simpler function t. This gives an error or "remainder term" of $\{t\}$, the fractional part. More precisely, we have

$$\pi(x) - \pi(\sqrt{x}) + 1 = x \sum_d \frac{\mu(d)}{d} + R = x \prod_{p \leqslant \sqrt{x}} \left(1 - \frac{1}{p}\right) + R$$

where d runs over the products of distinct primes which are less than or equal to $\sqrt{x}$ and where the "remainder" $R = R(x)$ is

$$R = -\sum_d \mu(d) \left\{\frac{x}{d}\right\}.$$

At first glance, the best we can expect to do is to use the trivial bound $\{t\} < 1$ which leads us to bound the remainder by

$$|R| \leqslant \sum_d 1 = 2^{\pi(\sqrt{x})},$$

which is absolutely enormous, much larger even than the number of integers $[x]$ with which we began. Of course, we have been particularly careless here, for example, fruitlessly sifting out multiples of d even for many integers d which exceed x. Hence, the above bound can certainly be improved somewhat. But not enough! In reality R is genuinely large. In fact, using old ideas due to Tchebyshev and to Mertens (see (2.21)), one knows that

$$\prod_{p \leqslant \sqrt{x}} \left(1 - \frac{1}{p}\right) \sim \frac{2e^{-\gamma}}{\log x},$$

so what we have been expecting to be our main term is actually wrong. Since, by the Prime Number Theorem,

$$\pi(x) \sim \frac{x}{\log x},$$

we see that the quantity R we have been referring to as the remainder has order of magnitude just as large as the main term.

1.2. Some Generality

At the moment we are in the rather depressing position of having a method which fails to give us good estimates for the number $\pi(x)$ of primes up to x, but worse yet, the only reason we even know that it must inevitably fail is because other techniques, coming from analytic number theory, succeed (in proving the Prime Number Theorem), thereby telling us so. What possibility remains for the value of the sieve is its capacity for generalization, giving some information in cases where the finer analytic machinery is lacking. Therefore, to consider the situation more generally is not merely a useful enterprise; it is the sieve's only excuse for being.

We thus consider a finite sequence of non-negative real numbers

$$\mathcal{A} = (a_n), \ n \leqslant x,$$

and a general set $\mathcal{P}$ of primes. For notational purposes it is convenient to introduce the product

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

Our goal will be to estimate the "sifting function"

$$S(\mathcal{A}, z) = \sum_{\substack{n \leqslant x \\ (n, P(z))=1}} a_n.$$

We shall proceed as in our original example, but in slightly different notation. We need the most basic property of the Möbius function

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases} \tag{1.1}$$

We also use the simple observation that δ divides both a and b if and only if δ divides (a, b), that is, the set of common divisors of two positive integers is just the same as the set of divisors of their greatest common divisor. Hence,

$$\sum_{\substack{d|n \\ d|P(z)}} \mu(d) = \begin{cases} 1 & \text{if } (n, P(z)) = 1, \\ 0 & \text{if } (n, P(z)) > 1. \end{cases}$$

Inserting this coprimality detector and then interchanging the order of summation, we obtain

$$\begin{aligned} S(\mathcal{A}, z) &= \sum_n a_n \sum_{d|(n,P(z))} \mu(d) = \sum_n a_n \sum_{\substack{d|n \\ d|P(z)}} \mu(d) \\ &= \sum_{d|P(z)} \mu(d) \sum_{n \equiv 0 \,(\mathrm{mod}\, d)} a_n = \sum_{d|P(z)} \mu(d) A_d(x), \end{aligned}$$

say. This is just once again, but in more general garb, the Legendre formula and here, as before, we need information about the "congruence sums"

$$A_d(x) = \sum_{\substack{n \leqslant x \\ n \equiv 0 \,(\mathrm{mod}\, d)}} a_n$$

