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

Author: Gideon Afriyie

Abstract:

This project was inspired by a course in abstract mathematics, where we explored foundational ideas around primes, including Mersenne primes, prime factorizations, and conjectures that attempt to describe the unpredictable behavior of prime numbers. That course ignited a personal curiosity about the possibility of identifying long-term trends in the distribution of primes, especially in the behavior of the gaps between them. Although formulating a definitive equation to predict primes remains elusive, this research represents an early step in examining the structural features of primes and assessing whether patterns or conjectural models offer insight into their growth.

The study began with the generation of the first 1,000 prime numbers using a custom Python script. These primes were saved to a CSV file and imported into Microsoft Excel for statistical analysis and visualization. The central object of analysis was the prime gap, defined as the difference between two consecutive primes. These gaps were calculated for all 999 consecutive pairs, revealing both frequent and rare patterns. The most common gap was 6, followed by gaps of 2 and 4. While smaller gaps occurred frequently among smaller primes, the gap sizes generally increased with larger primes. However, the increase was not smooth or uniform; instead, it was marked by occasional large jumps. The maximum gap observed in this data set was 34.

Using Excel, I calculated the mean with the mean gap falling around 17.05. The median gap was 17, and the mode, representing the most common value was 6. These findings suggest a central tendency in the prime gaps, despite the wide fluctuations. I visualized the prime gaps using scatter plots in Excel and Matplotlib, plotting the gaps against their prime indices. The result was a highly irregular but bounded distribution, which mirrors the known difficulty in predicting prime number patterns.

To ground the analysis, I compared the observed gap behavior to **Cramér's Conjecture**, which posits that the gap between consecutive primes is asymptotically bounded by the square of the logarithm of the prime number. Plotting both the actual gaps and the conjectured upper bound on the same graph revealed that all observed gaps fell well below this boundary, providing empirical support for the conjecture at least within the first 1,000 primes. Additionally, I examined the **Twin Prime Conjecture,** which posits that there are infinitely many pairs of primes that differ by two. Although this research does not attempt to prove or disprove the conjecture, I observed and counted numerous twin prime pairs, especially among smaller primes. The occurrence of such pairs diminished with increasing magnitude but continued to appear sporadically even at higher indices, aligning with the expectations of the conjecture.

Another angle of exploration involved plotting the natural logarithm of each prime number against its index. The **logarithmic growth** revealed a gradual flattening curve, consistent with the Prime Number Theorem, which suggests that primes become less frequent as numbers grow larger but still follow a discernible distribution. Further visualization involved plotting the gaps themselves against the logarithm of the primes to explore possible correlations. While no direct linear relationship emerged, the scatter was denser at lower logarithmic values and more dispersed at higher values, again reflecting increasing but bounded irregularity.

Although Mersenne primes were not directly analyzed in this project, they remain a notable example of rare and structurally unique primes. Their rapid growth and rarity highlight the exceptional behavior of primes of the form (2^n) - 1, which were frequently discussed in the abstract mathematics course that inspired this work. This connection serves as a conceptual bridge between structured mathematical forms and the more chaotic distribution seen in general primes.

This project represents an early-stage empirical investigation into prime number behavior. While it does not aim to resolve any open problems in number theory, it offers insight into how computational tools and statistical techniques can be used to visualize, analyze, and better understand one of mathematics' most enduring mysteries. Future directions include extending the dataset to the first 10,000 or 1M primes, exploring more sophisticated conjectures such as the **Riemann Hypothesis,** and building dynamic visual tools to interactively explore the primes.

The Python code used in this study can easily be scaled to generate larger datasets, and all data and figures can be found in the public GitHub repository associated with this project. This work stands as a demonstration of interdisciplinary thinking, linking mathematical curiosity, programming, and statistical reasoning in pursuit of patterns within one of the most enigmatic structures in mathematics: the prime numbers.