The following are lecture notes put together for a final presentation in 21849 - Arithmetic Statistics on Selberg Sieves from Kevin Ford's notes on Sieve Methods. Extra care has been put into the clarity of proof, especially for people with limited analytic number theory experience (me).

1 Preliminary Setup

1.1 Recall From Last Class...

- Relevant Notation:
  - $P(z)$: set of positive square free integers composed of only primes $\le z$.
  - $P^+(n)$: largest prime factor of n
  - $\omega(n)$: number of distinct prime factors of n
  - $\Omega(n)$: number of prime factors of n counted with multiplicity
  - $\mu(n)$: Möbius's function: $(-1)^{\omega(n)}$, supported on n squarefree
- A is our sieve problem, z is our 'sieving bound'
- We defined $A_d = \cap_{p|d} A_p = M \Pr\{A_d\}$
- By incl. excl.
$$S(A,z) = \sum_{d \in P(z)} \mu(d) A_d$$
  where P counts the number of square free integers made of primes $\le z$
- We approximate the sieve by choosing $X \sim M$, multiplicative function g st
$$A_d \approx Xg(d)$$
  where our error for each d is
$$r_d = A_d - Xg(d)$$
- The level of distribution for a sifting problem A is the largest D for which
$$\sum_{d \le D, d \in P(z)} |r_d| \le \varepsilon X V(z)$$
  for a chosen small $\varepsilon > 0$.
- An upper bound sieve is a sequence $\lambda^+ = (\lambda_d^+)$ supported on squarefree integers d which satisfies the conditions that $\lambda_1^+ = 1$ and $\sum_{d|m} \lambda_d^+ \ge 0$ for all $m > 1$.
- By Lemma 1.2, it follows that for any upper bound sieve, $S(A,z) \le \sum_{d \in P(z)} \lambda_d^+ A_d$

1.2 Setting up an Upperbound

The Selberg Sieve constructs an upperbound for $S(A,z)$ by careful construction of an upper bound sieve λ^+, starting with a general construction using any real-valued function λ s.t. $\lambda(1) = 1$.

1.2.1 Λ^2 Sieve

Let us construct our upper bound sieve. Given any λ real valued function, supported on squarefree numbers, where $\lambda(1) = 1$, consider the function $\lambda_d^+ := \sum_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2)$. Note this is also only supported on square free numbers as if d is not squarefree and $d = [d_1, d_2]$ for some d_1, d_2 then at least one of d_1, d_2 must also be not squarefree and thus every summand would be 0. Next, by construction, $\lambda_1^+ = \lambda(1)\lambda(1) = 1$. Finally, see that for every $m > 1$,

$$\begin{aligned}
\sum_{d|m} \lambda_d^+ &= \sum_{d|m} \sum_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2) \\
&= \sum_{d|m} \sum_{d_1,d_2} \mathbb{1}_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2) \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) \left[\sum_{d|m} \mathbb{1}_{[d_1,d_2]=d} \right] \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) \mathbb{1}_{[d_1,d_2]|m} \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) \mathbb{1}_{d_1,d_2|m} \\
&= \sum_{d_1,d_2|m} \lambda(d_1)\lambda(d_2) = \left(\sum_{d|m} \lambda(d) \right)^2 \geq 0
\end{aligned}$$

Thus we have found that

$$\boxed{\lambda_d^+ := \sum_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2)}$$

for *any* λ real valued function, supported on square free numbers, where $\lambda(1) = 1$, is an upper bound sieve. This is known as the Λ^2 Sieve.

2 Upperbounding $S(A, z)$

First, assume $(g), (r)$ exist as defined from the last class.

