Math 259: Introduction to Analytic Number Theory

The Selberg (quadratic) sieve and some applications

An elementary and indeed naïve approach to the distribution of primes is the following argument: an integer n is prime if and only if it is not divisible by the primes $\leq \sqrt{n}$; but half the integers are odd, 2/3 are not multiples of 3, 4/5 not multiples of 5, etc., and divisibility by any prime is independent of divisibility by finitely many other primes, so... Moreover, if n is restricted to an arithmetic progression $a \bmod q$ with $(a, q) = 1$ then the same factors $(p-1)/p$ arise except those for which $l|q$, from which we recover the factor $\prod_{p|q} p/(p-1) = q/\phi(q)$ in the asymptotic formula for $\pi(qx, a \bmod q)$.

The problem with estimating $\pi(x)$ etc. this way is that the divisibilities aren't quite independent. This is already implicit in our trial-division test for primality: if n is known to contain no primes $\leq \sqrt{n}$, the conditional probability that it be a multiple of some other prime $p \in (\sqrt{n}, n)$ is not $1/p$ but zero. Already for small p, the number of $n < x$ divisible by p is not quite x/p but $x/p + O(1)$, and similarly for n divisible by a product of distinct primes; so if we try to use the principle of inclusion and exclusion to recover the number of primes $n < x$, or even of n not divisible by r primes $p_1, \ldots, p_r$, we get an estimate of $x \prod_{i=1}^{r} (1 - \frac{1}{p_i})$ as expected, but with an "error term" $O(2^r)$ that swamps the estimate long before r can get usefully large.

This quandary is prototypical of "sieve" situations, in which we have a set S of A integers such that $\#(S \cap D\mathbf{Z})/A$ is approximated by a multiplicative function $\alpha(D)$ of the squarefree integer D, and are interested in the number $A(\prod_{p \in P} p)$ of $n \in S$ not divisible by any of the primes p in a given set P. (For instance, if S is an interval then $\alpha(D) = 1/D$; in general, α must be multiplicative for the divisibility of a random $n \in S$ by a prime p to be approximately independent of its divisibility by any other primes.) Several methods are now known for deriving "sieve inequalities", which are nontrivial upper bounds on $A(\prod_{p \in P} p)$. These inequalities use a variety of methods, but curiously the resulting bounds are similar in many important contexts, and often exceed the expected number by a factor asymptotic to 2. We shall develop one of the most general such inequalities, due to Selberg, and give some typical examples of its use in analytic number theory. While we state Selberg's sieve in the context of divisibility, in fact all that we are using is that each prime p sifts out a subset of S and that the events that a random $n \in S$ survives these tests for different p are approximately independent. Thus Selberg's sieve has a counterpart in the context of probability theory, for which see the final Exercise. Selberg's and many other sieves are collected in [Selberg 1969]; nice applications of sieve inequalities to other kinds of problems in number theory are interspersed throughout [Serre 1992].

Assume, then, that a_n ($n \in \mathbf{Z}$) are nonnegative real numbers with $\sum_{n \in \mathbf{Z}} a_n = A < \infty$, and that α is a multiplicative function satisfying $0 \leq \alpha(d) \leq 1$ for

each d (equivalently, for each prime d). For each squarefree $d > 0$ let

$$A_d := \sum_{m \in \mathbf{Z}} a_{md} = A\,\alpha(d) + r(d);$$

in any application, the $r(d)$ must be small compared to A. Let P be a finite set of primes, and D the squarefree integer $\prod_{p \in P} p$. We are interested in

$$A(D) := \sum_{(n,D)=1} a_n,$$

which is the number of $n \in A$ not divisible by any of the primes in P. We hope that $A(D)$ is approximately $A \prod_{p|D}(1 - \alpha(p))$, with an error that is usefully small if the $r(d)$ are. What we can show is:

Theorem (Selberg): *For each $z \geq 1$ we have*

$$A(D) \leq \frac{A}{S(D,z)} + R(D,z), \tag{1}$$

where S, R are defined by

$$S(D,z) := \sum_{\substack{d|D \\ d \leq z}} \prod_{p|d} \frac{\alpha(p)}{1-\alpha(p)}, \qquad R(D,z) := \sum_{\substack{d|D \\ d \leq z^2}} 3^{\omega(d)} |r(d)|$$

and $\omega(d) := \sum_{p|d} 1$, the number of distinct prime factors of d.

