POISSON IMITATORS AND SIEVE THEORY

ZARATHUSTRA BRADY

Abstract. Sieve theory is actually a question about probability distributions whose low moments agree with the low moments of Poisson distributions. In particular, we can derive Selberg's "parity problem" without using properties of the Möbius function or the Liouville function - instead, we use the fact that the alternating group forms a subgroup of the symmetric group.

1. Poisson Imitators

At a high level, sieve theory tries to bound the chance that a random number n is prime by showing that the naïve heuristic that each prime $p < \sqrt{n}$ independently has a $\frac{1}{p}$ chance of being a divisor of n is not too terribly wrong. If we disregard the very small primes (i.e., the primes less than $\ln(n)$), then any particular prime p is unlikely to divide n. If each random event $p \mid n$ was jointly independent of all other events $q \mid n$ for other primes q, then the number of prime divisors of n would be a sum of independent random variables in $\{0,1\}$, each of which is very unlikely to be 1, so we would expect the number of prime divisors of n to behave like a Poisson distribution. In reality, however, these events are only jointly independent in an approximate sense. While it is true that the random events $p \mid n$ and $q \mid n$ are approximately independent if $p, q < \sqrt{n}$, if we instead pick three primes p, q, r with $p \sim n^{1/2}$, $q \sim n^{1/3}$, $r \sim n^{1/4}$, then the random events $p \mid n$, $q \mid n$, $r \mid n$ are *not* jointly independent, since it is completely impossible for all three of p, q, r to divide n. The goal of sieve theory is to quantify how much this failure of independence could matter.

This leads us to consider the following setup. We have a given number k, Poisson rates $\nu_1, ..., \nu_k \geq 0$, and "widths" $\alpha_1, ..., \alpha_k \in [0,1]$. We wish to study probability distributions on k-tuples of natural numbers $(X_1, ..., X_k)$, such that the variable X_i approximately acts like a Poisson with rate ν_i, and such that for $i \neq j$, the variables X_i and X_j are approximately independent. The role of the widths α_i is that for α_i small, the approximations become more accurate - so the width α_i describes how much "wiggle room" the distribution of X_i has. We formalize this as follows.

Definition 1. We say that a k-tuple of random variables $(X_1, ..., X_k) \in \mathbb{N}^k$ is a *Poisson imitator* with parameters $(\nu_1, \alpha_1), ..., (\nu_k, \alpha_k)$ if the X_i satisfy the following moment constraints:

$$\forall m \in \mathbb{N}^k \text{ such that } \sum_i \alpha_i m_i \leq 1, \text{ we have } \mathbb{E}\Big[\binom{X_1}{m_1} \cdots \binom{X_k}{m_k}\Big] = \frac{\nu_1^{m_1}}{m_1!} \cdots \frac{\nu_k^{m_k}}{m_k!}.$$

In symbols, we will abbreviate this as $X \sim (\nu, \alpha)$.

To make the connection to Poisson distributions explicit, we point out the following description of the moments of a Poisson distribution.

Proposition 1. *A random variable $X \in \mathbb{N}$ is Poisson with rate ν (i.e., $\mathbb{P}[X = n] = e^{-\nu}\frac{\nu^n}{n!}$) if and only if X satisfies the moment conditions*

$$\forall m \in \mathbb{N}, \text{ we have } \mathbb{E}\Big[\binom{X}{m}\Big] = \frac{\nu^m}{m!}.$$

This material is based upon work supported by the NSF Mathematical Sciences Postdoctoral Research Fellowship under Grant No. (DMS-1705177).

In particular, if $X_1, ..., X_k$ are independent Poissons with rates $\nu_1, ..., \nu_k$, then $(X_1, ..., X_k)$ is a Poisson imitator with parameters $(\nu_1, 0), ..., (\nu_k, 0)$.

The main question we have about Poisson imitators is what we can say about the probability that $(X_1, ..., X_k) = (0, ..., 0)$. Note that in the special case where the X_i are independent Poissons with rates ν_i, we have

$$\mathbb{P}\big[(X_1, ..., X_k) = (0, ..., 0)\big] \;=\; e^{-\nu_1 - \cdots - \nu_k}.$$

So we will always compare other Poisson imitators to this baseline.

