A Statistical Study of the Twin Prime Conjecture in Number Theory

by Sydney Hoang and Sana Vosoughi, Irvine Valley College
Mentor: Prof. Lan Pham

Abstract

The twin prime conjecture states that there are infinitely many primes that differ by two such as three and five or seventeen and nineteen. Currently, the result is widely believed to be correct in the mathematical community with numerical simulation used for verification. Without a rigorous mathematical proof, the statement stands as a conjecture. In this study, we examine the problem with techniques from Statistics to help further provide evidence that the conjecture is correct. We examine various densities and distances among the twin prime pairs to find statistical trends in the hope that it will lead to a mathematical proof.

Introduction

Mathematicians have long studied prime numbers because they are the building blocks of integers. Prime numbers are numbers that are only divisible by itself and one. Two, five, seven, or eleven are examples of prime numbers. They are the building blocks of integers in that every integer is a product of prime numbers. For example, the number 21 is a product of prime numbers 3 and 7, or 100 is a product of 2, 2, 5, and 5. To understand properties of integers, mathematicians look to understand properties of prime numbers. One way to do this is to understand how dense the prime numbers are. Between 10 and 20, there are four prime numbers 11, 13,17, and 19. We say that the density of prime numbers is 4 out of 10 between 10 and 20. Another approach is to study how close the prime numbers are to each other. In particular, one can ask if there are prime numbers that are next to each other such as 13 and 15. These prime numbers are called twin primes. As numbers become large, natural questions to ask are (1) do twin primes still exist, and (2) if they do, how often do we encounter them? In 1846, a French mathematician conjectured that there are infinitely many twin primes which is now known as the twin prime conjecture. With the largest known twin primes on the order of ten to the power of sixteen, mathematicians believe that the conjecture is likely to be true. The consensus among mathematicians is that the conjecture is true, but the proof is out of reach with current techniques.

In this study, we examine the twin prime conjecture from the point of view of Statistics. We look to study various statistics of twin primes as they become large. A computer code was written to generate prime numbers and find twin prime numbers. We calculate various statistics such as the average distance between consecutive twin primes and the density of twin primes as they become large. The result is used to verify the prime gap distance of log(n) as n becomes large and to provide additional insights to verify the conjecture.

Current Work

Three recent works by Yitang Zhang, Marko V. Jankovic, and Giulio Morpurgo made progress towards a mathematical proof of the twin prime conjecture. Professor Yitang Zhang proved that there are infinitely many pairs where the distance between two consecutive prime numbers is less than 70 million. This is the first result that is very close to the twin prime conjecture in that a finite bound was placed between prime numbers that can eventually be lowered. The work by Marko Jankovic highlights patterns between twin prime pairs that significantly progressed the understanding of the problem. Jankovic found that a twin prime pair could be found by multiplying an integer with 6 and adding or subtracting 1 to find the pair. For example, when 6 is multiplied

© Think You?! The Proceedings of the Bay Honors Research Symposium, 2024. All Rights Reserved.

by 3, which is 18, 1 can be subtracted to find the prime number 17. To find the other twin prime pair of 17, 1 is added to 18 to solve for 19. However, this finding is not a definite proof to the twin prime conjecture, since this formula does not work for all combinations of 6 and any integer. The multiple of 4 and 6 does not work since adding 1 to 24, which gives us 25, is not a prime number. Despite the drawbacks to the finding, Jankovic made significant progress on the twin prime conjecture which led many other mathematicians to study on this finding. Recently, Giulio Morpurgo made predictions of the behavior of twin primes pairs on dePolignac's conjecture, which is connected to the Twin Prime conjecture. Morpugo observed that for an integer n, there are twin prime pairs between $(n-2)^2$ and n^2. Results were shown for integers up to 6.5×10^6.

Methodology

To find twin prime numbers, we wrote a computer code in C to find prime numbers and in the process, find twin pairs. The code finds and stores prime and twin prime numbers from some initial and final values read from an input file. The set of prime numbers and twin prime numbers are stored in separate files. Prime numbers already found are used to verify whether a number is prime. For example, to verify that 63 is a prime number, we check if it is divisible by prime numbers at most the square root of 63 which are 2, 3, 5, or 7. Since the prime numbers 2, 3, 5, and 7 are prime numbers already found and stored, we do not have to regenerate them. With this technique, the computer code generated prime numbers and twin prime pairs up to 4×10^9. We use R for statistical analysis and visualization.