Then, by the lemma from the previous class, and given $m = \prod_{p \leq z} p$, we have for *any* λ real valued function, supported on square free numbers, where $\lambda(1) = 1$, $P^+(n)$ is the largest prime factor of n,

$$\begin{aligned}
S(A, z) &\leq \sum_{d \in P(z)} \lambda_d^+ A_d \\
&= \sum_{d|m} \lambda_d^+ A_d \\
&= \sum_{d|m} \sum_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2) A_d \\
&= \sum_{d|m} \sum_{d_1,d_2} \mathbb{1}_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2) A_d \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) A_{[d_1,d_2]} \left[\sum_{d|m} \mathbb{1}_{[d_1,d_2]=d} \right] \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) A_{[d_1,d_2]} \mathbb{1}_{[d_1,d_2]|m} \\
&= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) A_{[d_1,d_2]} \mathbb{1}_{d_1,d_2|m}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{d_1,d_2 \mid m} \lambda(d_1)\lambda(d_2)A_{[d_1,d_2]} \\
&= \sum_{\substack{d_1,d_2 \\ P^+(d_1),P^+(d_2)\le z}} \lambda(d_1)\lambda(d_2)A_{[d_1,d_2]} \\
&= \sum_{\substack{d_1,d_2 \\ P^+(d_1d_2)\le z}} \lambda(d_1)\lambda(d_2)A_{[d_1,d_2]}
\end{aligned}
$$

Now, we use our previous framework approximating $S(A,z)$ through choice of X, g, whereby upperbounding the error,

$$
\begin{aligned}
\sum_{\substack{d_1,d_2 \\ P^+(d_1d_2)\le z}} \lambda(d_1)\lambda(d_2)A_{[d_1,d_2]} &\le X \sum_{\substack{d_1,d_2 \\ P^+(d_1d_2)\le z}} \lambda(d_1)\lambda(d_2)g([d_1,d_2]) + \sum_{\substack{d_1,d_2 \\ P^+(d_1d_2)\le z}} \left|\lambda(d_1)\lambda(d_2)r_{[d_1,d_2]}\right| \\
&=: XG + R
\end{aligned}
$$

and our understated goal is now to minimize $XG+R$, which is quite hard. To work towards this, we will limit the support of λ, which will make G much easier to minimize.

2.1 Restricting the support of λ

Our first observation for minimizing G is that we can ignore any $g(p)=0$. Recall that g is a multiplicative function over the primes, so any such values contribute nothing. Shorthand

$$
P = \prod_{p\le z, g(p)>0} p
$$

and then restrict the support of $\lambda(d)$ to $d \mid P$. Then, define

$$
h(d) = \begin{cases} \prod_{p\mid d} \frac{g(p)}{1-g(p)} & d \mid m \\ 0 & \text{otw.} \end{cases}
$$

and note that by construction, h is multiplicative and nonnegative. Inverting the above for $m \mid P$,

$$
\frac{1}{g(m)} = \prod_{p\mid m} \frac{1}{g(p)} = \prod_{p\mid m}\left(1+\frac{1-g(p)}{g(p)}\right) = \prod_{p\mid m}\left(1+\frac{1}{h(p)}\right) = \sum_{d\mid m} \frac{1}{\prod_{p\mid d} h(p)} = \sum_{d\mid m}\frac{1}{h(d)}
$$

Then, by the multiplicativity of g and the fact that for any $d_1, d_2 \mid P$, $(d_1,d_2) \mid P$ and so $g((d_1,d_2))>0$, we can write

$$
g([d_1,d_2])g((d_1,d_2)) = g(d_1)g(d_2) \Longrightarrow g([d_1,d_2]) = \frac{g(d_1)g(d_2)}{g((d_1,d_2))}
$$

