SOME REMARKS ON SELBERG'S SIEVE

Let

$$a : \mathbb{Z} \to \mathbb{R}^+, \tag{1}$$

$$A = \sum_n a(n) < \infty, \tag{2}$$

$$A_d = \sum_n a(dn), \tag{3}$$

and suppose that

$$A_d = f(d)X + R_d, \tag{4}$$

where

$$f \in \mathcal{M}, \tag{5,}$$

the set $\mathcal{M}$ of multiplicative functions, i.e $f(mn) = f(m)f(n)$ when $(m, n) = 1$ and $f(1) = 1$. It is also convenient to assume that $0 \leq f(p) < 1$ for each prime p.

$$A_d = f(d)X + R_d. \tag{4}$$

In principle we suppose that X is "large" and R_d is "small" compared with $f(d)X$ when d is relatively small.

Example 1. *Let $a(n) = 1$ when $Y < n \leq Y + X$ and $a(n) = 0$ otherwise. Then*

$$A_d = \frac{X}{d} + R_d, \quad |R_d| \leq 1.$$

There are many occasions when one is interested in the behaviour of expressions of the kind

$$S(A, P) = \sum_{\substack{n \\ (n,P)=1}} a(n) \tag{7}$$

where typically P is a product of primes.

Example 2. *If $P = \prod_{p \leq \sqrt{X}} p$, and a is as in Example 1, then*

$$\pi(X + Y) - \pi(Y) \leq \pi(\sqrt{X}) + S(A, P).$$

This is a formalisation of the sieve of Erathosthenes–Legendre.

Any method which deduces estimates for (7) from (4) is called a sieve.

Other examples.

Example 3 (Twin primes). *Let* $a(n) = 1$ *when* $n = m(m+2)$ *for some* $m \leq X$ *and* $a(n) = 0$ *otherwise and* P *as before. Then*

$$\sum_{\substack{p \leq X \\ p+2 \text{ prime}}} 1 \leq \pi(\sqrt{X}) + S(A, P).$$

It is easily verified that (4) holds with $f(d) = \rho(d)/d$, $\rho \in \mathcal{M}$, $\rho(2) = 1$, $\rho(p) = 2$ $(p > 2)$, *and with* $|R_d| \leq \rho(d)$.

Example 4 (Goldbach binary problem). *Let* X *be an even positive integer and let* $a(n) = \text{card}\{m : n = m(X-m), m < X\}$ *and* P *as before. Then*

$$\text{card}\{p < X : X - p \text{ prime}\} \leq 2\pi(\sqrt{X}) + S(A, P).$$

Again it is easily verified that (4) holds with $f(d) = \rho(d)/d$, $\rho \in \mathcal{M}$, $\rho(p) = 1$ *when* $p|X$, $\rho(p) = 2$ *when* $p \nmid X$, *and with* $|R_d| \leq \rho(d)$ *once more.*

Example 5. *Let $a(n) = 1$ when $n = m^2 + 1$ for some $m \le X$ and P as before. Then*

$$\text{card}\{m \le X : m^2 + 1 \text{ prime}\} \le \pi(\sqrt{X}) + S(A, P).$$

Also (4) holds with $f(d) = \rho(d)/d$ with $\rho \in \mathcal{M}$ and $\rho(2) = 1$, $\rho(p) = 2$ when $p \equiv 1 \pmod 4$ and $\rho(p) = 0$ otherwise, and $|R_d| \le \rho(d)$.

A more sophisticated version of Example 3 is

Example 6 (twin primes revisited). *Let $a(n) = 1$ when $n - 2$ is a prime $p \le Y$ and 0 otherwise and let $P = \prod_{p \le \sqrt{Y}} p$. Then*

$$\sum_{\substack{p \le Y \\ p+2 \text{ prime}}} 1 \le \pi(\sqrt{Y}) + S(A, P).$$

Now $A_d = \pi(Y; d, -2)$ and we have

$$A_d = f(d)X + R_d$$

where $f(d) = 0$ when d is even and $f(d) = \frac{1}{\phi(d)}$ when d is odd, and where now

$$X = \text{li}(Y) = \int_2^Y \frac{dt}{\log t}$$

and where R_d is relatively small ($\ll Y^{\frac{1}{2}+\varepsilon}$ on GRH).

