- MA3001 Analytic Number Theory -

Elementary sieve methods and Brun's theorem on twin primes

Tjerand Silde

April 23, 2014

Contents

Introduction

The purpose of this project is to study elementary sieve methods used to estimate the size of the finite subsets of the prime numbers less than or equal to a positive real number x. In addition, we will have a look at the twin primes. Concretely, we will study the theorems of the Norwegian mathematician Viggo Brun about the density of twin primes, and the reciprocal sum of twin primes converging to Brun's constant $\mathcal{B}$.

The Sieve of Eratosthenes

The most famous method to detect prime numbers is the sieve of Eratosthenes. The method starts by the first prime number, which is 2, and rule out all multiples of 2 up to a given number n. Then the next number to the right on the line of positive integers, which is not ruled out, is a prime. So it continues. Example $n = 25$:

2	3	$\cancel{4}$	5	$\cancel{6}$	7	$\cancel{8}$	9	$\cancel{10}$	11	$\cancel{12}$	13
$\cancel{14}$	15	$\cancel{16}$	17	$\cancel{18}$	19	$\cancel{20}$	21	$\cancel{22}$	23	$\cancel{24}$	25

2	**3**	$\cancel{4}$	5	$\cancel{6}$	7	$\cancel{8}$	$\cancel{9}$	$\cancel{10}$	11	$\cancel{12}$	13
$\cancel{14}$	$\cancel{15}$	$\cancel{16}$	17	$\cancel{18}$	19	$\cancel{20}$	$\cancel{21}$	$\cancel{22}$	23	$\cancel{24}$	25

2	**3**	$\cancel{4}$	**5**	$\cancel{6}$	7	$\cancel{8}$	$\cancel{9}$	$\cancel{10}$	11	$\cancel{12}$	13
$\cancel{14}$	$\cancel{15}$	$\cancel{16}$	17	$\cancel{18}$	19	$\cancel{20}$	$\cancel{21}$	$\cancel{22}$	23	$\cancel{24}$	$\cancel{25}$

All non-prime numbers n has at least one non-trivial factor less than or equal to the square root of n. For all the factors p greater than the square root of n, there mus exist a factor q less than the square root of n such that $p \cdot q = n$. It is then clear that all the numbers which reminds when ruled out all multiples of the primes less than or equal to the square root of n is prime, hence:

2	**3**	$\cancel{4}$	**5**	$\cancel{6}$	**7**	$\cancel{8}$	$\cancel{9}$	$\cancel{10}$	**11**	$\cancel{12}$	**13**
$\cancel{14}$	$\cancel{15}$	$\cancel{16}$	**17**	$\cancel{18}$	**19**	$\cancel{20}$	$\cancel{21}$	$\cancel{22}$	**23**	$\cancel{24}$	$\cancel{25}$.

Brun's theorem on twin primes

Definition 1. *A twin prime p is a prime where either $p - 2$ or $p + 2$ also is prime.*

We know that the sum of the reciprocals of all primes is divergent, and that there exist infinitely many prime numbers. However, this is not the case for twin primes. We don't know if there exist finite or infinitely many twin primes, but the Norwegian mathematician Viggo Brun proved that the sum of the reciprocals of all twin primes p is convergent, and converges to Brun's

constant $\mathcal{B} = 1.90216054$.

$$\sum_p \frac{1}{p} = \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{29} + \frac{1}{31} + \cdots = \mathcal{B}.$$

Remark that the prime 5 belongs to the two twin prime pairs $(3, 5)$ and $(5, 7)$. No other prime can have this property. This follows from the fact that one out of three consecutive odd numbers must be divisible by 3.

The goal of this section is to obtain Brun's theorem, and the process to obtain this result is divided into three parts.

Brun's Theorem. *There exists a positive constant C so that $\pi_2\ (x)$,the number of twin primes not exceeding x, satisfies, for $x > 3$,*

$$\pi_2(x) < C \cdot x \cdot \left(\tfrac{\log\log x}{\log x}\right)^2.$$

First step of Brun's method

The first step of Brun's method is to apply a double sieving to the sequence of the natural numbers. The purpose of this is to eliminate all those numbers n for which n or $n + 2$ is composite.

