1 Introduction

A sieve is a technique for bounding the size of a set after the elements with “undesirable properties” (usually of a number theoretic nature) have been removed. The undesirable properties could be divisibility by a prime from a given set, other multiplicative constraints (divisibility by a perfect square for example) or inclusion in a set of residue classes. The methods usually involve some kind of combinatorial reasoning where one “sees” the removal process going on in some form or another.

Sieve methods have also been used in the literature to describe procedures for attacking problems about detecting primes in sequences of integers, whether or not there is any kind of “removal process” going on or not.

The original sieve is, of course, the Sieve of Eratosthenes, the familiar process of creating a table of prime numbers by systematically removing those integers divisible by small primes (but keeping the primes themselves). The modern sieve was created by Viggo Brun in the period 1915-1922 as a way of attacking famous unsolved problems such as Golbach’s Conjecture and the Twin Prime problem (both, so far, unsuccessfully).

Sieve methods have since found enormous application in number theory, in particular to such problems as

- studying \( k \)-tuples of primes in special configurations (generalizations of twin primes);
- studying primes in short intervals;
- studying primes in arithmetic progressions;
- analyzing the multiplicative structure of integers (distribution of prime factors and divisors);
- analyzing the structure of shifted primes \( p - 1, p + 1 \) with application to the distribution of arithmetic functions such as \( \phi(n) \) and \( \sigma(n) \);
- studying the distribution of sizes of circles in Apollonian circle packings.

Estimates coming from sieve methods are often used as tools in many other types of problems, e.g. in studying Diophantine equations,

1.1 Notational conventions

\( \tau(n) \) is the number of positive divisors of \( n \)
\( \omega(n) \) is the number of distinct prime factors of \( n \)
\( \Omega(n) \) is the number of prime factors of \( n \) counted with multiplicity
\( \mu(n) \) is the Möbius’s function; \( \mu(n) = (-1)^{\omega(n)} \) if \( n \) is squarefree and \( \mu(n) = 0 \) otherwise.
\( P^+(n) \) is the largest prime factor of \( n \); \( P^+(1) = 0 \) by convention
$P^-(n)$ is the smallest prime factor of $n$; $P^-(1) = \infty$ by convention

$\Lambda(n)$ denotes the von Mangoldt function

### 1.2 Sieve notation

- $\mathcal{A} = (a_n)_{n \in \mathbb{Z}}$, a sequence of non-negative real numbers, often identified with the indicator function of a finite set $\mathcal{A}$ of integers;
- $\mathcal{P}$ is a set of primes;
- $P(z) = \prod_{p \leq z} p$;
- $S(\mathcal{A}, \mathcal{P}, z) = \sum_{(n, P(z)) = 1} a_n$, the sequence $\mathcal{A}$ sifted by the primes $\mathcal{P}$ up to $z$.

For example, if $a_n = 1_{\mathcal{A}}(n)$, the indicator function of a finite set $\mathcal{A}$, then $S(\mathcal{A}, \mathcal{P}, z)$ is the number of elements of $\mathcal{A}$ which are divisible by no prime $p \in \mathcal{P}$, $p \leq z$. Some specific examples:

(a) (primes) $a_n = 1$ for $1 \leq n \leq x$; $\mathcal{P}$ is the set of all primes; $z = \sqrt{x}$. Then $S(\mathcal{A}, \mathcal{P}, z) = \pi(x) - \pi(\sqrt{x}) + 1$, as the unsifted elements are the primes in $(\sqrt{x}, x]$ together with the number 1.

(b) (twin primes) $a_n = 1$ if $n = k(k + 2)$ for some positive integer $k \leq x$; $\mathcal{P}$ is the set of all primes; $z = \sqrt{x} + 2$. The unsifted numbers correspond to values of $k$ such that $n$ has no prime factor $\leq \sqrt{x} + 2$. That is, $S(\mathcal{A}, \mathcal{P}, z)$ counts $k \in (\sqrt{x} + 2, x]$ for which both $k$ and $k + 2$ are prime.

(c) (twin primes, 2nd version) $a_n = \Lambda(n - 2)$ for $3 \leq n \leq x$; $\mathcal{P}$ is the set of all primes; Then

$$S(\mathcal{A}, \mathcal{P}, \sqrt{x}) = \sum_{\sqrt{x} < p \leq x} \Lambda(p - 2),$$

a weighted sum over twin primes (plus a negligible sum over solutions of $p - 2 = q^b$ with $p, q$ prime and $b \geq 2$).

(d) (Goldbach) Let $N$ be an even, positive integer, put $a_n = \#\{k \in \mathbb{N} : n = k(N - k)\}$; $\mathcal{P}$ is the set of all primes. Then $S(\mathcal{A}, \mathcal{P}, \sqrt{N})$ counts numbers $k \in (\sqrt{N}, N]$ for which both $k$ and $N - k$ are prime. In particular, $S(\mathcal{A}, \mathcal{P}, \sqrt{N}) > 0$ implies that $N$ is the sum of two primes. If one could show this for all $N \geq 4$, one deduces Goldbach’s Conjecture.

(e) (primes in an arithmetic progression) Fix coprime positive integers $a$ and $q$, let $\mathcal{A} = \{1 \leq n \leq x : n \equiv a \pmod{q}\}$, $\mathcal{P}$ is the set of all primes. Then $S(\mathcal{A}, \mathcal{P}, \sqrt{x})$ counts primes in $(\sqrt{x}, x]$ that are in the arithmetic progression $a \mod q$.

(f) (sums of two squares) $a_n = 1$ for $1 \leq n \leq x$; $\mathcal{P}$ is the set of primes that are $\equiv 3 \pmod{4}$; $z = x$. Then $S(\mathcal{A}, \mathcal{P}, z)$ is the number of integers $n \leq x$, all of whose odd prime factors are $\equiv 1 \pmod{4}$. In particular, such numbers are the sum of two squares.

Later we will compare the merits of the two different approaches (b) and (c) to the twin prime problem.

The Möbius inversion formula

$$\sum_{d \mid n} \mu(d) = \begin{cases} 1 & n = 1 \\ 0 & n > 1 \end{cases}$$
implies the basic sieve identity

\[ S(\mathcal{A}, \mathcal{P}, z) = \sum_n a_n \sum_{d \mid (n, P(z))} \mu(d) = \sum_{d \mid P(z)} \mu(d) |\mathcal{A}_d|, \]

where

\[ |\mathcal{A}_d| := \sum_{d \mid n} a_n. \]

When \( a_n \) is the indicator function of a set \( \mathcal{A} \), \( |\mathcal{A}_d| \) is just the number of elements of \( \mathcal{A} \) which are divisible by \( d \). Identity (SI) makes it clear that good information on the behavior of the sequence \( (a_n) \) in progressions \( 0 \mod d \) is critical for bounding the sifting function.

**Example (a).** (Legendre, 1808) \( a_n = 1 \) for \( 1 \leq n \leq x \), \( \mathcal{P} \) is the set of all primes. \( S(\mathcal{A}, \mathcal{P}, z) \) counts at least the primes between \( z \) and \( x \) (together with possibly other numbers). Also, \( |\mathcal{A}_d| = \left\lfloor \frac{x}{d} \right\rfloor = \frac{x}{d} + O(1) \).

By (SI),

\[ \pi(x) - \pi(z) \leq S(\mathcal{A}, \mathcal{P}, z) = \sum_{d \mid P(z)} \mu(d) \left( \frac{x}{d} + O(1) \right) \]
\[ = x \sum_{d \mid P(z)} \frac{\mu(d)}{d} + O(\tau(P(z))) \]
\[ = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) + O \left( 2^{\pi(z)} \right). \]

Taking \( z = \log x \), and using the crude bound \( \pi(z) \leq z \) together with Mertens bound, we conclude that

\[ \pi(x) \leq z + \frac{x}{\log z} \left( e^{-\gamma} + o(1) \right) + O(2^{\log x}) = O \left( \frac{x}{\log \log x} \right). \]

**Remark:** We can do better by observing that in fact \( |\mathcal{A}_d| = 0 \) for \( d > x \) (cf. [4]), and thus restricting the sums to such \( d \).

### 1.3 General sieves

More generally, one can set up a sieve problem with a sequence \( \mathcal{A} \), a set \( \mathcal{P} \), and an arbitrary collection of properties \( B_p \), where \( B_p(n) \) is a logical variable (encoded as zero or one) for each natural number \( n \). Extending the definition to composite, squarefree \( d \) by

\[ B_d(n) = \prod_{p \mid d} B_p(n), \]

and defining

\[ S(\mathcal{A}, \mathcal{P}, z) = \sum_{n: B_p(n) = 0 \forall p \leq z} a_n, \quad |\mathcal{A}_d| := \sum_{n: B_d(n) = 1} a_n, \]

we have by inclusion-exclusion,

\[ S(\mathcal{A}, \mathcal{P}, z) = \sum_{d \mid P(z)} \mu(d) |\mathcal{A}_d|. \]
as well, for the quantities defined in (1.1). The traditional small sieve is given by the example $B_p(n) = 1$ if $p | n$ and 0 otherwise. One can think of removing from $\mathcal{A}$ elements in the residue classes $0 \mod p$ for primes $p \in \mathcal{P}, p \leq z$. A more general sieve problem is to remove from $\mathcal{A}$ the elements from a set $j(p)$ of residue classes modulo $p$ for $p \in \mathcal{P}, p \leq z$. Here $B_p(n) = 1$ if $n$ avoids the residue classes $j(p)$ modulo $p$ for every $p$. This is a problem of “large sieve” type, which we will encounter later in the course.

The twin prime problem may be recast in yet a third form (other than Examples (b) and (c) above) using this terminology: Let $\mathcal{A} = \{n : 1 \leq n \leq x\}$, $\mathcal{P}$ the set of all primes, $j(2) = \{0\}$ and $j(p) = \{0, 2\}$ for $p > 2$. The integers $n \leq x$ avoiding the residue classes $j(p)$ for $p \leq \sqrt{x}$ are odd numbers such that both $n$ and $n - 2$ are prime.

