Goldbach's Conjecture Project

by Tuan Bui

AMS 325

Project's title: Distribution of Goldbach's Conjecture and Goldbach's Weak Conjecture

Abstract:

This is my analysis project on Goldbach's strong and weak conjectures. By applying the Sieve of Eratosthenes to find prime numbers up to a given upper bound number, I can easily find Goldbach pairs in a reasonable time. In programming, using the Sieve method is a friendly memory way to generate prime numbers. The Goldbach pairs of both conjectures are visualized in the residue class of 3, and multiplication classes of 3, 5, and 7, which is also called Goldbach's comet.

Introduction:

1. **Goldbach's Conjecture**

It has been almost 280 years since Goldbach's Conjecture was first proposed. It is by far one of the most famous and enduring problems still unsolved in the world. In the letter, Christian Goldbach mentioned a conjecture that every natural even number greater than two is the sum of two prime numbers. According to Wikipedia, "T. Oliveira e Silva ran a distributed computer search that has verified the conjecture for $n \leq 4 \times 10^{18}$ (and double-checked up to 4×10^{17}) as of 2013."

2. Goldbach's Weak Conjecture:

In number theory, Goldbach's weak conjecture, also known as the odd Goldbach conjecture, the ternary Goldbach problem, or the 3-primes problem, states that "Every odd number greater than 5 can be expressed as the sum of three primes." In fact, it can be easy to prove if the strong conjecture is proven.

Proof. Suppose n is a number greater than 5, then $n - 3$ is greater than 2. Thus, if the strong conjecture is true, then $n - 3 = p + q$, for some prime numbers p, q. Therefore, if the strong conjecture holds, then the weak conjecture is also true.

Goldbach's weak conjecture was proven in 2013 by Harald Helfgott, a Peruvian mathematician. His proof has not yet appeared in a peer-reviewed publication, though was accepted for publication in the Annals of Mathematics Studies series in 2015, and has been undergoing further review and revision since.

Some state the conjecture as "Every odd number greater than 7 can be expressed as the sum of three odd primes." I use this statement for my project analysis.

3. The Sieve of Eratosthenes:

In mathematics, the Sieve of Eratosthenes is an ancient algorithm for finding all prime numbers up to any given limit N. It does so by iteratively marking as composite (i.e., not prime) the multiples of each prime, starting with the first prime number, 2. Once all the multiples of each discovered prime have been marked as composites, the remaining unmarked numbers are primes. Because a composite of a number n will be the product of integers smaller than or equal to $\sqrt{n}$, we only need to check for multiples of primes that are smaller than or equal to $\sqrt{n}$. Here is a proof for it:

Proof. Suppose a positive integer n is not prime, then n can be written as a product of two integer a and b, greater than 1 and $a \leq b$. Assume $a > \sqrt{n}$, then $ab \geq a^2 > n$, a contradiction, since $n = ab$. ■

Techniques:

The main data and plots in the project are generated by using Python. For a limitation of time, I used some data and plots generated by R, which were also done by myself since last year. Github is also used for version control, storing, and share code scripts, and data. I also used Jupyter Notebook for testing and debugging the codes. In my program, a few useful Python and R libraries are applied for analysis and comparison.

Algorithm:

1. **The Sieve of Eratosthenes:**

I created a (n+1)-element boolean list the initial value is True based on the list's Indices, from 0 to n. Because 0 and 1 are not prime, they are assigned False values. The program will begin with the first prime number, 2. Then, all the multiples of 2 are assigned False values. As I mentioned, the program will stop when it reaches $\sqrt{n}$. Finally, it will return the indices of all True elements left after the process, a prime list.

2. Goldbach’s conjecture:

The algorithm for the strong and weak conjecture are very similar. I will mention about the weak conjecture because there are 3 prime numbers in each partition. The strong conjecture is applied the same idea with it.

Suppose a set of Goldbach′s weak conjecture pairs:

$W = \{(num1,\ num2,\ num3) \mid num1,\ num2,\ num3 \text{ are prime } \&\ num1 \leq num2 \leq num3\}$

That implies $num1 \leq \frac{n}{3}$, for $n = num1 + num2 + num3$.

For $num1,\ num2$ in a prime list that was computed by Sieve method, check if $num3$ is also prime.

Results and Conclusion:

Using the Sieve of Eratosthenes method to find prime numbers is significantly time-consuming, and memory friendly. In my project, I use the isprime() function in sympy library as a counter-method against the Sieve of Eratosthenes to compare the elapsed running time. As expected, the running time using Sieve method is more efficiency. It is about 9 times better using isprime() function.