Definition 2. For $\nu \in \mathbb{R}^k_{\geq 0}$ and $\alpha \in [0,1]^k$, we define the *upper bound* function $F(\nu, \alpha)$ by

$$F(\nu, \alpha) \coloneqq \sup_{X \sim (\nu, \alpha)} e^{\sum_i \nu_i} \cdot \mathbb{P}[X = (0, ..., 0)].$$

Similarly, we define the *lower bound* function $f(\nu, \alpha)$ by

$$f(\nu, \alpha) \coloneqq \inf_{X \sim (\nu, \alpha)} e^{\sum_i \nu_i} \cdot \mathbb{P}[X = (0, ..., 0)].$$

Note that we always have

$$0 \;\leq\; f(\nu, \alpha) \;\leq\; 1 \;\leq\; F(\nu, \alpha) \;\leq\; e^{\sum_i \nu_i}.$$

A nice family of Poisson imitators can be constructed by considering the distribution of cycle sizes of a random permutation on the set $[n] = \{1, ..., n\}$.

Definition 3. Let σ be a permutation of $[n]$. Define X_i^σ to be the number of i-cycles which occur in σ.

Proposition 2. *If σ is a uniformly random permutation of $[n]$, then the random variables X_i^σ satisfy the moment conditions*

$$\forall m \in \mathbb{N}^k \text{ such that } \sum_i i m_i \leq n, \text{ we have } \mathbb{E}\Big[\binom{X_1^\sigma}{m_1} \cdots \binom{X_k^\sigma}{m_k}\Big] = \prod_i \frac{1}{i^{m_i} m_i!},$$

together with the extra condition

$$\sum_i i X_i^\sigma = n.$$

In particular, for any $k \leq n$ we see that $(X_1^\sigma, ..., X_k^\sigma)$ is a Poisson imitator with parameters $(1, 1/n), ..., (1/k, k/n)$.

Corollary 1. *For any $\frac{n}{2} \leq k < n$, we have*

$$F\big((1, 1/n), ..., (1/k, k/n)\big) \;\geq\; e^{1 + \cdots + 1/k} \cdot \frac{1}{n} \;\geq\; f\big((1, 1/n), ..., (1/k, k/n)\big).$$

For n large, we can use the approximation $1 + \cdots + 1/k \approx \ln(k) + \gamma$ to see that

$$e^{1 + \cdots + 1/k} \cdot \frac{1}{n} \approx e^\gamma \cdot \frac{k}{n}.$$

So we see that the constant e^γ, which is so common throughout sieve-theoretic arguments, comes up quite naturally in this framework. We also have another (pair of) families of Poisson imitators, analogous to the parity problem from sieve theory.

Proposition 3. *If σ is a uniformly random* even *permutation of $[n]$, then the random variables X_i^σ satisfy the moment conditions*

$$\forall m \in \mathbb{N}^k \text{ such that } \sum_i i m_i \leq n-2, \text{ we have } \mathbb{E}\Big[\binom{X_1^\sigma}{m_1} \cdots \binom{X_k^\sigma}{m_k}\Big] = \prod_i \frac{1}{i^{m_i} m_i!},$$

together with the extra conditions

$$\sum_i iX_i^\sigma = n,$$

$$\sum_i (i-1)X_i^\sigma \equiv 0 \pmod 2.$$

In particular, for any $k \le n$ we see that $(X_1^\sigma, ..., X_k^\sigma)$ is a Poisson imitator with parameters $(1, 1/(n-2)), ..., (1/k, k/(n-2))$. A similar statement holds if σ is taken to be a uniformly random odd *permutation of $[n]$.*

Corollary 2. *For any $k < n$, we have*

$$F\big((1, 1/(n-2)), ..., (1/k, k/(n-2))\big) \;\; \ge \;\; e^{1+\cdots+1/k} \cdot \frac{2}{n} \;\; \approx \;\; 2e^\gamma \cdot \frac{k}{n},$$

and for $\frac{n}{2} \le k < n$, we have

$$f\big((1, 1/(n-2)), ..., (1/k, k/(n-2))\big) = 0.$$

From these examples, we can construct many other examples with a few simple constructions:

- we can drop any X_i from a tuple of Poisson imitators (and drop the corresponding pair (ν_i, α_i)),
- we can increase any width α_i (without changing anything else),
- we can take independent copies of two Poisson imitators $(X_1, ..., X_k), (X_1', ..., X_l')$ and put them together to make the tuple $(X_1, ..., X_k, X_1', ..., X_l')$ (concatenating the lists of νs and αs),
- we can merge X_i and X_j within a tuple, replacing them with $X_{ij} = X_i + X_j$, and define the corresponding parameters to be $\nu_{ij} = \nu_i + \nu_j$, $\alpha_{ij} = \max(\alpha_i, \alpha_j)$.

