Introduction to Analytic Number Theory

Math 531 Lecture Notes, Fall 2005

A.J. Hildebrand
Department of Mathematics
University of Illinois

http://www.math.uiuc.edu/~hildebr/ant

Version 2013.01.07

Chapter 2

Arithmetic functions II: Asymptotic estimates

The values of most common arithmetic functions $f(n)$, such as the divisor function $d(n)$ or the Moebius function $\mu(n)$, depend heavily on the arithmetic nature of the argument n. As a result, such functions exhibit a seemingly chaotic behavior when plotted or tabulated as functions of n, and it does not make much sense to seek an "asymptotic formula" for $f(n)$.

However, it turns out that most natural arithmetic functions are very well behaved *on average*, in the sense that the arithmetic means $M_f(x) = (1/x)\sum_{n\le x} f(n)$, or, equivalently, the "summatory functions" $S_f(x) = \sum_{n\le x} f(n)$, behave smoothly as $x \to \infty$ and can often be estimated very accurately. In this chapter we discuss the principal methods to derive such estimates. Aside from the intrinsic interest of studying the behavior of $M_f(x)$ or $S_f(x)$, these quantities arise naturally in a variety of contexts, and having good estimates available is crucial for a number of applications. Here are some examples, all of which will be discussed in detail later in this chapter.

(1) The number of Farey fractions of order Q, i.e., the number of rationals in the interval $(0, 1)$ whose denominator in lowest terms is $\le Q$, is equal to $S_\phi(Q)$, where $S_\phi(x) = \sum_{n\le x} \phi(n)$ is the summatory function of the Euler phi function.

(2) The "probability" that two randomly chosen positive integers are coprime is equal to the limit $\lim_{x\to\infty} 2S_\phi(x)/x^2$, which turns to be $6/\pi^2$.

(3) The "probability" that a randomly chosen positive integer is squarefree

is equal to the "mean value" of the function $\mu^2(n)(= |\mu(n)|)$, i.e., the limit $\lim_{x\to\infty} M_{\mu^2}(x)$, which turns out to be $6/\pi^2$.

(4) More generally, if $f_A(n)$ is the characteristic function of a set $A \subset \mathbb{N}$, then the mean value $\lim_{x\to\infty} M_{f_A}(x)$ of f_A, if it exists, can be interpreted as the "density" of the set A, or the "probability" that a randomly chosen positive integer belongs to A.

(5) The Prime Number Theorem is equivalent to the relation $\lim_{x\to\infty} M_\Lambda(x) = 1$, which can be interpreted as saying that the function $\Lambda(n)$ is 1 on average. It is also equivalent to the relation $\lim_{x\to\infty} M_\mu(x) = 0$, which says essentially that a squarefree number is equally likely to have an even and an odd number of prime factors.

Notational conventions. Unless otherwise specified, x denotes a real number, and by $\sum_{n\le x}$ we mean a summation over all *positive* integers n that are less than or equal to x. (In those rare cases where we want to include the term $n = 0$ in the summation, we will indicate this explicitly by writing $\sum_{0\le n\le x}$ or $\sum_{n=0}^{[x]}$.)

Given an arithmetic function f, we let $S_f(x) = \sum_{n\le x} f(n)$ denote the associated summatory function, and $M_f(x) = S_f(x)/x$ the associated finite mean values. Note that if x is a positive integer, then $M_f(x)$ is just the arithmetic mean of the first $[x]$ values of f.

2.1 Big oh and small oh notations, asymptotic equivalence

2.1.1 Basic definitions

A very convenient set of notations in asymptotic analysis are the so-called "O" ("Big oh") and "o" ("small oh") notations, and their variants. The basic definitions are as follows, where $f(x)$ and $g(x)$ are functions defined for all sufficiently large x.

(1) **"Big Oh" estimate:** "$f(x) = O(g(x))$" means that there exist constants x_0 and c such that $|f(x)| \le c|g(x)|$ for all $x \ge x_0$.