Results

In our first analysis, we would like to understand how many twin primes are there among prime numbers. For example, in the interval [20,40] there are four prime numbers 23, 29, 31, and 37 with only one set of twin prime 29 and 31. We say that the density of twin primes among the primes is 1 out of 4 between 20 and 40. In Figure 1, we plot the ratio of twin primes and the number of primes for interval of size 2^{12}. The intervals are of the form $[a, a+2^{12}]$ where a

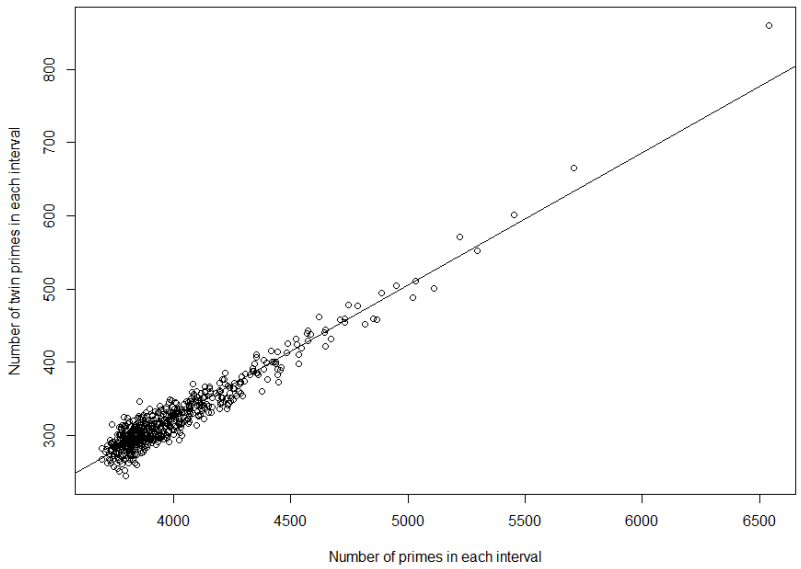

Figure 1: The graph models the number of primes in each interval to the

starts with 0, 2^{12}, 2^{13}, and so on. The x-axis is the number of primes in various intervals, and the y-axis is the number of twin primes. As a approaches infinity, we expect the number of primes to decrease. Consequently,

there is a cluster of points when for small values along the x-axis. To our surprise, we expect the ratio of twin primes to the number of primes to decrease. Instead, we observe that the ratio is constant as evident in the fact that the points cluster along a line instead of a curve of nonlinear function. The graph indicates that as the intervals gets farther out ($a \to \infty$), the number of twin primes is asymptotically of the form $C \cdot N(a)$ where $N(a)$ is the number of twin primes in the interval for integers from a to $a + 2^{12}$.

To better understand the percent of twin primes among the set of primes, we plot the ratio or the percentage of twin primes to the number of primes in Figure 2. The x-axis is the n^{th}-interval and the y-axis is the percentage. It is very surprising that the percent of twin primes hovers about 8% of the total of prime numbers in the interval even as the intervals are approaching infinity.