1.4 The small sieve

Legendre’s example illustrates the biggest shortcoming of (SI) for applications, namely the proliferation of summands, even for modest values of $z$. Viggo Brun’s fundamental idea was to replace the huge sum with a sum over a much smaller set of numbers $d$, at the expense of replacing the equality in (SI) with an inequality. Consider a finite sequence $\lambda = (\lambda_d)$ supported on positive, squarefree integers $d \leq D$, say, which replaces $\mu(d)$ in (SI). We will suppose that either $\lambda = \lambda^+ = (\lambda^+_d)_{1 \leq d \leq D}$ satisfies

$$\lambda^+_1 = 1, \sum_{d | m} \lambda^+_d \geq 0 \quad (m > 1)$$

or that $\lambda = \lambda^- = (\lambda^-_d)_{1 \leq d \leq D}$ if

$$\lambda^-_1 = 1, \sum_{d | m} \lambda^-_d \leq 0 \quad (m > 1).$$

We call $\lambda^+$ an upper bound sieve and $\lambda^-$ a lower bound sieve, since $(\lambda^+)$ and $(\lambda^-)$ imply

$$(S) \quad \sum_{d \leq D} \lambda^+_d |\mathcal{A}_d| \leq S(\mathcal{A}, \mathcal{P}, z) \leq \sum_{d \leq D} \lambda^-_d |\mathcal{A}_d|.$$

A major goal of sieve methods is to construct good sieve parameters $\lambda^\pm$, with $D$ as small as possible and with the sums in (S) mimicking $S(\mathcal{A}, \mathcal{P}, z)$ as closely as possible.

Basic assumptions about $\mathcal{A}$.

We assume that $|\mathcal{A}| := |\mathcal{A}_1|$ is approximated by a quantity $X$, and that there is a multiplicative function $g(n)$ satisfying

$$(g) \quad 0 \leq g(p) < 1 \quad (p \in \mathcal{P}), \quad g(p) = 0 \quad (p \notin \mathcal{P}),$$

so that $|\mathcal{A}_d| \approx Xg(d)$ for $d | P(z), d \leq D$; that is, we require that the “remainders”

$$(r) \quad r_d := |\mathcal{A}_d| - g(d)X, \quad d \leq D,$$

be “small on average”. We have in fact imposed no conditions whatsoever on $\mathcal{A}$ at this point, and that the choice of $X$ and $g$ appears arbitrary. However, for any given $\mathcal{A}$ there is usually a “canonical” best choice for $X$ and $g$; e.g. if $a_n = 1$ for $1 \leq n \leq x$, then $|\mathcal{A}_d| = [x/d]$ and one takes $X = x$ and $g(d) = 1/d$ so that $|r_d| = O(1)$.  

With this notation and \((r)\), we have
\[
\sum_{d \leq D} \lambda_d |\mathcal{A}_d| = X \sum_{d \leq D} g(d) \lambda_d + \sum_{d \leq D} \lambda_d r_d.
\]
As \(\lambda_d\) is a replacement for \(\mu(d)\), it is reasonable to suppose (if we have constructed our sieve well) that
\[
\sum_{d \leq D} g(d) \lambda_d \approx \sum_{d \in \mathcal{P}(z)} g(d) \mu(d) = \prod_{p \leq z} (1 - g(p)) =: V(z).
\]
That is, it is reasonable to guess that \(S(\mathcal{A}, \mathcal{P}, z) \sim XV(z)\). Let’s explore this in some detail, assuming that \(a_n\) is the indicator function of a finite set \(\mathcal{A}\). Then \(|\mathcal{A}_n|\) is the number of elements of \(\mathcal{A}\) which are divisible by a given prime \(p\), and \(g(p) \approx |\mathcal{A}_n|/|\mathcal{A}_1|\) can be thought of as the probability that a randomly selected element of \(\mathcal{A}\) is divisible by \(p\). For small \(p_1 \neq p_2\) the events \(\{p_1|n\}\) and \(\{p_2|n\}\) are close independent, and thus that probability that a randomly selected element of \(\mathcal{A}\) is divisible by none of the primes \(p \in \mathcal{P}, p \leq z\) ought to be about \(V(z)\).

The big caveat is that the events \(\{p_1|n\}\) and \(\{p_2|n\}\) are in fact highly dependent if \(p_1, p_2\) are large, e.g. if \(\mathcal{A} \subset [1, x]\) and \(p_1, p_2 > \sqrt{x}\) then in fact it is impossible for both \(p_1|n\) and \(p_2|n\). Thus, the above argument breaks down for large primes and we cannot expect that \(S(\mathcal{A}, \mathcal{P}, z) \sim XV(z)\).

**Example (a).** \(a_n = 1\) for \(1 \leq n \leq x\), \(\mathcal{P}\) is the set of all primes. As before, we see that \(S(\mathcal{A}, \mathcal{P}, \sqrt{x}) = \pi(x) - \pi(\sqrt{x}) + 1 \sim x/\log x\) by the Prime Number Theorem. However, Mertens’ theorem gives

\[
XV(\sqrt{x}) = x \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right) \sim 2e^{-\gamma} \frac{x}{\log x},
\]

with \(2e^{-\gamma} = 1.122\ldots\). The discrepancy between \(S(\mathcal{A}, \mathcal{P}, \sqrt{x})\) and \(XV(\sqrt{x})\) comes from the dependence among the events \(\{p|n\}\) for large primes \(p\).

### 1.5 Basic Goals

Broadly speaking, for sequences \(\mathcal{A}\) with structure (for some \(X\) and \(g\), the remainders \(r_d\) given by \(r\) are small on average), we will be able to prove the following:

- **(A)** (asymptotic) \(S(\mathcal{A}, \mathcal{P}, z) \sim XV(z)\) \((z \to \infty, z = X^{o(1)})\);
- **(U)** (upper bound) \(S(\mathcal{A}, \mathcal{P}, z) \ll XV(z)\) \((z \leq X)\);
- **(L)** (lower bound) \(S(\mathcal{A}, \mathcal{P}, z) \gg XV(z)\) \((z \leq X^{c})\),

where the constant \(c\) in (L) depends on the “dimension” of \(\mathcal{A}\) and also on the “level of distribution” of the remainders \(r_d\) (concepts which will be defined rigorously later). In plain language, we can get the expected asymptotic formula for small \(z\), an upper bound of the expected order for all \(z\), and a lower bound of the expected order for \(z\) being some (small) power of \(X\).

We have seen in Example (a) that the asymptotic (A) can fail if \(z = \sqrt{x}\). In fact, using the same sequence with \(z = x^c\) for fixed \(c > 0\) one has \(S(\mathcal{A}, \mathcal{P}, z) \sim w(c) XV(z)\) with \(w(c) \neq 1\) for almost all \(c \in (0, 1]\). Thus, in general the condition \(z = X^{o(1)}\) is necessary in order to conclude (A). The upper bound (U) is amazing in its generality, and it has enormous utility.
Example (b). \( \mathcal{A} = \{k(k+2) : 1 \leq k \leq x\} \), \( X = x \), \( \mathcal{P} \) is the set of all primes. \( S(\mathcal{A}, \mathcal{P}, \sqrt{x}+2) \) counts the number of twin prime pairs between \( \sqrt{x}+2 \) and \( x \). Hardy and Littlewood conjectured that the count of such pairs is \( \sim Cx/\log^2 x \) for a certain constant \( C \). By breaking up \([1, x]\) into subintervals of length \( d \), we easily derive

\[
|\mathcal{A}_d| = X \frac{\rho(d)}{d} + O(\rho(d)), \quad \rho(d) = \#\{0 \leq k \leq d-1 : k(k+2) \equiv 0 \pmod{d}\}.
\]

By the Chinese remainder theorem, \( \rho \) is multiplicative, \( \rho(2) = 1 \) and \( \rho(p) = 2 \) for \( p > 2 \). Putting \( g(d) = \rho(d)/d \) and applying Mertens, we get that

\[
(1.2) \quad V(z) = \frac{1}{2} \prod_{3 \leq p \leq z} \left( 1 - \frac{2}{p} \right) \sim \frac{c}{\log^2 z}, \quad c = \frac{e^{-2\gamma}}{2} \prod_{p>2} \frac{1 - 2/p}{(1 - 1/p)^2}.
\]

Sieve methods deliver an upper bound of the expected order

\[
S(\mathcal{A}, \mathcal{P}, \sqrt{x}+2) \ll XV(\sqrt{x}+2) \asymp \frac{x}{\log^2 x},
\]

but only a lower bound for somewhat smaller \( z \):

\[
S(\mathcal{A}, \mathcal{P}, x^{1/4}) \gg XV(x^{1/4}) \asymp \frac{x}{\log^2 x}.
\]

From the last estimate, we conclude that there are \( \gg x/\log^2 x \) values of \( k \leq x \) for which each of \( k \) and \( k+2 \) have at most 3 prime factors. This is a typical conclusion from a lower bound sieve.

