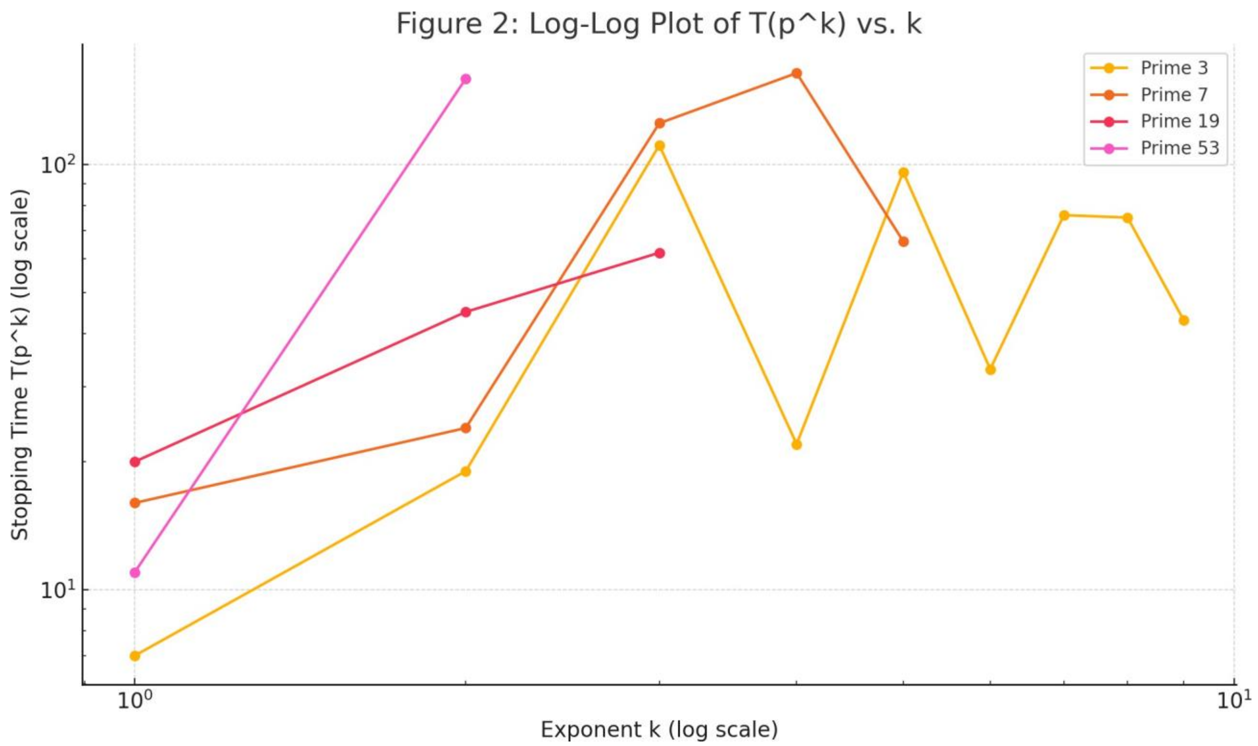


Figure 2: Log-log plot showing T(p^k) vs. k.