Let $\mathbf{T}(x)$ denote the number of first members n of pairs of twin primes for which $n \leq x$ and let $\mathbf{U}(x; y)$ be the number of odd numbers $n \leq x$ for which $n(n + 2)$ is not divisible by any of the odd primes $p_j \leq y$. The first observation we can make is the relationship between $\mathbf{T}$ and $\mathbf{U}$ when $p_j \leq y \leq \sqrt{x + 2}$,

$$\mathbf{T}(x) \leq r + \mathbf{U}(x; y),$$

where the r represent the number of primes $p_j \leq y$, as they may be twin primes as well. The important choice of y will be the work of step three. Now, as the left-hand side is the function which count all the first members of twin primes (by definition), we have to examine the right-hand side of the equation, or more precisely, $\mathbf{U}(x; y)$. This function can be rewritten as

$$\begin{aligned}\mathbf{U}(x; y) = \left[\frac{x+1}{2}\right] - \sum_i \mathbf{B}(x; p_i) + \sum_{i<j} \mathbf{B}(x; p_i \cdot p_j) \\ - \sum_{i<j<k} \mathbf{B}(x; p_i \cdot p_j \cdot p_k) + \cdots + (-1)^r \cdot \mathbf{B}(x; p_1 \cdot p_2 \cdots p_r),\end{aligned}$$

where $\mathbf{B}(x; p_1 \cdot p_2 \cdots p_r)$ count the number of odd numbers $n \leq x$ for which $n(n + 2)$ **is** divisible by the product $p_1 \cdot p_2 \cdots p_r$.

The first term on the right-hand side represent all the odd numbers between zero and x. Then we have counted all of them once, however, all the odd integers

n for which $n(n+2)$ is divisible by at least one of the primes $p \leq y$ is counted exactly zero times by the alternating sums of the function $\mathbf{B}$. Let $n(n+2)$ be divisible by the f primes $p_1, p_2, \cdots, p_f$, and only these. Then n is counted

$$\begin{aligned}
&1 && \textit{time in } \left[\frac{x+1}{2}\right] \\
&f && \textit{times in } \sum_i \mathbf{B}(x; p_i) \\
&\binom{f}{2} && \textit{times in } \sum_{i<j} \mathbf{B}(x; p_i \cdot p_j) \\
&\binom{f}{3} && \textit{times in } \sum_{i<j<k} \mathbf{B}(x; p_i \cdot p_j \cdot p_k) \\
&\vdots \\
&\binom{f}{f} && \textit{times in } \sum_{i<j<\ldots<f} \mathbf{B}(x; p_i \cdot p_j \cdots p_f),
\end{aligned}$$

as this sequence alternates, we get the sum

$$1 - \binom{f}{1} + \binom{f}{2} - + \cdots + (-1)^f \binom{f}{f} = (1-1)^f = 0.$$

Conclusion, we have only counted the odd numbers n such that $n(n+2)$ is not divisible by any of the primes $p_j \leq \sqrt{x+2}$, hence, this is an equality. Further, this is no less work than count each number itself, so we want to give an upper bound that is much simpler to calculate. We then choose an even number $m < r$, and by the following lemma

Lemma 1.

$$\sum_{\lambda=0}^{m} (-1)^\lambda \binom{f}{\lambda} = \begin{cases} = 0 \ \textit{for } m \geq f > 0 \\ > 0 \ \textit{for } m < f, m \textit{ even} \\ < 0 \ \textit{for } m < f, m \textit{ odd} \end{cases},$$

we can give the estimate

$$\mathbf{U}(x; y) < \left[\frac{x+1}{2}\right] + \sum_{f=1}^{m} (-1)^f \sum_{\rho^{(f)}} \mathbf{B}(x; p^{\rho^{(f)}}).$$

Lemma 1 is easily proved by symmetry. To improve this, we may replace $\mathbf{B}(x)$ by the result in the next lemma.