What we are really interested in are limiting problems, where the sizes of the tuples go to infinity while the sum of $\nu_i\alpha_i$ remains bounded. We define limiting functions $F_\kappa(s), f_\kappa(s)$ as follows.

Definition 4. For $\kappa : [0,1] \to \mathbb{R}_{\ge 0}$ well-behaved and $s > 1$, we define parameters $\nu^n \in \mathbb{R}_{\ge 0}^{n-1}$, $\alpha^n \in [0,1]^{n-1}$ by

$$\nu_i^n = \int_{\frac{i}{ns}}^{\frac{i+1}{ns}} \frac{\kappa(x)}{x}\, dx$$

and

$$\alpha_i^n = \frac{i+1}{ns}.$$

We define $F_\kappa(s)$ by

$$F_\kappa(s) = \lim_{n\to\infty} F\big((\nu_1^n, \alpha_1^n), ..., (\nu_{n-1}^n, \alpha_{n-1}^n)\big),$$

and similarly define $f_\kappa(s)$ by

$$f_\kappa(s) = \lim_{n\to\infty} f\big((\nu_1^n, \alpha_1^n), ..., (\nu_{n-1}^n, \alpha_{n-1}^n)\big).$$

To motivate the definition of this limiting problem, it may be helpful to think of each pair (ν, α) as describing a discrete measure μ on $[0,1]$ with finite support: the α_is describe the support of μ, while each ν_i describes the measure which μ assigns to the point α_i (assuming that the α_is are distinct). The limiting problem we are interested in involves approximating a measure μ on $[0,1]$ which satisfies

$$\int_0^1 x \; d\mu(x) < \infty$$

by a sequence of discrete measures of finite support. The function $\kappa : [0,1] \to \mathbb{R}_{\geq 0}$ corresponds to such a measure μ by the formula

$$\kappa(x) = x\frac{d\mu(x)}{dx},$$

and the parameter $s > 1$ is used to indicate that we wish to truncate the measure μ to the interval $[0, 1/s]$. In order to check that the limiting problem defined above behaves sensibly, we need an analogue of the Fundamental Lemma of sieve theory, which we will prove later in the paper.

In most cases we consider, the function κ will be a constant function. Taking the limit of our permutation examples, we have the abstract parity problem.

Theorem 1 (Parity problem)**.** *For $1 < s < 2$, we have $f_1(s) = 0$, and for all $s > 1$, we have $F_1(s) \geq \frac{2e^\gamma}{s}$.*

2. Sieve weights

How do we actually compute the functions $F(\nu, \alpha), f(\nu, \alpha)$? Let's focus on $F(\nu, \alpha)$. The standard technique for bounding a probability is to apply Markov's inequality: we choose a function

$$\theta : \mathbb{N}^k \to \mathbb{R}_{\geq 0}$$

such that

$$\theta(0, ..., 0) = 1,$$

and apply the bound

$$\mathbb{P}\big[X = (0, ..., 0)\big] \leq \mathbb{E}[\theta(X)].$$

In order to compute the right hand side, we choose θ to be a linear combination of the moments of X which we have values for. Thus, we take θ of the form

$$\theta(x) = \sum_{m \in \mathbb{N}^k} \lambda_m \prod_i \binom{x_i}{m_i},$$

with the λ_ms supported on m such that $\alpha \cdot m \leq 1$. In order to ensure that $\theta(0) = 1$, we take $\lambda_0 = 1$.

Definition 5. A pair of functions $\lambda, \theta : \mathbb{N}^k \to \mathbb{R}$ is called a *system of sieve weights* if

$$\theta(x) = \sum_{m \in \mathbb{N}^k} \lambda_m \prod_i \binom{x_i}{m_i}$$

and

$$\theta(0, ..., 0) = \lambda_{(0,...,0)} = 1.$$

We say that (λ, θ) is *compatible* with the widths α if

$$\lambda_m \neq 0 \quad \Longrightarrow \quad \alpha \cdot m \leq 1,$$

and we write such a system as $(\lambda, \theta)_\alpha$.

