On Holey Kitchen Devices and their uses in Number Theory

This notelet is a virtually unedited version of the mini-essay I wrote in December 2007 for Tim Browning's TCC Analytic Number Theory course assignment, which asked for:

An overview of the well-known topic: the Selberg sieve method and the inequality

$$\pi(x+y) - \pi(x) \leqslant \frac{2y}{\log y}\left(1 + O(1/\log y)\right).$$

Ever since the first baker tried to bake the first cake using slightly clumpy flour, sieves have fascinated mankind. But they have fascinated mathematicians too. And not just mathematicians who bake cakes. Mathematical sieves have their origin in antiquity. These sieves deal with the mathematical equivalent of clumpy flour - the prime numbers.

The general idea of sieves is to estimate the number of elements in a set $A \subset \mathbb{N}$ that are not divisible by any element of a set P of primes. This is done by "sifting out" certain residue classes of primes p that are in P. The sieve of Eratosthenes is perhaps the most well known example of such a sieve. At its simplest it gives an algorithm for finding all the prime numbers in the interval $[N, N^2)$, by crossing off the multiples of all primes up to and including N, then whatever remains uncrossed on the list must be prime, since any composite number $\leqslant N^2$ must have a prime factor $\leqslant N$. Thus, simply by knowing the primes up to N we can 'sift' the interval $[N, N^2)$ and be left with only prime numbers[1].

The sieve of Eratosthenes is over two thousand years old, so it is not surprising that newer, better sieves have been invented in the mean time. The science of sieves is notoriously heavy going. A lot of notation and hard work is required to achieve results that are relatively simple to verify using other techniques. Sieves do give results that other techniques cannot attain, though, including approximations to notoriously tricky problems like the twin primes and Golbach conjectures.

As previously mentioned they require a veritable kitchenful of notation, so we'll get that out of the way now. We'll let A and P be the sets mentioned at the beginning, we then set

$$A_d = \{a \in A : a \equiv 0 \pmod{d}\}.$$

So in particular $A_1 = A$. Since we are estimating the sizes of these sets we introduce $X > 1$ which approximates $|A|$, and let r_1 be the remainder

$$r_1 = |A| - X.$$

[1] Another simple but jolly nice example is to take a set A of positive square-free integers and sift out all the zero residue classes (i.e. multiples) of primes p such that $p \equiv 3 \pmod 4$. What is left is the set of all integers in S that can be written as the sum of two squares.

We also introduce a multiplicative function ω_0 where $\omega_0(p) \geqslant 0$ is chosen so that $\frac{\omega_0(p)}{p}X$ approximates $|A_p|$, and we go on to define more remainder terms

$$r_p = |A_p| - \frac{\omega_0(p)}{p}X.$$

Since ω_0 is a positive multiplicative function we have that $\omega_0(1) = 1$, and that

$$\omega_0(d) = \prod_{p|d} \omega_0(p)$$

for square-free d. Imaginatively we define the remainder terms

$$r_d = |A_d| - \frac{\omega_0(d)}{d}X.$$

We finally characterise all the primes we're interested in sifting out – the ones in P – with the number

$$P(z) = \prod_{\substack{p<z \\ p\in P}} p.$$

The so called *Sieve problem*, then, is to estimate the value of the sifting function

$$S(A; P, z) = |\{a \in A : (a, P(z)) = 1\}|,$$

or, more generally,

$$S(A_q; P, z) = |\{a \in A_q : (a, P(z)) = 1\}|,$$

where q is a square-free integer and $(q, P(z)) = \left(q, P^C\right) = 1$, where $\left(q, P^C\right) = 1$ means that q is coprime with every prime number not in P. To achieve this we need to fiddle with ω_0 slightly so that it better reflects the set P. We define the multiplicative function ω by

$$\omega(p) = \begin{cases} \omega_0(p) & \text{if } p \in P \\ 0 & \text{if } p \in P^C. \end{cases}$$

and $\omega(1) = 1$. Then $\omega(d) = \prod_{p|d} \omega(p)$ for square-free d, and we can define our ultimate remainder

$$R_d = |A_d| - \frac{\omega(d)}{d}X.$$

There will be more notation later, but this suffices for now.

Sieves with all the above trimmings were slowly brought into the modern world by Legendre and, more recently, Viggo Brun. It was Atle Selberg who came up with one of the better sieves of the twentieth century, though. His work, as with many of the other twentieth century sieves, relied on the following seemingly simple observation.

Let λ_d a sequence of real numbers satisfying $\lambda_1 = 1$, then

$$S(A; P, z) \leqslant \sum_{a\in A} \left(\sum_{\substack{d|a \\ d|P(z)}} \lambda_d \right)^2,$$

for if $a \in A$ and $(a, P_z) = 1$ then the sum in parentheses is simply 1, and there will be exactly $S(A; P, z)$ such values of a. All the other cases add non-negative terms to the sum, hence the inequality follows. Eratosthenes' sieve corresponds to a choice of λ_d using the μ-function, but we can no better estimate the resulting sum than we can estimate $S(A; P, z)$. The above series can be easily rewritten as

$$\sum_{a \in A} \Bigg(\sum_{\substack{d|a \\ d|P(z)}} \lambda_d \Bigg)^2 = \sum_{d_1, d_2 | P(z)} \lambda_{d_1} \lambda_{d_2} \sum_{\substack{a \in A \\ a \equiv 0 \pmod{D}}} 1,$$

where $D = \operatorname{lcm}(d_1, d_2) = [d_1, d_2]$. But clearly

$$\sum_{\substack{a \in A \\ a \equiv 0 \pmod{D}}} 1 = |A_D| = \frac{\omega(D)}{D} X + R_D.$$