Example (g) (Selberg). The values of \( c \) for which sieve method deliver lower bounds (L) are invariably smaller than we would like. There is a fundamental barrier at work which “explains” this, known as the “parity barrier”. Roughly speaking, the small sieve works with inputs \( X \), the function \( g \) and estimates for the remainders \( r_d \). However, even if “super-good” estimates for \( r_d \) are available, the sieve fundamentally cannot distinguish between numbers with an odd number of prime factors from those with an even number of prime factors. Consider two sequences defined as follows. Recall the Liouville function \( \lambda(n) = (-1)^{\Omega(n)} \) (the completely multiplicative function which is -1 at all primes); (we temporarily accept a slight clash of notation with the symbol for sieves). Let

\[
\mathcal{A}^+ = \{1 \leq n \leq x : \lambda(n) = 1\}, \quad \mathcal{A}^- = \{1 \leq n \leq x : \lambda(n) = -1\},
\]

\( \mathcal{P} \) the set of all primes, \( z = \sqrt{x} \). The prime number theorem (with classical error term) implies

\[
\sum_{n \leq x} \lambda(n) = O(xe^{-c' \sqrt{\log x}}), \text{ some } c' > 0.
\]

Therefore,

\[
|\mathcal{A}_d^\pm| = \sum_{m \leq x/d} \frac{1 \pm \lambda(dm)}{2} = \frac{1}{2} \left| \frac{x}{d} \right| \pm \frac{1}{2} \sum_{m \leq x/d} \lambda(m) = \frac{x}{2d} + O\left(\frac{x}{d}e^{-c' \sqrt{\log(x/d)}}\right).
\]
Let \( X = \frac{x}{2}, \) \( g(d) = \frac{1}{2}. \) For both sequences, \( r_d \) is uniformly small for \( d \leq x^{1-\varepsilon} \) (pretty much the best range that one can hope for), and
\[
XV(\sqrt{x}) \sim \frac{e^{-\gamma}x}{\log x}.
\]
However, numbers \( n \leq x \) that have no prime factor \( \leq \sqrt{x} \) are prime or 1, thus
\[
S(\mathcal{A}^+, \mathcal{P}, \sqrt{x}) = 1, \quad \text{while} \quad S(\mathcal{A}^-, \mathcal{P}, \sqrt{x}) \sim \frac{x}{\log x}.
\]

### 1.6 Legendre’s sieve

The simplest of all sieves is Legendre’s sieve, where one works directly with the identity (SI) (or the more general (SI')) rather than with inequalities (S). Adopting the notation \((g)\) and \((r)\) (with \(D = P(z)\)), we have
\[
S(A, P, z) = X \sum_{d|P(z)} \mu(d)g(d) + \sum_{d|P(z)} \mu(d)r_d = X \prod_{\substack{p \in \mathcal{P} \\ p \leq z}} (1 - g(p)) + O\left( \sum_{d|P(z)} |r_d| \right).
\]
Although this sieve suffers from the large number of remainder summands, it is useful in situations where the “densities” \( g(p) \) are rather small on average, and so the product on the right hand side of (1.3) captures the true behavior of set of interest for relatively small \( z \). A prominent example of the use of Legendre’s sieve was given by C. Hooley [3], who deduced Artin’s primitive root conjecture from the Generalized Riemann Hypothesis (for Dedekind zeta functions of certain number fields). Details will be given later in Section 3.

### 1.7 Brun’s pure sieve (simple version)

We begin by re-interpreting (SI) combinatorially as a form of inclusion-exclusion:
\[
S(\mathcal{A}, \mathcal{P}, z) = \omega(P(z)) \sum_{j=0}^{\omega(P(z))} (-1)^j \sum_{d|P(z) \atop \omega(d) = j} |\mathcal{A}_d|.
\]
Brun’s idea is to truncate the outer sum to \( j \leq k \), observing that if \( k \) is even the resulting sum is an upper bound for \( S(\mathcal{A}, \mathcal{P}, z) \) and if \( k \) is odd then the sum is a lower bound for \( S(\mathcal{A}, \mathcal{P}, z) \) (these are known as the “Bonferonni inequalities” in probabilistic combinatorics). In our general notation, the sieve is
\[
\lambda_d = \begin{cases} 
\mu(d) & \text{if } \omega(d) \leq k \\
0 & \text{if } \omega(d) > k.
\end{cases}
\]
For any odd positive integer \( k_{\text{odd}} \) and even positive integer \( k_{\text{even}} \), we have
\[
\sum_{d|m \atop \omega(d) \leq k_{\text{odd}}} \mu(d) \leq \sum_{d|m \atop \omega(d) \leq k_{\text{even}}} \mu(d) \quad (m > 1),
\]
that is, the sieve satisfies (\( \lambda^- \)) for odd \( k \) and (\( \lambda^+ \)) for even \( k \). Therefore by (S),
\[
\sum_{d|P(z) \atop \omega(d) \leq k_{\text{odd}}} \mu(d)|\mathcal{A}_d| \leq S(\mathcal{A}, \mathcal{P}, z) \leq \sum_{d|P(z) \atop \omega(d) \leq k_{\text{even}}} \mu(d)|\mathcal{A}_d|.
\]
The following lemma and its corollary tell us precisely the error we make in approximating “full sums” with truncated sums of this type. It is a special case of a more general “fundamental sieve identity” (see Ch. 2, §1 of [6] or Lemma 2 of [1]).

**Lemma 1.1.** Let \( \mathcal{P} \) be a finite set of primes with product \( P \), let \( f \) be a multiplicative function and \( k \) a non-negative integer. Then

\[
\sum_{d \mid P, \omega(d) \leq k} \mu(d) f(d) = \prod_{p \in \mathcal{P}} (1 - f(p)) + (-1)^k \sum_{\ell \mid P, \omega(\ell) = k + 1} f(\ell) \prod_{p \in \mathcal{P}} (1 - f(p)).
\]

**Proof.** First, write

\[
\prod_{p \in \mathcal{P}} (1 - f(p)) = \sum_{d \mid P} \mu(d) f(d)
\]

and consider the summands on the right side with \( \omega(d) \geq k + 1 \) and \( \mu(d) \neq 0 \). Let \( \ell \) be the product of the \( k + 1 \) largest prime factors of \( d \) and put \( h = d/\ell \). Then \( h \mid Q_\ell \), where

\[
Q_\ell = \prod_{p \in \mathcal{P}, \, p < P^{-}(\ell)} p,
\]

and we obtain

\[
\prod_{p \in \mathcal{P}} (1 - f(p)) - \sum_{d \mid P, \omega(d) \leq k} \mu(d) f(d) = \sum_{d \mid P, \omega(d) \geq k + 1} \mu(d) f(d)
\]

\[
\sum_{\ell \mid P, \omega(\ell) = k + 1} f(\ell) \sum_{h \mid Q_\ell} \mu(h) f(h)
\]

\[
= (-1)^{k + 1} \sum_{\ell \mid P, \omega(\ell) = k + 1} f(\ell) \prod_{p \in \mathcal{P}} (1 - f(p)). \quad \Box
\]

**Example.** Let \( m > 1 \), \( \mathcal{P} \) the set of primes dividing \( m \), and \( f(p) = 1 \) for all \( p \). We see that Brun’s inequalities \((B_0)\) follow as a special case of Lemma 1.1.

**Corollary VW.** Let \( \mathcal{P} \) be a finite set of primes with product \( P \), let \( f \) be a multiplicative function with \( 0 \leq f(p) < 1 \) for \( p \in \mathcal{P} \), and let \( k \) be a non-negative integer. Then

\[
\sum_{d \mid P, \omega(d) \leq k} \mu(d) f(d) = V + (-1)^k W,
\]

where

\[
V = \prod_{p \in \mathcal{P}} (1 - f(p)), \quad 0 \leq W \leq \frac{(\log \frac{1}{V})^{k+1}}{(k + 1)!}.
\]
Proof. Apply Lemma 1.1 and an old Erdős trick. We have

\[ 0 \leq W \leq \sum_{\ell \nmid P \atop \omega(\ell) = k+1} f(\ell) \leq \frac{1}{(k+1)!} \left( \sum_{p \in \mathcal{P}} f(p) \right)^{k+1} \]
\[ \leq \frac{1}{(k+1)!} \left( \sum_{p \in \mathcal{P}} - \log(1 - f(p)) \right)^{k+1} \]
\[ = \left( \frac{\log \frac{1}{V(z)}}{(k+1)!} \right)^{k+1}. \]

□

Remarks. The inequalities in the proof of Corollary VW are relatively sharp if the quantities \( f(p) \) are small. This will be the case in most applications, at least of “small sieve” type.

Now adopt the notation \((g)\) and \((r)\). We have

\[ \sum_{d \mid P(z) \atop \omega(d) \leq k} \mu(d) |\mathcal{A}_d| = XU_k + R_k, \]

where

\[ U_k = \sum_{d \mid P(z) \atop \omega(d) \leq k} \mu(d) g(d), \quad |R_k| \leq \sum_{d \mid P(z) \atop d \leq z^k} |r_d|. \]

As before, define

\[(V) \quad V(z) := \prod_{p \in \mathcal{P} \atop g(p) = 1} (1 - g(p)). \]

By Corollary VW and the elementary inequality \( e^k \geq k^k/k! \),

\[ |U_k - V(z)| \leq \left( \frac{\log \frac{1}{V(z)}}{(k+1)!} \right)^{k+1} \leq \left( \frac{e \log \frac{1}{V(z)}}{k+1} \right)^{k+1}. \]

Consequently, if \( k \geq 4 \log(1/V(z)) - 1 \), then

\[ |U_k - V(z)| \leq \left( \frac{e}{4} \right)^{4 \log(1/V(z))} = V(z)^{4 \log 4 - 4} \leq V(z)^{3/2}. \]

Now apply \((B_1)\) with the cutoffs \( k = \left\lfloor 4 \log \frac{1}{V(z)} \right\rfloor \) and \( k = \left\lceil 4 \log \frac{1}{V(z)} \right\rceil + 1 \) (one of which is even, the other odd), and we arrive at the following simple, but very general first sieve bound.

