When the sieve works

Dimitris Koukoulopoulos
joint work with Andrew Granville and Kaisa Matomäki

Centre de recherches mathématiques
Université de Montréal

Canadian Number Theory Association XII Meeting
University of Lethbridge
22 June 2012

The general sieve problem

Given $A \subset \mathbb{N}$ and a set of primes P, what is the size of

$$S(A, P) := \#\{a \in A : p|a \Rightarrow p \notin P\} \quad ?$$

Examples:

- Taking $A = \mathbb{N} \cap [1, x]$, $P = \{p \leq \sqrt{x}\}$ we count primes.
- Taking $A = \{n(n+2) : n \leq x\}$, $P = \{p \leq \sqrt{x}\}$ we count twin primes.
- Taking $A = \{n \leq x : n \equiv 1 \pmod 4\}$, $P = \{p \leq \sqrt{x} : p \equiv 3 \pmod 4\}$ we count (a dense subset of) numbers that can be written as the sum of two squares (Iwaniec).
- Taking $A = \mathbb{N} \cap [1, x]$, $P = \{p > y\}$ we count y-smooth/friable numbers.

Goal of classical sieve methods: Given A, estimate $S(A, P)$ for $P \subset \{p \leq y\}$ with y as large as possible (ideally, with $y^2 \approx \max\{p | \prod_{a \in A} a\}$).

A heuristic argument

We focus on the case when $A = \mathbb{N} \cap [1, x]$, $P \subset \{p \leq x\}$. We let

$$S(x, P) = \#\{n \leq x : p|n \Rightarrow p \notin P\}.$$

Heuristically, for a prime p

$$\mathbf{Prob}\,(n \leq x : p|n) = \frac{\lfloor x/p \rfloor}{\lfloor x \rfloor} \approx \frac{1}{p}.$$

In general, for primes $p_1 < p_2 < \cdots < p_r$

$$\mathbf{Prob}\,(n \leq x : p_1 \cdots p_r | n) = \frac{\lfloor x/(p_1 \cdots p_r) \rfloor}{\lfloor x \rfloor} \approx \frac{1}{p_1} \cdots \frac{1}{p_r}.$$

So, we expect that

$$\frac{S(x, P)}{\lfloor x \rfloor} = \mathbf{Prob}\,(n \leq x : p \nmid n \,\forall p \in P) \approx \prod_{p \in P} \left(1 - \frac{1}{p}\right).$$

Expectations and reality

We know that $\#\{p \leq x\} \sim x/\log x$. However, the heuristic predicts that

$$\frac{\#\{p \leq x\}}{x} \sim \frac{S(x, \{p \leq \sqrt{x}\})}{x} \sim \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right) \sim \frac{2e^{-\gamma}}{\log x}; \quad 2e^{-\gamma} > 1.$$

In general,

$$S(x; P) \ll x \prod_{p \in P} \left(1 - \frac{1}{p}\right).$$

Also, if $\max P \leq x^{1/2-\epsilon}$, then

$$S(x; P) \asymp_{\epsilon} x \prod_{p \in P} \left(1 - \frac{1}{p}\right).$$

But if $P = \{x^{1/u} < p \leq x\}$, then $S(A, P) = x/u^{(1+o(1))u}$, whereas the prediction is that $S(A, P) \approx x/u$.

When does the sieve work?

Question

When does the sieve work or, more precisely, when is it true that

$$S(x,P) \asymp x \prod_{p\in P}\left(1-\frac{1}{p}\right) \quad ? \qquad (*)$$

Hildebrand showed that the smooth primes are the extreme example: Let $u \geq 1$ and $P \subset \{p \leq x\}$.

$$\sum_{p\in P}\frac{1}{p} \lesssim \log u \quad \Longrightarrow \quad S(x,P) \gtrsim S(x,\{x^{1/u} < p \leq x\}) = \frac{x}{u^{(1+o(1))u}}.$$

It is generally expected that if P^c contains enough many big primes, then $(*)$ should hold.