Remark: Given D and α, the series $S(D,z)$ and $R(D,z)$ are increasing functions of z because they accumulate more positive terms as z grows. For $z = 1$ we have $S(D,z) = 1$ and $R(D,z)$=0, so (1) is the trivial inequality $A(D) \leq A(1) = A$. For $z \geq D$ we have

$$S(D,z) = S(D,D) = \sum_{d|D} \prod_{p|d} \frac{\alpha(p)}{1-\alpha(p)} = \prod_{p|D} \left(1 + \frac{\alpha(p)}{1-\alpha(p)}\right) = \prod_{p|D} \frac{1}{1-\alpha(p)},$$

so $1/S(D,z)$ is the expected factor $\prod_{p|D}(1-\alpha(p))$, and (1) is implied by the inclusion-exclusion estimate $\big|A(D) - A\prod_{d|D}(1-\alpha(p))\big| \leq \sum_{d|D} |r(d)|$. Thus Selberg's inequality may be regarded as an interpolation between inclusion-exclusion and the trivial $A(D) \leq A$. Note that (1) is only an upper bound: we do *not* claim that $|A(D) - A/S(D,z)| \ll R(D,z)$.

Proof: Let λ_d $(d|D)$ be arbitrary real parameters with $\lambda_1 = 1$ (and eventually $\lambda_d = 0$ once $d > z$). Then

$$A(D) \leq \sum_n a_n \Big(\sum_{d|(n,D)} \lambda_d \Big)^2 = \sum_{d_1,d_2|D} \sum \lambda_{d_1} \lambda_{d_2} \sum_{[d_1,d_2]|n} a_n,$$

where $[d_1,d_2] := \mathrm{lcm}(d_1,d_2)$. The inner sum is just $A_{[d_1,d_2]}$, so we have

$$A(D) \leq \sum_{d_1,d_2|D} \sum \lambda_{d_1}\lambda_{d_2} \big(A\alpha([d_1,d_2]) + r([d_1,d_2]) \big) \leq AQ + R,$$

where Q is the quadratic form

$$Q := \sum_{d_1,d_2|D}\sum \alpha([d_1,d_2])\,\lambda_{d_1}\lambda_{d_2}$$

in the λ_d, and

$$R := \sum_{d_1,d_2|D}\sum \big|\lambda_{d_1}\lambda_{d_2} r([d_1,d_2])\big|.$$

Now for $d|D$ the number of pairs d_1, d_2 such that $d = [d_1, d_2]$ is $3^{\omega(d)}$ (why?); thus (1) will follow from the following

Lemma: *The minimum of the quadratic form Q subject to the conditions $\lambda_1 = 1$ and $d > z \Rightarrow \lambda_d = 0$ is $1/S(D,z)$, and is attained by λ_d with $|\lambda_d| \le 1$.*

Proof of Lemma: By continuity we may assume that $0 < \alpha(p) < 1$ for all $p \in P$. (In fact, for our purpose we can exclude from the start the possibilities $\alpha(p) = 0$ or 1 — do you see why?) Since $[d_1,d_2]\gcd(d_1,d_2) = d_1 d_2$ and α is multiplicative, we have

$$Q = \sum_{d_1,d_2|D}\sum \frac{\alpha(d_1)\lambda_{d_1}\cdot\alpha(d_2)\lambda_{d_2}}{\alpha(\gcd(d_1,d_2))}.$$

Selberg's key insight is that this quadratic form is diagonalized by introducing coefficients $\delta(e)$ for $e|D$, determined by

$$\frac{1}{\alpha(d)} = \sum_{e|d}\delta(e).$$

Then

$$Q = \sum_{e|D}\delta(e)\Big[\sum_{e|d}\alpha(d)\lambda_d\Big]^2.$$

Let $x(e)$, then, be defined by

$$x(e) := \sum_{e|d}\lambda_d\alpha(d).$$

By Möbius inversion we find

$$\delta(e) = \prod_{p|e}\frac{1-\alpha(p)}{\alpha(p)}, \qquad \lambda_d = \frac{1}{\alpha(d)}\sum_{d|e}\mu(e/d)x(e).$$

Our conditions on the λ_d then become

$$\sum_{e|D}\mu(e)x(e) = \alpha(1)\lambda_1 = 1, \qquad e > z \Rightarrow x(e) = 0.$$

By the Schwarz inequality, the minimum of Q subject to these conditions is

$$\Big[\sum_{e|D,\,e\le z}\frac{1}{\delta(e)}\Big]^{-1} = 1/S(D,z),$$

and is attained at $x(e) = \mu(e)/(\delta(e)S(D,z))$. This yields

$$S(D,z)\lambda_d = \frac{\mu(d)}{\alpha(d)} \sum_{d|e\le z} \frac{1}{\delta(e)} = \frac{\mu(d)}{\alpha(d)\delta(d)} \sum_{\substack{f|(D/d)\\ f\le z/d}} \frac{1}{\delta(f)}.$$