One cannot get very far in sieve theory without meeting the Möbius function μ defined by $\mu(n) = (-1)^k$ when n is the product of k different primes and to be 0 when n has a repeated prime factor. The fundamental property of μ is

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & (n = 1), \\ 0 & (n > 1). \end{cases}$$

One way of seeing this is to observe that $\mu \in \mathcal{M}$ and (*via* the Euler products) that for $\Re s > 1$,

$$\zeta(s) \sum_{d=1}^{\infty} \frac{\mu(d)}{d^s} = 1.$$

$$S(A,P) = \sum_{\substack{n \\ (n,P)=1}} a(n) \tag{7}$$

The condition $(n,P)=1$ can be rewritten as

$$\sum_{d|(n,P)} \mu(d)$$

so that

$$S(A,P) = \sum_{d|P} \mu(d) f(d) X + \sum_{d|P} \mu(d) R_d.$$

The basic problem with this is that the "error term" has far too many terms in it. For example, with $P = \prod_{p \le \sqrt{X}} p$ there would be $2^{\pi(\sqrt{X})}$ terms.

$$S(A,P) = \sum_{\substack{n \\ (n,P)=1}} a(n) \tag{7}$$

Modern sieve theory attempts to overcome this problem by seeking functions $\lambda_d^{\pm}$ such that

$$\sum_{d|m} \lambda_d^- \leq \sum_{d|m} \mu(d) \leq \sum_{d|m} \lambda_d^+$$

but the support for the $\lambda_d^{\pm}$ is restricted. We will not be concerned with lower bound sieves, where the theory is more delicate.

Selberg introduced a very simple and elegant upper bound sieve which is very effective in many situations, and also has the merit of great flexibility. Let

$$\lambda_1 = 1 \tag{8}$$

and suppose that the $\lambda_q \in \mathbb{R}$ are otherwise at our disposal. Then

$$\sum_{d|m} \mu(d) \leq \left(\sum_{d|m} \lambda_d \right)^2 .$$

In order to retain some structure we suppose that the support $\mathcal{D}$ of the λ_d is a divisor closed set of squarefree numbers. Thus for each $d \in \mathcal{D}$, $\mu(d) \neq 0$ and if $q|d$, then $q \in \mathcal{D}$.

Example 7. *$\mathcal{D} = \{d|P : d \leq D\}$ where*

$$P = \prod_{p \leq \sqrt{X}} p.$$

We recall that

$$S(A,P) = \sum_{\substack{n \\ (n,P)=1}} a(n) \text{ and } \lambda_1 = 1.$$

Thus

$$\begin{aligned} S(A,P) &\leq \sum_n a(n) \left(\sum_{d|n} \lambda_d \right)^2 \\ &= \sum_d \sum_e \lambda_d \lambda_e \sum_m a(m[d,e]) \\ &= X \sum_d \sum_e \lambda_d \lambda_e f([d,e]) + R \end{aligned}$$

where

$$R = \sum_d \sum_e \lambda_d \lambda_e R_{[d,e]}.$$

$$S(A,P) \le X \sum_d \sum_e \lambda_d \lambda_e f([d,e]) + R$$

where

$$R = \sum_d \sum_e \lambda_d \lambda_e R_{[d,e]}.$$

Example 8. *Consider Example 1,* $a(n) = 1$ *iff* $n \in (Y, Y+X]$ *with* $\mathcal{D}$ *as in Example 7. Then*

$$|R| \le \left(\sum_d |\lambda_d| \right)^2 \le D^2 \|\lambda\|_\infty^2.$$

The interesting part is the main term XM where

$$M = \sum_d \sum_e \lambda_d \lambda_e f([d,e]).$$

We want to minimise this subject to the condition $\lambda_1 = 1$. It is helpful to view M as a quadratic form in the $\boldsymbol{\lambda}$. Our first objective is to diagonalise this form, and this can be done quite easily. It is also useful to assume that $\mathcal{D}$ is such that $f(d) \neq 0$ when $d \in \mathcal{D}$.

