Lecture 5

The Sieve of Eratosthenes - Legendre

The sieve of Eratosthenes allows one to determine the primes not exceeding x assuming only knowledge of the primes not exceeding $\sqrt{x}$.

Recall that in this process, we write down all the numbers from 1 to x. Cross out 1 and for each prime $p \le \sqrt{x}$, we cross out all multiples of p on this list. The numbers remaining are the primes in $(\sqrt{x}, x]$.

Example: Finding all primes between 1 and 100, sifting by 2, 3, 5 & 7

~~1~~	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	83	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~

$\Rightarrow$ Primes in $(10, 100]$: 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

This proceedure is remarkable not only insofar as it gives a fast algorithm for listing primes, but also that it suggests the useful viewpoint of the primes as the integers surviving a "sieving process".

The aim of sieve theory is to construct estimates for the number of integers that remain in a set after members of certain arithmetic progressions are removed.

General situation

Given a finite sequence $\mathcal{A} = \{a_i\}$ of natural numbers, a set $\mathcal{P}$ of prime numbers (not necessarily all primes, but usually all), and $z > 0$ we consider

$$S(\mathcal{A}, \mathcal{P}, z) := \#\{a \in A : \underbrace{p \mid a \ \& \ p \in \mathcal{P} \Rightarrow p > z}\}$$

In other words a is <u>not</u> divisible by any $p \leq z$ with $p \in \mathcal{P}$.

Examples:

1. <u>Primes in an interval</u>: Take $\mathcal{A} = \{x_0 < n \leq x_0 + x\}$ and $\mathcal{P} = \{\text{all primes}\}$.

   For all $z > 0$

$$\pi(x_0 + x) - \pi(x_0) \leq z + S(\mathcal{A}, \mathcal{P}, z)$$

2. <u>Twin Primes</u>: Take $\mathcal{A} = \{n(n+2) : n \leq x\}$ and $\mathcal{P} = \{\text{all primes}\}$.

   For all $z > 0$

$$\pi_2(x) \leq z + S(\mathcal{A}, \mathcal{P}, z)$$

\# twin primes $\leq x$.

Note that if $P(z) := \prod_{\substack{p \in \mathcal{P} \\ p \le z}} p$, then

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

$$S(\mathcal{A}, \mathcal{P}, z) = \#\{a \in \mathcal{A} : (a, P(z)) = 1\} = \sum_{a \in \mathcal{A}} \sum_{d | (a, P(z))} \mu(d)$$

Since $d | (a, P(z)) \Leftrightarrow d | a$ & $d | P(z)$ it follows, after changing

Exercise 1 (easy!)

the order of summation, that

$$S(\mathcal{A}, \mathcal{P}, z) = \sum_{d | P(z)} \mu(d) A_d \qquad (*)$$

where
$$A_d := \#\{a \in \mathcal{A} : d | a\}.$$

Theorem 1 (Sieve of Eratosthenes - Legendre)

If we have a non-negative multiplicative function ν such that

$$A_d =: X \frac{\nu(d)}{d} + r_d \qquad (d | P(z))$$

where X is an approximation to the size of $\mathcal{A}$, then

$$S(\mathcal{A}, \mathcal{P}, z) = X \prod_{\substack{p \in \mathcal{P} \\ p \le z}} \left(1 - \frac{\nu(p)}{p}\right) + \sum_{d | P(z)} \mu(d) r_d$$

Equal!

Proof: It follow from $(*)$ and our assumptions on A_d that

$$S(\mathcal{A}, \mathcal{P}, z) = X \sum_{d | P(z)} \mu(d) \frac{\nu(d)}{d} + \sum_{d | P(z)} \mu(d) r_d \qquad \square$$

Revisiting our Examples:

1. Primes in an interval: If $\mathcal{A} = \{x_0 < n \le x_0 + x\}$ and $\mathcal{P} = \{\text{all primes}\}$

then
$$A_d = \left\lfloor \frac{x_0+x}{d} \right\rfloor - \left\lfloor \frac{x_0}{d} \right\rfloor = \frac{x}{d} + r_d$$

with $|r_d| \le 1$. Theorem 1 therefore implies that

$$S(\mathcal{A},\mathcal{P},z) = x \prod_{p\le z} \left(1-\frac{1}{p}\right) + O\left(2^{\pi(z)}\right).$$