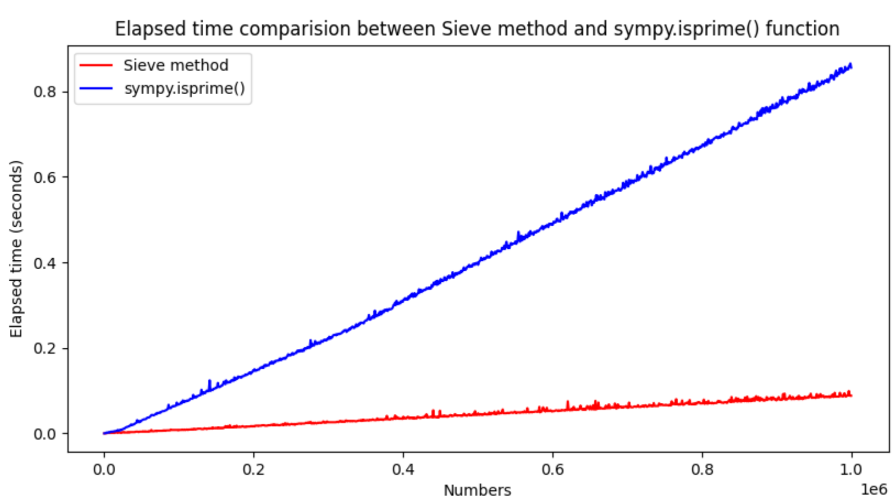

Fig 1. The running time between the Sieve of Eratosthenes and sympy.isprime(), generated by Python 3.9.5

Based on the plot above, the slope of the sympy.isprime() observations is larger than the slope of the Sieve method's, which indicates that the elapsed time is significantly increased as a number getting larger. Hence, the Sieve of Eratosthenes is significantly using in computing massive prime set.

According to the Prime number theory, every prime number can be written in the form of 6m $\pm$ 1. Assume that the conjecture holds from the first 10^6 natural even numbers:

$$(6a + 1) + (6b - 1) = 6(a + b) + 0 \equiv 0 \; mod \; 6 \equiv 0 \; mod \; 3$$

$$(6a - 1) + (6b + 1) = 6(a + b) + 0 \equiv 0 \; mod \; 6 \equiv 0 \; mod \; 3$$

$$(6a + 1) + (6b + 1) = 6(a + b) + 2 \equiv 2 \; mod \; 6 \equiv 2 \; mod \; 3$$

$$(6a - 1) + 3 = 6a + 2 \equiv 2 \; mod \; 6 \equiv 2 \; mod \; 3$$

$$(6a - 1) + (6b - 1) = 6(a + b) - 2 \equiv 4 \; mod \; 6 \equiv 1 \; mod \; 3$$

$$2 + 2 \equiv 4 \; mod \; 6 \equiv 1 \; mod \; 3$$

$$(6a + 1) + 3 = 6a + 4 \equiv 4 \; mod \; 6 \equiv 1 \; mod \; 3$$

From the result, I can conclude that every even integer number can written in the form of 3n $\pm$ 1. Consequently, I analyzed the relation between how the observations distribute and the congruence class of 3 as the plot below.

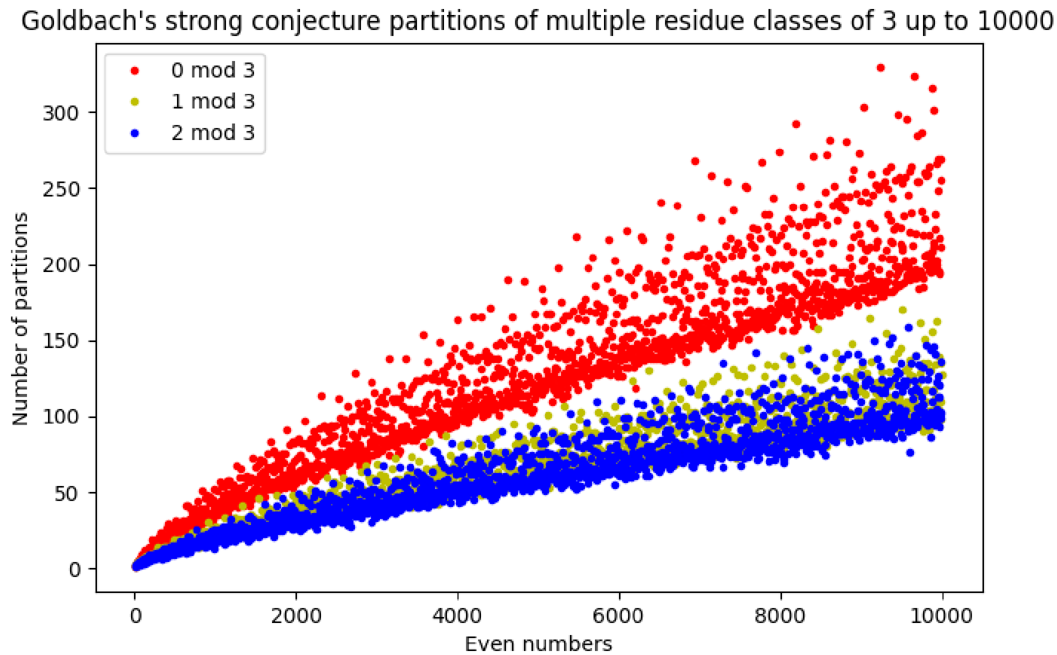

Fig 2. Goldbach partitions of the first 10000 natural even numbers,
generated by Python 3.9.5