$$M = \sum_d \sum_e \lambda_d \lambda_e f([d,e]).$$

Write $(d,e) = m$, $d = qm$, $e = rm$, so that $(q,r) = 1$. Since $f \in \mathcal{M}$ and qrm is squarefree we have $f([d,e]) = f(qrm) = f(qm)f(rm)/f(m)$ and

$$M = \sum_m f(m)^{-1} \sum_q \sum_{\substack{r \\ (q,r)=1}} \lambda_{qm}\lambda_{rm} f(qm) f(rm).$$

Now we use the Möbius function to remove the condition $(q,r) = 1$. Thus

$$M = \sum_m f(m)^{-1} \sum_l \mu(l) \left(\sum_d \lambda_{dlm} f(dlm) \right)^2.$$

Now we collect together the terms with $lm = n$ and observe that by multiplicativity we have

$$\sum_{\substack{l,m \\ lm=n}} f(m)^{-1}\mu(l) = \prod_{p|n} \frac{1-f(p)}{f(p)}.$$

Denoting this expression by $g(n)^{-1}$ we have

$$M = \sum_n g(n)^{-1} \left(\sum_d \lambda_{dn} f(dn) \right)^2.$$

$$g(n) = \prod_{p|n} \frac{f(p)}{1 - f(p)}$$

$$M = \sum_n g(n)^{-1} \left(\sum_d \lambda_{dn} f(dn) \right)^2 .$$

Let

$$\omega_n = \sum_d \lambda_{dn} f(dn) \quad (n \in \mathcal{D}).$$

There is a bijection between the $\boldsymbol{\lambda}$ and the $\boldsymbol{\omega}$. In fact we could view the transformation from the one to the other as being by an upper triangular matrix, which is obviously invertible. However there is a standard number theoretic way of expressing the inversion. We have

$$\sum_n \omega_{nm} \mu(n) = \sum_n \sum_d \lambda_{dnm} f(dnm) \mu(n)$$

and collecting together the terms with $nd = q$ this becomes for $m \in \mathcal{D}$

$$\sum_n \omega_{nm} \mu(n) = \sum_q \lambda_{qm} f(qm) \sum_{n|q} \mu(n) = \lambda_m f(m).$$

$$g(n) = \prod_{p|n} \frac{f(p)}{1 - f(p)} \tag{9}$$

$$M = \sum_n g(n)^{-1} \left(\sum_d \lambda_{dn} f(dn) \right)^2 \tag{10}$$

$$\omega_n = \sum_d \lambda_{dn} f(dn) \quad (n \in \mathcal{D}) \tag{11}$$

$$\lambda_m f(m) = \sum_n \omega_{nm} \mu(n) \quad (m \in \mathcal{D}) \tag{12}$$

Thus we are seeking to minimise

$$M = \sum_n g(n)^{-1} \omega_n^2 \text{ under } \sum_n \omega_n \mu(n) = \lambda_1 = 1.$$

Let $\lambda = 1/\sum_{n \in \mathcal{D}} g(n)$. Then

$$\begin{aligned} M &= \sum_{n \in \mathcal{D}} \frac{(\omega_n - \lambda \mu(n) g(n))^2}{g(n)} + 2\lambda \sum_n \omega_n \mu(n) - \lambda \\ &= \sum_{n \in \mathcal{D}} \frac{(\omega_n - \lambda \mu(n) g(n))^2}{g(n)} + \lambda. \end{aligned}$$

$$\lambda_m f(m) = \sum_n \omega_{nm}\mu(n) \quad (m \in \mathcal{D}) \tag{12}$$

$$\lambda = 1/\sum_{n\in\mathcal{D}} g(n)$$

$$M = \sum_{n\in\mathcal{D}} \frac{(\omega_n - \lambda\mu(n)g(n))^2}{g(n)} + \lambda$$