which give the mass of the subsequence running over the multiples of d, in other words, over the subsequence $\mathcal{A}_d = (a_m)\,,\ m \equiv 0 \,(\mathrm{mod}\, d)$, and which, in our beginning example, was just $[x/d]$. Specifically, we need a useful approximation formula. We assume we can write this sum in the form

$$A_d(x) = g(d)X + r_d(x)\ , \tag{1.2}$$

where X is a handy approximation to

$$A(x) = A_1(x) = \sum_{n \leqslant x} a_n\ ,$$

the total mass of our sequence, where $g(d)$, the "density function", has several nice properties ($g(d) = 1/d$ in our example) and $r_d(x)$ is a "remainder term" which is small, at least on average over d (this was $r_d(x) = -\{x/d\} \ll 1$ in our example). In other words, we need to count integers before we can count primes, and how well we count the first determines how well we count the second.

On insertion of our approximation formula (1.2) the sifting function becomes

$$S(\mathcal{A}, z) = X \sum_{d|P(z)} \mu(d) g(d) + \sum_{d|P(z)} \mu(d) r_d(x)\ .$$

The function $g(d)$ acts like a probability, approximating the fraction of the total mass which resides in the multiples of d. (It is useful to remember $g(d) = 1/d$ as being the prototype for such a function.) Because of this it is natural that we assume that $g(1) = 1$ and that, for each $d > 1$, we have $0 \leqslant g(d) < 1$. This last inequality will be needed in some places and is also natural in our situation. Indeed, if for some $d > 1$ we had $g(d) = 1$, then virtually everything would be a multiple of d. There would not be much point in looking for primes in such a sequence. We shall also assume that g is a multiplicative function, that is, whenever $(d_1, d_2) = 1$ we have

$$g(d_1 d_2) = g(d_1)g(d_2).$$

The essence of this is that we are assuming that divisibility by two relatively prime integers are independent events. In practice, this is true only to a quite limited extent and this fact is in large measure responsible for the failure of the sieve to achieve more than it has.

Very often we only use (1.2) for d squarefree, but sometimes it helps to assume (1.2) for all d. Because $\mathcal{A}_d$ is a subsequence of $\mathcal{A}$ it is natural to assume that

$$g(d_1) \leqslant g(d_2) \tag{1.3}$$

if $d_2|d_1$. In particular, $g(p^\ell)$ is non-increasing in ℓ for any p.

1.3. Some Examples

We consider some examples. In many of the most basic examples the sequence $\mathcal{A}$ is just the characteristic function of an interesting set of integers. In such a case we shall frequently not bother to distinguish between the function and the set on which it is supported.

Example 1.1. We begin with a slight extension of our original example to the set of integers in an interval. Thus, we consider

$$\begin{aligned} \mathcal{A} &= \{m \mid x - y < m \leqslant x\}, \qquad \mathcal{P} = \{\text{all primes}\}, \\ A_d(x) &= \left[\frac{x}{d}\right] - \left[\frac{x-y}{d}\right], \qquad X = y, \\ g(d) &= \frac{1}{d}, \quad r_d(x) = -\left\{\frac{x}{d}\right\} + \left\{\frac{x-y}{d}\right\}, \quad |r_d(x)| \leqslant 1. \end{aligned}$$

Example 1.2. Now, for a little more variety, consider

$$\begin{aligned} \mathcal{A} &= \left\{m^2 + 1 \leqslant x\right\}, \quad \mathcal{P} = \{p;\ p \not\equiv 3 \pmod 4\}, \\ A(x) &= \left[\sqrt{x-1}\right], \qquad X = \sqrt{x}, \\ g(p) &= \begin{cases} 2/p, & p \equiv 1 \pmod 4, \\ 1/2, & p = 2, \end{cases} \qquad |r_d(x)| \leqslant 2^{\nu(d)}, \end{aligned}$$

this last estimate following from the bound $|r_p(x)| \leqslant 2$ and the Chinese Remainder Theorem. Here, there is no need to sieve by the primes congruent to three modulo four since none of the integers in our set is divisible by such a prime. Equivalently, we could achieve the same results sifting by the set of all primes and simply setting $g(p) = 0$ for the additional primes. In this example, were we able to get a positive lower bound for $S(\mathcal{A}, \sqrt{x})$ we would be producing primes of the form m^2+1. A proof

that there are infinitely many such primes would settle an outstanding problem in the subject.

