Patterns and Gaps in Prime Numbers: A Statistical and Computational Investigation

Gideon Afriyie

Abstract

This project was inspired by an undergraduate course I took in abstract mathematics, where we explored foundational ideas about prime numbers, such as Mersenne primes, prime factorizations, and open problems in number theory. That course ignited a personal curiosity about the long-term behavior of prime numbers, particularly the gaps between them.

While it is well known that prime numbers appear to be distributed “randomly,” many mathematicians have proposed conjectures suggesting deep underlying structure. This research aimed to investigate such patterns through empirical methods, specifically by generating, analyzing, and visualizing the first 1,000 prime numbers.

Using Python, I designed a prime number generator, computed prime gaps, visualized their distribution, and analyzed them using statistical techniques. I compared my results to classical conjectures, such as Cramér’s Conjecture and the Prime Number Theorem, using both raw and logarithmic transformations. All code, data, and visuals were produced using Python and Excel and are published in the public GitHub repository for transparency and reproducibility.

Research Motivation

The motivation behind this project was rooted in an intellectual desire to reconcile the seeming chaos of prime distributions with the mathematical order suggested by number theory. Are there observable regularities in how prime numbers appear? Can we detect statistical boundaries or expected behaviors in their growth?

I was drawn to these questions:

- Do prime gaps increase in a predictable way?
- Are conjectured bounds, like Cramér’s, consistent with observed data?
- What kinds of statistical tendencies do prime gaps exhibit?

- Are there hidden structures, such as twin primes, that persist even in small datasets?

Methodology

1. Generating Prime Numbers

Using Python, I implemented a simple but efficient algorithm to generate the first 1,000 prime numbers. These were exported as a CSV file and imported into Excel for auxiliary computations and visuals.

2. Computing Prime Gaps

Prime gaps were calculated as the difference between two consecutive primes:

$$\text{Gap}_n = p_{n+1} - p_n$$

This yielded 999 prime gaps from the original list of 1,000 primes.

3. Statistical Analysis

Key descriptive statistics were calculated using Excel and Python:

- Mean of gaps: ≈ 17.05
- Median of gaps: 17
- Mode of gaps: 6
- Maximum gap: 34

These values reveal a central tendency and variance in gap sizes. The frequent recurrence of small gaps (especially 6, 2, and 4) aligns with known behaviors of smaller primes, while larger gaps increase irregularly but remain bounded.

Visualizations

4. Plotting Prime Gaps

Using Matplotlib and Excel, I visualized:

- Gaps vs. Prime Index (scatter plot)

- Frequency distribution (histogram)
- Prime size vs. gap size (line and scatter)

The resulting plots illustrated:

- High frequency of small gaps at early indices
- Occasional large gaps interrupting clusters of small ones
- A general upward trend in gap size, though not smooth

5. Logarithmic Analysis

To compare with theoretical bounds, I applied:

$$\log(p), \quad \log^2(p)$$

Plots of gaps vs. $\log^2(p)$ showed that all actual gaps fell well below the Cramér bound, providing limited but consistent empirical support for the conjecture in the range studied.

Additionally, plotting $\log(p)$ vs. index revealed a flattening growth curve, in line with the Prime Number Theorem.

Exploring Conjectures

6. Cramér's Conjecture

Cramér suggests:

$$p_{n+1} - p_n = \mathcal{O}((\log p_n)^2)$$

My visualization confirmed this bound held true for all 1,000 primes.

7. Twin Prime Conjecture

Pairs like $(p, p+2)$ were manually observed and counted. Many such pairs occurred early on, but they diminished in frequency as primes increased, again aligning with the expectations of the conjecture.

Conceptual Exploration

Although the main project focused on general primes, we reflected on our earlier coursework on Mersenne primes, primes of the form:

$$2^n - 1$$

These are rare and grow exponentially. Their unique form contrasts with the irregular yet measurable behavior of general primes and inspired me to think about structured vs. unstructured number families.

Conclusions

This study does not attempt to solve any open mathematical problem. However, it demonstrates that computational and statistical methods can be effective in visualizing and understanding key behaviors in prime number theory, even in small datasets.

Key Takeaways:

- Prime gaps show a nonrandom structure with clear central tendencies.
- The gaps grow roughly, but not precisely, with the size of primes.
- Classical conjectures hold for small prime sets-offering grounds for deeper investigation.

Future Work

- Extend dataset to first 10,000 or 1 million primes
- Automate analysis pipeline for gap bounds, conjecture comparisons, and cluster detection
- Integrate more advanced math libraries (e.g., `sympy`, `numpy`) for symbolic and numerical validation
- Explore modulo-based patterns, Goldbach pairings, or the Riemann Hypothesis
- Build interactive visual dashboards for educational use or presentation

Tools Used

- **Python**: Prime generation, analysis, matplotlib visualization
- **Excel**: Statistical summaries, histogram, scatter plots
- **Matplotlib**: Plots of gap behavior, theoretical bounds
- **GitHub**: Public repository for code, data, visuals
- **Mathematical Concepts**: Cramér's Conjecture, Prime Number Theorem, Twin Prime Conjecture

GitHub Repository

Link to GitHub: `https://github.com/GideonAfriyie/prime-number-analysis`

References

[1] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arithmetica, vol. 2, 1937, pp. 23–46.

[2] J. Hadamard and C.J. de la Vallée-Poussin, *Prime Number Theorem*, 1896.

[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 Mathematica, vol. 44, 1923, pp. 1–70.