Lemma 2. *Let ρ be an odd number and $\nu(\rho)$ the number of its different prime factors. Then the number $\boldsymbol{B}(x; \rho)$ of odd numbers $n \leq x$ for which $n(n+2)$ is divisible by ρ is*

$$\boldsymbol{B}(x;\rho) = 2^{\nu(\rho)}\left\{\tfrac{x}{2\rho} + \Theta\right\}, |\Theta| \leq 1.$$

Here, Θ is either 0 or 1. This follows from the Chinese remainder theorem, by taking a closer look at the two congruences

$$\begin{aligned} n &\equiv 0 \mod p_\alpha \cdots p_\beta \\ n &\equiv 0 \mod p_\delta \cdots p_\epsilon, \end{aligned}$$

where ρ is divided into products of primes which either divides n or $n+2$. Then there exists a unique solution, and by counting all the solutions for all ρ, we get the lemma. Θ depends if x is even or odd, and hence is either 0 or 1.

We observe that for ρ big, the lemma is useless, because Θ is dominating. This was one of Brun's brilliant observartions. That's the reason that we chose an even number $m < r$ to stop the sum. Let $\rho^{(f)}$ denote a number which is the product of f different prime factors taken from the primes $p_j \leq u$, then the first step of Brun's method results in

$$\mathbf{T}(x) \leq r + \frac{x}{2}\sum_{f=0}^{m}(-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} + \sum_{f=0}^{m}\sum_{\rho^{(f)}} 2^f.$$

Second step of Brun's method

Next, we want to give an estimate which only depends on x and y. We will evaluate each of the sums in the result in last section. Starting by the first term, we may split it into two parts and evaluate each of them as follows

$$\begin{aligned} &= \sum_{f=0}^{m}(-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} \\ &= \sum_{f=0}^{r}(-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} - \sum_{f=m+1}^{r}(-1)^f \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} \\ &= \prod_{j=1}^{r}(1 - \frac{2}{p_j}) + \sum_{f=m+1}^{r}(-1)^{f-1} \cdot s_f, \end{aligned}$$

for

$$s_f = \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}}, s_f \leq \frac{s_1^f}{f!}.$$

Here, s_f is the f'th principal sum. Then we have

$$s_1 = \sum_{\rho^{(f)}} \frac{2^f}{\rho^{(f)}} = 2 \cdot \sum_{3 \leq p \leq y} \frac{1}{p} = 2 \cdot \log\log y + O(1),$$

where the last equivalence for primes p is shown in lecture, such that

$$\prod_{j=1}^{r}(1-\frac{2}{p_j}) < e^{-2\cdot\sum_{j=1}^{r}\frac{1}{p_j}} = e^{-s_1}$$

and

$$\sum_{f=m+1}^{r}(-1)^{f-1}\cdot s_f \leq s_{m+1} \leq \frac{s_1^{m+1}}{(m+1)!} \leq \left(\frac{es_1}{m+1}\right) \leq (\frac{1}{e})^{m+1} \leq e^{-e^2 s_1} \leq e^{-s_1},$$

by Sterlings formula $n! > \left(\frac{n}{e}\right)$ if we choose $e^2 s_1 < m+1 < 9s_1$. The second term is more straight forward, by computations, we have

$$\sum_{f=0}^{m}\sum_{\rho^{(f)}} 2^f = \sum_{f=0}^{m}\binom{r}{f}2^f < \sum_{f=0}^{m}(2r)^f < \frac{(2r)^{m+1}}{2r-1} \leq (2r)^{m+1}.$$

This results in the equation, by the trivial estimate $y > 2r$,

$$\mathbf{T}(x) \leq y + xe^{-s_1} + y^{9s_1}.$$

Third step of Brun's method

In this last section we have to choose a suitable y. For some positive constant c, we have

$$2\cdot\log\log y - c < s_1 < 3\cdot\log\log y.$$

Thus, $\mathbf{T}$(x) may be written as

$$\mathbf{T}(x) \leq y + e^c\frac{x}{(\log y)^2} + y^{27\cdot\log\log y}.$$

As we want the middle term to grow fastest, we choose

$$y = x^\gamma, 0 < \gamma \leq \frac{1}{2},$$

hence

$$\mathbf{T}(x) \leq x^{1/2} + e^c\frac{x}{(\gamma\log x)^2} + y^{27\cdot\log\log y}.$$