Recall that for all $z \ge 2$
$$\prod_{p \le z}\left(1-\frac{1}{p}\right) = \frac{e^{-\gamma}}{\log z} + O\left(\frac{1}{(\log z)^2}\right) \qquad \text{(Mertens)}.$$

Taking $z = \log x$, it follows that $2^{\pi(z)} \le 2^z = x^{\log 2}$ & hence

$$S(\mathcal{A},\mathcal{P},z) = e^{-\gamma}\frac{x}{\log z} + O\left(\frac{x}{(\log z)^2}\right) \quad \text{as } x \to \infty.$$

$$\Rightarrow \quad \pi(x_0+x) - \pi(x) \le \left(e^{-\gamma} + \varepsilon(x)\right)\frac{x}{\log\log x}$$

where $\varepsilon(x) \to 0$ as $x \to \infty$.

Note: Although this bound is very weak, it is uniform in x_0.

Exercise 2: Show that if $z \ge (\log x)^{1+\varepsilon}$ for some $\varepsilon > 0$ (in particular $z = \sqrt{x}$), then $2^{\pi(z)}$ (and hence the "error term") is in fact bigger than any power of x.

2. Twin Primes: If $\mathcal{A}=\{n(n+2): n\le x\}$ and $\mathcal{P}=\{\text{all primes}\}$, then

$$A_d = \#\{n\le x : \underbrace{n(n+2)\equiv 0 \bmod d}\}$$

$$\Updownarrow$$

$$\underset{(2|d)}{n\equiv 0 \bmod 2} \quad \& \quad \underset{(p|d,\, p\neq 2)}{n\equiv 0 \text{ or } -2 \bmod p}$$

By the Chinese Remainder Theorem there are therefore $\nu(d)$ solutions mod d, where ν is the multiplicative function defined by

$$\nu(p)=\begin{cases}1 & \text{if } p=2\\ 2 & \text{if } p>2.\end{cases}$$

Each interval of length d contains $\nu(d)$ integers n counted in A_d. We can therefore write

$$A_d = x\,\frac{\nu(d)}{d} + r_d \quad , \text{with } |r_d|\le \nu(d).$$

and Theorem 1 implies that

$$\sum_{d|P(z)} \nu(d) = \prod_{p\le z}(1+\nu(p)) \le 3^{\pi(z)}.$$

$$S(\mathcal{A},\mathcal{P},z) = x\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right) + O\left(3^{\pi(z)}\right)$$

Taking $z=\frac{1}{2}\log x$, it follows that $3^{\pi(z)}\le 3^z = x^{\log 3/2}$ & hence

$$S(\mathcal{A},\mathcal{P},z)\le x\; 2\prod_{p\le z}\left(1-\frac{1}{p}\right)^2 + O\left(x^{\log 3/2}\right)$$

$$\Rightarrow \quad \pi_2(x) \ll \frac{x}{(\log\log x)^2} \quad \text{as } x\to\infty.$$

Another Example: Primes of the form $\square+1$.

Define
$$\pi_{\square+1}(x) = \{n \le x : n^2+1 \text{ is prime}\}.$$

Let
$$\mathcal{A} = \{n^2+1 : n \le x\} \quad \text{and} \quad \mathcal{P} = \{\text{all primes}\}.$$

Then
$$A_d = \#\{n \le x : n^2+1 \equiv 0 \bmod d\}$$

Exercise (3): Show that $A_d = x\,\frac{\nu(d)}{d} + r_d$ where w is a multiplicative function defined by
$$\nu(p) = \begin{cases} 1 & \text{if } p=2 \\ 2 & \text{if } p\equiv 1 \bmod 4 \\ 0 & \text{if } p \equiv 3 \bmod 4 \end{cases}$$
and $|r_d| \le \nu(d)$.