We say that a system of sieve weights (λ, θ) forms an *upper bound sieve*, written $(\lambda, \theta) \geq 0$, if θ satisfies

$$x \in \mathbb{N}^k \quad \Longrightarrow \quad \theta(x) \geq 0,$$

and we say that (λ, θ) forms a *lower bound sieve*, written $(\lambda, \theta) \leq 0$, if θ satisfies

$$x \in \mathbb{N}^k \setminus \{(0, ..., 0)\} \quad \Longrightarrow \quad \theta(x) \leq 0.$$

If (λ,θ) forms an upper bound sieve which is compatible with α, then for any $X \sim (\nu,\alpha)$ we have

$$\mathbb{P}\big[X = (0, ..., 0)\big] \le \mathbb{E}[\theta(X)] = \sum_{m\in\mathbb{N}^k} \lambda_m \prod_i \frac{\nu_i^{m_i}}{m_i!}.$$

Alternatively, since the right hand side of the above comes out the same regardless of which Poisson imitators we take for X, we can evaluate it in the case where the X_i are independent Poissons, to get the formula

$$\sum_{m\in\mathbb{N}^k} \lambda_m \prod_i \frac{\nu_i^{m_i}}{m_i!} \quad = \quad e^{-\sum_i \nu_i} \cdot \sum_{x\in\mathbb{N}^k} \theta(x) \prod_i \frac{\nu_i^{x_i}}{x_i!}.$$

This formula is also easy to prove directly. As a consequence, we see that for an upper bound sieve (λ,θ) which is compatible with α, we have

$$F(\nu,\alpha) \le \sum_{x\in\mathbb{N}^k} \theta(x) \prod_i \frac{\nu_i^{x_i}}{x_i!}.$$

Is this approach the best we can do? In fact it is!

Theorem 2. *For any rates $\nu \in \mathbb{R}^k_{\ge 0}$ and widths $\alpha \in (0,1]^k$, we have*

$$F(\nu,\alpha) = \min_{(\lambda,\theta)_\alpha \ge 0} \sum_{x\in\mathbb{N}^k} \theta(x) \prod_i \frac{\nu_i^{x_i}}{x_i!},$$

where the min *is over upper bound sieves (λ,θ) which are compatible with α, and*

$$f(\nu,\alpha) = \max\Big(0,\ \max_{(\lambda,\theta)_\alpha \le 0} \sum_{x\in\mathbb{N}^k} \theta(x) \prod_i \frac{\nu_i^{x_i}}{x_i!}\Big).$$

This is a special case of the following general result from the theory of convex optimization.

Theorem 3. *Let S be a countable set, let $s_0 \in S$, let $M_0, ..., M_k : S \to \mathbb{R}_{\ge 0}$ be any nonnegative functions with $M_0 = 1$, and let $\mu_0, ..., \mu_k \in \mathbb{R}_{\ge 0}$. If $\mathcal{X}$ denotes the set of random variables X taking values in S such that for each i we have*

$$\mathbb{E}[M_i(X)] = \mu_i,$$

and if $\mathcal{Y}^\pm$ denotes the set of tuples of weights $(\lambda_0, ..., \lambda_k)$ such that

$$x \in S \quad \Longrightarrow \quad \pm\sum_i \lambda_i M_i(x) \ge \begin{cases} \pm 1 & x = s_0, \\ 0 & x \ne s_0, \end{cases}$$

then as long as $\mathcal{X}$ contains a random variable X_0 with full support S we have

$$\begin{aligned} \sup_{X\in\mathcal{X}} \mathbb{P}[X = s_0] \quad &= \quad \min_{\lambda\in\mathcal{Y}^+} \sum_i \lambda_i \mu_i, \\ \inf_{X\in\mathcal{X}} \mathbb{P}[X = s_0] \quad &= \quad \max_{\lambda\in\mathcal{Y}^-} \sum_i \lambda_i \mu_i. \end{aligned}$$

Proof. We will only prove the formula for $\sup_{\mathcal{X}} \mathbb{P}[X = s_0]$, as the other formula is proved similarly. We may assume without loss of generality that $M_0, ..., M_k$ are linearly independent functions.

Let $M_0' : S \to \mathbb{R}$ be the function with $M_0'(s_0) = -1$ and $M_0'(x) = 0$ for $x \ne s_0$. For any $x \in S$, we define $M(x)$ by

$$M(x) = \big(M_0'(x), M_1(x), ..., M_k(x)\big) \in \mathbb{R}^{k+1}.$$

Then for every $C > \sup_{\mathcal{X}} \mathbb{P}[X = s_0]$ we must have

$$(-C, \mu_1, ..., \mu_k) \notin \operatorname{Conv}\{M(x) \mid x \in S\}.$$

Thus there must be some separating hyperplane, that is, there is some nonzero vector $v = (v_0, ..., v_k)$ such that