Then, we can rewrite G, where $m=(d_1,d_2)$ as

$$
\begin{aligned}
G &= \sum_{P^+(d_1d_2)\le z} \lambda(d_1)\lambda(d_2)g([d_1,d_2]) \\
&= \sum_{d_1,d_2\mid P} \lambda(d_1)\lambda(d_2)\frac{g(d_1)g(d_2)}{g((d_1,d_2))} \\
&= \sum_{d_1,d_2\mid P} \lambda(d_1)\lambda(d_2)g(d_1)g(d_2) \sum_{d\mid (d_1,d_2)} \frac{1}{h(d)}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{d_1,d_2|P} \lambda(d_1)\lambda(d_2)g(d_1)g(d_2) \sum_{d|P} \mathbb{1}_{d|(d_1,d_2)} \frac{1}{h(d)} \\
&= \sum_{d|P} \frac{1}{h(d)} \sum_{d_1,d_2|P} \lambda(d_1)\lambda(d_2)g(d_1)g(d_2)\mathbb{1}_{d|(d_1,d_2)} \\
&= \sum_{d|P} \frac{1}{h(d)} \sum_{\substack{d_1,d_2|P \\ d|(d_1,d_2)}} \lambda(d_1)\lambda(d_2)g(d_1)g(d_2) \\
&= \sum_{d|P} \frac{1}{h(d)} \sum_{d|d_1,d_2} \lambda(d_1)\lambda(d_2)g(d_1)g(d_2) \\
&= \sum_{d|P} \frac{1}{h(d)} \left(\sum_{d|m} \lambda(m)g(m) \right)^2
\end{aligned}
$$

Recall that the support of $\lambda(d)$ is only for $d \mid P$. Now define,

$$\xi(d) = \frac{\mu(d)}{h(d)} \sum_{d|m} \lambda(m)g(m)$$

supported on $d \mid P, d \leq \sqrt{D}$ so that we can write G as a quadratic in $\lambda(d)$ as

$$G = \sum_{d|P} \frac{h(d)}{\mu(d)^2} \xi(d)^2 = \sum_{d|P} h(d)\xi(d)^2$$

The inclusion of μ is for inversion purposes. First, let us start with proving a simple identity about the Möbius function. Note we can express for all $n \geq 1$,

$$1 = \sum_{d|n} \mathbb{1}_{d=1}$$

and thus by Möbius inversion, $\mathbb{1}_{m=1} = \sum_{d|m} \mu(d)$. Now see that for $\ell \mid P$,

$$
\begin{aligned}
\sum_k h(k\ell)\xi(k\ell) &= \sum_k \mu(k\ell) \sum_{k\ell|m} \lambda(m)g(m) \\
&= \mu(\ell) \sum_k \mu(k) \sum_{k\ell|m} \lambda(m)g(m) \\
&= \mu(\ell) \sum_{\ell|m} \lambda(m)g(m) \left[\sum_{k|m/\ell} \mu(k) \right] \\
&= \mu(\ell) \sum_{\ell|m} \lambda(m)g(m)\mathbb{1}_{m=\ell} \\
&= \mu(\ell)\lambda(\ell)g(\ell)
\end{aligned}
$$

Then, for $\ell \mid P$,

$$\lambda(\ell) = \frac{1}{\mu(\ell)g(\ell)} \sum_k h(k\ell)\xi(k\ell) = \frac{\mu(\ell)}{g(\ell)} \sum_{\ell|d} h(d)\xi(d)$$

We use the Cauchy-Schwarz Inequality to minimize G,

$$1 = \lambda(1)^2 = \left(\sum_{d\leq\sqrt{D}} h(d)^{1/2}\xi(d)h(d)^{1/2} \right)^2 \leq \left(\sum_{d\leq\sqrt{D}} h(d)\xi(d)^2 \right) J = GJ$$

for $\boxed{J = \sum_{d \le \sqrt{D}} h(d)}$ where equality holds iff $\xi(d) = \frac{1}{J}$ for every $d \le \sqrt{D}$, $d \mid P$. That is, $\xi(d)$ is constant.

Then,

$$\begin{aligned}
\lambda(\ell) &= \frac{\mu(\ell)}{g(\ell)} \sum_{\ell \mid d} h(d)\xi(d) \\
&= \frac{\mu(\ell)}{Jg(\ell)} \sum_{\ell \mid d \le \sqrt{D}} h(d) = \frac{\mu(\ell)h(\ell)}{Jg(\ell)} \sum_{m \le \sqrt{D}/\ell, (m,\ell)=1} h(m)
\end{aligned}$$

Thus, we get $GJ = 1 \iff G = \frac{1}{J}$ and so

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

2.2 Handling R

If we can upperbound $\lambda(\ell)$, then we upperbound R wrt $|r_\ell|$.

