Forward-Gap Parameterization of Primes and Its Statistical Structure

Jian Hu

Abstract

We investigate a structural reparameterization of the prime sequence via normalized forward gaps, assigning to each odd prime half the distance to its successor. Algebraic identities induced by this labeling are examined together with statistical properties of the resulting integer sequence for primes up to 10^9. The induced representation yields affine geometric families under a difference-of-squares transformation and exhibits pronounced oscillatory structure, slow growth of distinct values consistent with Cramér's heuristic, and relatively low empirical Shannon entropy. Observed gap frequencies are compared with predictions of the Hardy–Littlewood prime pair conjecture and display close numerical agreement within the tested computational range. These results provide a descriptive framework for the forward-gap sequence.

Keywords: prime gaps, Cramér model, Hardy–Littlewood conjecture, empirical distribution, Shannon entropy.

1 Introduction

Let (p_k) denote the increasing sequence of primes. Define the forward gap

$$g_k := p_{k+1} - p_k.$$

For $k \geq 2$, g_k is even.

Definition 1. *For $k \geq 2$, define*

$$h_k := \frac{p_{k+1} - p_k}{2}.$$

Each odd prime is uniquely labeled by the integer h_k. We analyze algebraic structure and statistical properties induced by this parameterization.

2 Difference-of-Squares Identity

Define

$$\ell_k := p_{k+1}^2 - p_k^2.$$

The quantity ℓ_k measures the interval between consecutive prime squares and encodes both local gap size and global prime magnitude.

Algebraically,

$$\ell_k = (p_{k+1} - p_k)(p_{k+1} + p_k) = 2h_k(2p_k + 2h_k).$$

Proposition 1. *For $k \geq 2$,*

$$p_k = \frac{\ell_k}{4h_k} - h_k.$$

Proof. Substituting $p_{k+1} = p_k + 2h_k$ into

$$\ell_k = (p_{k+1} - p_k)(p_{k+1} + p_k)$$

and simplifying yields the identity. □

Remark 1. *For fixed h, primes satisfying $h_k = h$ lie on the affine line*

$$p = \frac{1}{4h}\ell - h.$$

Thus the forward-gap labeling partitions the prime sequence into affine geometric families indexed by h, as illustrated in Figure 1.

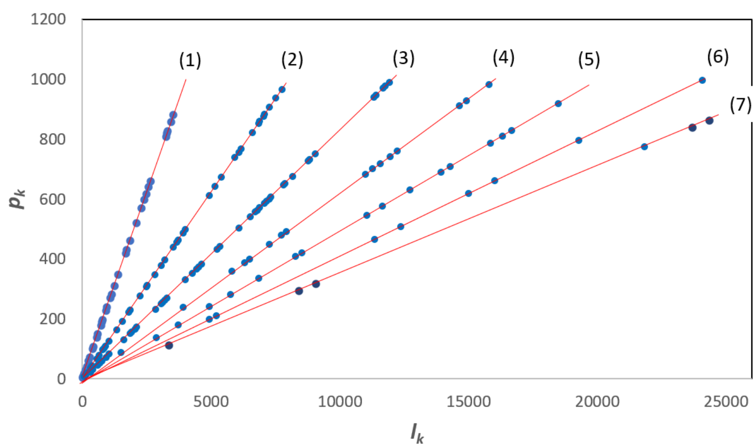

Figure 1: Plot of (ℓ_k, p_k) for initial primes (p_k from 3 to 997). Points cluster along distinct affine lines corresponding to fixed h.

3 Statistical Structure of the Forward-Gap Sequence

For $x > 0$, define the empirical frequency

$$f_x(h) := \frac{\#\{k : p_k \leq x,\ h_k = h\}}{\pi(x)}, \tag{1}$$

where π(x) denotes the prime counting function. This section examines statistical properties of the induced sequence (h_k) for $x \leq 10^9$.

3.1 Empirical Frequency Distribution

All primes up to 10^9 were computed, yielding

$$\pi(10^9) = 50{,}847{,}534.$$

Within this range,

$$1 \leq h \leq 141.$$

Despite more than fifty million primes, only 141 distinct values occur.

3.2 Frequency of Small Gaps

Table 1 lists frequencies for small values of h.

Table 1: Observed frequencies of small forward-gap values up to 10^9.

h	Gap ($2h$)	Frequency	Ratio
1	2	3422970	6.73%
2	4	3423456	6.73%
3	6	6087747	11.97%
4	8	2694851	5.30%
5	10	3483719	6.85%
6	12	4467679	8.79%
7	14	2464425	4.85%
8	16	1845671	3.63%
9	18	3350760	6.59%
10	20	1823532	3.59%

Small values dominate strongly, with h=3 typically the most frequent within the tested range [1].

Figure 2 displays the global distribution $f_x(h)$.