And so

$$S(A; P, z) \leqslant X \sum_{d_1, d_2 | P(z)} \lambda_{d_1} \lambda_{d_2} \frac{\omega(D)}{D} + \sum_{d_1, d_2 | P(z)} |\lambda_{d_1} \lambda_{d_2} R_D| =: X\Sigma_1 + \Sigma_2.$$

Selberg's masterplan was to minimise the right hand side above by choosing the right sequence λ_d.[2] Alack and alas! Minimising the series proved to be far too hard in any generality, so to simplify matters Selberg threw away many of the λ_d, in particular he chose to set $\lambda_d = 0$ for any $d \geqslant z$. That left him with $[z] - 2$ values to choose in order to minimise the quadratic form Σ_1.

Minimising Σ_1 is all well and good, but it is done at the risk of ignoring Σ_2. This is okay though, since the condition that $\lambda_d = 0$ for $d \geqslant z$ ensures that the sum Σ_2 doesn't have too many terms, moreover the sum is made up of remainder terms and as Selberg (might have) put it himself: "a remainder term is just for Christmas, not for life."

Using Lagrangian multipliers, Selberg managed to minimise Σ_1 and found that the best values to take for the λ_d was

$$\lambda_d = \frac{\mu(d)}{\prod_{p|d} \left(1 - \frac{\omega(p)}{p}\right)} \frac{G_d(z/d)}{G(z)}$$

where

$$G_k(x) = \sum_{\substack{d < x \\ (d,k)=1}} \mu^2(d) g(d)$$

for a certain multiplicative function g. Plugging $d = 1$ into the above expression does indeed give $\lambda_1 = 1$, and moreover if $d \geqslant z$ then $G_d(z/d)$ is an empty sum

[2] If you prefer cooking analogies then the prime numbers can be thought of as the clumpier lumps of flour, while composite numbers are finer particles. Eratosthenes' sieve has perfectly sized holes so that only the lumpy prime numbers are left in the sieve. But Eratosthenes' sieve is also sitting over an active volcano, so after sifting we can't get a good idea of just how many prime numbers we have left. (Unless we count the flour particles one by one, not a fun job, especially over a volcano.) Selberg's sieve allows us to see better what gets left in the sieve, but at the cost of having slightly smaller holes that don't sift out all the composite numbers.

and hence we get $\lambda_d = 0$. Plugging these values of λ_d into our expression for Σ_1 may seem like an utterly terrifying task, but with a few pages of algebraic jiggery pokery we arrive at the very simple

$$\Sigma_1 = \frac{1}{G(z)}.$$

It's relatively straightforward to show that

$$G_d(z/d) \leqslant \prod_{p|d} \left(1 - \frac{\omega(p)}{p}\right) G(z)$$

so that $|\lambda_d| \leqslant 1$. Thence

$$\Sigma_2 \leqslant \sum_{\substack{d_1,d_2<z \\ d_1,d_2|P(z)}} |R_{[d_1,d_2]}|.$$

If we now ask ourselves how many different values of d_1, d_2 will proffer the same value of $d = [d_1, d_2]$ then after a few minutes of thought we should answer ourselves "why, it's $3^{\Omega(d)}$, where $\Omega(d)$ is the number of prime divisors of d, self." This is because if our desired value of d has prime factorisation $p_1 \cdots p_n$, say, then each p_i must appear as a factor of either d_1, or of d_2, or of both. So our estimate on Σ_2 becomes

$$\Sigma_2 \leqslant \sum_{\substack{d<z^2 \\ (d,P^C)=1}} \mu^2(d) 3^{\Omega(d)} |R_d|.$$

Putting everything together we get Selberg's upper bound for the sieve function:

$$S(A; P, z) \leqslant \frac{X}{G(z)} + \sum_{\substack{d<z^2 \\ (d,P^C)=1}} \mu^2(d) 3^{\Omega(d)} |R_d|.$$

If we choose our X and ω_0 correctly then this inequality can give powerful results. One of these results is known as the Brun-Titchmarsh inequality, named by Yu V. Linnik in 1961 after the pioneering sieve-meister Brun, as well as the mathematician, Fellow of the Royal Society, and everyone's favourite horticulturist, Titchmarsh[3]. If we write

$$\pi(x; k, \ell) = \#\{p \leqslant x : p \equiv \ell \pmod{k}\}$$

then, first of all, we can note that $\pi(x; 1, 0) = \pi(x)$. Using Selberg's work it is possible to show that

$$\pi(x+y; k, \ell) - \pi(x; k, \ell) \leqslant \frac{y}{\phi(k) \log \sqrt{\frac{y}{k}}} \left(1 + \frac{4}{\log \sqrt{\frac{y}{k}}}\right).$$

[3] Apologies to Edward C. Titchmarsh.

Setting $k = 1, \ell = 0$ we get

$$\begin{aligned}\pi(x+y) - \pi(x) &\leqslant \frac{y}{\log \sqrt{y}} \left(1 + \frac{4}{\log \sqrt{y}}\right) \\ &= \frac{2y}{\log y} \left(1 + \frac{8}{\log y}\right) \\ &= \frac{2y}{\log y} \left(1 + \mathcal{O}\left(\frac{1}{\log y}\right)\right).\end{aligned}$$