The graph can probably be explained as a consequence of how the number of partitions varies between different congruence classes of 3. It just expresses the number of partitions up to 10^5, but it is clear that the number of partitions is steadily rising. That would make me assume that the probability of finding a Goldbach pair for large numbers is more than the smaller ones', which means conjecture holds.

I used regression analysis to see the relations between the number of partitions and the natural even number, in form of $y = \beta_0 + \beta_1 x + \beta_i * (r = 0)$. r is a remainder when a number is divided by 3, and set as a binary variable.

```
Call:
lm(formula = y ~ ., data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-327.66  -81.93    4.36   66.49 1027.92 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.286e+02  1.399e+00   234.9   <2e-16 ***
x            8.642e-03  1.978e-05   436.8   <2e-16 ***
r2          -3.834e+02  1.399e+00  -274.1   <2e-16 ***
r1          -3.765e+02  1.399e+00  -269.1   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 127.7 on 49995 degrees of freedom
Multiple R-squared:  0.8526,	Adjusted R-squared:  0.8526 
F-statistic: 9.638e+04 on 3 and 49995 DF,  p-value: < 2.2e-16
```

Fig 3. Summary of the regression method fit lines,
generated by R 4.2.0

The square of correlation coefficient for this model is 0.8526, which indicates that the model fits the observations very well. Regression fit function is found as:

$$y = \beta_0 + \beta_1 x + \beta_2(r = 1) + \beta_3(r = 2)$$

with:

$$\beta_0 = 3.286\mathrm{e} + 02, \beta_1 = 8.642\mathrm{e} - 03, \beta_2 = -3.765\mathrm{e} + 02, \text{and } \beta_3 = -\ 3.834\mathrm{e} + 02$$

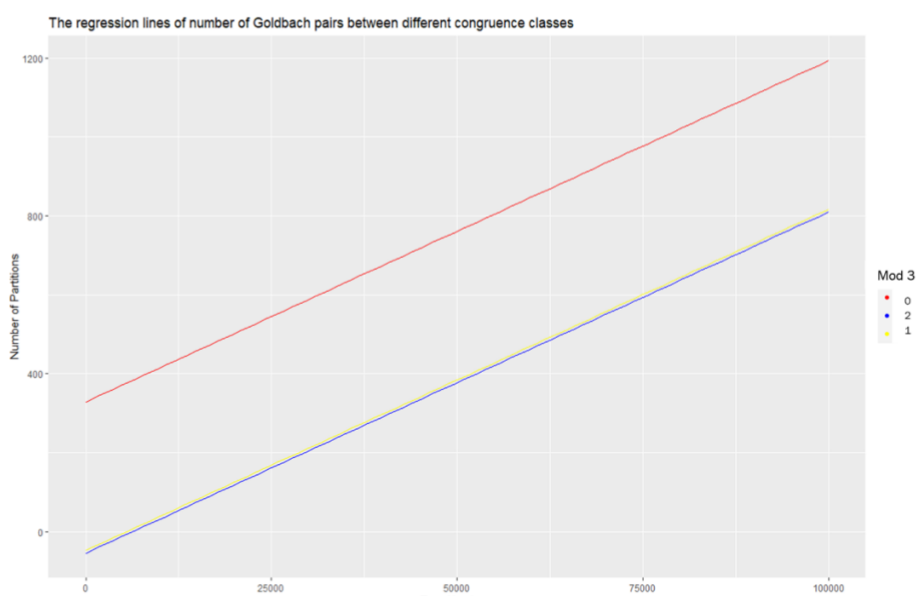

Fig 4. The regression method fit lines of different residue classes of 3,
generated by R 4.2.0

As expected, the even numbers in group 0 mod 3 have more Goldbach pairs than the ones in groups 1 and 2 mod 3. I also see that the slope of the 0 mod 3 group fit line is over twice the slope of the others. The slope of group 1 and 2 mod 3 fit lines are likely close.

References:

1. "Goldbach's conjecture" Wikipedia, Wikimedia Foundation,

   https://en.wikipedia.org/wiki/Goldbach%27s_conjecture.

2. "Goldbach's weak conjecture" Wikipedia, Wikimedia Foundation,

   https://en.wikipedia.org/wiki/Goldbach%27s_weak_conjecture.

3. "Sieve of Eratosthenes" Wikipedia, Wikimedia Foundation,

   https://en.wikipedia.org/wiki/Sieve_of_Eratosthenes.