The last step is to choose

$$\gamma = \frac{1}{30\cdot\log\log x}, x \geq 3.$$

From the new estimate

$$\mathbf{T}(x) \leq x^{1/2} + 900\cdot e^c\cdot x\cdot\left(\frac{\log\log x}{\log x}\right)^2 + x^{9/10}$$

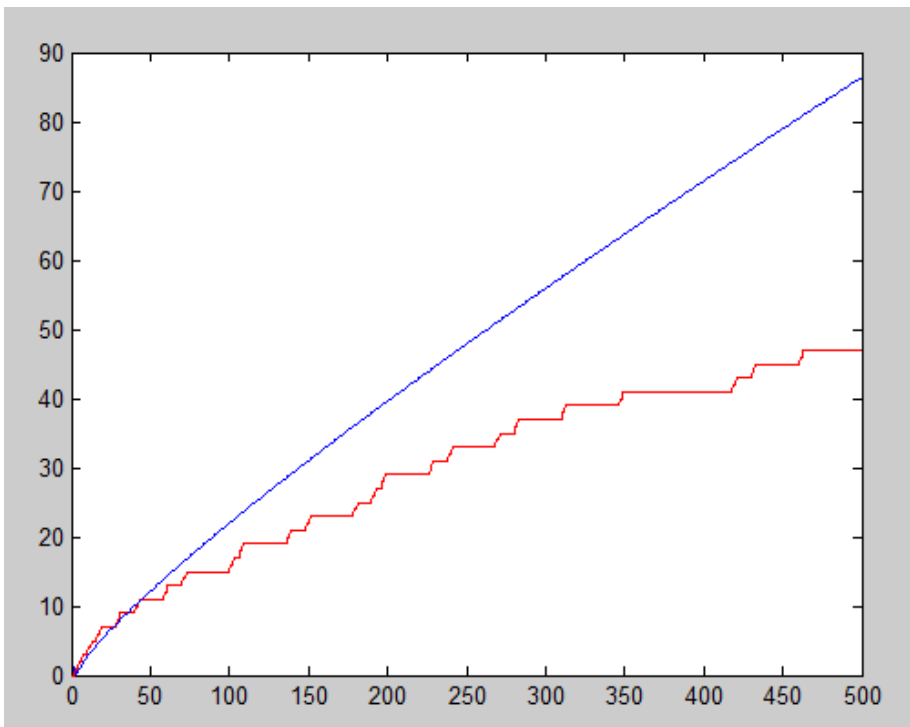
and the fact that $\pi_2 \leq 2\mathbf{T}(x)$, we have obtained Brun's theorem

Brun's Theorem. *There exists a positive constant C so that $\pi_2\ (x)$,the number of twin primes not exceeding x, satisfies, for $x > 3$,*

$$\pi_2\ (x) < C \cdot x \cdot \left(\frac{\log \log x}{\log x}\right)^2.$$

Plot Brun's theorem

Plot for $C = 2$:



Analysing the theorem

The partial sums of Bruns theorem is

$$\begin{aligned}
S(x) = & \sum_{twin\ prime\ p \leq x} \frac{1}{p} \\
= & \sum_{3 \leq odd\ n \leq x} \frac{1}{n}(\pi_2(n) - \pi_2(n-2)) \\
= & 2 \cdot \sum_{3 \leq odd\ n \leq x} \pi_2(n) \frac{1}{n(n+2)} + \mathcal{O}\left(\frac{1}{n}\left[\frac{\log\log x}{\log x}\right]^2\right) \\
= & \mathcal{O}\left(\sum_{3 \leq n \leq x} \frac{1}{n}\left[\frac{\log\log x}{\log x}\right]^2\right).
\end{aligned}$$