EXAMPLE 1.3. Moving to another famous conjecture, we consider the following:

$$\mathcal{A} = \{m(m+2) \leqslant x\}, \qquad \mathcal{P} = \{\text{all primes}\},$$
$$g(p) = \begin{cases} 2/p\ , & p \text{ odd}, \\ 1/2\ , & p = 2, \end{cases} \qquad |r_d(x)| \leqslant 2^{\nu(d)}.$$

Here, were we to give a positive lower bound for $S\left(\mathcal{A}, x^{1/4}\right)$ we would be producing integers $m(m+2)$ where both factors are prime and differ by two. The twin prime conjecture predicts that there are infinitely many such pairs.

EXAMPLE 1.4. As an alternative approach via the sieve to attack the twin prime conjecture, we consider the sequence:

$$\mathcal{A} = \{p-2;\ p \leqslant x\}, \qquad \mathcal{P} = \{\text{odd primes}\},$$
$$A_d(x) = \pi(x; d, 2), \qquad X = \pi(x), \qquad g(d) = \frac{1}{\varphi(d)},$$

where $\pi(x; d, a)$ is the number of primes up to x which are congruent to a modulo d and where $\varphi(d)$, the Euler function, counts the number of reduced residue classes modulo d. This sequence offers some advantages over the previous one for studying the twin prime problem and it gives stronger results in that direction, although this was not so in the earliest results. The most obvious advantage is that we are starting from the beginning with the knowledge that one of our two numbers, namely p, is a prime. On the other hand, the remainder term is more complicated, namely $r_d(x) = \pi(x; d, 2) - \pi(x)/\varphi(d)$, and it is much more difficult to bound it successfully. In the current state of knowledge, a reasonably good bound can only be given on average over d; the most powerful bound of this type being the Bombieri–Vinogradov theorem which we shall prove in Section 9.18. Again in this example, if we were to be successful in giving a positive lower bound, this time for $S(\mathcal{A}, \sqrt{x})$, then we would be producing twin primes.

There are considerable generalizations to all of the above examples. One may take a polynomial with integer coefficients, say in one variable (although not necessarily so), and consider $\mathcal{A}$ to be the sequence of its values as the variable runs through the integers in a segment, or the primes in a segment, or the primes in a segment of an arithmetic progression. It is possible to give many other cases wherein well–known problems concerning primes, for instance the Goldbach conjecture that every even integer exceeding 2 is the sum of two primes, can be phrased in such a manner as to follow from sufficiently strong sieve-theoretic estimates. Formulating them this way is, however, by far the easier part of the problem; producing successful estimates is a very stern challenge indeed!

Not always will the target of the sieve be a set of primes. Perhaps the simplest case is the following.

EXAMPLE 1.5. Let

$$\mathcal{A} = \{m \leqslant x\}, \quad \mathcal{P} = \{p;\ p \mid q\},$$

where q is a given positive integer. Now the target is the set of integers $m \leqslant x$ with $(m, q) = 1$. By the inclusion-exclusion argument in Section 1.1 the number of such integers is

$$\sum_{d|q} \mu(d)\Big[\frac{x}{d}\Big] = \frac{\varphi(q)}{q}x + R$$

where

$$R = -\sum_{d|q} \mu(d)\Big\{\frac{x}{d}\Big\}$$

so $|R| \leqslant \tau(q)$. Note that, in the event that x is an integral multiple of q, the remainder terms all vanish so we have an exact formula. We shall never be so lucky again.