(2) **"Small Oh" estimate:** "$f(x) = o(g(x))$" means that $g(x) \ne 0$ for sufficiently large x and $\lim_{x\to\infty} f(x)/g(x) = 0$.

(3) **Asymptotic equivalence:** "$f(x) \sim g(x)$" means that $g(x) \neq 0$ for sufficiently large x and $\lim_{x\to\infty} f(x)/g(x) = 1$.

(4) **Vinogradov's $\ll$ notation:** "$f(x) \ll g(x)$" is equivalent to "$f(x) = O(g(x))$"; we also write $f(x) \gg g(x)$ if $g(x) \ll f(x)$.

(5) **Order of magnitude estimate:** "$f(x) \asymp g(x)$" means that both $f(x) \ll g(x)$ and $f(x) \ll g(x)$ hold; this is easily seen to be equivalent to the existence of positive constants c_1 and c_2 and a constant x_0 such that $c_1|g(x)| \leq |f(x)| \leq c_2|g(x)|$ holds for $x \geq x_0$. If $f(x) \asymp g(x)$, we say that f and g have **the same order of magnitude**.

An **asymptotic formula for** $f(x)$ means a relation of the form $f(x) \sim g(x)$. An **asymptotic estimate for** $f(x)$ is any estimate of the form $f(x) = g(x)+O(R(x))$ (which is to be interpreted as "$f(x)-g(x) = O(R(x)$"), where $g(x)$ is a main term and $O(R(x))$ an error term.

Note that an asymptotic formula $f(x) \sim g(x)$ is only meaningful if the approximating function $g(x)$ is, in some sense, "simpler" than the function $f(x)$ one seeks to approximate. Similarly, an asymptotic estimate $f(x) = g(x)+O(R(x))$ is only meaningful if the error term, $R(x)$, is of smaller order than the main term $g(x)$ in the approximation to $f(x)$, in the sense that $R(x)$ satisfies $R(x) = o(g(x))$. (If $R(x)$ were of the same, or larger, order than $g(x)$, then the above estimate would be equivalent to $f(x) = O(R(x))$, and there would be no point in leaving $g(x)$ in this estimate.)

2.1.2 Extensions and remarks

There exists several extensions and variants of the basic above notations defined above:

More general ranges and limits. "$f(x) = O(g(x))$ $(0 \leq x \leq 1)$" means that the estimate holds in the range $0 \leq x \leq 1$, rather than a range of the form $x \geq x_0$. Similarly, "$f(x) = o(g(x))$ $(x \to 0)$" means that the limit in the definition of the "oh" notation is taken when $x \to 0$ rather than $x \to \infty$. By convention, if no explicit range is given, the range is understood to be of the form $x \geq x_0$, for some x_0. Similarly, the limit implicit in a o-estimate is understood to be the limit as $x \to \infty$, unless a different limit is explicitly given.

Dependence on parameters. If f or g depend on some parameter λ, then the notation "$f(x) = O_\lambda(g(x))$", or, equivalently, "$f(x) \ll_\lambda g(x)$", indicates that the constants x_0 and c implicit in the estimate may depend on λ. If the constants can be chosen independently of λ, for λ in some range, then the estimate is said to hold **uniformly** in that range. Dependence on several parameters is indicated in an analogous way, as in "$f(x) \ll_{\lambda,\alpha} g(x)$".

Oh and oh expressions in equations. A term $O(g(x))$ in an equation stands for a function $R(x)$ satisfying the estimate $R(x) = O(g(x))$. For example, the notation "$f(x) = x(1 + O(1/\log x))$" means that there exist constants x_0 and c and a function $\delta(x)$ defined for $x \geq x_0$ and satisfying $|\delta(x)| \leq c/\log x$ for $x \geq x_0$, such that $f(x) = x(1 + \delta(x))$ for $x \geq x_0$.