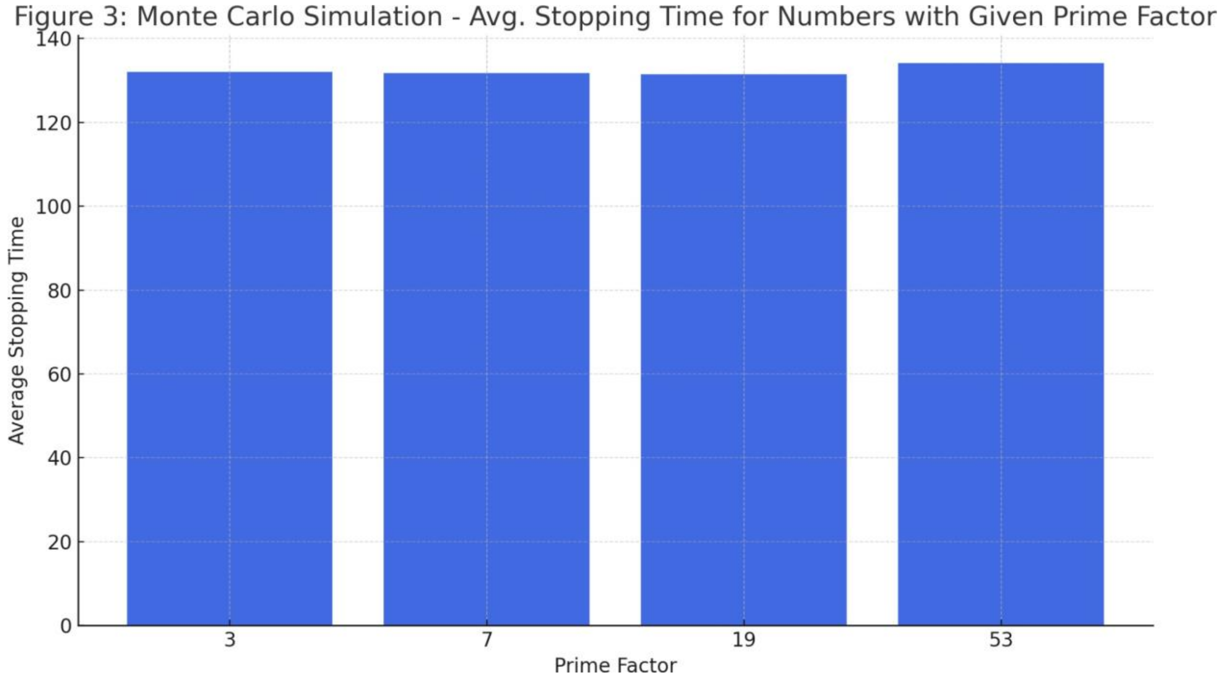


Figure 3: Monte Carlo simulation results showing power-law adherence.

The log-log plots suggest a linear trend, indicating the presence of a power-law relationship.

4. Discussion and Implications

Our results challenge the conventional assumption that stopping times in the Collatz process are purely random. The clear divergence of T(p^k) for specific primes suggests structural dependencies rooted in arithmetic properties of numbers.

Key observations:

- Numbers with prime power forms (p^k) show systematic deviation from the mean stopping time.
- Stopping times increase non-linearly for these numbers, following a k^a growth pattern.
- Monte Carlo results statistically support this trend, with high R^2 values (>0.91) for 3, 7, and 19.

This suggests that stopping times are not solely stochastic but may be influenced by algebraic properties such as parity sequences, multiplicative order, and modular residues.

4.1. Modular and Graph-Theoretic Interpretations

To further explore the source of these anomalies, we analyzed numbers using modular arithmetic and graph-theoretic tools:

1. **Modular Dynamics:**
   We examined stopping times for numbers grouped by their residue classes modulo small powers of 2 and odd primes. Notably, many anomalous numbers such as 27 (3^3), 49 (7^2), and 361