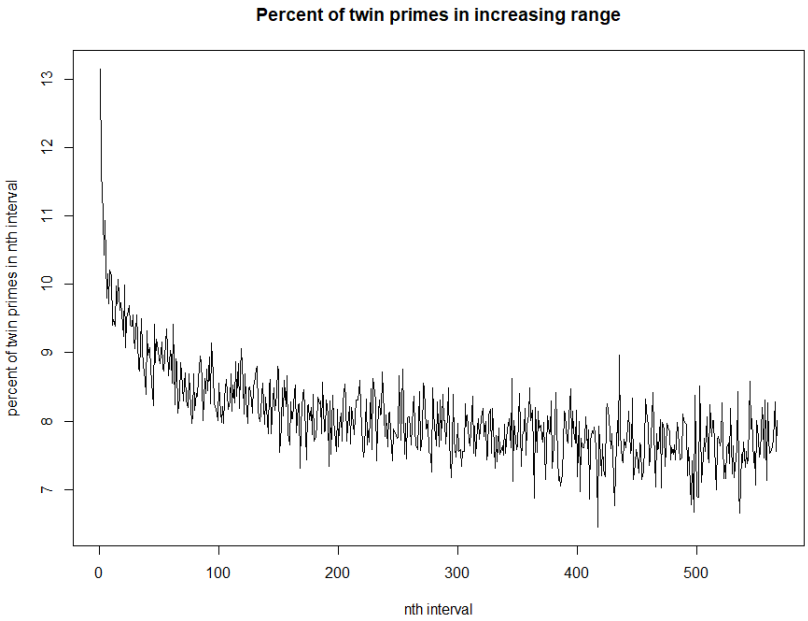

Figure 2: This models on the percentages of twin primes (in Figure 1) in each nth interval.

In the last part of the study, we examine how rare twin prime numbers become for large integers. We study the midpoint of consecutive intervals that contain a fixed number of twin primes. Starting with 40 primes, we computed the first interval and determined an average midpoint. Then, we determined the next interval that contains another set of 40 twin primes and found another average midpoint. In Figure 3, we plotted the logarithm of approximate midpoints versus the logarithm of average gap or size of the interval. As expected, the size of the interval is increasing as the number of twin primes spread out and becomes rarer. We observed a linear relationship among the logarithm of the midpoints and the logarithm of interval sizes.

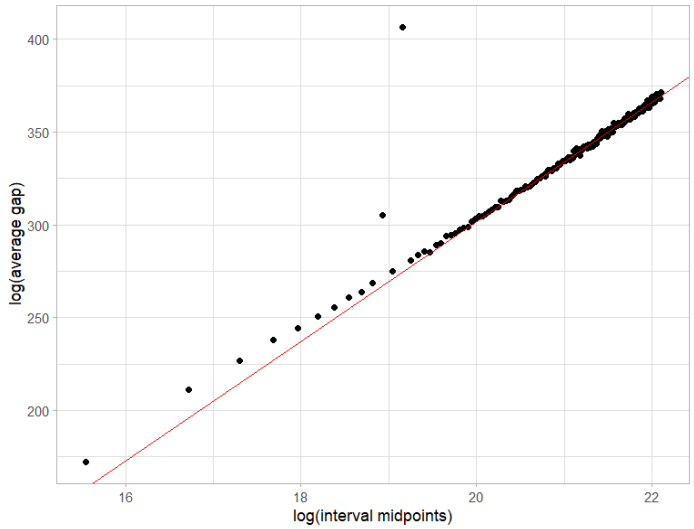

Figure 3: The relationship between interval midpoints and average prime gap with a log scale is modeled.

Conclusion

It is clear that there is some underlying mathematical relationship between the distribution of twin prime numbers in the set of prime numbers from the results we obtained. The results are surprising since we did not expect a functional relationship to exist. The linear relationship we obtained was even more surprising. With the limitation of the study restricted to only the first 4×10^9 integers, it is possible that the linear trend disappears for larger data sets. For future studies, we plan to increase the size of the data sets and investigate reasons as to why a functional relationship between the number of twin primes in a set of prime numbers should exist.

References

Balasubramanian, K., & Carbo-Dorca, R. (2022). Three conjectures on extended twin primes and the existence of isoboolean and singular primes inspired by relativistic quantum computing. https://doi.org/10.1007/s10910-022-01364-5

Gensel, Berndt. (2020) An Elementary Proof of the Twin Prime Conjecture. https://doi.org/10.48550/arXiv.1909.07975

Morpurgo, G. (2023). On a Stricter Twin Primes Conjecture, and on the Polignac's Conjecture in general. https://doi.org/10.48550/arXiv.2210.15487

Sellers, J. K. (2021). A step towards proving de Polignac's Conjecture. https://doi.org/10.48550/arXiv.2108.13834

Mothebe, M. F., Kagiso, D. N., & Modise B. T. (2021). A proof of cases of de Polignac's conjecture. https://doi.org/10.48550/arXiv.1912.09290

Jankovic, M.V. (2022). Proof of the Twin Prime Conjecture (Together with the Proof of Polignac's Conjecture for Cousin Primes). https://doi.org/10.48550/arXiv.1903.05410

Zhang, Yitang. (2014) Bounded gaps between primes. https://doi.org/10.4007/annals.2014.179.3.7