For this reason, we use the complementary notation

$$Q = \{p \leq x\} \setminus P, \quad \Psi(x;Q) = S(x,P) = \#\{n \leq x : p|n \Rightarrow p \in Q\}.$$

The effect of the big primes

Proposition

If $Q \subset \{p \le x^{1-\epsilon}\}$, $u \in [1, \log x]$ and $\kappa = \sum_{q\in Q\cap[x^{1/u},x]} 1/q$, then

$$\frac{\Psi(x;Q)}{x} \ll_\epsilon \left(\kappa + u^{-\epsilon u/2} + x^{-1/10}\right) \cdot \prod_{p\le x,\, p\notin Q}\left(1-\frac{1}{p}\right).$$

Proposition

If $\epsilon > 0$, $Q \subset \{p \le x\}$ and $u \in [1, \log x]$ are such that

$$\sum_{q\in Q,\, x^{1/u}<q\le x} \frac{1}{q} > \epsilon,$$

then $\exists t \in [x^{1/u}, x]$: $\quad \dfrac{\Psi(t;Q)}{t} \gg \dfrac{\epsilon \min\{1, \epsilon u\}}{\log u} \displaystyle\prod_{p\le t,\, p\notin Q}\left(1-\frac{1}{p}\right).$

A different extremal example

The key is how big $\sum_{q \in Q \cap [x^{1/u}, x]} 1/q$ is. Consider

$$Q = \bigcup_{m=1}^{N-1} \left\{ x^{\frac{m}{N+1}} < p < x^{\frac{m}{N}} \right\},$$

If $n \leq x$ has all its prime factors in Q, then $n \in \bigcup_{m=1}^{N} \left(x^{\frac{m}{N+1}}, x^{\frac{m}{N}} \right)$.

$$\Psi(x; Q) = O(x^{1-1/N}) + \sum_{\substack{x^{\frac{N}{N+1}} < n \leq x \\ p|n \Rightarrow p \in Q}} 1 \ll_N \frac{x}{\log^2 x}.$$

Note that

$$\sum_{q \in Q} \frac{1}{q} = (N-1) \log \frac{N+1}{N} \sim 1 - \frac{3/2 + o(1)}{N} < 1.$$

A problem in additive combinatorics

Estimating $\Psi(x; Q)$ is essentially equivalent to finding solutions to

$$\log p_1 + \cdots + \log p_r = \log x + O(1) \quad (r \in \mathbb{N},\ p_1, \ldots, p_r \in Q).$$

Theorem (Bleichenbacher)

Let $T \subset (0,1)$ be open. If $\int_T dt/t > 1$, then there are $t_1, \cdots, t_k \in T$ such that $t_1 + \cdots + t_k = 1$. This is optimal, as the example $T = \bigcup_{m=1}^{N} \left(\frac{m}{N+1}, \frac{m}{N}\right)$ shows.

Corollary (Lenstra-Pomerance)

Let $Q \subset \{p \leq x\}$, $u \geq 1$.

$$\sum_{q \in Q,\, x^{1/u} < q \leq x} \frac{1}{q} > 1+\epsilon \quad \Rightarrow \quad \frac{\Psi(x; Q)}{x} \gg_\epsilon \frac{e^{-O(u)}}{(\log x)^{u-1}} \prod_{p \leq x,\, p \notin P} \left(1 - \frac{1}{p}\right).$$

Quantitative Bleichenbacher

The main defect of Bleichenbacher's theorem is that it does not say anything about how many solutions there are to $e_1 + \cdots + e_k = 1$ other than that there is at least 1.

It is easier to look at the discrete analogue of this problem: Given $A \subset [1, N] \cap \mathbb{N}$ with $\sum_{a \in A} 1/a > 1$, how many solutions are there to $a_1 + \cdots + a_k = N + O(1)$ with $k \in \mathbb{N}$, $a_1, \ldots, a_k \in A$?

Theorem (Granville-K-Matomäki)

$\exists \lambda > 1, c > 0$ such that if $1 \leq u \leq c\sqrt{N}$, $A \subset [N/u, N] \cap \mathbb{N}$ satisfy $\sum_{a \in A} 1/a \geq \lambda$, then $\exists k \in \mathbb{N}, n \in [N - k, N]$ such that

$$\sum_{\substack{(a_1,\ldots,a_k) \in A^k \\ a_1 + \cdots + a_k = n}} \frac{1}{a_1 \cdots a_k} \gg \frac{u^{-O(u)}}{N} \left(\sum_{a \in A} \frac{1}{a} \right)^k .$$

Application to the sieve

Corollary

$\exists\lambda' > 1, c' > 0$ such that if $Q \subset \{p \le x\}$, $1 \le u \le c'\sqrt{\log x}$, then

$$\sum_{q\in Q,\, x^{1/u}<q\le x} \frac{1}{q} \ge \lambda' \quad \Rightarrow \quad \frac{\Psi(x;Q)}{x} \gg \frac{1}{u^{O(u)}} \prod_{p\le x,\, p\notin Q} \left(1-\frac{1}{p}\right).$$

Motivated by Bleichenbacher's theorem, we conjecture that this result holds for any $\lambda' > 1$.

