Selberg's Sieve I

Peiran Tao

Department of Mathematics
University of Waterloo

July 23rd, 2024

Overview

1. Notations
2. Sieve of Eratosthenes
3. Selberg's Sieve

Notations

1. $\mathbb{N}$ = the set of natural numbers (positive integers).
2. $\mathbb{P}$ = the set of all prime numbers.
3. For $x > 0$, let:

$$\pi(x) = \# \text{ of prime numbers } \leq x$$

   to be the prime counting function.
4. For nonzero $a, b \in \mathbb{N}$, denote:

$$(a, b) := \gcd(a, b) \quad \text{and} \quad [a, b] := \operatorname{lcm}(a, b)$$

Sieve Method

Let A be an arbitrary set (usually a subset of $\mathbb{N}$).

Sieve Methods are techniques used to estimate the size of A after elements with some undesirable property have been removed.

Sieve of Eratosthenes

A classic application of sieve method is to estimate $\pi(x)$.

To estimate $\pi(x)$ is the same as estimating the size of $[1, x] \cap \mathbb{P}$.

Using the language of sieve method, let $A = [1, x] \cap \mathbb{N}$. To find all primes, we want to estimate A after removing 1 and all composite numbers.

Characterize composite numbers

Theorem (1.1)

Let $x \geq 2$ be a real number. Let $n \in \mathbb{N}$ with $2 \leq n \leq x$. If n is composite, then n has a prime factor p with $p \leq \sqrt{x}$.

Proof: Suppose the result is not true. Since n is composite, it must have at least two prime factors p, q (not necessarily distinct). Then $p, q > \sqrt{x}$, so:

$$n \geq pq > \sqrt{x}\sqrt{x} = x$$

which is a contradiction. $\square$

Sieve of Eratosthenes

So, to remove all composite numbers, it suffices to remove all integers in A that do not satisfy the property in Lemma 1.1.

For $x \geq 2$, if we remove all the multiplies of primes $\leq \sqrt{x}$ in A, the numbers that remain are primes numbers in $(\sqrt{x}, x]$ and the number 1, thus:

$$\pi(x) - \pi(\sqrt{x}) + 1 = \pi(x, \sqrt{x}) \tag{1.1}$$

Here $\pi(x, \sqrt{x})$ denote the number of $n \leq x$ with no prime factors $\leq \sqrt{x}$.

Generalization

Instead of removing multiples of primes $\leq \sqrt{x}$, we can replace $\sqrt{x}$ with an arbitrary $z > 0$.

Definition

Let $A \subseteq \mathbb{N}$ be a finite subset of $\mathbb{N}$. Let $P \subseteq \mathbb{P}$ be a set of prime numbers and let $z > 0$. Define:

$$S(A, P, z) = \#\ \textit{of } a \in A \textit{ that is not divisible by any } p < z \textit{ with } p \in P$$

Generalization

If we define:

$$P_z = \prod_{\substack{p \in P \\ p < z}} p$$

For $p \in P$ and $p < z$, we have $p \mid a$ if and only if $(a, P_z) > 1$.

Therefore, we can rewrite $S(A, P, z)$ as:

$$S(A, P, z) = \sum_{\substack{a \in A \\ (a, P_z) = 1}} 1 = \sum_{a \in A} F(a)$$

where:

$$F(a) = \begin{cases} 1 & \text{if } (a, P_z) = 1 \\ 0 & \text{if } (a, P_z) > 1 \end{cases}$$

Generalization

So, is there a nice function that behave like F?

Let $n \in \mathbb{N}$. Define the **Möbius function**:

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n \text{ is not squarefree} \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ is squarefree} \end{cases}$$

Lemma (1.2)

Let μ denote the Möbius function, then:

$$I(n) := \sum_{d|n} \mu(d) = \begin{cases} 1 & \textit{if } n = 1 \\ 0 & \textit{if } n > 1 \end{cases}$$

Generalization

Proof: If $n = 1$, trivial. Otherwise, write $n = p_1^{e_1} \cdots p_r^{e_r}$. Since $\mu(d) = 0$ if d is not squarefree, we have:

$$\begin{aligned}
I(n) &= \sum_{\substack{d \mid n \\ d \text{ squarefree}}} \mu(d) \\
&= \sum_{(s_1, \cdots, s_r) \in \{0,1\}^r} \mu(p_1^{s_1} \cdots p_r^{s_r}) \\
&= \sum_{(s_1, \cdots, s_r) \in \{0,1\}^r} (-1)^{s_1 + \cdots + s_r} \\
&= \left(\sum_{s_1 \in \{0,1\}} (-1)^{s_1} \right) \cdots \left(\sum_{s_r \in \{0,1\}} (-1)^{s_r} \right) \\
&= 0
\end{aligned}$$

As desired. $\square$

Generalization

By the lemma, we have:

$$I((a, P_z)) = \sum_{d|(a,P_z)} \mu(d) = \begin{cases} 1 & \text{if } (a, P_z) = 1 \\ 0 & \text{if } (a, P_z) > 1 \end{cases}$$

Hence, we have:

$$S(A, P, z) = \sum_{a \in A} \sum_{d|(a,P_z)} \mu(d) \tag{1.2}$$

Generalization

If we directly analyze the sum in (1.2), we can get the general Sieve of Eratosthenes, called the Legendre's Sieve.