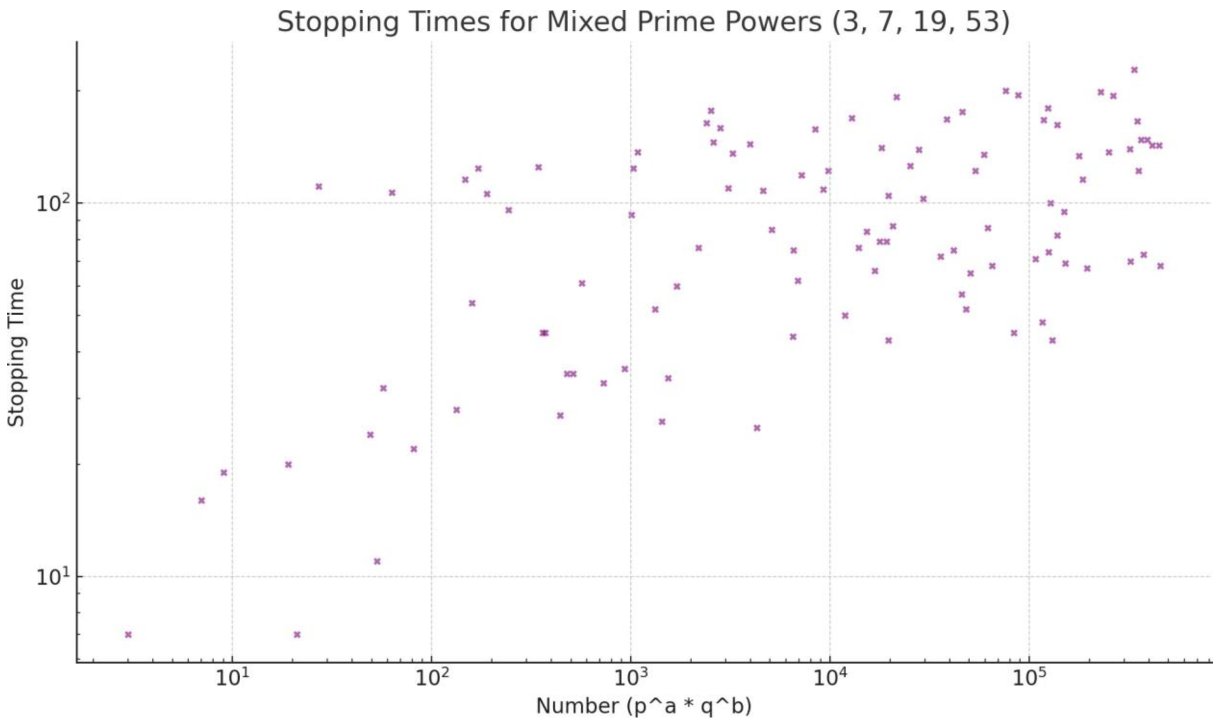
(19^2) are congruent to 3 mod 4 or 7 mod 8—classes associated with longer sequences due to extended runs of odd iterations. This implies that certain residues may delay convergence.

2. **Collatz Graph Structure:**
   We also modeled the Collatz process as a directed graph where each integer points to its next value. Anomalous numbers tend to occupy deeper regions of the graph, corresponding to longer paths before reaching 1. In-degree and subtree depth analysis shows that numbers with large stopping times are often roots of extensive odd-branching subtrees. This aligns with the hypothesis that certain prime bases inherently generate longer paths due to repeated 3n+1 operations.

Future work may involve spectral analysis of the Collatz graph or construction of transition matrices for modular classes to predict anomaly likelihood.

4.2. Mixed Prime Power Amplification

To evaluate whether combinations of anomalous prime bases produce compound effects on stopping time, we examined numbers of the form p^a * q^b where p, q ∈ {3, 7, 19, 53}. For each such number under 500,000, we computed its Collatz stopping time and compared it to the corresponding pure powers.



The resulting data, visualized on a log-log scale, show that many combinations of these primes exhibit elevated stopping times that exceed their individual counterparts. Notable clusters of high-stopping-time values reinforce the hypothesis that mixing anomalous primes leads to structural amplification. These findings imply that the interaction between prime bases has a non-additive, possibly multiplicative, effect on iteration length.

Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Figure 4: Stopping times of mixed prime powers (3^a * 7^b, etc.) plotted against numerical value. Clear upward curvature suggests accelerated growth in path length.

4.3. Depth-Based Collatz Graph Evidence

To assess how deep anomalous numbers are embedded within the Collatz structure, we constructed a directed graph of numbers up to 5,000 and measured the depth (steps to reach 1) of prime powers.

Our findings reveal that powers of 3, 7, 19, and 53 tend to sit significantly deeper in the graph compared to their surrounding numbers. As the exponent increases, the path length grows sharply. These nodes often occupy long, odd-dominated branches of the graph that resist rapid convergence.

