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

Author: Gideon Afriyie

Abstract

This research investigates the behavior of prime number gaps through a computational and statistical lens, building on foundational number-theoretic concepts such as Mersenne primes, twin primes, and key conjectures in analytic number theory. Inspired by a coursework in abstract mathematics, this study aims to empirically explore whether discernible patterns exist in the distribution and growth of prime numbers, particularly in the spacing between them.

Using Python, the first 1,000 prime numbers were generated and stored in a structured dataset. The focus of the analysis was on prime gaps, defined as the differences between consecutive primes. Statistical examination of these gaps revealed both expected and irregular trends. Common gaps such as 6, 2, and 4 were frequent among lower primes, while gaps tended to widen with increasing prime index. The maximum observed gap was 34. Descriptive metrics indicated a mean gap of 17.05, a median of 17, and a modal gap of 6, suggesting a central tendency amidst growing irregularity.

To assess theoretical alignment, the data was compared against Cramér's Conjecture, which posits that prime gaps are asymptotically bounded by (log p)^2. Overlaying the empirical gaps with Cramér's predicted bound confirmed that all observed gaps remained well below the conjectured ceiling, reinforcing its plausibility within this range. The presence of twin primes (gaps of 2) was also examined. Consistent with the Twin Prime Conjecture, such pairs appeared frequently at low indices and sporadically at higher ones, suggesting a non-vanishing occurrence rate.

The study also incorporated logarithmic transformations. By plotting the natural logarithm of each prime against its index, the analysis captured a visibly flattening curve, consistent with the

Prime Number Theorem, which states that primes become less dense but follow a predictable distribution. Further visualizations of prime gaps against log p highlighted increasing dispersion and variability as prime magnitude grew.

While Mersenne primes were not directly computed, their conceptual inclusion provided a comparative framework, showcasing structured subsets of primes amid broader statistical randomness. These primes, of the form (2^n) - 1, remain a key object of theoretical interest due to their rarity and algebraic significance.

Conclusion & Future Work

This project serves as a bridge between theoretical number theory and empirical data science. While it does not attempt to resolve unsolved problems, it demonstrates how **computational methods,** and **statistical tools** can illuminate patterns within a classically unpredictable domain. Future work will extend the dataset to at least 10,000 primes, apply deeper number-theoretic models such as those connected to the **Riemann Hypothesis**, and develop interactive visual tools for real-time exploration. The entire codebase, datasets, and visualizations are available on the project's GitHub repository, emphasizing reproducibility and public engagement.

This investigation exemplifies the interdisciplinary power of mathematics, programming, and data analysis in exploring one of the most enduring mysteries in mathematics: the prime numbers.