SELBERG'S SIEVE
COURSE NOTES, 2015

1. Bounding the number of twin primes

We recall that we have given a bound for the following sieving problem. Given:

- $\mathcal{A} \subset \mathbb{Z}$ with $\#\mathcal{A} = X$;
- $\mathcal{P}$ a set of primes and

$$P(z) = \prod_{\substack{p \le z \\ p \in \mathcal{P}}} p;$$

- We defined the set of elements of $\mathcal{A}$ not divisible by all primes $p \in \mathcal{P}$, $p \le z$:

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, z) = \#\{n \in \mathcal{A} : \gcd(n, P(z)) = 1\}$$

and wish to estimate $\mathcal{S}(\mathcal{A}, \mathcal{P}, z)$.

For each square-free d such that $p|d \Rightarrow p \in \mathcal{P}$ define

$$\mathcal{A}_d = \{n \in \mathcal{A} : d|n\}.$$

and assume

$$\#\mathcal{A}_d = \frac{X}{f(d)} + R_d$$

where f is a multiplicative function. We proved:

Theorem 1.1. *In the notation as above, we have*

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, z) \le \frac{X}{S(z)} + R(z),$$

where

$$S(z) = \sum_{\substack{d \le z \\ d | P(z)}} \frac{\mu^2(d)}{(\mu * f)(d)}\,, \quad R(z) = \sum_{\substack{d_1, d_2 \le z \\ d_1, d_2 | P(z)}} \left| R_{[d_1, d_2]} \right|\,.$$

Date: May 7, 2015.

1.1. **Twin primes.** We will now apply the above to the problem of giving an upper bound to the number $\pi_2(X)$ of twin primes up to X. To do so, we take

$$\mathcal{A} = \{m(m+2) : m \leq X\}$$

and $\mathcal{P}$ to be all primes. Then

$$\pi_2(X) - \pi_2(z) \leq \mathcal{S}(\mathcal{A}, \mathcal{P}, z)$$

and it suffices to give an upper bound for $S(\mathcal{A}, \mathcal{P}, z)$. By Selberg's sieve,

$$\mathcal{S}(\mathcal{A}, \mathcal{P}, z) \leq \frac{X}{S(z)} + R(z)$$

and we need a lower bound for S and an upper bound for R. We will show that $S(z) \gg (\log z)^2$ and $R(z) \ll (z \log z)^2$, and taking say $z = X^{1/4}$ we find

$$\frac{X}{S(z)} + R(z) \ll \frac{X}{(\log z)^2} + (z \log z)^2 \ll \frac{X}{(\log X)^2}$$

Thus we find

Theorem 1.2.

$$\pi_2(X) \ll \frac{X}{(\log X)^2}$$

We start by finding the multiplicative function $f(d)$: In previous classes we saw

Lemma 1.3. *Let*

$$\rho(d) = \#\{c \bmod d : c(c+2) = 0 \bmod d\}$$

Then

$$\#\mathcal{A}_d = X\frac{\rho(d)}{d} + O(\rho(d))$$

Hence $f(d) = d/\rho(d)$ and $R_d = \rho(d)$. We can compute that for p prime,

$$\rho(p) = \begin{cases} 1, & p = 2 \\ 2, & p \neq 2 \end{cases}$$

1.2. **An upper bound for R.** We claim that

$$R(z) \ll (z \log z)^2 \tag{1}$$

We have $R_d = \rho(d) \leq 2^{\omega(d)}$ for d squarefree (where $\omega(d)$ denotes the number of distinct prime factors of d), and hence

$$R(z) = \sum_{\substack{d_1, d_2 \leq z \\ \text{squarefree}}} \rho([d_1, d_2]) \leq \sum_{\substack{d_1, d_2 \leq z \\ \text{squarefree}}} 2^{\omega([d_1, d_2])}$$

Now

$$\omega([d_1, d_2]) \leq \omega(d_1 d_2) \leq \omega(d_1) + \omega(d_2)$$

and therefore

$$R(z) \le \sum_{\substack{d_1,d_2\le z\\ \text{squarefree}}} 2^{\omega(d_1)+\omega(d_2)} = \Big(\sum_{\substack{d\le z\\ \text{squarefree}}} 2^{\omega(d)}\Big)^2$$

For d squarefree, $2^{\omega(d)} = \tau(d)$ and therefore

$$\sum_{\substack{d\le z\\ \text{squarefree}}} 2^{\omega(d)} \le \sum_{n\le z} \tau(n) \ll z\log z$$

which gives $R(z) \ll (z\log z)^2$ as claimed.