But this talk is not called the Legendre's Sieve, so by contrapositive we are not going to analyze the sum directly.

Selberg's trick

Look at the sum (1.2):

$$S(A, P, z) = \sum_{a \in A} \sum_{d|(a,P_z)} \mu(d)$$

Note that $\sum_{d|(a,P_z)} \mu(d)$ is either 1 or 0, so:

$$\sum_{d|(a,P_z)} \mu(d) \leq \left(\sum_{d|(a,P_z)} \lambda_d \right)^2 \tag{2.1}$$

for any sequence $(\lambda_d) \subseteq \mathbb{R}$ with $\lambda_1 = 1$.

Selberg's trick

But obviously, we cannot choose (λ_d) to be an arbitrary sequence. We need to choose it so that the quadratic form with indeterminates λ_d:

$$\left(\sum_{d|(a,P_z)} \lambda_d\right)^2 = \sum_{d_1,d_2|(a,P_z)} \lambda_{d_1}\lambda_{d_2}$$

is minimal. Otherwise, our upper bound is too big, then this trick is useless.

Selberg's Sieve

Let A, P, z as in the setting in the above definition.

For $d \mid P_z$, we define:

$$A_d = \{a \in A : d \mid a\} = \sum_{\substack{a \in A \\ d \mid a}} 1$$

Suppose there is a multiplicative function f with $f(p) > 1$ for all prime $p \in P$ such that:

$$|A_d| = \frac{X}{f(d)} + R_d \tag{2.2}$$

1. Think of X as an estimation of $|A|$.
2. Think of (2.2) as an estimation of $|A_d|$, with $1/f(d)$ the 'density' of A_d in A, and R_d as the error term to the estimation.

Möbius Inversion

Lemma (2.1)

Let $f, F : \mathbb{N} \to \mathbb{C}$. *Then:*

$$F(n) = \sum_{d|n} f(d) \iff f(n) = \sum_{d|n} F(d)\mu\left(\frac{n}{d}\right)$$

This is known as the **Möbius Inversion Formula**.

Selberg's Sieve

Now we can start the derivation for Selberg's Sieve.

$$
\begin{aligned}
S(A, P, z) = \sum_{\substack{a \in A \\ (a, P_z) = 1}} 1 &= \sum_{a \in A} \sum_{d | (a, P_z)} \mu(d) \\
&\leq \sum_{a \in A} \left(\sum_{d | (a, P_z)} \lambda_d \right)^2 \\
&= \sum_{a \in A} \sum_{d_1, d_2 | (a, P_z)} \lambda_{d_1} \lambda_{d_2}
\end{aligned}
$$

Selberg's Sieve

Note that:

$$d \mid (a,b) \iff d \mid a \text{ and } d \mid b$$
$$[a,b] \mid \ell \iff a \mid \ell \text{ and } b \mid \ell$$

Therefore:

$$\begin{aligned} S(A,P,z) &\leq \sum_{a \in A} \sum_{\substack{d_1,d_2 \mid a \\ d_1,d_2 \mid P_z}} \lambda_{d_1}\lambda_{d_2} \\ &= \sum_{d_1,d_2 \mid P_z} \lambda_{d_1}\lambda_{d_2} \sum_{\substack{a \in A \\ d_1,d_2 \mid a}} 1 \\ &= \sum_{d_1,d_2 \mid P_z} \lambda_{d_1}\lambda_{d_2} \sum_{\substack{a \in A \\ [d_1,d_2] \mid a}} 1 \end{aligned}$$

Selberg's Sieve

$$S(A, P, z) \leq \sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} |A_{[d_1, d_2]}|$$

Recall that:

$$|A_d| = \frac{X}{f(d)} + R_d$$

We get:

$$\begin{aligned} S(A, P, z) &\leq \sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} \left(\frac{X}{f([d_1, d_2])} + R_{[d_1, d_2]} \right) \\ &= X \underbrace{\sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{f([d_1, d_2])}}_{T} + \underbrace{\sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} R_{[d_1, d_2]}}_{R} \end{aligned}$$

Selberg's Sieve

Hence we get:

$$S(A, P, z) \leq XT + R$$

Remember, our goal is to minimize this upper bound by choosing (λ_d) optimally.