Sketch of the proof

For sets of integers C, D, let $C + D = \{c + d : c \in C,\ d \in D\}$.
$\exists v \in [1, u]$ such that the set $B = A \cap [1, N/v]$ has $\geq \frac{\lambda N}{2v^2}$ elements. We will show that

$$\exists k : \quad \#\{(b_1, \ldots, b_k) \in B^k : b_1 + \cdots + b_k \in [N - k, N]\} \geq \frac{|B|^k}{u^{O(u)} N}.$$

Varying Ruzsa-Chang: if $|B + B| \leq 4|B|$, then $B + B + B$ contains a GAP $P = \{a_0 + a_1 k_1 + \cdots + a_d k_d : |k_j| \leq K_j\}$ of size $|P| \gg |B|$ and rank $d \ll 1$. Also, $r_{B+B+B}(n) \gg |B|^2\ \forall n \in P$. So (*) follows.

If $|B + B| > 4|B|$, replace B with $2B = B + B$ and repeat.

$$2B \subset [1, 2N/v] = [1, N/(v/2)] \quad \text{and} \quad |2B| > 4 \cdot \frac{\lambda N}{2v^2} = \frac{\lambda N}{2(v/2)^2}.$$

Apply induction; this process terminates at some k with $2^k \leq 2v/\lambda$.

Problem: We need to keep track of the representations!
Use instead restricted sumsets $\{n \in B + B : r_{B+B}(n) \geq \eta |B|\}$ (ideas from Balog-Szemeredi-Gowers theorem).

An application

Let f be a Hecke eigencuspform for $SZ_2(\mathbb{Z})$ of weight k. It has $k/12 + O(1)$ zeroes on the upper half plane $\mathbb{H}$, which are equidistributed by QUE (Rudnick).

Ghosh and Sarnak initiated the study of "real" zeroes of f, i.e. zeroes on the geodesics

$$\delta_1 = \{z \in \mathbb{H} : \Re(z) = 0\}, \quad \delta_2 = \{z \in \mathbb{H} : \Re(z) = 1/2\}$$
$$\delta_3 = \{z \in \mathbb{H} : |z| = 1, 0 \leq \Re(z) \leq 1/2\}.$$

They showed that

$$N(f) := \#\{z \in \delta_1 \cup \delta_2 : f(z) = 0\} \gg_\epsilon k^{1/4-1/80-\epsilon}.$$

Matokäki, using methods described before, showed that $N(f) \gg_\epsilon k^{1/4-\epsilon}$.

How sieve methods enter the picture

$$f(z) = \sum_{n=1}^{\infty} \lambda(n) n^{(k-1)/2} e^{2n\pi i z}.$$

Ghosh-Sarnak: If $C \leq m \leq \epsilon\sqrt{k/\log k}$, $\alpha \in \mathbb{R}$, and $y_m = \frac{k-1}{4\pi m}$, then

$$\left(\frac{e}{m}\right)^{(k-1)/2} f(\alpha + iy_m) = \lambda(m) e^{2m\pi i\alpha} + O(k^{-\delta}).$$

So if m_1 is even, m_2 is odd, and $|\lambda(m_1)|, |\lambda(m_2)| \geq k^{-\delta/2}$, then f has a zero in the line segment connecting $\alpha + iy_{m_1}$ and $\alpha + iy_{m_2}$ for $\alpha = 0$ or $\alpha = 1/2$, i.e in $\delta_1 \cup \delta_2$.

Since $\lambda(p)^2 = \lambda(p^2) + 1$, we have that $\max\{|\lambda(p)|, |\lambda(p^2)|\} \geq 1/2$.

So we need to show that $N_1 \cup N_2$ contains many integers, where

$$N_j = \{n \in \mathbb{N} : n \text{ square-free and odd, } p|n \Rightarrow |\lambda(p^j)| \geq 1/2\} \quad (j = 1, 2).$$

Even though we don't have much control over the location of the primes in $P_j = \{p > 2 : |\lambda(p^j)| \geq 1/2\}$ for $j = 1, 2$, the methods described before are general enough that can handle this problem.

Thank you for your attention!