$$v \cdot M(x) \geq -v_0 C + \sum_{i>0} v_i \mu_i$$

for all $x \in S$ and $C > \sup_{\mathcal{X}} \mathbb{P}[X = s_0]$. Since no linear combination of $M_1, ..., M_k$ is constant and since there is some $X_0 \in \mathcal{X}$ with full support, we must have $v_0 \neq 0$. Since we may take C arbitrarily large, we must have $v_0 > 0$. If we set $\lambda_i = \frac{v_i}{v_0}$ for $i > 0$, then we see that

$$M_0'(x) + \sum_{i>0} \lambda_i M_i(x) \geq -\sup_{X \in \mathcal{X}} \mathbb{P}[X = s_0] + \sum_{i>0} \lambda_i \mu_i,$$

so we can take $\lambda_0 = \sup_{\mathcal{X}} \mathbb{P}[X = s_0] - \sum_{i>0} \lambda_i \mu_i$ to complete the proof. □

Thus we can approximate $F(\nu, \alpha)$ from above by finding an upper bound sieve $(\lambda, \theta)_\alpha$, and we can approximate it from below by finding a Poisson imitator $X \sim (\nu, \alpha)$. If we happen to find an optimal upper bound sieve and an optimal Poisson imitator simultaneously, then the inequality $\theta(m) \geq 0$ must have equality whenever $\mathbb{P}[X = m]$ is positive.

Proposition 4. *An upper bound sieve $(\lambda, \theta)_\alpha \geq 0$ and a Poisson imitator $X \sim (\nu, \alpha)$ will satisfy*

$$\mathbb{P}\big[X = (0, ..., 0)\big] \quad = \quad e^{-\sum_i \nu_i} F(\nu, \alpha) \quad = \quad \sum_{m \in \mathbb{N}^k} \lambda_m \prod_i \frac{\nu_i^{m_i}}{m_i!}$$

if and only if they satisfy the complementary slackness *condition*

$$\forall m \in \mathbb{N}^k \setminus \{(0, ..., 0)\}, \;\; \theta(m) \cdot \mathbb{P}[X = m] = 0.$$

Proof. We have

$$\sum_{m \in \mathbb{N}^k} \lambda_m \prod_i \frac{\nu_i^{m_i}}{m_i!} = \mathbb{E}[\theta(X)]$$

since $X \sim (\nu, \alpha)$, and

$$\mathbb{E}[\theta(X)] - \mathbb{P}\big[X = (0, ..., 0)\big] = \sum_{m \in \mathbb{N}^k \setminus \{(0, ..., 0)\}} \theta(m) \cdot \mathbb{P}[X = m]. \qquad □$$

Let's go back to the case $\nu_i = \frac{1}{i}, \alpha_i = \frac{i}{n-2}$, where i runs from 1 to k. We can use a discretized version of Selberg's sieve to show that for n even and $\frac{n-2}{2} \leq k < n$, the upper bound parity problem construction is best possible. We define an upper bound sieve by

$$\theta(x) = \Big(\sum_{m \leq x} (-1)^{\sum_i m_i} \Big(1 - 2 \sum_i \frac{i}{n} m_i \Big)_+ \prod_i \binom{x_i}{m_i} \Big)^2,$$

where $(1 - 2 \sum_i \frac{i}{n} m_i)_+$ refers to the positive part of $1 - 2 \sum_i \frac{i}{n} m_i$, i.e.

$$a_+ := \begin{cases} a & a \geq 0, \\ 0 & a \leq 0. \end{cases}$$

That this sieve is compatible with α follows from the fact that the corresponding λs are given by

$$\lambda_m = \sum_{a,b \leq m \leq a+b} (-1)^{\sum_i a_i + b_i} \Big(1 - 2 \sum_i \frac{i}{n} a_i \Big)_+ \Big(1 - 2 \sum_i \frac{i}{n} b_i \Big)_+ \prod_i \binom{m_i}{m_i - a_i, m_i - b_i, a_i + b_i - m_i},$$

and the only way for any summand to be nonzero is if we have $\sum_i i a_i, \sum_i i b_i \leq \frac{n}{2} - 1$ (this is where we use the assumption that n is even), so $\sum_i i m_i \leq \sum_i i(a_i + b_i) \leq n - 2$.