Indeed, first observe that for any $\ell \ge 1$, we can partition J into a double sum based on (d, ℓ):

$$J = \sum_{k \mid \ell} \sum_{d \le \sqrt{D}, (d,l)=k} h(d) = \sum_{k \mid l} h(k) \sum_{m \le \sqrt{D}/k, (m,l/k)=1} h(m)$$

we can lower bound this sum by restricting the inner sum further for $m \le \sqrt{D}/\ell$ and $(m, \ell) = 1$, which also removes the dependence on the outer sum.

$$\begin{aligned}
\sum_{k \mid l} h(k) \sum_{\substack{m \le \sqrt{D}/k \\ (m,l/k)=1}} h(m) &\ge \left(\sum_{k \mid \ell} h(k) \right) \sum_{\substack{m \le \sqrt{D}/\ell \\ (m,\ell)=1}} h(m) \\
&= \left(\sum_{k \mid \ell} h(\frac{\ell}{k}) \right) \sum_{\substack{m \le \sqrt{D}/\ell \\ (m,\ell)=1}} h(m) \\
&= h(\ell) \left(\sum_{k \mid \ell} \frac{1}{h(k)} \right) \sum_{\substack{m \le \sqrt{D}/\ell \\ (m,\ell)=1}} h(m) \\
&= \frac{h(\ell)}{g(\ell)} \sum_{m} h(m)
\end{aligned}$$

By plugging this lower bound into our definition for $\lambda(\ell) = \frac{\mu(\ell)h(\ell)}{Jg(\ell)} \sum_m h(m)$, we will get that $\lambda(\ell) = c\mu(\ell)$ for a constant $0 \le c \le 1$, and thus $|\lambda(\ell)| \le 1$ for all ℓ. It follows that

$$R \le \sum_{d_1, d_2 \le \sqrt{D}, P^+(d_1 d_2) \le z} \left| r_{[d_1,d_2]} \right| \le \sum_{d \le D, P^+(d) \le z} 3^{\omega(d)} |r_d|$$

where $\omega(d)$ is the number of distinct prime factors of d. The last inequality holds by our choice to decompose d's factors into d_1, d_2, where each factor is assigned as d_1, d_2, or both.

Our minimization is complete, and we have the following result,

Theorem 2.1: Selberg's sieve

Let A be a sieve problem, and assume $(g), (r)$. Let $z \geq 2$ and $D \geq 1$. Then,

$$S(A, z) \leq \frac{X}{J} + \sum_{d \leq \sqrt{D}, P^{+}(d) \leq z} 3^{\omega(d)} |r_d|$$

where $J = \sum_{n \leq \sqrt{D}} \mu^2(n) h(n)$, and h is the multiplicative function on primes $p \leq z$ by $h(p) = \frac{g(p)}{1-g(p)}$

3 Tight Selberg for General Sieves

Lemma 3.1: Selberg's sieve is best possible in dimension 1

For general sieve problems, Selberg's upperbound sieve cannot be improved.

Proof. Recall Selberg's example

$$A = \{n \leq x : \lambda(n) = -1\}$$

where $\lambda(n) = (-1)^{\Omega(n)}$ and set $X = x/2$

Previously we saw by the Prime Number Theorem,

$$A_d = \sum_{d \leq x} \frac{1 - \lambda(d)}{2} = \frac{x/2}{d} + O\Big(\frac{x}{d} e^{-c\sqrt{\lg(x/d)}}\Big) \qquad (c > 0)$$

We treat the big O term here as our error term, where we bound

$$|r_d| \ll \frac{x}{d} e^{-c\sqrt{\lg(x)}}$$

for every $d \leq x$.

We want to make use of Selberg's sieve, so put $g(d) = \frac{1}{d}$ for $d \in P(z)$.