The precision in the above example comes from the fact that the sifting set of primes is fixed. If instead we allow q to grow the problem again becomes difficult. In two very basic situations we may take q to be the product of the primes $p < z$ in which case we have simply rephrased our original problem or we may take q to be the product of primes $p > z$ in which case we are counting integers without large prime factors. We shall touch on the asymptotics for these examples much later, in Section 12.2.

Another important sequence that appears as the target of a sieve is the set of squarefree numbers. In this scenario, rather than sieving by a set of primes we sift by squares of primes. For this purpose the basic formula is

$$\sum_{d^2|n} \mu(d) = \begin{cases} 1 & \text{if } n \text{ is squarefree,} \\ 0 & \text{otherwise.} \end{cases}$$

Sieving them from the set of all positive integers is easy, so we consider a somewhat more ambitious problem.

Example 1.6. Let

$$\mathcal{A} = \big\{m^2 + 1 \leqslant x\big\}\ , \qquad \mathcal{P} = \big\{p \equiv 1 \,(\bmod\, 4)\big\}\ .$$

Then, by the corresponding Legendre formula we have

$$\Big|\big\{m^2 + 1 \leqslant x;\ m^2 + 1 \text{ squarefree}\big\}\Big| = \sum_{d|P(x)} \mu(d) A_{d^2}(x)\ .$$

Splitting into classes $m \equiv \nu \,(\bmod\, d^2)$ we find that

$$A_{d^2}(x) = \frac{\rho(d^2)}{d^2}\sqrt{x} + O\big(\rho(d^2)\big)$$

where $\rho(q)$ is the number of solutions to the congruence $\nu^2 + 1 \equiv 0 \,(\bmod\, q)$.

We have $\rho(d^2) = \rho(d) = \tau(d)$ for $d \mid P(x)$, so the above approximation is only good for $d \leqslant x^{\frac{1}{4}}$. For larger d the remainder term is too large. In reality $\mathcal{A}_{d^2}$ is frequently empty. To treat the contribution from larger d, say $d > D$, we change the role of the variables. We write

$$m^2 + 1 = d^2 k \leqslant x\ ,$$

and we estimate the number of solutions in m, d for every given $k \leqslant K = xD^{-2}$. For fixed k this reduces to the counting of units in the real quadratic field $\mathbb{Q}(\sqrt{k})$.

Since the units (solutions to Pell's equation) grow exponentially, the number of these in the relevant range is $O(\log x)$. Therefore,

$$\sum_{d|P(x)} \mu(d) A_{d^2}(x) = \sum_{\substack{d|P(x)\\ d\leqslant D}} \mu(d) \left\{ \frac{\tau(d)}{d^2}\sqrt{x} + O\big(\tau(d)\big) \right\} + O(K \log x)$$
$$= \sqrt{x} \prod_{p\in\mathcal{P}} \Big(1 - \frac{2}{p^2}\Big) + O\left(\Big(D + \frac{x}{D^2}\Big)\log x\right) .$$

Choosing $D = x^{\frac{1}{3}}$ we conclude that

$$\Big|\big\{m^2 + 1 \leqslant x;\ m^2 + 1 \text{ squarefree}\big\}\Big| = cx^{\frac{1}{2}} + O\big(x^{\frac{1}{3}} \log x\big)$$

where c is a positive constant given by

$$c = \prod_{p\equiv 1 (\mathrm{mod}\, 4)} \Big(1 - \frac{2}{p^2}\Big) .$$

We remark that the more general asymptotic formula for $\ell-$free values of degree ℓ polynomials was established by G. Ricci [**139**] in 1933.

In our next example the target is the set of integers which can be written as the sum of two squares. The result is cleaner if we restrict to odd integers $a^2 + b^2$ with a, b relatively prime.