"Big oh" versus "small oh". "Big oh" estimates provide more information than "small oh" estimates or asymptotic formulas, since they give explicit bounds for the error terms involved. An o-estimate, or an asymptotic formula, only shows that a certain function tends to zero, but does not provide any information for the rate at which this function tends to zero. For example, the asymptotic relation $f(x) \sim g(x)$, or equivalently, the o-estimate $f(x) = g(x) + o(g(x))$, means that the ratio $(g(x) - f(x))/g(x)$ tends to zero, whereas a corresponding O-estimate, such as $f(x) = g(x) + O(\epsilon(x)g(x))$, with an explicit function $\epsilon(x)$ (e.g., $\epsilon(x) = 1/\log x$), provides additional information on the speed of convergence.

O-estimates are also easier to work with and to manipulate than o-estimates. For example, O's can be "pulled out" of integrals or sums provided the functions involved are nonnegative, whereas such manipulations are in general not allowed with o's.

For the above reasons, O-estimates are much more useful than o-estimates, and one should therefore try to state and prove results in terms of O-estimates. It is very rare that one can prove a o-estimate, without getting an explicit bound for the o-term, and hence a more precise O-estimate, by the same argument.

2.1.3 Examples

Examples from analysis

(1) $x^\alpha = O_{\alpha,c}(e^{cx})$ for any fixed real numbers α and $c > 0$.

(2) $\exp(x^\alpha) \sim \exp((x+1)^\alpha)$ if $\alpha < 1$.

(3) $\log x = O_\epsilon(x^\epsilon)$ for any fixed $\epsilon > 0$.

(4) $\log(1+x) = O(x)$ for $|x| \le 1/2$ (and, more generally, for $|x| \le c$, for any constant $c < 1$).

(5) $\cos x = 1 + O(x^2)$ for $|x| \le 1$ (in fact, for all real x).

(6) $1/(1+x) = 1 + O(x)$ for $|x| \le 1/2$ (say).

(7) Let $0 < \alpha < 1$ be fixed. Then, for any constants $A > 0$ and $\epsilon > 0$,

$$x^\epsilon \ll_{\epsilon,\alpha} \exp\left(-(\log x)^\alpha\right) \ll_{A,\alpha} (\log x)^{-A}).$$

The proofs of such estimates are usually straightforward exercises at the calculus level. To illustrate some typical arguments, we give the proofs of (1), (2), and (5):

Proof of (1). Let α and $c > 0$ be given. We need to show that there exist constants C and x_0 such that $x^\alpha \le Ce^{cx}$ for $x \ge x_0$. Setting $f(x) = x^\alpha e^{-cx}$, this is equivalent to showing that $f(x)$ is bounded for sufficiently large x. This, however, follows immediately on noting that (i) $f(x)$ tends to zero as $x \to \infty$ (which can be seen by l'Hopital's rule) and (ii) $f(x)$ is continuous on the positive real axis, and hence must attain a maximal value on the interval $[1, \infty)$. An alternative argument, which yields explicit values for the constants C and x_0 runs as follows:

If $\alpha \le 0$, then for $x \ge 1$ we have $f(x) = x^\alpha e^{-cx} \le 1$, so the desired bound holds with constants $C = x_0 = 1$. Assume therefore that $0 < \alpha < 1$. We have $\log f(x) = \alpha \log x - cx$ and hence $f'(x)/f(x) = (\log f(x))' = \alpha/x - c$. Hence $f'(x) \le 0$ for $x \ge \alpha/c$, and setting $x_0 = \alpha/c$ and $C = x_0^\alpha e^{-\alpha}$ it follows that $f(x) \le f(x_0) = x_0^\alpha e^{-cx_0} = C$ for $x \ge x_0$. □