Theorem B (Brun’s sieve, simple version). Assume \((g)\) and \((r)\). Then

\[ S(\mathcal{A}, \mathcal{P}, z) = X \left( V(z) + O(V(z)^{3/2}) \right) + O\left( \sum_{d \mid P(z) \atop d \leq D} |r_d| \right), \]

where \( D = z^{\left\lfloor 4 \log(1/V(z)) \right\rfloor + 1} \), and \( V(z) \) is given by \((V)\).
Remarks. We have imposed virtually no hypotheses on the sequence $\mathcal{A}$ in this theorem. In applications, one has typically $V(z) \asymp (\log z)^{-\kappa}$ for some fixed $\kappa$, and thus we have applied $(B_1)$ with $k \approx 4\kappa \log \log z$. Also, typically in the sum in identity (SI) and the sums in Brun’s inequalities $(B_1)$, the summands corresponding to $d > X$ are negligible (or identically zero). Hardy and Ramanujan in 1917 showed that most integers $d \leq X$ have about $\log \log X$ prime factors, and thus it is natural to choose $k$ somewhat larger than $\log \log X$ in order to capture the bulk of the sum.

Example (b). Counting twin primes. $\mathcal{A} = \{k(k + 2) : 1 \leq k \leq x\}$, $\mathcal{P}$ is the set of all primes,

$$g(p) = \frac{\rho(p)}{p}, \quad \rho(2) = 1, \quad \rho(p) = 2 \quad (p > 2)$$

and

$$|r_d| \leq \rho(d) \leq \tau(d).$$

By an earlier calculation (1.2), $V(z) \sim c(\log z)^{-2}$ for some constant $c$ (that is, we have $\kappa = 2$). Take

$$z = x^{16 \log \log x},$$

so that

$$D = z^{\lfloor 4 \log(1/V(z)) \rfloor + 1} = z^{8 \log \log z + O(1)} \ll x^{1/2}.$$

The remainder terms are handled easily:

$$\sum_{d \mid P(z), d \leq D} |r_d| \leq \sum_{d \leq D} \tau(d) \ll D \log D \ll x^{1/2} \log x.$$

By Theorem B,

$$S(\mathcal{A}, \mathcal{P}, z) \sim X V(z) \sim \frac{cx}{\log^2 z} \sim 256cx \left(\frac{\log \log x}{\log x}\right)^2.$$

As $S(\mathcal{A}, \mathcal{P}, z) \geq \#\{z < k \leq x : k, k + 2 \text{ both prime}\}$, we see that

$$\#\{k \leq x : k, k + 2 \text{ both prime}\} \ll x \left(\frac{\log \log x}{\log x}\right)^2,$$

which misses the conjectured order by a factor $(\log \log x)^2$. Applying partial summation gives an immediate corollary.

Corollary 1 (V. Brun, 1917). We have

$$\sum_{p \mid p, p + 2 \text{ prime}} \frac{1}{p} < \infty.$$

Remarks. Applying Theorem B to Example (a) yields $\pi(x) \ll x \frac{\log \log x}{\log x}$, missing the true order of $\pi(x)$ by a factor $\log \log x$.

Brun later gave much more complicated versions of his sieve, where the simple truncation $(B_1)$ is replaced by sieves where one considers the summation over integers $d$ with restricted prime factors of various sizes. A greatly simplified version of this idea was found by Hooley, and which will be the subject of the following section.
2 Hooley’s variation of Brun’s sieve

In this section, we present a variation of Brun’s pure sieve due to C. Hooley [5], which is very simple, from a combinatorial viewpoint, and yet powerful enough to deliver quickly each of the desired sieve bounds (U), (L) and (A).

2.1 Upper bound

The fundamental idea is to partition \( \mathcal{P} \cap \left[ 2, z \right] \) into sets \( \mathcal{P}_1, \ldots, \mathcal{P}_t \) and so every \( d|P(z) \) has a unique factorization as

\[
  d = d_1 \cdots d_t, \quad d_i|P_i, \quad P_i := \prod_{p \in \mathcal{P}_i} p.
\]

Using \((B_0)\), we have the upper bound

\[
  \sum_{d|(n,P(z))} \mu(d) = \prod_{j=1}^t \sum_{d_j|(n,P_j)} \mu(d_j) \leq \prod_{j=1}^t \sum_{\omega(d_j) \leq k_j} \mu(d_j),
\]

where \(k_1, \ldots, k_t\) are arbitrary positive, even, integers. That is, we apply an upper bound sieve

\[
  \lambda_d^+ = \begin{cases} 
    \mu(d) & \text{if } \omega(d_j) \leq k_j \ (1 \leq j \leq t) \\
    0 & \text{otherwise.}
  \end{cases}
\]

The motivation for this comes from looking at the statistical distribution of prime factors of integers \( \leq X \). Not only does a typical integer have about \( \log \log X \) prime factors, but the prime factors themselves are typically uniformly distributed on a \( \log \log \)-scale; that is, there are about \( \log \log t \) prime factors \( \leq t \), uniformly for \( t \leq X \). With \((B_0)\), we can restrict the number of large prime factors of the summands \( d \) while still retaining almost all summands. Thus, we can keep the inequality in \((B_0)\) relatively sharp, while at the same time keeping the size of the summands \( d \) relatively small; no larger than a fixed power of \( z \).

Before making specific choices for \( \mathcal{P}_j \) and \( k_j \), we will work out a very general inequality for the sifting function from \((g)\) and \((r)\), analogous to the estimate in Theorem B. Generically we will take each \( \mathcal{P}_j \) to be the set of primes in successive intervals:

\[
  z_{t+1} < 2 \leq z_t < z_{t-1} < \cdots < z_1 = z, \quad \mathcal{P}_j = \mathcal{P} \cap \left( z_{j+1}, z_j \right].
\]

Summing \((B_0)\) over \( n \leq x \) gives

\[
  S(\mathcal{A}, \mathcal{P}, z) \leq \sum_{d_1|P_1 \atop \omega(d_1) \leq k_1} \cdots \sum_{d_t|P_t \atop \omega(d_t) \leq k_t} \mu(d_1) \cdots \mu(d_t) |\mathcal{A}_{d_1 \cdots d_t}| = XU_1 \cdots U_t + R,
\]

where

\[
  U_j = \sum_{d_j|P_j \atop \omega(d_j) \leq k_j} \mu(d_j) g(d_j), \quad R = \sum_{d_1|P_1 \atop \omega(d_1) \leq k_1} \cdots \sum_{d_t|P_t \atop \omega(d_t) \leq k_t} \mu(d_1 \cdots d_t) r_{d_1 \cdots d_t}.
\]
Define

\[(V_j L_j) \quad V_j = \prod_{p \in P_j} (1 - g(p)), \quad L_j = \log \frac{1}{V_j}.\]

By Corollary VW,

\[U_j \leq V_j + \frac{L_j^{k_j+1}}{(k_j + 1)!} = V_j \left(1 + e^{L_j} \frac{L_j^{k_j+1}}{(k_j + 1)!}\right) \leq V_j \exp \left\{e^{L_j} \frac{L_j^{k_j+1}}{(k_j + 1)!}\right\}.\]

In the sum defining \(R\), we have

\[d_1 \cdots d_t \leq D := z_1^{k_1} \cdots z_t^{k_t}.\]

We conclude that

\[(B\,U1) \quad S(\mathcal{A}, \mathcal{P}, z) \leq XV(z)e^E + R', \quad E := \sum_{j=1}^{t} e^{L_j} \frac{L_j^{k_j+1}}{(k_j + 1)!}, \quad R' := \sum_{d | P(z), d \leq D} |r_d|.\]

In order to provide estimates which are readily usable in practice, we will impose typical hypotheses on the function \(g(p)\) and on the remainders \(r_d\). First, we will assume the one-sided bound

\[(\Omega) \quad \prod_{y \leq p \leq w} (1 - g(p))^{-1} \leq \left(\frac{\log w}{\log y}\right)^{\kappa} \exp \left(\frac{B}{\log y}\right) \quad (2 \leq y < w)\]

for some constants \(B > 0, \kappa \geq 0\); recall that \(g(p) = 0\) for \(p \notin \mathcal{P}\). Although \(B, \kappa\) are not uniquely defined by \((\Omega)\), the smallest admissible value of \(\kappa\) (with \(B\) remaining bounded) is sometimes referred to as the “dimension” or “sifting density”. Think of \(\kappa\) as the average number of residue classes modulo \(p\) which are removed in the sieving process; e.g. the average of \(j(p)\) in the general sieve described at the end of subsection 1.2. Inequality \((\Omega)\), called the Iwaniec condition in the literature, says that \(g(p) \leq \kappa/p\) on average. In particular, we have

**Proposition g.** Suppose that \(k \geq 0\) is a real number, \(g(p) \leq k/p\) for all \(p \geq p_0\), and \(g(p) \leq 1 - c\) for all \(p\), where \(c > 0\) is constant. Then \((\Omega)\) holds with \(\kappa = k\), and \(B\) depending only on \(k, p_0\) and \(c\).

**Proof.** Let \(p_1 = \max(p_0, 2k)\). First, if \(y > p_1\), then

\[
\prod_{y \leq p \leq w} (1 - g(p))^{-1} \leq \prod_{y \leq p \leq w} \left(1 - \frac{k}{p}\right)^{-1} = \prod_{y \leq p \leq w} \left(1 - \frac{1}{p}\right)^{-k} \left(1 + O_k \left(\frac{1}{p^2}\right)\right) \leq \left(\frac{\log w}{\log y}\right)^{k} \left(1 + O_k \left(\frac{1}{\log y}\right)\right).
\]
If \( y < p_1 \), then
\[
\prod_{y \leq p \leq w} (1 - g(p))^{-1} \leq \prod_{y \leq p_1} c^{-1} \prod_{p_1 < p \leq w} (1 - k/p)^{-1}
\]
\[
= \left( \prod_{y \leq p \leq p_1} c^{-1} \right) \left( \frac{\log w}{\log p_1} \right)^k \left( 1 + O_k \left( \frac{1}{\log p_1} \right) \right)
\]
\[
\leq \left( \frac{\log w}{\log y} \right)^k \exp \left( \frac{B}{\log y} \right)
\]
by Mertens’ bounds, where \( B \) depends only on \( k, p_0, \) and \( c \).