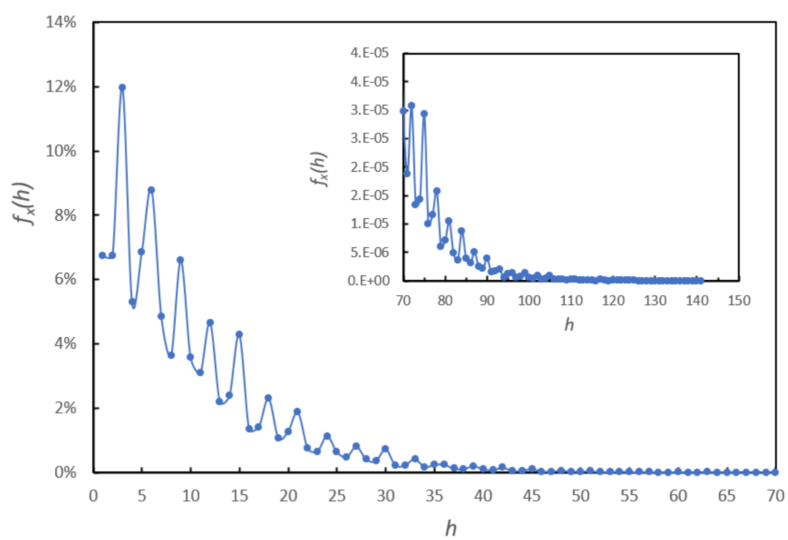

Figure 2: Empirical distribution $f_x(h)$ for $x \leq 10^9$. Main plot: $h = 1$ to 70. Inset: $h = 70$ to 141.

The distribution exhibits structured oscillatory behavior rather than monotone decay.

3.3 Oscillatory Behavior

The alternating peak structure shown in Figure 2 reflects residue-class constraints imposed by small primes.

An exponential envelope

$$y \approx Ae^{-\lambda h} \tag{2}$$

provides a good empirical fit over moderate ranges of h.

Figure 3 shows the distribution together with the fitted envelope.

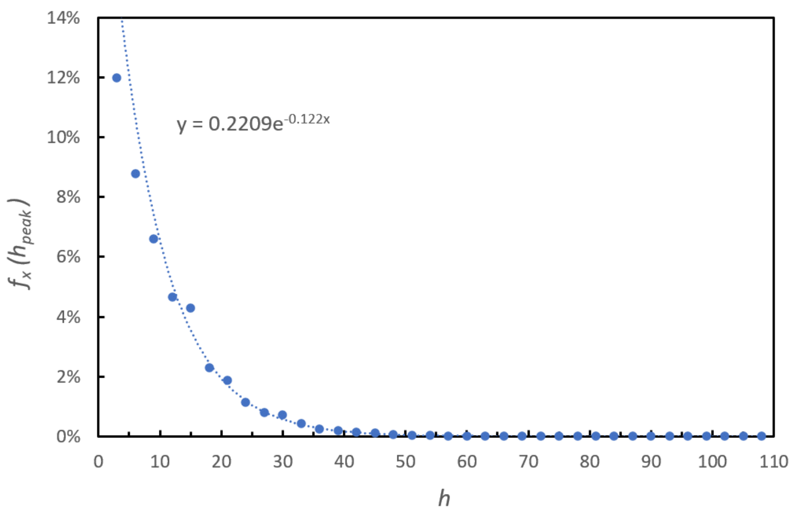

Figure 3: Oscillatory structure with exponential envelope fit. Data beyond $h \approx 108$ are omitted due to sparse frequency and increased statistical noise. The dashed line is a least-squares exponential fit with $R^2 = 0.9982$.

The exponential envelope is purely descriptive; no asymptotic law is claimed.

3.4 Growth of Distinct Values

Define the maximal observed forward parameter

$$h_{\max}(x) := \max\{h_k : p_k \leq x\},$$

and the number of distinct attained values

$$D(x) := \#\{h_k : p_k \leq x\}.$$

Computational data up to $x = 10^9$ indicate that the attained values of h_k are nearly contiguous from 1 up to $h_{\max}(x)$, with only occasional omissions in restricted ranges. Consequently, $D(x)$ and $h_{\max}(x)$ grow at comparable rates.

Empirically, we observe the approximate relation

$$h_{\max}(x) \approx 0.49(\log_2 \pi(x))^2 - 0.88 \log_2 \pi(x) + 1.85, \tag{3}$$

This quadratic growth in $\log \pi(x)$ is consistent with Cramér's heuristic model, under which maximal prime gaps are expected to be of order $(\log x)^2$ [2]. Since $h_k = g_k/2$ and $\log x \sim \log \pi(x)$, this heuristic predicts

$$h_{\max}(x) = O((\log \pi(x))^2), \tag{4}$$

in agreement with the observed behavior.