Obviously $M \geq \lambda$. Moreover the choice

$$\omega_n = \lambda\mu(n)g(n)$$

gives

$$\sum_n \omega_n\mu(n) = 1 \text{ and } M = \lambda.$$

Also

$$\begin{aligned}\lambda_m &= \frac{\lambda}{f(m)} \sum_n g(mn)\mu(mn)\mu(n) \\ &= \lambda\mu(m)\frac{g(m)}{f(m)} \sum_{\substack{n \\ nm\in\mathcal{D}}} g(n).\end{aligned}$$

$$\lambda_m = \lambda\mu(m)\frac{g(m)}{f(m)} \sum_{\substack{n \\ nm\in\mathcal{D}}} g(n)$$

Theorem (Selberg). *Suppose that* $a : \mathbb{Z} \to \mathbb{R}^+$, $A_d = \sum_n a(dn)$ *and that* $A_d = f(d)X + R_d$ *where* $f \in \mathcal{M}$ *and* $0 \le f(p) < 1$. *Let* $P \in \mathbb{N}$ *be squarefree and* $\mathcal{D}$ *be a divisor closed subset of the divisors of* P. *Then*

$$S(A,P) \le \frac{X}{\sum_{n\in\mathcal{D}} g(n)} + \sum_{d\in\mathcal{D}}\sum_{e\in\mathcal{D}} \lambda_d\lambda_e R_{[d,e]}$$

where $g(n) = \prod_{p|n} \frac{f(p)}{1-f(p)}$. *Moreover*

$$|\lambda_d| \le 1.$$

To see the last statement, write $\frac{g(m)}{f(m)} = \prod_{p|m} \frac{1}{1-f(p)}$ $= \prod_{p|m}(1+g(p)) = \sum_{d|m} g(d)$. Then $|\lambda_m| \le$

$$\lambda \sum_{d|m} g(d) \sum_{\substack{n \\ nd\in\mathcal{D} \\ (n,m/d)=1}} g(n) = \lambda \sum_{d|m} \sum_{\substack{k \\ (k,m)=d}} g(k) = 1$$

Theorem (Selberg). *Suppose that* $a : \mathbb{Z} \to \mathbb{R}^+$, $A_d = \sum_n a(dn)$ *and that* $A_d = f(d)X + R_d$ *where* $f \in \mathcal{M}$ *and* $0 \leq f(p) < 1$. *Let* $P \in \mathbb{N}$ *be squarefree and* $\mathcal{D}$ *be a divisor closed subset of the divisors of* P. *Then*

$$S(A, P) \leq \frac{X}{\sum_{n \in \mathcal{D}} g(n)} + \sum_{d \in \mathcal{D}} \sum_{e \in \mathcal{D}} \lambda_d \lambda_e R_{[d,e]}$$

where $g(n) = \prod_{p|n} \frac{f(p)}{1-f(p)}$. *Moreover*

$$|\lambda_d| \leq 1.$$

Example 9. *Following examples 1 and 2,*

$$\pi(X + Y) - \pi(Y) \leq \pi(\sqrt{X}) + S(A, P)$$

where $a(n) = 1$ *when* $Y < n \leq Y + X$ *and* $a(n) = 0$ *otherwise,* $P = \prod_{p \leq \sqrt{X}} p$, $A_d = \frac{X}{d} + R_d$, $|R_d| \leq 1$. *Thus* $f(d) = 1/d$. *Let* $\mathcal{D} = \{d | P : d \leq D\}$ *with* $D \leq \sqrt{X}$. *Then*

$$\sum_{n \in \mathcal{D}} g(n) = \sum_{d \leq D} \frac{\mu(d)^2}{\phi(d)}$$

and $|\lambda_d| \leq 1$.