Next, we choose the \( k_j \) and \( z_j \). There is great flexibility to choose these parameters optimally for the problem at hand, however a good all-purpose choice is given by

\[(z_j) \quad z_j = z^{2^j} (1 \leq j \leq t + 1), \quad \text{where } z_{t+1} < 2 \leq z_t, \quad \mathcal{P}_j = \mathcal{P} \cap (z_{j+1}, z_j)\]

and

\[(k_j) \quad k_j = b + 2(j - 1) \quad (1 \leq j \leq t), \quad b \in 2\mathbb{N}.
\]

Inserting hypotheses \((\Omega)\) and \((z_j)\) into the definition \((V_jL_j)\), we have

\[(L_j) \quad L_j = \log \prod_{z_{j+1} < p \leq z_j} (1 - g(p))^{-1} \leq \log(2^\kappa) + \frac{B}{\log \max(2, z_{j+1})}
\]

\[\leq \log(2^\kappa) + \frac{B}{\log 2} =: L \quad (1 \leq j \leq t).
\]

Using both \((z_j)\) and \((k_j)\) in the definition of \( E \) from \((BH.U1)\), we see that

\[\text{(D)} \quad D = z_1^{k_1} \cdots z_t^{k_t} \leq z^{b + \frac{1}{2}(b+2) + \frac{1}{4}(b+4) + \cdots} \leq z^{2b+4}\]

and

\[\text{(E)} \quad E \leq e^L \sum_{j=1}^{t} \frac{L^{b+2j-1}}{(b+2j-1)!} \leq e^L \frac{L^b}{b!} \sum_{j=1}^{\infty} \frac{L^{2j-1}}{(2j-1)!} \leq e^{2L} \frac{L^b}{b!}.
\]

We conclude from \((BH.U1)\) the following.

**Theorem BH.1.** Assume \((g), (r)\) and \((\Omega)\). Let \( b \) be an even, positive integer. Then

\[S(\mathcal{A}, \mathcal{P}, z) \leq C_1(\kappa, B, b)XV(z) + \sum_{d | P(z)} |r_d|,
\]

where

\[V(z) = \prod_{p \in \mathcal{P}} (1 - g(p)), \quad C_1(\kappa, B, b) = \exp \left\{ 2^{2\kappa} e^{2B/\log 2} \left( \frac{\log(2^\kappa) + B}{\log 2} \right)^b \right\}.
\]
Remark. We have made no real effort to optimize the size of the constant $C_1(\kappa, B, b)$. What is important is that it depends only on the parameters $\kappa, B$ from ($\Omega$) and the free parameter $b$.

Finally, we must handle the remainders. A typical hypothesis is the following:

$$ (R(\theta, A)) \quad \sum_{d \mid P(z)} |r_d| \leq C_2 \frac{X}{(\log X) \theta}.$$  

In the literature, one often encounters the notion of “level of distribution” of sequences $\mathcal{A}$. With respect to the choice of $g$ and $X$, a sequence $\mathcal{A}$ has level of distribution $\theta$ if $(R(\theta, A))$ holds for every $A > 0$, the constant $C_2$ depending only on $A$ and $\theta$.

**Theorem BH.2.** Assume $(g), (r), (\Omega)$ and $(R(\theta, \kappa))$ for some $\theta > 0$. If $2 \leq z \leq X$, then

$$ S(\mathcal{A}, \mathcal{P}, z) \ll_{\kappa, B, C_2, \theta} XV(z). $$

**Proof.** First assume that $2 \leq z \leq X^{\theta/8}$. We apply Theorem BH.1 with $b = 2$, observing that $z^{2b+4} = z^8 \leq X^\theta$ and that

$$ V(z) \geq e^{-B/\log 2} \left( \frac{\log 2}{\log z} \right)^\kappa \geq_{B, \kappa} \frac{1}{(\log X)^\kappa} $$

by ($\Omega$). Therefore,

$$ S(\mathcal{A}, \mathcal{P}, z) \leq C_1(\kappa, B, 2) XV(z) + C_2 \frac{X}{(\log X)^\kappa} \ll_{\kappa, B, C_2} XV(z). $$

Now assume $X^{\theta/8} \leq z \leq X$. By ($\Omega$),

$$ \frac{V(X^{\theta/8})}{V(z)} = \prod_{X^{\theta/8} < p \leq z} (1 - g(p))^{-1} \leq e^{B/\log 2} \left( \frac{\log z}{\log(X^{\theta/8})} \right)^\kappa \leq e^{B/\log 2}(8/\theta)^\kappa. $$

Hence,

$$ S(\mathcal{A}, \mathcal{P}, z) \leq S(\mathcal{A}, \mathcal{P}, X^{\theta/8}) \ll_{\kappa, B, C_2} XV(X^{\theta/8}) \ll_{\kappa, B, C_2, \theta} XV(z). $$

This gives a general purpose upper bound of type $(U)$.

**Application: generalized twin primes and Goldbach partitions**

Theorem BH.2 is very easy to apply in practice. All one needs to verify is that the sifting function $g(p)$ has regular behavior ($\Omega$) (say, by verifying the hypotheses of Proposition $g$) and that the remainders $r_d$ are small on average in some reasonable range $(R(\theta, A))$.

**Theorem TP.** For nonzero integers $a$ and $h$, let $N_{a,h}(x)$ be the number of integers $n \leq x$ for which both $n$ and $an + h$ are both prime. Uniformly for $x \geq 2$, $a$ and $h$, we have

$$ N_{a,h}(x) \ll \frac{|ah|}{\varphi(|ah|)} \frac{x}{\log^2 x}. $$
Proof. To avoid trivial cases, we may assume that $2 \nmid (a + h)$ (if $2 | (a + h)$, then $2 | (an + h)$ whenever $n$ is odd, hence $N_{a,h}(x) \ll 1$) and that $(a, h) = 1$. Let $a_n = \# \{ 1 \leq k \leq x : n = k(a + h) \}$, $X = x$, $z = \sqrt{X}$, and let $\mathcal{P}$ be the set of all primes. For squarefree $d$, let $\rho(d)$ be the number of solutions $k \mod d$ of the congruence

$$k(a + h) \equiv 0 \pmod{d}.$$ 

By the Chinese remainder theorem, $\rho$ is multiplicative, and for prime $q$, we have

$$\rho(q) = \begin{cases} 1 & \text{if } q | ah \ \text{or} \ q \nmid ah. \end{cases}$$

Setting $g(p) = \frac{\rho(p)}{p}$, we have $g(p) \leq \frac{2}{p}$ for all $p$ and thus by Proposition g, $(\Omega)$ holds with $\kappa = 2$ and $B$ independent of $a, h$. Moreover,

$$|\mathcal{A}_d| = X \frac{\rho(d)}{d} + r_d, \quad |r_d| \leq \rho(d) \quad (\mu^2(d) = 1),$$

and thus

$$\sum_{d | P(z)} |r_d| \leq \sum_{d \leq X^\theta} \tau(d) \ll X^\theta \log X.$$ 

In particular, $(R(0.9, 2))$ holds, with $C_2$ independent of $a, h$. By Theorem BH.2,

$$S(\mathcal{A}, \mathcal{P}, z) \ll xV(z).$$

Here (using that $2 | ah$),

$$V(z) = \prod_{p \leq z \atop p | ah} \left( 1 - \frac{1}{p} \right) \prod_{p \leq z \atop p \nmid ah} \left( 1 - \frac{2}{p} \right) = \frac{1}{2} \prod_{3 \leq p \leq z} \left( 1 - \frac{2}{p} \right) \prod_{3 \leq p \leq z \atop p | ah} \frac{p - 1}{p - 2} \ll \prod_{3 \leq p \leq z} \left( 1 - \frac{1}{p} \right)^2 \prod_{p | ah} \frac{p}{p - 1} \ll \frac{1}{\log^2 z} \frac{|ah|}{\phi(|ah|)}.$$ 

the implied constants being absolute. Finally, there are at most $\sqrt{x}$ values of $k \leq x$ such that $k$ and $a + h$ are prime and $\leq \sqrt{x}$. Thus,

$$N_{a,h}(x) \leq \sqrt{x} + S(\mathcal{A}, \mathcal{P}, \sqrt{x}) \ll \sqrt{x} + \frac{|ah|}{\phi(|ah|)} \frac{x}{\log^2 x} \ll \frac{|ah|}{\phi(|ah|)} \frac{x}{\log^2 x}.$$ 

□

Some special cases of Theorem TP include counts of twin primes $(a = 1, h = 2)$, near twins $(a = 1, h = 4)$, Sophie Germain primes $(a = 2, h = 1)$, many other examples of generalized twin primes. Another application is to Goldbach’s Conjecture, by taking an even integer $2M$ and taking $(a = -1, h = 2M)$; here $N_{a,h}(M)$ counts the number of ways that $2M$ can be written as the sum of two primes.
Corollary 2. We have
\[
\#\{p \leq x : p, p + 2 \text{ both prime}\} \ll \frac{x}{\log^2 x},
\]
\[
\#\{p \leq x : 2p + 1 \text{ both prime}\} \ll \frac{x}{\log^2 x},
\]
\[
\#\{p, q \text{ prime} : p + q = 2M\} \ll \frac{M}{\phi(M)} \frac{M}{\log^2 M}.
\]

Application: The Brun-Titchmarsh inequality

Theorem BT (Brun-Titchmarsh inequality). There is a constant \(C\) so that if \(1 \leq a \leq q, (a, q) = 1\) and \(q < y \leq x\) then
\[
\#\{p : x - y < p \leq x, p \equiv a \pmod{q}\} \leq C \frac{y}{\phi(q) \log(y/q)}.
\]

Remarks. The case \(a = q = 1\) is due to Hardy-Littlewood (1923), while the case \(y = x\) was shown by Titchmarsh (1930). The constant \(C = 2\) is admissible (Montgomery-Vaughan, 1973). The utility of this bound is its uniformity in \(x, y, a, q\). If \(\pi(x; q, a) := \#\{p \leq x : p \equiv a \pmod{q}\}\), the prime number theorem for arithmetic progressions implies that
\[
\pi(x; q, a) \sim \frac{x}{\phi(q) \log x}
\]
for every individual \(q, a\). Issues of uniformity are crucially important, and are intimately connected to questions of the existence of zeros of \(L\)-functions lying off the critical line (especially so-called Siegel zeros, real zeros lying very close to 1). If one assumes the Generalized Riemann Hypothesis, then the above asymptotic holds uniformly for \(q \leq \sqrt{x}\) or so, and in a smaller range if one considers primes in “short” intervals \((x - y, x]\). On the other hand, the Brun-Titchmarsh inequality gives an upper bound of the correct order even for very large moduli \(q = x^{1-\varepsilon}\) for fixed and very short intervals \(y/q = x^\varepsilon\) (\(\varepsilon > 0\) fixed).