We know that

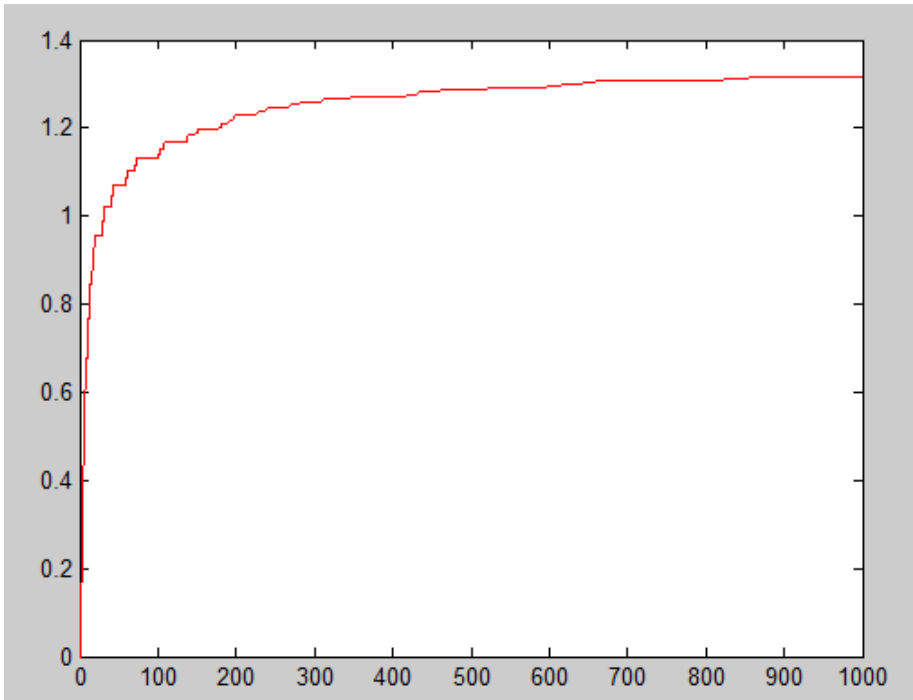
$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^{1+\epsilon}}$$

is convergent for $\epsilon > 0$. Hence

$$\sum_{n=3}^{\infty} \frac{(\log\log n)^2}{n(\log n)^2}$$

is convergent. This proves the that the sum of the reciprocals of all twin primes p is convergent.

Plot Brun's theorem



We observe that the sum is converging very very slowly, and the result after 1000 iterations is far from the estimate expected. To get a better estimate we need to evaluate a product which runs over a lot more twin primes, but as these calculations require a lot of computer-power, the reader have to believe that the sum converges to $\mathcal{B} = 1.90216054$.

Python code

As it may be interesting for the reader to test the estimates given when following Brun's three steps, the functions programmed in Python is attached.

```
import math

#function isPrime(n)
#input: some positive integer n
#output: 1 if the number is prime,
#0 otherwise
def isPrime(n):
        return not (n < 2 or any(n % i == 0 for i in range(2, int(n ** 0.5) + 1)))

#function T1(x)
#input: some positive integer x
#output: number of twinprimes n
#such that both n and n+2 is prime where n <= x
def T1(x):
        return len([i for i in range(x)if isPrime(i)and isPrime(i+2)])

#function T2(x)
#input: some positive integer x
#output: number of twinprimes n
#such that both n and either n+2
#or n-2 is prime, where n <= x
def T2(x):
        return len([i for i in range(x)if isPrime(i)
                and (isPrime(i-2) or isPrime(i+2))])

#function U(x,primes)
#input: some positive integer x
#and a list of odd primes
#output: the number of odd numbers
#n such that n <= x and n*(n+2) is
#not divisible by any of the primes in the list
def U(x, primes):
        return len([n for n in range(math.ceil(x/2.0))
                if len([div for div in primes if
                div<math.sqrt(x+2) and (2*n+1)*((2*n+1)+2)%div==0])==0])

#function B(x,primeproduct)
#input: some positive integer x
#and a product of odd primes
#output: the number of odd numbers
#n such that n <= x and n*(n+2) is
#divisible by the product of the odd primes
def B(x, pp):
        return len([n for n in range(math.ceil(x/2.0))
                if (2*n+1)*((2*n+1)+2)%pp==0])
```

```
#function v(p)
#input: some positive odd integer p
#output: the number of p's different
#prime factors
def v(p):
        return len([(i+1) for i in range(p)if p%(i+1)==0
                and isPrime(i+1)])
```

Bibliography

[1] Hans Rademacher; Lectures on elementary number theory.
Blasidell Publishing Company, 1st Edition, 1964.