Since the upper bound parity problem Poisson imitator X^σ was supported on tuples with

$$\sum_i iX_i^\sigma = n,$$
$$\sum_i X_i^\sigma \equiv 1 \pmod{2},$$

we just have to check that $\theta(x) = 0$ when $(x_1, ..., x_k)$ can be extended to a sequence $(x_1, ..., x_n)$ such that $\sum_{i\le n} ix_i = n$, $\sum_{i\le n} x_i$ is odd, and $\sum_{i\le k} x_i \ge 1$. A proof of this fact, in a slightly different setting, can be found in Proposition 45 of [1]. We have proved the following result.

Theorem 4 (Selberg's upper bound is optimal for $\kappa = 1$ and $s < 2$)**.** *For n even and $\frac{n-2}{2} \le k < n$, we have*

$$F\big((1, 1/(n-2)), ..., (1/k, k/(n-2))\big) = e^{1+\cdots+1/k} \cdot \frac{2}{n} \approx 2e^\gamma \frac{k}{n}.$$

Taking the limit as $n \to \infty$, we have

$$F_1(s) = \frac{2e^\gamma}{s}$$

for $1 < s < 2$.

3. Connection to sieve theory

First we will give a slightly informal description of how the Poisson imitator problem is relevant to sieve theory. Suppose that we start with an interval $[x, x+y) \subseteq \mathbb{Z}$ of length y, and then for each prime $p < z = y^{1/s}$ we delete all numbers from this interval which lie in any of κ_p "bad" congruence classes modulo p. We wish to know how many elements remain after this process.

Assuming that there is a well-behaved function $\kappa : [0, 1] \to \mathbb{R}_{\ge 0}$ such that

$$\sum_{y^a<p<y^b} \frac{\kappa_p}{p} \approx \int_a^b \frac{\kappa(x)}{x}\, dx,$$

then for any non-overlapping intervals $[a_1, b_1], ..., [a_k, b_k] \subseteq (0, 1/s)$, if we pick a random number $n \in [x, x+y)$, then the expected number of ways to choose a set of m_1 distinct primes $p \in [y^{a_1}, y^{b_1}]$, another set of m_2 distinct primes $p \in [y^{a_2}, y^{b_2}]$, ..., and a set of m_k distinct primes $p \in [y^{a_k}, y^{b_k}]$, such that n is in one of the bad congruence classes for *each* of these primes is

$$\approx \prod_{i=1}^k \frac{1}{m_i!}\Big(\int_{a_i}^{b_i} \frac{\kappa(x)}{x}\, dx\Big)^{m_i} + O\Big(\frac{\prod_i y^{b_i m_i}}{y}\Big)$$

by the Chinese Remainder Theorem.

Thus if we define random variables X_i to be the number of primes $p \in [y^{a_i}, y^{b_i}]$ such that our random $n \in [x, x+y)$ is in a bad congruence class for p, we see that the tuple $(X_1, ..., X_k)$ forms a Poisson imitator with rates $\int_{a_i}^{b_i} \frac{\kappa(x)}{x}\, dx$ and widths b_i.

The general sifting setup is as follows. We assume we have a set A of size y, and for each prime $p < z = y^{1/s}$ we have a "bad" set $A_p \subseteq A$ of size roughly $\frac{\kappa_p}{p} \cdot y$, such that if we extend the notation multiplicatively by

$$A_d = \bigcap_{p|d} A_p$$

and

$$\kappa_d = \prod_{p|d} \kappa_p,$$

then we have

$$|A_d| = \frac{\kappa_d}{d} \cdot y + O(\kappa_d)$$

for every squarefree d with all prime factors bounded by z. Under these assumptions (and no additional assumptions!), we wish to find the optimal upper and lower bounds for

$$\mathcal{S}(A, z) := \Big|A \setminus \Big(\bigcup_{p<z} A_p\Big)\Big|.$$

Theorem 5 (Selberg [2])**.** *In the above setup, if the function $\kappa : [0,1] \to \mathbb{R}_{\geq 0}$ is sufficiently well-behaved, then the asymptotically best (as $y \to \infty$, with s held fixed) upper and lower bounds for $\mathcal{S}(A, z)$ are given by*

$$f_\kappa(s) + o(1) \leq \frac{\mathcal{S}(A, z)}{y \prod_{p<z}(1 - \frac{\kappa_p}{p})} \leq F_\kappa(s) + o(1).$$

Furthermore, there is an effective procedure which takes κ, s, and any $\epsilon > 0$, and returns parameters (ν, α) such that

$$|F_\kappa(s) - F(\nu, \alpha)|,\ |f_\kappa(s) - f(\nu, \alpha)| \ < \ \epsilon.$$

The proof of Selberg's result is based on several clever ideas. We briefly summarize them below.