Example 9 continued.

$$\pi(X+Y) - \pi(Y) \leq \pi(\sqrt{X}) + S(A,P)$$

$$S(A,P) \leq \frac{X}{\sum_{d \leq D} \frac{\mu(d)^2}{\phi(d)}} + D^2$$

Let $s(q)$ denote the squarefree kernel of q, $s(q) = \prod_{p|q} p$. Then

$$\begin{aligned}
\sum_{d \leq D} \frac{\mu(d)^2}{\phi(d)} &= \sum_{d \leq D} \frac{\mu(d)^2}{d} \prod_{p|d} \sum_{k=1}^{\infty} p^{-k} \\
&= \sum_{\substack{q \in \mathbb{N} \\ s(q) \leq D}} \frac{1}{q} \\
&\geq \sum_{q \leq D} \frac{1}{q} \\
&\geq \log D
\end{aligned}$$

$$\pi(X+Y)-\pi(Y) \le \frac{X}{\log D} + \sqrt{X} + D^2$$

The choice $D = \frac{\sqrt{X}}{\log X}$ gives the Brun–Titchmarsh theorem

$$\pi(X+Y)-\pi(Y) \le \frac{2X}{\log X} + O\left(\frac{X \log\log X}{\log^2 X}\right).$$

By working harder the error term can be removed altogether, but this does not concern us here today. However there are some observations I should make here. The optimising choice of λ_m in the proof of the Brun–Titchmarsh theorem is

$$\lambda_m = \mu(m) m \phi(m)^{-1} \frac{\sum_{\substack{n \le D/m \\ (n,m)=1}} \frac{\mu(n)^2}{\phi(n)}}{\sum_{n \le D} \frac{\mu(n)^2}{\phi(n)}}.$$

The sum in the denominator is asymptotically $\log D$ and, at least when m is not too close to D, the sum in the numerator ought to be asymptotically $\phi(m)m^{-1}\log(D/m)$. Thus λ_m should be close to

$$\lambda_m^* = \mu(m)\frac{\log R/m}{\log R}.$$

$$\lambda_m = \mu(m) m \phi(m)^{-1} \frac{\sum_{n \le D/m \atop (n,m)=1} \frac{\mu(n)^2}{\phi(n)}}{\sum_{n \le D} \frac{\mu(n)^2}{\phi(n)}}. \qquad (13)$$

$$\lambda_m^* = \mu(m) \frac{\log R/m}{\log R}.$$

Indeed λ_m^* can be used instead of the optimal choice, although there is more work involved in the analysis to push things through. We will see later situations where the optimal choice is not known but a choice of this kind is still effective.

Returning to (13) we have

$$|\lambda_m| \le m \phi(m)^{-1} \frac{\sum_{n \le D/m} \frac{\mu(n)^2}{\phi(n)}}{\sum_{n \le D} \frac{\mu(n)^2}{\phi(n)}}.$$

We have seen that the denominator is $\ge \log D$. The numerator is

$$\le \prod_{p \le D/m} \frac{p}{p-1} \ll \log(2D/m)$$

by elementary prime number theory.

$$|\lambda_m| \ll \frac{m \log(2D/m)}{\phi(m) \log D},$$

whence

$$R \ll \left(\sum_{m \le D} \frac{m \log(2D/m)}{\phi(m) \log D} \right)^2 \ll D^2 \log^{-2} D.$$

Now one can take $D = \sqrt{X}$ in the Brun–Titchmarsh theorem and obtain

$$\pi(X + Y) - \pi(Y) \le \frac{2X}{\log X} + O\left(\frac{X}{\log^2 X} \right).$$

There is a new example which I need to mention.