But we have

$$\frac{1}{\alpha(d)\delta(d)} = \sum_{e|d} \frac{1}{\delta(e)} :$$

both sides of the equation are multiplicative functions of the squarefree integer d (since α, δ are both multiplicative), and the equation holds for prime d by our above formula for $\delta(e)$. Thus we have

$$S(D,z)\lambda_d = \mu(d) \sum\sum_{e,f} \frac{1}{\delta(ef)},$$

with each $ef \le z$ and no ef values repeated. Thus the sum has absolute value at most $S(D,z)$, so $|\lambda_d| \le 1$ as claimed. This concludes the proof of the Lemma, and thus also of Selberg's inequality (1). □□

Typically we will let $D = D(y) = \prod_{p\le y} p$. For instance, we show:[1]

Corollary. *Fix q. For all a, x_0, A such that $\gcd(a,q) = 1$, we have*

$$\pi(x_0 + Aq, a \bmod q) - \pi(x_0, a \bmod q) < \Big(\frac{2q}{\phi(q)} + O(\frac{\log\log A}{\log A})\Big)\frac{A}{\log A}. \qquad (2)$$

Proof: Let a_n be the characteristic function of the arithmetic progression

$$\{n | n \equiv a \bmod q,\ 0 < n - x_0 < Aq\}.$$

Then $A(D(y))$ is an upper bound on $\pi(x_0+Aq, a \bmod q) - \pi(x_0, a \bmod q) - \pi(y)$. We take $\alpha(n) = 1/n$ if $\gcd(n,q) = 1$ and $\alpha(n) = 0$ otherwise. Then $|r(d)| \le 1$ for each d, and so $R(D,z)$ is bounded by the sum of the n^{-s} coefficients of $\zeta^3(s)$ for $n \le z^2$, so is $\ll (z\log z)^2$. [An equivalent and more elementary way to handle $\sum_{n\le x} 3^{\omega(n)}$ is to note that $3^{\omega(n)}$ is at most the number of representations $n = n_1n_2n_3$ of n as a product of three positive integers.] As to $S(D,z)$, we take $z = y$ and expand $\alpha/(1-\alpha)$ in a geometric series to find

$$S(D,z) > \sum_{\substack{n\le z\\ (n,q)=1}} \frac{1}{n} = \frac{\phi(q)}{q}\log z + O(1). \qquad (3)$$

Thus Selberg's bound (1) is $(q/\phi(q))A/\log z + O(z^2\log^2 z)$. We choose $y = z = A^{1/2}/\log^2 A$, and deduce the upper bound (2), absorbing the correction $\pi(y)$ into the error term since $\pi(y) < y < A^{1/2}$. □

[1] This result in fact predates Selberg, because this choice of a_n is regular enough to be treated with earlier sieve inequalities.

In particular, we may to obtain an elementary upper bound on $\pi(Aq, a \bmod q)$ by taking $x_0 = 0$. The implied O-constant in (2) depends on q, but tractably and effectively so, without invoking zeros of L-functions and the like. The only issue is the dependence on q of the $O(1)$ error in (3). We may write

$$\sum_{\substack{n \le z \\ (n,q)=1}} \frac{1}{n} = \sum_{d|q} \mu(q/d) \sum_{n=1}^{\lfloor z/d \rfloor} \frac{1}{dn} = \sum_{d|q} \frac{\mu(q/d)}{d} (\log z + O(1 + \log d)).$$

Thus the error in (3) is bounded by

$$\sum_{d|q} |\mu(q/d)| \frac{1 + \log d}{d}.$$

For instance, we readily deduce that for all ϵ there exists an effective $q_0(\epsilon)$ such that if $q > q_0(\epsilon)$ then

$$\pi(x_0 + Aq, a \bmod q) - \pi(x_0, a \bmod q) < (2 + \epsilon) \frac{q}{\phi(q)} \frac{A}{\log A}$$

for all A, x_0, a with $A > q$. If the coefficient 2 were any smaller, this upper bound would be enough to banish the Siegel-Landau zero!

Exercises

1. What are the λ_d if $z \ge D$? Explain.

2. Complete the two proofs outlined above that $\sum_{n \le x} 3^{\omega(n)} \ll x(\log x)^2$ (one by comparison with the coefficients of ζ^3, the other by counting solutions of $n_1 n_2 n_3 \le x$). Can you prove that in fact $\sum_{n \le x} 3^{\omega(n)} \sim Cx(\log x)^2$ for some constant $C > 0$, and numerically compute C?