Proof. Let \(\mathcal{A} = \{x - y < n \leq x : n \equiv a \pmod{q}\}\), \(a_n = 1_{\mathcal{A}}(n)\), \(\mathcal{P}\) the set of all primes. Then \(\mathcal{A}\) has about \(X = y/q\) elements. We may assume that \(y \geq 2q\) (that is, \(X \geq 2\)), for if \(y < 2q\) then
\[
\pi(x; q, a) - \pi(x - y; q, a) \leq 2 \leq \frac{y}{\log(y/q)} \leq \frac{y}{\phi(q) \log(y/q)},
\]
as desired (using that the function \(x/\log x\) has a local minimum of \(e\) as \(x = e\)). Next, for any squarefree \(d\),
\[
|\mathcal{A}_d| = \frac{\rho(d)}{d} X + r_d, \quad |r_d| \ll \rho(d),
\]
where \(\rho(d)\) is the number of solutions \(n\) of the congruence \(qn + a \equiv 0 \pmod{d}\). We have \(\rho(p) = 1\) if \(p \nmid q\) and \(\rho(p) = 0\) if \(p|q\). Since \(g(p) := \rho(p)/p \leq 1/p\) for all \(p\), Proposition \(g\) implies that \((\Omega)\) holds with \(\kappa = 1\) and \(B\) independent of \(x, y, a, q\). Let \(\theta = 1/2\). The remainders are very easy:
\[
\sum_{d \leq X^\theta} |r_d| \ll X^\theta \ll \frac{X}{\log X}.
\]
Take \(z = X^{1/2}\). By Theorem BH.2,
\[
S(\mathcal{A}, \mathcal{P}, z) \ll XV(z).
\]
Also, as primes less than \( z \) are sifted out of \( S(\mathcal{A}, \mathcal{P}, z) \),

\[
S(\mathcal{A}, \mathcal{P}, z) \geq \pi(x; q, a) - \pi(x - y; q, a) - \pi(z; q, a) \geq \pi(x; q, a) - \pi(x - y; q, a) - z/q - 1.
\]

Finally,

\[
V(z) = \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \leq \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \prod_{p|q} \left( 1 - \frac{1}{p} \right)^{-1} = \frac{q}{\phi(q)} \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \ll \frac{q}{\phi(q) \log z} \ll \frac{q}{\phi(q) \log X}.
\]

We conclude that

\[
\pi(x; q, a) - \pi(x - y; q, a) \ll \frac{\sqrt{X}}{q} + 1 + X \frac{q}{\phi(q) \log X} \ll \frac{y}{\phi(q) \log(y/q)}.
\]

### 2.2 The lower bound

One cannot, by analogy with (BH.U), multiply together the sums on the left side of Brun’s inequality \((B_0)\), since these sums may be negative. Instead, we begin with the inequality (BH.U) used to prove the upper bound for \( S(\mathcal{A}, \mathcal{P}, z) \), and develop a corresponding lower bound from it by subtracting off appropriate quantities. We use the following simple inequality:

**Lemma 2.1.** Suppose that \( 0 \leq x_j \leq y_j \) for \( 1 \leq j \leq t \). Then

\[
x_1 \cdots x_t \geq y_1 \cdots y_t - \sum_{\ell=1}^{t} (y_{\ell} - x_{\ell}) \prod_{j=1, j \neq \ell}^{t} y_j.
\]

**Proof.** The inequality is an equality when \( t = 1 \), and follows by induction on \( t \) using

\[
y_1 \cdots y_t - x_1 \cdots x_t = (y_1 \cdots y_{t-1} - x_1 \cdots x_{t-1})y_t + x_1 \cdots x_{t-1}(y_t - x_t) \leq (y_1 \cdots y_{t-1} - x_1 \cdots x_{t-1})y_t + x_1 \cdots y_{t-1}(y_t - x_t).StoredProcedure
\]

We apply Lemma 2.1 with

\[
x_j = \sum_{d|n,P_j} \mu(d), \quad y_j = \sum_{d|n,P_j, \omega(d) \leq k_j} \mu(d), \quad (1 \leq j \leq t),
\]

where \( k_1, \ldots, k_t \) are even, positive integers. From Lemma 1.1 (with \( f(p) = 1 \) for each \( p \)) we have that

\[
0 \leq y_t - x_t \leq \sum_{d|n,P_\ell, \omega(d) = k_\ell + 1} 1,
\]

hence by (BH.U) and Lemma 2.1,

\[
(BH.L) \quad \sum_{d|n,P(z)} \mu(d) \geq \prod_{j=1}^{t} \sum_{d_j|n,P_j, \omega(d_j) \leq k_j} \mu(d_j) - \prod_{\ell=1}^{t} \left( \sum_{d_\ell|n,P_\ell, \omega(d_\ell) = k_\ell + 1} 1 \right) \prod_{j=1}^{t} \sum_{d_j|n,P_j, j \neq \ell, \omega(d_j) \leq k_j} \mu(d_j).
\]
Multiplying by \(a_n\) and summing over \(n\), we get

\[
S(\mathcal{A}, \mathcal{P}, z) \geq \sum_{d_1, \ldots, d_t} \mu(d_1) \cdots \mu(d_t) |\mathcal{A}_{d_1 \cdots d_t}| - \sum_{\ell=1}^t \sum_{d_1, \ldots, d_t} \mu \left( \frac{d_1 \cdots d_t}{d_\ell} \right) |\mathcal{A}_{d_1 \cdots d_t}|.
\]

Observe that in the above expression, the products \(d_1 \cdots d_\ell\) are all distinct. Assume a prime partition of type \((z_j)\) as in the upper bound. Inserting the approximations \((g)\) and \((r)\), and adopting our previous notation \((UR)\) for quantities \(U_j\), we can rewrite this as

\[
S(\mathcal{A}, \mathcal{P}, z) \geq XU_1 \cdots U_t \left( 1 - \sum_{\ell=1}^t \frac{1}{U_\ell} \sum_{d_\ell | P_\ell} \frac{g(d_\ell)}{\omega(d_\ell) = \omega(d_\ell) | P_\ell} \right) - R'',
\]

where, defining \(D\) as before by \((D)\),

\[
|R''| \leq \sum_{d \in \mathcal{D}_z} |r_d|.
\]

By Corollary VW, \(U_i \geq V_i\) for every \(i\) \((V_i\) defined in \((V_jL_j)\)), thus in particular \(U_1 \cdots U_t \geq V(z)\) and \(1/U_\ell \leq 1/V_\ell = e^{k_\ell}\). Another invocation of the Erdős trick (cf., the proof of Corollary VW) gives

\[
\sum_{d_\ell | P_\ell, \omega(d_\ell) = k_\ell+1} g(d_\ell) \leq \frac{1}{(k_\ell + 1)!} \left( \sum_{p \in P_\ell} g(p) \right)^{k_\ell+1} \leq \frac{L_{k_\ell+1}^{k_\ell+1}}{(k_\ell + 1)!}.
\]

Thus, recalling the definition of \(E\) from \((BH.U1)\), we get

\[
(BH.L1) \quad S(\mathcal{A}, \mathcal{P}, z) \geq XV(z)(1 - E) - R''.
\]

As with the upper bound, assume the average bound \((\Omega)\) for the function \(g\), and define the even integer cutoffs \(k_j\) by \((k_j)\), where \(b\) is an arbitrary positive even integer. We have the same bound \((D)\) for \(D\). Also, by \((L_j)\),

\[
L_j \leq \log(2^\kappa) + \frac{B}{\log z_{j+1}} = \log(2^\kappa) + \frac{2jB}{\log z}.
\]

We bound \(E\) by separating the sum appearing in \((BH.U1)\) into those summands with \(2^j \leq \sqrt{\log z}\) and those with \(2^j > \sqrt{\log z}\). The former summands contribute

\[
\leq 2^\kappa e^{B/\sqrt{\log z}} \sum_{j=1}^\infty \frac{(\log(2^\kappa) + B/\sqrt{\log z})^{b+2j-1}}{(b+2j-1)!} = 2^\kappa \sum_{j=1}^\infty \frac{(\log(2^\kappa))^{b+2j-1}}{(b+2j-1)!} + O_{b,B,\kappa} \left( \frac{1}{\sqrt{\log z}} \right),
\]

while the latter contribute

\[
\leq 2^\kappa e^{B/\log 2} \sum_{2^j > \sqrt{\log z}} \frac{(\log(2^\kappa) + B/\log 2)^{2j+1}}{(2j+1)!} \leq B_{\kappa} \frac{1}{\log z}.
\]