Given this exercise, it follows from Theorem 1 that
$$S(\mathcal{A},\mathcal{P},z) = x\,\frac{1}{2}\prod_{\substack{p\le z\\ p\equiv 1 \bmod 4}}\left(1-\frac{2}{p}\right) + O\left(3^{\pi(z)}\right)$$
$$\le x\,\frac{1}{2}\exp\Big(-2\sum_{\substack{p\le z\\ p\equiv 1 \bmod 4}}\frac{1}{p}\Big) + O\left(3^{\pi(z)}\right).$$

Exercise (4): Show that $\displaystyle\sum_{\substack{p\le z\\ p\equiv 1 \bmod 4}}\frac{1}{p} = \frac{1}{2}\log\log z + O(1)$.

Since $\pi_{\square+1}(x) \le S(\mathcal{A},\mathcal{P},z) + z^{1/2}$ for all $z>0$, it follows that taking $z = \frac{1}{2}\log x \Rightarrow \pi_{\square+1}(x) \ll \dfrac{x}{\log\log x}$.

The following simple consequence of Theorem 1 is often useful:

<u>Corollary 1:</u> Let $\mathcal{P}$ be a set of primes and

$$M(x) = \{ n \le x : p \nmid n \text{ for all } p \in \mathcal{P} \}$$

then

$$\lim_{x \to \infty} \frac{1}{x} M(x) = \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right).$$

In particular

$$\lim_{x \to \infty} \frac{1}{x} M(x) = 0 \underset{(*)}{\iff} \sum_{p \in \mathcal{P}} \frac{1}{p} \text{ diverges}.$$

<u>Proof:</u> Let $\mathcal{A} = \{ n \le x \}$. It is clear that

$$M(x) \le S(\mathcal{A}, \mathcal{P}, z)$$

for any $z > 0$ and that

$$A_d = \frac{x}{d} + r_d \quad \text{with } |r_d| \le 1 \qquad (d \mid P(z)).$$

Theorem 1 (with $z = \log x$) implies that

$$S(\mathcal{A}, \mathcal{P}, z) = x \prod_{\substack{p \in \mathcal{P} \\ p \le \log x}} \left(1 - \frac{1}{p}\right) + O\left(2^{\log x}\right)$$

Why?

$$= x \left(\prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) + \varepsilon(x) \right)$$

where $\varepsilon(x) \to 0$ as $x \to \infty$. This gives $(*)$ ($\Leftarrow$ only).

Now if $\sum_{p \in \mathcal{P}} \frac{1}{p} < \infty$, then

$\varepsilon'(x) \to 0$ as $x \to \infty$

$$M(x) \ge S(\mathcal{A}, \mathcal{P}, z) - \sum_{\substack{p \in \mathcal{P} \\ p > z}} \frac{x}{p} = x \left(\prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) + \varepsilon'(x) \right). \qquad \square$$

A quick application:

Theorem 2:

$$\lim_{x\to\infty} \frac{\#\{n\le x : \frac{4}{n} = \frac{1}{a}+\frac{1}{b}+\frac{1}{c} \text{ has \underline{no} solution with } a,b,c\in\mathbb{N}\}}{x} = 0$$

Remark: Erdős & Straus famously conjectured that $\frac{4}{n}$ can <u>always</u> be written as the sum of 3 unit fractions.

Proof: If $n=(4k-1)q$ with $4k-1$ prime, then

$$\frac{4}{n} = \frac{4}{q(4k-1)} = \frac{1}{2qk} + \frac{1}{2qk} + \frac{1}{qk(4k-1)}$$

It therefore suffices to show that

$$\lim_{x\to\infty} \frac{\#\{n\le x : p\nmid n \text{ for all } p\equiv 3 \bmod 4\}}{x} = 0$$

but this follows immediately from Corollary 1 and the fact that

$$\sum_{\substack{p\le x\\ p\equiv 3\bmod 4}} \frac{1}{p} = \frac{1}{2}\log\log x + O(1).$$ $\square$

Exercise 5: Show that

$$\lim_{x\to\infty} \frac{\#\{n\le x : n=a^2+b^2 \text{ for some } a,b\in\mathbb{N}\}}{x} = 0$$

Remark: In fact

$$\#\{n\le x : n=a^2+b^2\} \sim \frac{1}{\sqrt{2}} \prod_{p\equiv 3\bmod 4} \left(1-\frac{1}{p}\right)^{-1/2} \frac{x}{\sqrt{\log x}}.$$