- The optimal bound for any particular y must be based on a system of sieve weights, by a similar argument to Theorem 3.
- The Fundamental Lemma of sieve theory is sufficiently strong that we can handle the small primes (i.e. less than y^ϵ) in a generic way, without losing much in the asymptotics.
- We can handle the very large primes (i.e. above $z^{1-\epsilon}$) by a union bound, without losing much in the asymptotics.
- We can combine "good enough" sieves for the small and very large primes together with a nearly sharp sieve for the medium primes, and get a nearly sharp sieve overall.
- We can bucket the medium and large primes into intervals $[y^{a_i}, y^{b_i}]$, and show that if we average over the buckets, the sieve weights λ_d must be bounded on average if the sieve bound is any good, by showing that the average value of $\theta(d) = \sum_{k|d} \lambda_k$ must be bounded and using Möbius inversion.
- We can take a subsequence $y \to \infty$ such that the averaged sieve weights have limiting values, by a compactness argument.
- We can take these limiting values of the averaged sieve weights, and apply them instead to primes "one interval down", to avoid issues with the discretization, without losing much in the asymptotics. To make this step work out nicely, Selberg chooses the intervals such that the integral $\int \frac{\kappa(x)}{x}\, dx$ over each interval has the same size.

The interested reader can find the full statement and proof in Selberg's "Lectures on Sieves" [2], or (in the special case where κ is constant) in Chapter 6 of the author's thesis [1].

4. Selberg's upper bound sieve within this framework

Selberg's upper bound sieve takes the form of a polynomial

$$\theta(x) = \Big(\sum_{\alpha \cdot m \leq \frac{1}{2}} \ell_m \binom{x}{m}\Big)^2,$$

where we have adopted the muliplicative notation

$$\binom{x}{m} := \prod_i \binom{x_i}{m_i}.$$

In order to ensure that $\theta(0) = 1$, we must take $\ell_0 = 1$. We may then choose the other ℓ_ms freely, in order to minimize the sum

$$\sum_{x\in\mathbb{N}^k} \theta(x)\frac{\nu^x}{x!} = e^\nu \sum_{\alpha\cdot m\le 1} \lambda_m \frac{\nu^m}{m!},$$

where we have again adopted the multiplicative notations $\nu^x := \prod_i \nu_i^{x_i}$ and $x! = \prod_i x_i!$, and the sieve weights λ_m are given by the formula

$$\lambda_m = \sum_{a,b\le m\le a+b} \frac{m!}{(m-a)!(m-b)!(a+b-m)!}\ell_a\ell_b.$$

Rewriting the sum we wish to minimize as a quadratic form in the ℓs, it becomes

$$e^\nu \sum_{a,b} \ell_a\ell_b \sum_{a,b\le m\le a+b} \frac{\nu^m}{(m-a)!(m-b)!(a+b-m)!}.$$

In an attempt to separate the variables a bit, we set $k = a + b - m$, and rewrite this as

$$e^\nu \sum_k \sum_{a,b\ge k} \ell_a\ell_b \frac{\nu^{a+b-k}}{(b-k)!(a-k)!k!} \;=\; e^\nu \sum_k \frac{\nu^k}{k!}\Big(\sum_{a\ge k} \ell_a \frac{\nu^{a-k}}{(a-k)!}\Big)^2.$$

Now we've diagonalized our quadratic form. Setting

$$\xi_k = (-1)^k \sum_{a\ge k} \ell_a \frac{\nu^{a-k}}{(a-k)!},$$

we have an analogue of the Möbius inversion formula:

$$\begin{aligned}\ell_b &= \sum_{a\ge b} \ell_a \frac{\nu^{a-b}}{(a-b)!} \sum_{a\ge k\ge b} \binom{a-b}{a-k}(-1)^{k-b} \\ &= (-1)^b \sum_{k\ge b} \xi_k \frac{\nu^{k-b}}{(k-b)!}.\end{aligned}$$

Thus we wish to minimize the quadratic form

$$e^\nu \sum_k \xi_k^2 \frac{\nu^k}{k!}$$

subject to the requirement that the ξ_ks are supported on $\alpha\cdot k \le \frac{1}{2}$ and the linear constraint

$$\ell_0 = \sum_k \xi_k \frac{\nu^k}{k!} = 1.$$

The Cauchy-Schwartz inequality gives

$$\Big(\sum_{\alpha\cdot k\le\frac{1}{2}} \xi_k^2 \frac{\nu^k}{k!}\Big)\Big(\sum_{\alpha\cdot k\le\frac{1}{2}} \frac{\nu^k}{k!}\Big) \ge \Big(\sum_{\alpha\cdot k\le\frac{1}{2}} \xi_k \frac{\nu^k}{k!}\Big)^2 = \ell_0^2 = 1,$$

with equality when the ξ_k are all equal.