Then, $h(p) = \frac{1}{p-1}$, $p \leq z$, and so take $z = \sqrt{D} = x^{\frac{1}{2} - \varepsilon(x)}$ where $\varepsilon(x) = \frac{1}{\lg \lg x}$. For these parameters, the error is bounded by

$$\begin{aligned}
\sum_{d \leq D} \mu^2(d) 3^{\omega(d)} |r_d| &\ll \sum_{d \leq D} \frac{x}{d} e^{-c\sqrt{2\varepsilon(x) \lg(x)}} \mu^2(d) 3^{\omega(d)} \\
&= x e^{-c\sqrt{2\varepsilon(x) \lg(x)}} \sum_{d \leq D} \frac{\mu^2(d) 3^{\omega(d)}}{d} \\
&\ll \frac{x}{\lg^{10} x} \prod_{p \leq x} (1 + 3/p) \\
&\ll \frac{x}{\lg^{10} x} \Big(\prod_{p \leq x} (1 + 1/p)\Big)^3 = \frac{x}{\lg^7 x}
\end{aligned}$$

where the second to last inequality holds by the multiplicative representation of picking each prime in the $3^{\omega}(d)$ term and $D \ll x$.

Now, the value of J,

$$J = \sum_{n \leq \sqrt{D}} \mu^2(n) h(n)$$

$$= \sum_{n \le \sqrt{D}} \frac{\mu^2(n)}{n-1} = \sum_{n \le \sqrt{D}} \frac{\mu^2(n)}{\phi(n)}$$

It is proven earlier in the section that

$$\sum_{n \le \sqrt{D}} \frac{\mu^2(n)}{\phi(n)} \ge \frac{1}{2} \lg x$$

by seeing that

$$\sum_{n \le x} \frac{1}{n-1} \ge \int_1^{\lfloor x+1 \rfloor} \frac{dt}{t} \ge \lg x$$

and by Selberg,

$$S(A, z) \le \frac{x}{(1 - 2\varepsilon(x)) \lg x} + O\left(\frac{x}{\lg^2 x}\right) \sim \frac{x}{\lg x} \qquad (x \to \infty)$$

For general sieves, we know

$$S(A, \sqrt{x}) = \pi(x) - \pi(\sqrt{x}) + 1 \sim \frac{x}{\lg x}$$

so this upperbound cannot be improved with different dimension. □

4 Applications

We briefly describe some upperbounds on sieves we derive with Selberg's results.

4.1 Twin Primes

Theorem 4.1: Twin Prime Upperbound

Let k be an even, positive integer. Then, uniformly in $k \le x$,

$$\#\{p \le x : p + k \text{ prime}\} \le C \prod_{2 < p | k} \left(\frac{p-1}{p-2}\right) \frac{x}{\log^2 x} \left(4 + O\left(\frac{\lg_2 x}{\log x}\right)\right)$$

where

$$C = 2 \prod_p \left(1 - \frac{1}{(p-1)^2}\right) \approx 1.32$$

is the *twin prime constant.*

Proof. (setup)
Our problem is $A = \{p + k : p \le x\}$, where we choose $D = x^{\frac{1}{2}} \lg^{-B} x$ for sufficiently large B.
We set $g(p) = \frac{1}{p-1}$ for $p \not| k$, so $h(p) = \frac{1}{p-2}$.
Then, apply the Selberg Sieve. □

4.2 Brun-Titchmarch Inequality

Recall the previously derived upperbound:

Theorem 4.2: Brun-Titchmarch Inequality V.1

There is a constant C so that uniformly, for $1 \leq a \leq q < y \leq x$ and $(a, q) = 1$,

$$\pi(x; q, a) - \pi(x - y; q, a) \leq C \frac{y}{\phi(q) \log(y/q)}$$

With Sieve we find a finer grain bound.

Theorem 4.3: Brun-Titichmarch Inequality V.2

For $x \geq y \geq q \geq 1$ and $(a, q) = 1$, we have

$$\pi(x; q, a) - \pi(x - y; q, a) \leq \frac{2y}{\phi(q) \log(5y/q)} \left(1 + O\left(\frac{\lg(5y/q)}{\log(5y/q)}\right)\right)$$

where $\pi(x; q, a)$ is the number of primes $p \leq x$ st $p \cong a \bmod q$