Let us analyze T first.

The Main Term

By Möbius Inversion, there is $f_1 : \mathbb{N} \to \mathbb{C}$ such that:

$$f(n) = \sum_{d|n} f_1(n)$$

Explicitly, we define:

$$f_1(n) = \sum_{d|n} f(d)\mu\left(\frac{n}{d}\right)$$

For $n = p$ a prime, we get:

$$f_1(p) = \sum_{d|p} f(d)\mu\left(\frac{p}{d}\right) = f(1)\mu(p) + f(p)\mu(1) > 0$$

The Main Term

Lemma (2.2)

If f is multiplicative, then we have:

$$f([d_1, d_2])f((d_1, d_2)) = f(d_1)f(d_2)$$

We have:

$$\begin{aligned} T &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1}\lambda_{d_2}}{f([d_1, d_2])} \\ &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1}\lambda_{d_2}}{f(d_1)f(d_2)} f((d_1, d_2)) \\ &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1}\lambda_{d_2}}{f(d_1)f(d_2)} \sum_{\delta | (d_1, d_2)} f_1(\delta) \end{aligned}$$

The Main Term

Now, we choose $\lambda_d = 0$ for $d > z$. We have:

$$
\begin{aligned}
T &= \sum_{\substack{d_1, d_2 \le z \\ d_1, d_2 \mid P_z}} \frac{\lambda_{d_1}\lambda_{d_2}}{f(d_1)f(d_2)} \sum_{\delta \mid (d_1, d_2)} f_1(\delta) \\
&= \sum_{\substack{\delta \le z \\ \delta \mid P_z}} f_1(\delta) \sum_{\substack{d_1, d_2 \le z \\ d_1, d_2 \mid P_z \\ \delta \mid (d_1, d_2)}} \frac{\lambda_{d_1}\lambda_{d_2}}{f(d_1)f(d_2)} \\
&= \sum_{\substack{\delta \le z \\ \delta \mid P_z}} f_1(\delta) \left(\sum_{\substack{d \le z \\ d \mid P_z \\ \delta \mid d}} \frac{\lambda_d}{f(d)} \right)^2
\end{aligned}
$$

The Main Term

Define:

$$u_\delta = \sum_{\substack{d \le z \\ d | P_z \\ \delta | d}} \frac{\lambda_d}{f(d)}$$

Hence we get:

$$T = \sum_{\substack{\delta \le z \\ \delta | P_z}} f_1(\delta) u_\delta^2$$

The Main Term

It turns out, by another Inversion formula, we have:

$$\frac{\lambda_d}{f(d)} = \sum_{\substack{\delta | P_z \\ d | \delta}} \mu\left(\frac{\delta}{d}\right) u_\delta \tag{2.3}$$

Plug in $d = 1$ yields:

$$1 = \frac{\lambda_1}{f(1)} = \sum_{\delta | P_z} \mu\left(\delta\right) u_\delta$$

To choose λ_d, it suffices to choose u_δ.

The Main Term

It turns out that:

$$T = \sum_{\substack{\delta \leq z \\ \delta | P_z}} f_1(\delta) \left(u_\delta - \frac{\mu(\delta)}{f_1(\delta) V(z)} \right)^2 + \frac{1}{V(z)}$$

where:

$$V(z) = \sum_{\substack{\delta \leq z \\ d | P_z}} \frac{\mu^2(\delta)}{f_1(\delta)}$$

The Main Term

The first sum is non-negative as $f_1(p) > 1$ for all p.

So, T is minimized when:

$$u_\delta = \frac{\mu(\delta)}{f_1(\delta)V(z)}$$

So we can choose:

$$\lambda_d = f(d) \sum_{\substack{\delta | P_z \\ d | \delta}} \mu\left(\frac{\delta}{d}\right) u_\delta \tag{2.4}$$

Therefore, we have:

$$T = \frac{1}{V(z)}$$

The Error Term

The error term depends on λ_d. It turns out that, given:

$$\lambda_d = f(d) \sum_{\substack{\delta | P_z \\ d | \delta}} \mu\left(\frac{\delta}{d}\right) u_\delta$$

we must have $|\lambda_d| \leq 1$ for all d. Hence:

$$R \leq \left| \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} \lambda_{d_1} \lambda_{d_2} R_{[d_1, d_2]} \right| \leq \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}|$$

The final result

$$S(A, P, z) \leq \frac{X}{V(z)} + \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}| \tag{2.5}$$

Given a problem, if we want to apply Selberg's Sivee, we need to:

1. Find suitable A, P, z.
2. Estimate $|A_d|$ for $d \mid P_z$.
3. Find a lower bound for $V(z)$.