Example 10 (prime k–tuples). *Let* $\mathbf{h} = h_1, h_2, \ldots, h_k$ *be a* k*–tuple of distinct positive integers and we are interested in the number* $\pi_k(X; \mathbf{h})$ *of* $m \leq X$ *such that the* $m + h_j$ *are simultaneously prime. Let* $a(n) = \text{card}\{m \leq X : (m + h_1) \ldots (m + h_k) = n\}$. *Then with* $P = \prod_{p \leq \sqrt{X}} p$ *we have*

$$\pi_k(X; \mathbf{h}) \leq \pi(\sqrt{X}) + S(A, P).$$

Now $A_d = f(d)X + R_d$ where $f(d) = \rho(d)/d$, $|R_d| \leq \rho(d)$ and $\rho(d)$ is the number of solutions of $(x + h_1) \ldots (x + h_k) \equiv 0 \pmod{d}$. Then $\rho \in \mathcal{M}$, $\rho(p) \leq k$ and when $p \nmid \Delta = \prod_{1 \leq i < j \leq k} |h_j - h_i|$ we have $\rho(p) = k$. This is an example pf a k–dimensional sieving situation. If the $\mathbf{h}$ give a complete set of residues modulo p for some prime p, then obviously there are not many k-tuples which are simultaneously prime. Thus a natural condition for the existence of many prime k-tuples is that $\rho(p) < p$ for all primes p, i.e $f(p) < 1$.

Because

$$\sum_{p\leq z}\frac{g(p)}{p}\sim k\log\log z\text{ as }z\to\infty$$

we expect that the optimal λ_m are quite close to

$$\lambda_m^*=\mu(m)\frac{\log^k D/m}{\log^k D}$$

and it is this which at least in part will motivate some choices in the work of Goldston, Pintz and Yıldırım. I should add also, since it will come up later, that following Hardy & Littlewood we expect that

$$\pi_k(X,\mathbf{h})\sim\mathfrak{S}_k(\mathbf{h})\frac{X}{\log^k X}$$

where

$$\mathfrak{S}_k(\mathbf{h})=\prod_p\big(1-\rho(p)/p\big)(1-1/p)^{-k}.$$

and the Selberg sieve will give

$$\pi_k(X,\mathbf{h})\ll\mathfrak{S}_k(\mathbf{h})\frac{X}{\log^k X}$$

Now consider Example 6 once more.

Example 6 (twin primes revisited). *We found*

$$\sum_{\substack{p\le Y\\ p+2 \text{ prime}}} 1 \le \pi(\sqrt{Y}) + S(A,P)$$

where $A_d = \pi(Y;d,-2) = f(d)X + R_d$ *with* $f(d) = 0$ *when* d *is even and* $f(d) = \frac{1}{\phi(d)}$ *when* d *is odd, and where* $X = \mathrm{li}(Y) = \int_2^Y \frac{dt}{\log t}$.

By the Selberg sieve

$$S(A,P) \le \frac{\mathrm{li}(Y)}{\sum_{\substack{d\le D\\ 2\nmid d}} \mu(d)^2 \prod_{p\mid d} \frac{1}{p-2}} + R$$

$$R = \sum_{d\le D}\sum_{e\le D} h(d)h(e)\left|\pi(Y;[d,e],-2) - \frac{\mathrm{li}(Y)}{\phi([d,e])}\right|$$

and $h(d) = 0$ $(2|D)$, $h(d) = \mu(d)^2 \prod_{p\mid d} \frac{p-1}{p-2}$ $(2 \nmid d)$. The sum in the main term is asymptotically $C \log D$ for large D. How big can we make D and yet control the error term? The Siegel–Walfisz theorem gives a rather small choice for D, $D \le (\log Y)^A$.

$$R = \sum_{d \leq D} \sum_{e \leq D} h(d)h(e) \left| \pi(Y; [d, e], -2) - \frac{\mathrm{li}(Y)}{\phi([d, e])} \right|$$

As we are averaging over d and e we can make use of the celebrated Bombieri–A. I. Vinogradov theorem, which in essence says that we have GRH on average. Thus it is possible to take D close to $Y^{1/4}$. The Bombieri–Vinogradov theorem, and the theory of the large sieve which underpins it will be the subject of my next two lectures.