Theorem 6 (Selberg's upper bound sieve)**.** *If we define ℓ_m by*

$$\ell_m = (-1)^m \frac{\sum_{\alpha\cdot(k+m)\le\frac{1}{2}} \frac{\nu^k}{k!}}{\sum_{\alpha\cdot k\le\frac{1}{2}} \frac{\nu^k}{k!}}$$

and

$$\theta(x) = \Big(\sum_{\alpha\cdot m\le\frac{1}{2}} \ell_m \binom{x}{m}\Big)^2,$$

then

$$F(\nu,\alpha) \le \sum_{x\in\mathbb{N}^k} \theta(x)\frac{\nu^x}{x!} \;=\; \frac{e^\nu}{\sum_{\alpha\cdot k\le\frac{1}{2}} \frac{\nu^k}{k!}}.$$

Taking a limit, if we define $\sigma_\kappa(s)$ by

$$\sigma_\kappa(s) \coloneqq \lim_{\epsilon\to 0} e^{-\int_\epsilon^{\frac{1}{s}} \frac{\kappa(x)}{x}dx} \sum_k \frac{1}{k!}\int_\epsilon^{\frac{1}{s}} \cdots \int_\epsilon^{\frac{1}{s}} 1_{\{\sum_i x_i\le\frac{1}{2}\}} \prod_{i\le k}\frac{\kappa(x_i)}{x_i}\; dx_1\cdots dx_k,$$

then

$$F_\kappa(s) \le \frac{1}{\sigma_\kappa(s)}.$$

The sums which occur in Selberg's sieve have a probabilistic interpretation: if the X_i are independent Poissons with parameters ν_i, then

$$\mathbb{P}[\alpha\cdot X\le\tfrac{1}{2}] \;=\; e^{-\nu}\sum_{\alpha\cdot k\le\frac{1}{2}}\frac{\nu^k}{k!}.$$

If we only want a quick-and-dirty estimate in the region where $\alpha\cdot\nu$ and $\frac{1}{s} \coloneqq \max\alpha_i$ are small, we can use Markov's inequality to get

$$\begin{aligned}\mathbb{P}[\alpha\cdot X\le\tfrac{1}{2}] \;&\ge\; 1-e^{-t/2}\mathbb{E}[e^{t\alpha\cdot X}]\\ &=\; 1-e^{-t/2}\prod_i e^{\nu_i(e^{t\alpha_i}-1)}\end{aligned}$$

for any $t\ge 0$. Since $s\alpha_i\le 1$, we can simplify this with the bound

$$e^{t\alpha_i}-1 \le s\alpha_i(e^{t/s}-1)$$

to get

$$\mathbb{P}[\alpha\cdot X\le\tfrac{1}{2}] \;\ge\; 1-e^{-t/2+s(\alpha\cdot\nu)(e^{t/s}-1)}.$$

The best choice for t is $s\ln(1/(2\alpha\cdot\nu))$, which will be ≥ 0 as long as $\alpha\cdot\nu\le\frac{1}{2}$.

Corollary 3 (Fundamental Lemma for upper bound sieves)**.** *For any ν,α with $\alpha\cdot\nu\le\frac{1}{2}$ and $\max\alpha_i\le\frac{1}{s}$, we have*

$$F(\nu,\alpha) \le \frac{1}{1-e^{-\frac{s}{2}\left(\ln(1/(2\alpha\cdot\nu))+2(\alpha\cdot\nu)-1\right)}}.$$

Taking a limit, we see that for any κ, s with $\int_0^{\frac{1}{s}}\kappa(x)\;dx\le\frac{1}{2}$ we have

$$F_\kappa(s) \le \frac{1}{1-e^{-\frac{s}{2}\left(\ln(1/(2\int_0^{1/s}\kappa(x)\;dx))+2\int_0^{1/s}\kappa(x)\;dx-1\right)}}.$$

References

[1] Zarathustra Elessar Brady. *Sieves and iteration rules*. PhD thesis, Stanford University, 2017.

[2] Atle Selberg. *Collected papers. Vol. II*. Springer-Verlag, Berlin, 1991. With a foreword by K. Chandrasekharan.

Email address: `notzeb@gmail.com`