Therefore, we have the following:
Theorem BH.3. Assume \((g), (r)\) and \((\Omega)\), and let \(b\) be a positive, even integer. Then
\[
S(\mathcal{A}, \mathcal{P}, z) \geq XV(z) \left(1 - 2^a \sum_{j=1}^{\infty} \frac{\log(2^j)}{(b + 2j - 1)!} + O_{b,B,\kappa} \left(\frac{1}{\sqrt{\log z}}\right)\right) - \sum_{d \mid P(z), d \leq z^{2b+5}} |r_d|.
\]

Application: twin primes

First, as in Example (b), take \(\mathcal{A} = \{k(k+2) : 1 \leq k \leq x\}\), so that \(X = x\), \(g(p) = \rho(p)/p\) with \(\rho(2) = 1\), \(\rho(p) = 2\) for \(p > 2\), \((\Omega)\) holds with \(\kappa = 2\), \(V(z) \asymp 1/\log^2 z\). Choosing \(b = 4\), we see that
\[
1 - 4 \sum_{j=1}^{\infty} \frac{(\log 4)^{2j+3}}{(2j+3)!} = 1 - 4 \left(\sinh(\log 4) - \log 4 - \frac{(\log 4)^3}{6}\right) \geq 0.82,
\]
and therefore by Theorem BH.3, if \(z\) is large then
\[
S(\mathcal{A}, \mathcal{P}, z) \geq 0.8xV(z) - \sum_{d \leq z^{13}} \tau(d) \gg \frac{x}{\log^2 z} - O(z^{13}\log z).
\]

If we take \(z = x^{1/13-\epsilon}\), where \(\epsilon > 0\) is very small, we get \(S(\mathcal{A}, \mathcal{P}, z) \gg x/\log^4 x\), a result of type (L). We conclude that there are \(\gg x/\log^2 x\) integers \(n \leq x\), such that each of \(n\) and \(n+2\) have at most 13 prime factors.

Now we approach the problem as in Example (c) (this is A. R\'enyi’s approach), taking
\[
\mathcal{A} = \{p + 2 : p \text{ prime} \leq x - 2\},
\]
\(X = \text{li}(x)\), \(\mathcal{P}\) the set of all primes. Here
\[
|\mathcal{A}_d| = \pi(x; d, -2) = \frac{\text{li}(x)}{\phi(d)} + r_d,
\]
where \(r_d\) is expected to be small by the prime number theorem for arithmetic progressions. Taking \(g(d) = 1/\phi(d)\), we easily verify the \((\Omega)\) holds with \(\kappa = 1\). Indeed, for \(y \geq 3\), Mertens’ bounds give
\[
\prod_{y \leq p \leq w} (1 - g(p))^{-1} = \prod_{y \leq p \leq w} \left(1 - \frac{1}{p-1}\right)^{-1} = \prod_{y \leq p \leq w} \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{1}{(p-1)^2}\right)^{-1}
\leq \frac{\log w}{\log y} \left(1 + O\left(\frac{1}{\log y}\right)\right).
\]

For the error terms, we use the famous theorem of Bombieri and Vinogradov:

Theorem BV (Bombieri-A.I.Vinogradov, 1965). For every \(A > 0\) there is a \(B > 0\) so that
\[
\sum_{q \leq x^{1/2} \text{ (log x)}^{-B}} \max_{y \leq x} \max_{(a,q) = 1} \left|\pi(y; q, a) - \frac{\text{li}(y)}{\phi(q)}\right| \ll \frac{x}{(\log x)^A}.
\]

Take \(b = 2\) and \(z = x^{1/13-\epsilon}\), where \(\epsilon > 0\) is very small. Since \(z^{2b+5} = x^{1/2-9\epsilon}\), Theorem BV implies
\[
\sum_{d \leq z^{2b+5}} |r_d| \ll \frac{x}{(\log x)^4}.
\]
Also,

\[ 1 - 2 \sum_{j=1}^{\infty} \frac{(\log 2)^{2j+1}}{(2j+1)!} = 1 - 2(\sinh(\log 2) - \log 2) \geq 0.88, \]

and we conclude that for large \( x \),

\[ S(\mathcal{A}, \mathcal{P}, z) \geq 0.8XV(z) \gg \frac{x}{\log^2 x}. \]

Finally, we observe that \( S(\mathcal{A}, \mathcal{P}, z) \) counts primes \( p \) such that \( p + 2 \) has no prime factor \( \leq z \); in particular, \( p + 2 \) has at most 18 prime factors.

### 2.3 Asymptotic formula for the sifting function (Fundamental Lemma)

We combine the upper and lower bound sieve theorems to achieve an asymptotic formula for the sifting function when \( z \) is small compared with \( x \).

**Theorem BH.4** (Fundamental Lemma). Assume (\( g \)), (\( r \)) and that (\( \Omega \)) holds for some constants \( \kappa \geq 0 \) and \( B > 0 \). For any \( 2 \leq z \leq D \), we have

\[ S(\mathcal{A}, \mathcal{P}, z) = XV(z) \left( 1 + O_{\kappa, B} \left( s^{-s/3} \right) \right) + \Delta \sum_{d \leq z^{2b+4}/d!} |r_d|, \quad s = \frac{\log D}{\log z}, \quad |\Delta| \leq 1. \]

**Proof.** Let \( L = \log(2^\kappa) + B/\log 2 \), and \( s_0 = s_0(\kappa, B) = \max(81, (3eL)^4) \). When \( 1 \leq s \leq s_0 \), the lower bound is trivial, and (BH.U1) together with (E) and (D) give the upper bound

\[ S(\mathcal{A}, \mathcal{P}, z) \leq XV(z)e^E + \sum_{d \leq z^{2b+4}/d!} |r_d|, \]

where \( E \leq e^{2L}L^b/b! \). Taking \( b = 2 \) establishes the desired bound for \( 8 \leq s \leq s_0 \) since \( z^{2b+4} = z^8 \leq D \). When \( 1 \leq s < 8 \), we use the inequality \( S(\mathcal{A}, \mathcal{P}, z) \leq S(\mathcal{A}, \mathcal{P}, D^{1/8}) \) and the case \( s = 8 \) just proved.

Now assume that \( s > s_0 \). Apply (BH.U1) and (BH.L1) with \( b = 2 \left\lfloor \frac{1}{4}(s - 5) \right\rfloor \) (so that \( z^{2b+5} \leq z^s = D \)), together with (E) and (D). We obtain

\[ E \leq e^{2L}L^b/b! \leq \kappa, B \left( \frac{eL}{b} \right)^b \leq \left( \frac{2eL}{s - 9} \right)^{(s-9)/2} \leq \left( \frac{3eL}{s} \right)^{4s/9} \leq s^{-s/3} \]

and therefore

\[ S(\mathcal{A}, \mathcal{P}, z) = XV(z) \left( 1 + O(s^{-s/3}) \right) + \Delta \sum_{d \leq z^{2b+5}/d!} |r_d|, \]

where \( |\Delta| \leq 1. \)
3 Further Applications

3.1 Prime values of polynomials

Generalizing greatly the study of twin primes, we can use the sieve to study prime values of arbitrary polynomials.

**Theorem 1.** Let \( F_1, \ldots, F_k \) be distinct, irreducible polynomials in \( \mathbb{Z}[x] \), each with positive leading coefficient. Put \( F = F_1 \cdots F_k \), \( \ell = \deg(F) \). Suppose \( F \) has no fixed prime factor (i.e., for every prime \( p \) there is an \( n \) so that \( p \nmid F(n) \)). Then

\[
\# \{ n \leq x : F_i(n) \text{ prime for every } i \} \ll_F \frac{x}{\log^k x}.
\]

Further, there is an integer \( m \), depending only on \( \ell \), such that for large \( x \),

\[
\# \{ n \leq x : \Omega(F_i(n)) \leq m \ (1 \leq i \leq k) \} \gg_F \frac{x}{\log^k x}.
\]

**Proof.** We will show the upper bound (3.1), leaving (3.2) as an exercise. Let \( x \) be large and

\[ a_n = \# \{ 1 \leq k \leq x : F(k) = n \} \]

(i.e., \( a_n \) is like the “indicator” function of the multiset \( \mathcal{A} = \{ F(k) : 1 \leq k \leq x \} \)), \( \mathcal{P} \) the set of all primes, \( X = x, z = \sqrt{x} \). Generically write

\[
\rho_G(d) = \# \{ 0 \leq n < d : G(n) \equiv 0 \pmod{d} \},
\]

where \( G \) is any polynomial. As with twin primes, we have

\[
|\mathcal{A}_d| = \frac{\rho_F(d)}{d} X + r_d, \quad |r_d| \leq \rho_F(d),
\]

and so we set \( g(d) = \rho_F(d)/d \). Now \( \rho_F \) is multiplicative by the Chinese remainder theorem. Also,

\[
\rho_F(p) < p \quad \text{by hypothesis},
\]

\[
\rho_F(p) \leq \ell \quad \text{by Lagrange’s theorem}.
\]

Thus, if \( \theta < 1 \) is fixed, then

\[
\sum_{d \leq x^\theta} \sum_{d \mid P(z)} |r_d| \leq \sum_{d \leq x^\theta} d g(d) \leq x^\theta \sum_{d \mid P(z)} g(d) = x^\theta \prod_{p \leq z} (1 + g(p)) \leq x^\theta \prod_{p \leq z} \left(1 + \frac{\ell}{p}\right) \ll \ell \frac{x^\theta (\log x)^\ell}{(\log x)^{p\ell}}.
\]