Proof of (2). Let $f(x) = \exp(x^\alpha)$. We need to show that $\lim_{x\to\infty} f(x)/f(x+1) = 1$. Now, $f(x)/f(x+1) = \exp\{x^\alpha - (x+1)^\alpha\}$, so the desired relation is equivalent to $x^\alpha - (x+1)^\alpha \to 0$ as $x \to \infty$. The latter relation holds since

$$|x^\alpha - (x+1)^\alpha| \le \max_{x \le y \le x+1} \alpha y^{\alpha - 1}$$

by the mean value theorem of calculus, and since the expression on the right here tends to zero as $x \to \infty$, as $\alpha < 1$. □

Proof of (5). In the range $|x| \le 1$, the estimate $\cos x - 1 = O(x^2)$ with 1 as O-constant follows immediately from the mean value theorem (or, what amounts to the same argument, Taylor's series for $\cos x$ truncated after the first term and with an explicit error term):

$$|\cos x - 1| = |\cos 0 - \cos x| \le |x| \max_{0 \le |y| \le |x|} |\sin y| \le x^2,$$

where in the last step we used the fact that $|\sin y|$ is increasing and bounded by $|y|$ on the interval $[-1, 1]$. To extend the range for this estimate to all of $\mathbb{R}$, we only need to observe that, since $|\cos x| \le 1$ for all x we have $|\cos x - 1| \le 2$ for all x, and hence $|\cos x - 1| \le 2x^2$ for $|x| \ge 1$. Thus, the desired estimate holds for all x with O-constant 2. □

Examples from number theory

(1) The prime number theorem (PNT) is the statement that $\pi(x) \sim x/\log x$, or, equivalently, $\pi(x) = x/\log x + o(x/\log x))$. Here $\pi(x)$ is the number of primes not exceeding x.

(2) A sharper form of the PNT asserts that $\pi(x) = x/\log x + O(x/(\log x)^2)$. Factoring out the main term $x/\log x$, this estimate can also be written as $\pi(x) = (x/\log x)(1 + O(1/\log x))$, which shows that $O(1/\log x)$ is the *relative error* in the approximation of $\pi(x)$ by $x/\log x$.

(3) A still sharper form involves the approximation $\mathrm{Li}(x) = \int_2^x (1/\log t) dt$. The currently best known estimate for $\pi(x)$ is of the form $\pi(x) = \mathrm{Li}(x) + O_\alpha(x \exp(-(\log x)^\alpha))$, where α is any fixed real number $< 3/5$. By Example (7) above the error term here is of smaller order than $x(\log x)^{-A}$ for any fixed constant A, but of larger order than $x^{1-\epsilon}$, for any fixed $\epsilon > 0$.

(4) The Riemann Hypothesis is equivalent to the statement that $\pi(x) = \mathrm{Li}(x) + O_\epsilon(x^{1/2+\epsilon})$ for any fixed $\epsilon > 0$.

Additional examples and remarks

The following examples and remarks illustrate common uses of the O- and o-notations. The proofs are immediate consequences of the definitions.

(1) A commonly seen O-estimate is $f(x) = O(1)$. This simply means that $f(x)$ is bounded for sufficiently large x (or for all x in a given range). Similarly $f(x) = o(1)$ means that $f(x)$ tends to 0 as $x \to \infty$.

(2) If C is a positive constant, then the estimate $f(x) = O(Cg(x))$ is equivalent to $f(x) = O(g(x))$. In particular, the estimate $f(x) = O(C)$ is equivalent to $f(x) = O(1)$. The same holds for o-estimates.

(3) O-estimates are transitive, in the sense that if $f(x) = O(g(x))$ and $g(x) = O(h(x))$, then $f(x) = O(h(x))$.

(4) As an application of this transitivity and the basic estimates above we have, for example, $\log(1+O(f(x))) = O(f(x))$ and $1/(1+O(f(x))) = 1 + O(f(x))$ whenever $f(x) \to 0$ as $x \to \infty$. (The latter condition ensures that the function represented by the term $O(f(x))$ is bounded by $\leq 1/2$ for sufficiently large x, so the estimates $\log(1 + y) = O(y)$ and $(1 + y)^{-1} = 1 + O(y)$ are applicable with y being the function represented by $O(f(x))$.)