1.3. **A lower bound for S.** To give a lower for for $S(z)$, we replace $S(z)$ by a somewhat more convenient function. Let $\tilde f$ be the completely multiplicative function whose values at primes is $\tilde f(p) = f(p)$, that is

$$\tilde f(n) = \prod_p \tilde f(p)^{k_p}, \quad n = \prod p^{k_p}$$

Lemma 1.4.

$$S(z) = \sum_{\substack{d\le z\\ d|P(z)}} \frac{1}{(f*\mu)(d)} \ge \sum_{\substack{n\le z\\ p|n\Rightarrow p\in\mathcal{P}}} \frac{1}{\tilde f(n)}$$

Proof. We have for d squarefree

$$(f*\mu)(d) = \prod_{p|d}(f(p)-1)$$

and hence

$$\frac{1}{(f*\mu)(d)} = \frac{1/f(p)}{1-1/f(p)} = \sum_{k\ge1}\frac{1}{f(p)^k} = \prod_{p|d}\sum_{k\ge1}\frac{1}{f(p)^k} = \sum_{n\in\mathcal{N}(d)}\frac{1}{\tilde f(n)}$$

where

$$\mathcal{N}(d) = \{n : p \mid n \Leftrightarrow p \mid d\}\ .$$

Note that the sets of $\mathcal{N}(d)$ are disjoint for different squarefree d's and that every integer $d \le z$ belongs to exactly one of the sets $\mathcal{N}(d)$ (d squarefree), namely for $d = \mathrm{rad}(n) := \prod_{p|n} p$.

We have

$$S(z) = \sum_{\substack{d\le z\\ d|P(z)}} \frac{1}{(f*\mu)(d)} = \sum_{\substack{d\le z\\ d|P(z)}} \sum_{n\in\mathcal{N}(d)} \frac{1}{\tilde f(n)}$$

Since $\tilde f > 0$, we can drop all terms with $n > z$ without increasing the result, retaining only those $n \le z$ which are divisible only by primes from some $d \mid P(z)$, that is that are divisible only by primes from $\mathcal{P}$. Hence

$$S(z) \ge \sum_{\substack{n\le z\\ p|n\Rightarrow p\in\mathcal{P}}} \frac{1}{\tilde f(n)}$$

as claimed. ∎

Proposition 1.5.

$$S(z) \gg (\log z)^2$$

Proof. We use Lemma 1.4 to obtain

$$S(z) \geq \sum_{n \leq z} \frac{\tilde{\rho}(n)}{n}$$

Here

$$\tilde{\rho}(n) = \prod_{p|n} \rho(p)^{k_p} = \prod_{\substack{p|n \\ p \neq 2}} 2^{k_p} = 2^{\Omega_{\text{odd}}(n)}$$

where $\Omega_{\text{odd}}(n)$ is the number of odd prime powers dividing n. We have

$$2^{\Omega_{\text{odd}}(n)} \geq \tau_{\text{odd}}(n)$$

where

$$\tau_{\text{odd}}(n) = \sum_{\substack{d|n \\ d \text{ odd}}} 1$$

is the number of odd divisors of n. Indeed, since both functions are multiplicative, it suffices to check this for $n = p^k$ a prime power, and then for $p \neq 2$ we check that

$$2^{\Omega_{\text{odd}}(p^k)} = 2^k \geq k + 1 = \tau_{\text{odd}}(p^k)$$

while for $p = 2$ both sides are 1.

Hence we find

$$S(z) \geq \sum_{n \leq z} \frac{\tau_{\text{odd}}(n)}{n}$$

Lemma 1.6.

$$D_{\text{odd}}(x) := \sum_{n \leq x} \tau_{\text{odd}}(n) = \frac{1}{2} x \log x + O(x)$$

Proof. We have

$$\begin{aligned} D_{\text{odd}}(x) &= \sum_{n \leq x} \tau_{\text{odd}}(n) = \sum_{n \leq x} \sum_{\substack{d|n \\ d \text{ odd}}} 1 \\ &= \sum_{\substack{d \leq x \\ d \text{ odd}}} \sum_{\substack{n \leq x \\ d|n}} 1 = \sum_{\substack{d \leq x \\ d \text{ odd}}} \frac{x}{d} + O(1) = x \sum_{\substack{d \leq x \\ d \text{ odd}}} \frac{1}{d} + O(x) \ . \end{aligned}$$

Now

$$\begin{aligned} \sum_{\substack{d \leq x \\ d \text{ odd}}} \frac{1}{d} &= \sum_{d \leq x} \frac{1}{d} - \sum_{c \leq x/2} \frac{1}{2c} \\ &= \log x + O(1) - \frac{1}{2} \log \frac{x}{2} + O(1) = \frac{1}{2} \log x + O(1) \end{aligned}$$

and therefore

$$\sum_{n\le x}\tau_{\text{odd}}(n) = x\Big(\frac{1}{2}\log x + O(1)\Big) + O(x) = \frac{1}{2}x\log x + O(x)$$

as claimed. □

Now to show that $S(z) \ge \frac{1}{4}(\log z)^2 + O(\log z)$: We saw that

$$S(z) \ge \sum_{n\le z}\frac{\tau_{\text{odd}}(n)}{n}$$

Using summation by parts

$$\begin{aligned}\sum_{n\le z}\frac{\tau_{\text{odd}}(n)}{n} &= \frac{D_{\text{odd}}(z)}{z} + \int_1^z \frac{D_{\text{odd}}(t)}{t^2}dt\\ &= O(\log z) + \int_1^z \frac{\frac{1}{2}t\log t + O(t)}{t^2}dt\\ &= \frac{1}{4}(\log z)^2 + O(\log z)\end{aligned}$$

Hence we find $S(z) \ge \frac{1}{4}(\log z)^2 + O(\log z)$ as claimed. □