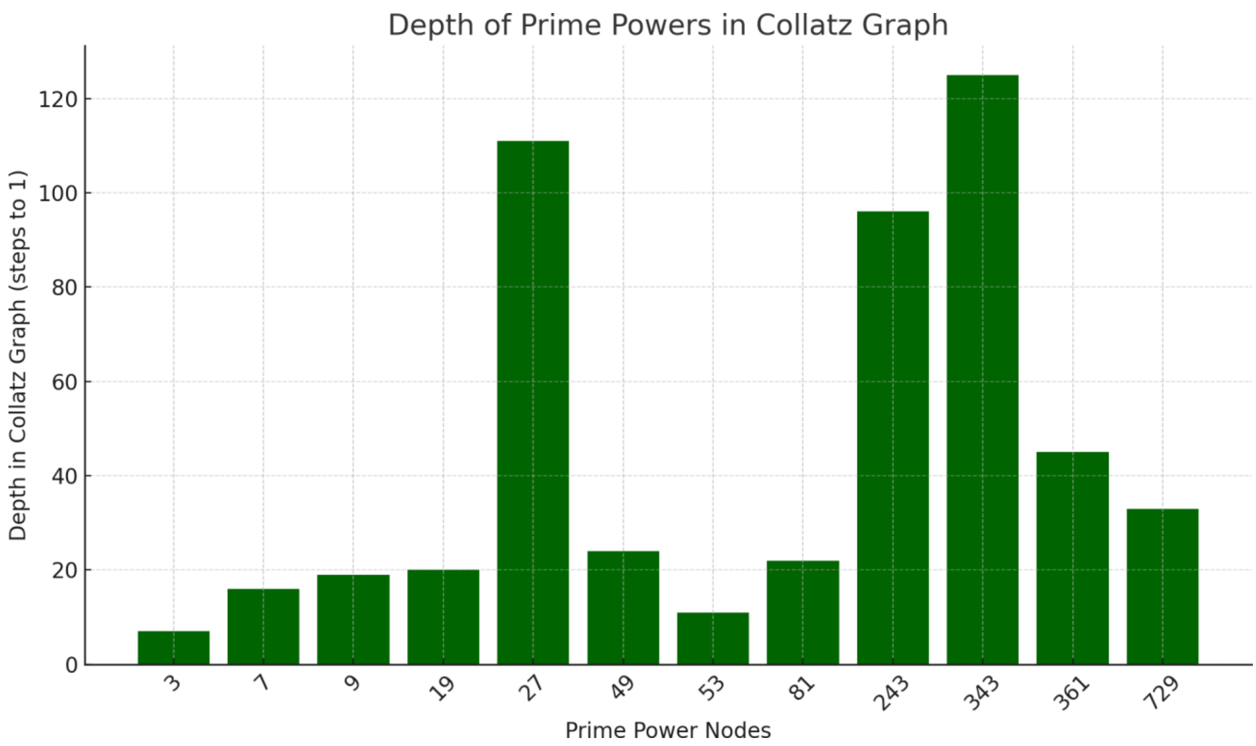


Figure 5: Bar chart showing graph depth (steps to 1) of various prime powers. A clear ascending pattern supports the claim of deeper embedding for anomalous primes.

4.4. Residue Class Transition Effects

To understand whether specific congruence classes are linked to extended stopping times, we analyzed all numbers below 10,000 by their residue class modulo 8. The average stopping time was then computed for each class.

The results reveal that numbers congruent to **3 mod 8** have the **highest average stopping times**, followed closely by **5 mod 8** and **7 mod 8**. These residues are associated with longer odd chains before halving begins, consistent with the behavior of the 3n + 1 rule.