Therefore, \((R(\theta, \ell)) \) holds. Since \( g(p) \leq \ell/p \) for all \( p \), by Proposition g, \((\Omega) \) holds with \( \kappa = \ell \) and some constant \( B \) (which depends only on \( \ell \), although we make no use of this fact). By Theorem BH.2,

\[
S(\mathcal{A}, \mathcal{P}, z) \ll XV(z).
\]
As before, the left side is at least as large as the count of \( k \leq x \) such that each \( F_i(k) \) is prime and \( > z \) (and there are \( O_F(\sqrt{x}) \) values of \( k \) with each of \( F_i(k) \) prime, and one of them is \( \leq z \)). It remains to bound \( V(z) \) from above. To do this, we need the following two facts:

\[
\rho_F(p) = \rho_{F_1}(p) + \cdots + \rho_{F_k}(p) \quad \text{for all but finitely many } p, \\
\sum_{p \leq x} \frac{\rho_{F_i}(p)}{p} = \log \log x + O_F(1) \quad (1 \leq i \leq k).
\]

From (3.3) and (3.4) we deduce that

\[
\log V(z) = \sum_{p \leq z} \log \left( 1 - \frac{\rho_{F}(p)}{p} \right) = O_F(1) - \sum_{p \leq z} \frac{\rho_{F}(p)}{p} = -k \log \log z + O_F(1),
\]

and thus \( V(z) \sim_F (\log z)^{-k} \). In conclusion,

\[
\# \{ k \leq x : F_i(k) \text{ prime for every } i \} \leq O_F(\sqrt{x}) + S(\mathcal{A}, \mathcal{P}, z) \ll_F \frac{x}{(\log x)^k}.
\]

Proof of (3.3): Suppose the equation in (3.3) does not hold. Then there are \( i \neq j \) and some \( n \) with \( p|F_i(n) \) and \( p|F_j(n) \), i.e., \( p|(F_i(n), F_j(n)) \). As \( F_i \) and \( F_j \) are distinct, irreducible and have no fixed prime factor (in particular, neither is a multiple of the other), \( (F_i, F_j) = 1 \) over \( \mathbb{Q}[x] \). Hence, there are \( G, H \in \mathbb{Q}[x] \) such that \( F_iG + F_jH = 1 \). Clearing denominators gives \( \tilde{G}, \tilde{H} \in \mathbb{Z}[x] \) with \( F_i\tilde{G} + F_j\tilde{H} = C_{ij}, \) where \( C_{ij} \in \mathbb{Z} \). It follows that \( p|C_{ij} \). As there are only finitely many pairs \( i, j \), there are finitely many possible \( p \).

Proof of (3.4): this is a consequence of the Prime Ideal Theorem of Landau. In some special cases, it follows from Mertens theorem for primes in arithmetic progressions, e.g. if \( F(x) = x^2 + 1 \), then

\[
\sum_{p \leq x} \frac{\rho_{F}(p)}{p} = 1 + 2 \sum_{p \leq x \ (p=1 \mod 4)} \frac{1}{p} = \log \log x + O(1).
\]

\[
3.2 \text{ Buchstab’s function}
\]

Buchstab’s function

\[
\Phi(x, z) = \# \{ n \leq x : P^-(n) > z \}
\]

is perhaps the most basic sieve function.

**Theorem Bu.** (i) Uniformly for \( x \geq z \geq 2 \), we have

\[
\Phi(x, z) \ll \frac{x}{\log z}.
\]

(ii) Uniformly for \( x \geq 2z \geq 4 \), we have

\[
\Phi(x, z) \gg \frac{x}{\log z}.
\]
(iii) We have
\[ \Phi(x, z) \sim e^{-\gamma} \frac{x}{\log z} \quad (z \to \infty, \ z = x^{o(1)}) \]

Proof. Exercise. \qed

Later we will revisit Buchstab’s function and establish an asymptotic (using different methods) when \( \log x/\log z \) is bounded.

3.3 Primitive roots of primes

Are there infinitely many primes \( p \) for which the fraction \( 1/p \) has period \( p - 1 \) in base 10? Equivalently, is 10 a primitive root of \( p \) for infinitely many primes \( p \). This question was addressed by Gauss in his Disquisitiones Arithmeticae (1801). In 1927, E. Artin conjectured an asymptotic formula for the number of primes \( p \leq x \) for which a given squarefree integer \( a \) is a primitive root; his formula was later discovered not to hold up against numerical data by D. H. Lehmer, and a revised conjecture was later proposed by H. Heilbronn and others. In 1967, C. Hooley deduced the revised conjecture from the Generalized Riemann Hypothesis for certain Dedekind zeta functions [3]; see also [4, Ch. 3]. The theorem makes use of sieve methods, in particular the general sieve from §1.3. Here we will specialize to the case \( a = 2 \).

Theorem H (Hooley [3]). Assume the Generalized Riemann Hypothesis for \( \zeta_K(s) \) for the number fields \( K = \mathbb{Q}(\sqrt[4]{2}, e^{2\pi i/k}), k \) running over all squarefree integers. Then
\[ \#\{p \leq x : 2 \text{ is a primitive root of } p\} \sim C \pi(x), \quad C = \prod_{q \text{ prime}} \left( 1 - \frac{1}{q(q-1)} \right). \]

Proof. We set this up as a general sieve problem as follows. For primes \( p \) and \( q \), let \( B_q(p) = 1 \) if
\[ q \mid p - 1 \quad \text{and} \quad 2^{(p-1)/q} \equiv 1 \pmod{p}, \]
and \( B_q(p) = 0 \) otherwise. As in §1.3, we extend the definition by defining
\[ B_d(p) = \prod_{q \mid d} B_q(p) \quad (\text{squarefree } d). \]
Clearly, 2 is a primitive root of \( p \) (order \( p - 1 \)) if and only if \( B_q(p) = 0 \) for all primes \( q \) (and it suffices to check for primes \( q \leq p \)). Let
\[ \mathcal{A} = \{p \leq x : p \text{ prime}\}, \quad \mathcal{P} = \text{ all primes}. \]
We need good, uniform bounds for \( |\mathcal{A}_d| \), where \( \mathcal{A}_d = \{p \in \mathcal{A} : B_d(p) = 1\} \). Bounds for individual \( d \), but with poor uniformity, are available unconditionally by the analog of the prime number theorem for the number fields \( K \) (the prime ideal theorem of Landau), but the uniformity is very poor, especially for \( x^{o(1)} < d \leq \sqrt{x} \), and not good enough for our application.

Lemma 3.1 (Hooley). Assume the Generalized Riemann Hypothesis for \( \zeta_K(s) \) for the number fields \( K = \mathbb{Q}(\sqrt[4]{2}, e^{2\pi i/k}), k \) running over all squarefree integers. Then, uniformly for squarefree \( d \leq x \),
\[ |\mathcal{A}_d| = \frac{\text{li}(x)}{d\phi(d)} + O(\sqrt{x} \log x). \]
Define the cutoffs
\[
z_1 = \frac{1}{6} \log x, \quad z_2 = \sqrt{x (\log x)}^{-2}, \quad z_3 = \sqrt{x \log x}.
\]
Evidently,

\[
S(\mathcal{A}, \mathcal{P}, z_1) \geq \# \{ p \leq x : 2 \text{ is a primitive root of } p \} \geq S(\mathcal{A}, \mathcal{P}, z_1) - S_1 - S_2 - T,
\]
where
\[
S_1 = \sum_{z_1 < q \leq z_2} |\mathcal{A}_q|, \quad S_2 = \sum_{z_2 < q \leq z_3} |\mathcal{A}_q|, \quad T = \# \{ p \leq x : B_q(p) = 1 \text{ for some } q > z_3 \}.
\]

First, by the Legendre sieve (SI') and that \( P(z_1) = x^{1/6+o(1)} \leq x, \)
\[
S(\mathcal{A}, \mathcal{P}, z_1) = \sum_{d | P(z_1)} \mu(d)|\mathcal{A}_d|
= \sum_{d | P(z_1)} \mu(d) \left( \frac{\text{li}(x)}{d \phi(d)} + O(\sqrt{x \log x}) \right)
= \text{li}(x) \prod_{q \leq z_1} \left( 1 - \frac{1}{q(q-1)} \right) + O(\sqrt{x (\log x) 2^{\pi(z_1)})}
= \left( C + O \left( \frac{1}{\log x} \right) \right) \text{li}(x) = \left( C + O \left( \frac{1}{\log x} \right) \right) \pi(x).
\]

For \( S_1, \) we again use Lemma 3.1 and obtain
\[
S_1 = \sum_{z_1 < q \leq z_2} \left( \frac{\text{li}(x)}{q(q-1)} + O(\sqrt{x \log x}) \right)
\ll \frac{\text{li}(x)}{\log x} + \sqrt{x (\log x) \pi(z_2)} \ll \frac{\text{li}(x)}{\log x}.
\]

For \( S_2, \) the bounds from Lemma 3.1 are too poor, but we do better by observing that \( |\mathcal{A}_q| \leq \pi(x; q, 1) \) and using Theorem BT (Brun-Titchmarsh inequality):
\[
S_2 \ll \sum_{z_2 < q \leq z_3} \pi(x; q, 1) \ll \sum_{z_2 < q \leq z_3} \frac{x}{q \log x} \ll \frac{x}{\log^2 x} \sum_{z_2 < q \leq z_3} \frac{\log q}{q} \ll \frac{x \log \log x}{\log^2 x}.
\]

Finally, if \( B_q(p) = 1 \) for some \( q > z_3, \) then \( p | (2^m - 1) \) for some positive integer \( m \leq \sqrt{x} / \log x. \) Thus, \( p | M, \) where
\[
M = \prod_{m \leq \sqrt{x} / \log x} (2^m - 1).
\]
Thus,
\[
T \leq \# \{ p | M : p \text{ prime} \} \leq \frac{\log M}{\log 2} \leq \left( \frac{\sqrt{x}}{\log x} \right)^2 = \frac{x}{\log^2 x}.
\]
Combining the estimates for \( S(\mathcal{A}, \mathcal{P}, z_1), S_1, S_2 \) and \( T \) with (3.5) proves the theorem. \( \square \)
References


Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

E-mail address: ford@math.uiuc.edu