Example 1.7. Here we take

$$\mathcal{A} = \big\{n \leqslant x;\ n \equiv 1 \pmod 4\big\},$$
$$\mathcal{P} = \big\{p;\ p \equiv 3 \pmod 4\big\},$$

getting

$$S(\mathcal{A}, \mathcal{P}, \sqrt{x}) = \Big|\big\{n \leqslant x;\ n = a^2 + b^2 \text{ odd with } (a, b) = 1\big\}\Big| .$$

This problem (of estimating $S(\mathcal{A}, \mathcal{P}, z)$) is intermediate in difficulty between the squarefree numbers and the primes. Whereas the first was easy enough to do right here, the second we shall solve in Chapter 14, but only with techniques that are well advanced.

As we have seen in the last few examples, there are many variations on our original sieve format. An important one deals with the situation where we want to sift out many residue classes for each prime in our set $\mathcal{P}$. In essence, Example 1.3 deals with two residue classes for each odd prime. Now we present a problem in which we wish to remove a great many more.

Example 1.8. Let

$$\mathcal{A} = \big\{n \leqslant x\big\},$$
$$\mathcal{P} = \text{the set of all primes},$$
$$\Omega_p = \left\{\omega \,(\mathrm{mod}\, p);\ \omega = 0 \text{ or } \Big(\frac{\omega}{p}\Big) = 1\right\} .$$

Note that the number of classes to be removed $\omega(p) = \frac{1}{2}(p + 1)$, if $p > 2$, is very large. The problem is to estimate

$$S(\mathcal{A}, \mathcal{P}, \Omega) = \Big|\big\{n \leqslant x;\ n \,(\mathrm{mod}\, p) \notin \Omega_p \text{ for each } p \leqslant \sqrt{x}\big\}\Big| .$$

This counts the integers $n \leqslant x$ which are quadratic non-residues for all primes $p \leqslant \sqrt{x}$. This is the problem which gave rise to the large sieve. We shall develop the large sieve theory in Chapter 9 and apply it to obtain the upper bound

$$S(\mathcal{A}, \mathcal{P}, \Omega) \ll \sqrt{x} \ .$$

In slightly more general form this idea will be used to estimate the least quadratic non-residue.

Finally, we just mention:

EXAMPLE 1.9. Let

$$F = F_1 \dots F_r$$

be the product of irreducible polynomials with integer coefficients and take $\mathcal{A}$ to be the sequence of values $F(n)$ or $F(p)$. In this case, if $r > 1$ there is no chance to find primes, but it might be possible to find integers with r prime factors, and it is interesting to see how close one can come to this goal. Here, in general, it is not so simple to describe precisely the density function g.

1.4. A Model of a Sifting Sequence for a Given Density

In this section we are going to create a sequence $\mathcal{B} = \big(b_n\big)$ which satisfies the sieve axioms for a given density function g of dimension $\kappa = 1$; see Section 5.5. Naturally, we assume that $g(d)$ is multiplicative with

$$0 \leqslant g(p) < 1 \tag{1.4}$$

and

$$g(p^\alpha) \geqslant g(p^{\alpha+1}) \geqslant 0 \ , \tag{1.5}$$

for any prime p and any $\alpha \geqslant 0$. Given such a function g, we define the companion function h, which is multiplicative with

$$h(p^\alpha) = \frac{g(p^\alpha) - g(p^{\alpha+1})}{1 - g(p)} \ . \tag{1.6}$$

We call h the relative density function. Note that if $g(d)$ is completely multiplicative then $h(d) = g(d)$. If $g(d)$ is supported on squarefree numbers, then so is $h(d)$ and

$$h(p) = \frac{g(p)}{1 - g(p)} \ , \tag{1.7}$$

hence the name "relative density function".

For simplicity we assume that the function

$$f(n) = h(n)n \tag{1.8}$$

satisfies the conditions of Lemma A.15, that is, (A.78), (A.82) and (A.83). Hence we obtain

$$\sum_{\substack{n \leqslant x \\ n \equiv 0 (\mathrm{mod}\, d)}} f(n) = \hat{f}(d)x + O\Big(\tilde{f}(d)(x/d)^\theta\Big) \ , \tag{1.9}$$

where $\hat{f}$ is given by (A.91) and $\tilde{f}$ by (A.90).