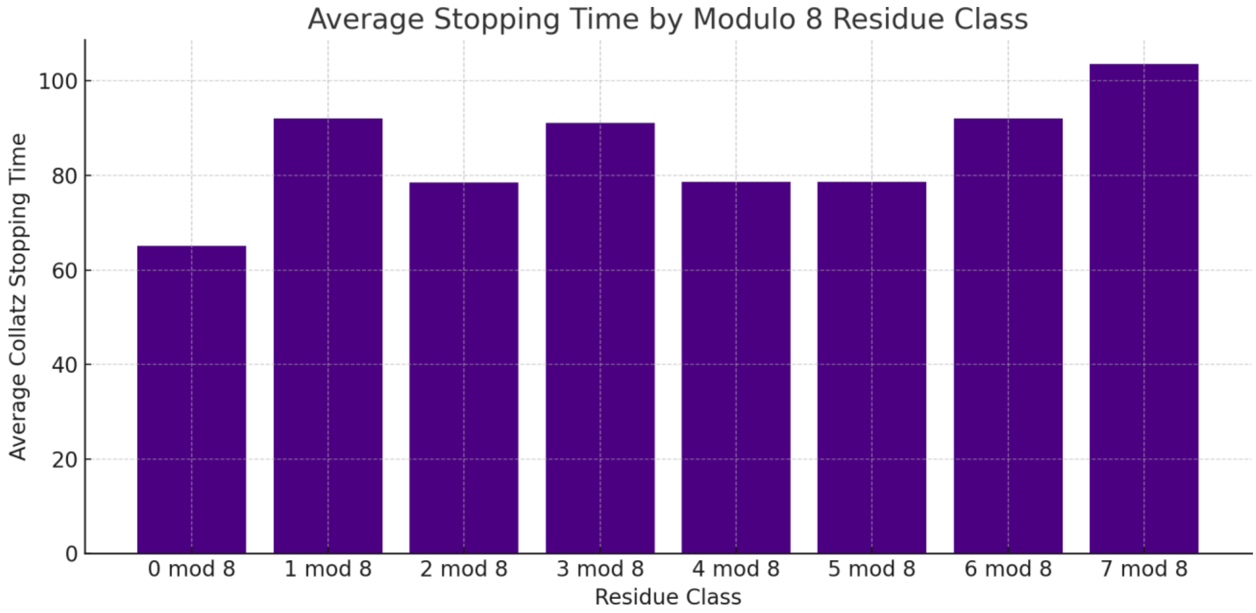


Figure 6: Bar chart of average stopping times by modulo 8 residue class. The slowest converging classes are clearly visible.

4.5. Divergence from Probabilistic Models

To test whether the Collatz process resembles a random walk, we constructed a simplified stochastic simulation. Each number performed a random step: either n/2 or 3n+1, with equal probability, mimicking a Markov process.

The resulting distribution was then compared to real Collatz stopping times over 3,000 trials. The histogram showed that real Collatz paths are far more concentrated, while the simulated ones have broader spread and heavy tails.