4 Comparison with Hardy–Littlewood Predictions

The Hardy–Littlewood (HL) prime pair conjecture [3] predicts the asymptotic count of prime pairs separated by $2h$ below x, without isolation requirements.

To compare with consecutive prime gaps, non-isolated contributions were removed using a heuristic adjustment.

Table 2 compares observed frequencies with Hardy–Littlewood predictions for selected values of h.

Table 2: Observed frequencies versus Hardy–Littlewood heuristic predictions.

x	Obs($h = 1$)	HL($h = 1$)	Obs($h = 3$)	HL($h = 3$)	Obs($h = 5$)	HL($h = 5$)
10^6	8170	8247	13549	14930	7079	7022
10^7	58981	58752	99987	108327	54431	54287
10^8	438776	440367	766708	821974	428968	430150
10^9	3422970	3425309	6087747	6451136	3483719	3484461

Agreement improves as x increases, consistent with the heuristic nature of the conjecture.

The relative error between the observed frequency and Hardy-Littlewood predictions is defined as

$$\mathrm{RelErr}(h, x) := \frac{\left|f_x(h) - f_x^{HL}(h)\right|}{f_x^{HL}(h)}.$$

where $f_x(h)$ is empirical frequency (1), and $f_x^{HL}(h)$ the Hardy–Littlewood predicted frequency. Figure 4 shows variation of the relative errors with x for h=1,3,5 across the computational range of 10^6-10^9.

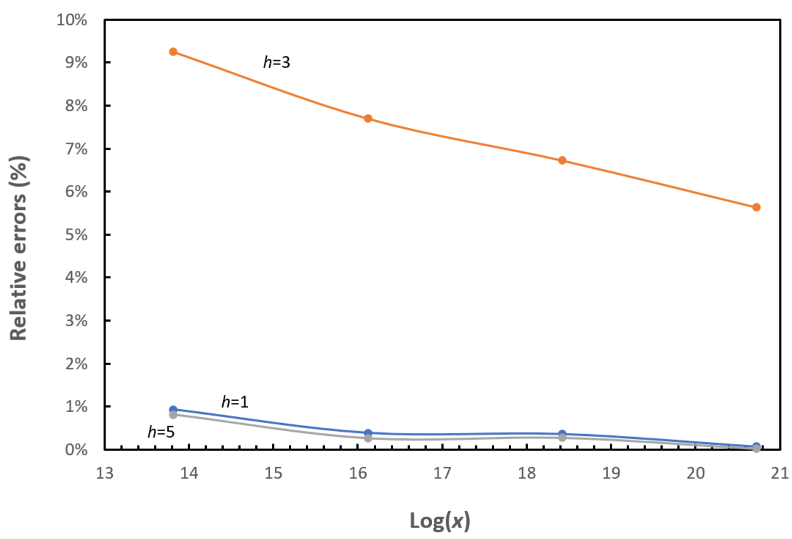

Figure 4: Variation of relative error between observed counts and Hardy–Littlewood predictions for $h = 1, 3, 5$ (solid lines are included as guides to the eye). For the values of h considered, the HL prediction is strictly positive.

The larger relative error observed for $h=3$ likely reflects stronger residue-class interactions at moderate scales, given the comparatively large Hardy–Littlewood constant associated with this gap. Extrapolation of the observed trend suggests decreasing relative error for larger x, consistent with the heuristic nature of the conjecture.

5 Shannon Entropy

We define the Shannon entropy [4] of the empirical distribution f_x by

$$H(x) := -\sum_h f_x(h) \log_2 f_x(h),$$

where $f_x(h)$ is defined in (1). This quantity measures the information-theoretic spread of the forward-gap distribution.

For $x = 10^9$,

$$H(10^9) \approx 4.6 \text{ bits}.$$

For comparison, a uniform distribution over $\{1, \ldots, 141\}$ would have entropy $\log_2 141 \approx 7.14$ bits.

The observed entropy is substantially smaller, reflecting strong concentration in small gap values.

Equivalently, the effective support size

$$2^{H(10^9)} \approx 24$$

is far below the full observed range.

No asymptotic entropy behavior is asserted.

6 Conclusion

The forward-gap parameterization induces linear geometric structure and structured statistical behavior within the computational range $x \leq 10^9$. The induced sequence (h_k) exhibits structured oscillatory distribution, slow support growth consistent with classical gap heuristics, and relatively low empirical entropy. Observed frequencies display close numerical agreement with predictions of the Hardy–Littlewood prime pair conjecture within the tested range. These observations provide a descriptive structural framework for the forward-gap sequence.

References

[1] A. Odlyzko, M. Rubinstein, and M. Wolf, *Jumping champions*, Experiment. Math. 8 (1999), 107–118.

[2] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. 2 (1936), 23–46.

[3] 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 Math. 44 (1923), 1–70.

[4] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Technical Journal. 27 (1948), 379–423, 623–656.