3. Prove that for each integer $n > 0$ the number of primes $p < x$ such that $p + 2n$ is also prime is $O_n(x/\log^2 x)$. In particular, conclude that the sum

$$\frac{1}{3} + \frac{1}{5} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{29} + \frac{1}{31} + \frac{1}{41} + \frac{1}{43} + \cdots$$

of the reciprocals of twin primes converges. (This result was first obtained by Brun [1919] using his less powerful sieve inequality. The sum may be considered "convergent" if it is finite, that is, if the twin-prime conjecture is false.)

4. Prove that for all $\epsilon > 0$ there exists an effective constant $x_0(\epsilon)$ such that, for each $x > x_0(\epsilon)$, there are at most $((8/\pi) + \epsilon)x/\log x$ integers $n < x$ such that $n^2 + 1$ is a prime. Generalize.

It is of course a famous open problem to find a similar *lower* bound on the number of such n, or even to prove that it is unbounded as $x \to \infty$ — that is, to prove that there are infinitely many primes of the form $n^2 + 1$. More generally, one conjectures that for every irreducible polynomial $P \in \mathbf{Z}[X]$ there exist infinitely many integers n such that $P(n)$ is prime, provided that for each prime p there exists at least one $n \in \mathbf{Z}$ such that $P(n) \not\equiv 0 \bmod p$. This, in turn, is the special case $k = 1$ of "Hypothesis H" of Schinzel

and Sierpiński, which asserts that for irreducible polynomials $P_1, \ldots, P_k \in \mathbf{Z}[X]$ there are infinitely many $n \in \mathbf{Z}$ such that each $P_i(n)$ is prime, provided that for each prime p there exists at least one $n \in \mathbf{Z}$ such that $P_i(n) \not\equiv 0 \bmod p$ for each i. Hypothesis H is also a generalization of a conjecture of Dickson on the simultaneous primality of $a_i n + b_i$, which itself generalizes the twin prime conjecture. In each case one expects that in fact the number of $n < x$ such that each $P_i(n)$ is prime is asymptotic to $cx/(\log x)^k$ for some constant c given by an infinite product over p depending on the polynomials P_i (assuming that $P_i \neq \pm P_j$ for distinct i, j; this is the Bateman-Horn conjecture). For more information on these various conjectures, see Chapter 6 of [Ribenboim 1996], particularly pages 372, 391, and 409. The only case of any of these conjectures that has been proved is the case of a single linear polynomial, which is Dirichlet's theorem. Sieve methods, including Selberg's, yield an upper bound with the same asymptotic behavior but a larger c.

As usual, one can formulate analogous problems over polynomial rings such as $\mathbf{F}_q[T]$ in place of $\mathbf{Z}$. For instance, fix $P \in \mathbf{F}_q[T, X]$, and ask whether there exist infinitely many polynomials $n(T)$ such that $P(T, n(T))$ is irreducible. A necessary condition is that P be irreducible as a polynomial of two variables and that for each nonconstant $p \in \mathbf{F}_q[T]$ there exist $n \in \mathbf{F}_q[T]$ such that $P(T, n(T))$ is not a multiple of p. One might be tempted to conjecture that again this necessary condition is also sufficient; but here this conjecture is false! Explicit families of counterexamples are given in [CCG 2003]. In any event, one can use the same sieve inequalities to give an upper bound $O(q^d/d)$ on the number of $n(T)$ of degree at most d for which $P(T, n(T))$ is irreducible.

5. Let p_i ($i \in [m] := \{1, 2, \ldots, m\}$) be probabilities, i.e., real numbers in $[0, 1]$; and let $E_1, \ldots, E_m$ be events approximating independent events with those probabilities, i.e., such that for each $I \subseteq [m]$ the probability that E_i occurs for all $i \in I$ is $\prod_{i \in I} p_i + r(I)$. Obtain upper bounds on the probability that *none* of the E_i occurs, bounds which correspond to and/or generalize Selberg's (1). (See for instance [Chow 1998], where an even further generalization is proposed.)

References

[Brun 1919] Brun, V.: La série $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \ldots$, où les dénominateurs sont nombres premiers jumeaux est convergente ou finie, *Bull. des Sci. Math.* **43** (1919), 100–104 and 124–128.

[Chow 1998] Chow, T.: The Combinatorics behind Number-Theoretic Sieves, *Adv. in Math.* **138** (1998), 293–305.

[CCG 2003] Conrad, B., Conrad, K., Gross, R.: Hardy-Littlewood Conjecture for function fields. Preprint, 2003.

[Ribenboim 1996] Ribenboim, P.: *The New Book of Prime Number Records*, New York: Springer 1996.

[Selberg 1969] Selberg, A.: Lectures on Sieves, pages 66–247 of his *Collected Papers II* [O 9.89.2 (II)]

[Serre 1992] Serre, J.-P.: *Topics in Galois Theory.* Boston: Jones and Bartlett 1992. [BB 9.92.12 / QA214.S47]