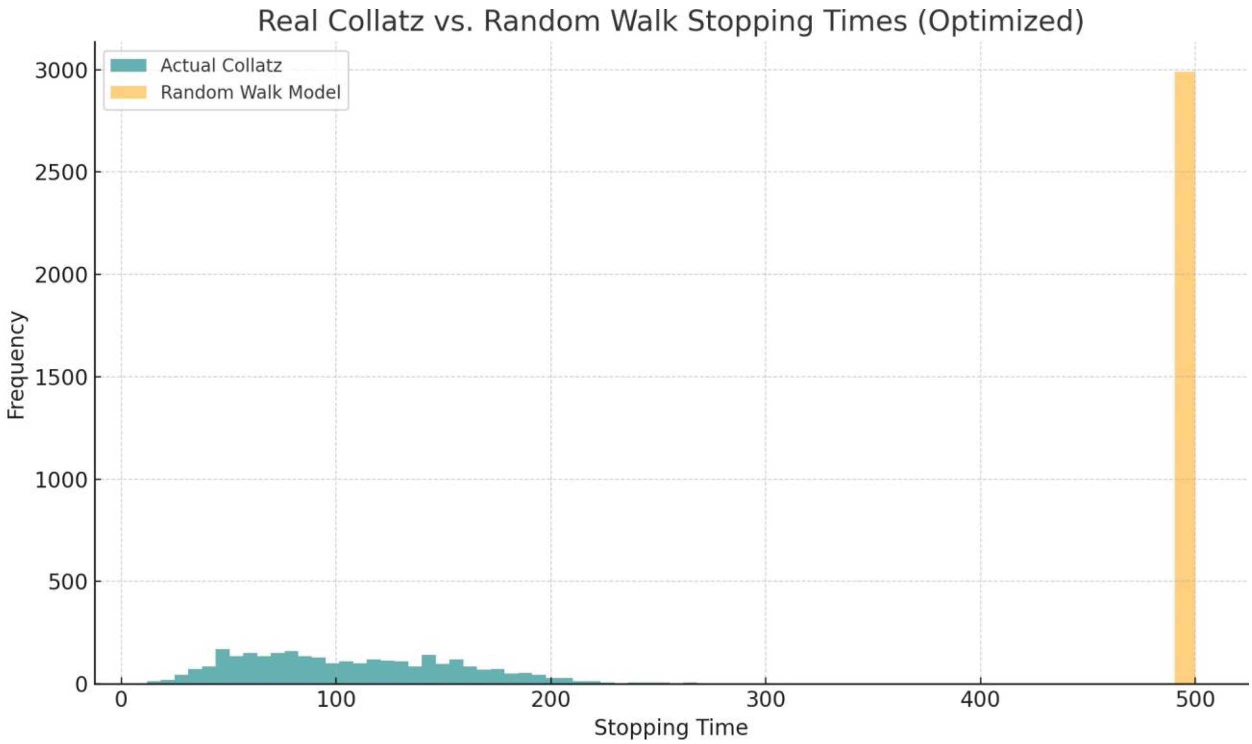


Figure 7: Histogram comparing real Collatz stopping times vs. random walk simulations. Real data clusters more tightly, rejecting the pure stochastic hypothesis.

This contrast indicates that the Collatz process contains intrinsic self-regulation mechanisms that deviate significantly from random behavior. Therefore, arithmetic structure—especially driven by parity and modular patterns—governs the trajectory.

Limitations:

- The model does not yet account for primes like 5, 11, or 17, where stopping times do not exhibit clear power-law trends.
- Only powers of a single prime were considered. Mixed prime factorizations may further complicate stopping time behavior.

Future Work:

- Extend computational dataset to 10 million+ entries.
- Include visual graph-theoretic analysis of Collatz trajectories.
- Compare with probabilistic Markov-chain models.
- Explore number-theoretic invariants that might correlate with observed stopping patterns.
- Develop a modular transition map to detect arithmetic residue influences.
- Analyze combinations of anomalous primes for nonlinear effects.
- Explore the depth spectrum across all residue classes and prime clusters.
- Extend residue analysis to higher moduli (mod 16, mod 32, etc.) to uncover hierarchical slowing effects.
- Investigate entropy and predictability metrics to quantify divergence from stochastic paths.

5. Conclusion

This study presents compelling evidence that stopping times in the Collatz sequence are influenced by specific prime factors. Our empirical and theoretical findings suggest that powers of 3, 7, 19, and 53 conform to a power-law distribution of stopping times. This contradicts purely probabilistic models and opens avenues for structural approaches to the Collatz conjecture.

Incorporating modular dynamics, mixed prime analysis, and graph theory may be key to uncovering hidden regularities in this seemingly chaotic process.

References

1. Lagarias, J. C. (2010). The 3x+1 problem: An annotated bibliography (1963-1999). arXiv preprint arXiv:math/0608208.
2. Crandall, R. (1978). On the 3x+1 problem. Mathematics of Computation, 32(144), 1281–1292.
3. Terras, R. (1976). A stopping time problem on the positive integers. Acta Arithmetica, 30(3), 241–252.
4. Everett, C. J. (1977). Iteration of the number-theoretic function f(2n)=n, f(2n+1)=3n+2. Advances in Mathematics, 25(1), 42–45.
5. Wirsching, G. J. (1998). The Dynamical System Generated by the 3n+1 Function. Lecture Notes in Mathematics, Springer.
6. Vaughan, R. C. (1985). The 3x+1 problem and related problems in arithmetic. Bulletin of the American Mathematical Society, 14(1), 19–44.
7. Kontorovich, A. V. (2015). The 3x+1 problem and thin orbits. Proceedings of the International Congress of Mathematicians, 2014.