(5) If $f(x) = g(x) + O(1)$, then $e^{f(x)} \asymp e^{g(x)}$, and vice versa.

(6) If $f(x) = g(x) + o(1)$, then $e^{f(x)} \sim e^{g(x)}$, and vice versa.

(7) "O's" can be pulled out of sums or integrals *provided the function inside the O-term is nonnegative.* For example, if $F(x) = \int_0^x f(y)dy$ and $f(y) = O(g(y))$ for $y \geq 0$, where g is a nonnegative function, then $F(x) = O\left(\int_0^x g(y)dy\right)$. (This does not hold without the nonnegativity condition, nor does an analogous result hold for o-estimates; for counterexamples see the exercises.)

(8) According to our convention, an asymptotic estimate for a function of x without an explicitly given range is understood to hold for $x \geq x_0$ for a suitable x_0. This is convenient as many estimates (e.g., $\log\log x = O(\sqrt{\log x})$) do not hold, or do not make sense, for small values of x, and the convention allows one to just ignore those issues. However, there are applications in which it is desirable to have an estimate involving a simple explicit range for x, such as $x \geq 1$, instead of an unspecified range like $x \geq x_0$ with a "sufficiently large" x_0. This can often be accomplished in two steps as follows: First establish the desired estimate for $x \geq x_0$, with a certain x_0. Then use direct (and usually trivial) arguments to show that the estimate also holds for $1 \leq x \leq x_0$. For example, one form of the PNT states that $(*)$

$\pi(x) = \mathrm{Li}(x) + O(x(\log x)^{-2})$. Suppose we have established $(*)$ for $x \geq x_0$. To show that $(*)$ in fact holds for $x \geq 2$, we can argue as follows: In the range $2 \leq x \leq x_0$ the functions $\pi(x)$ and $\mathrm{Li}(x)$ are bounded from above, say $|\pi(x)|, |\mathrm{Li}(x)| \leq A$ for $2 \leq x \leq x_0$ and some constant A (depending on x_0). On the other hand, the function in the error term, $x(\log x)^{-2}$, is bounded from below by a positive constant, say $\delta > 0$, in this range. (For example, we can take $\delta = 2(\log x_0)^{-2}$.) Hence, for $2 \leq x \leq x_0$ we have

$$|\pi(x) - \mathrm{Li}(x)| \leq 2A \leq \frac{2A}{\delta} x(\log x)^{-2} \quad (2 \leq x \leq x_0),$$

so $(*)$ holds for $2 \leq x \leq x_0$ with $c = 2A/\delta$ as O-constant.

2.1.4 The logarithmic integral

The *logarithmic integral* is the function $\mathrm{Li}(x)$ defined by

$$\mathrm{Li}(x) = \int_2^x (\log t)^{-1} dt \quad (x \geq 2).$$

This integral is important in number theory as it represents the best known approximation to the prime counting function $\pi(x)$. The integral cannot be evaluated *exactly* (in terms of elementary functions), but the following theorem gives (a sequence of) asymptotic estimates for the integral in terms of elementary functions.

Theorem 2.1 (The logarithmic integral). *For any fixed positive integer k we have*

$$\mathrm{Li}(x) = \frac{x}{\log x}\left(\sum_{i=0}^{k-1} \frac{i!}{(\log x)^i} + O_k\left(\frac{1}{(\log x)^k}\right)\right) \quad (x \geq 2).$$

In particular, we have $\mathrm{Li}(x) = x/\log x + O(x/(\log x)^2)$ *for* $x \geq 2$.

To prove the result we require a crude estimate for a generalized version of the logarithmic integral, namely

$$\mathrm{Li}_k(x) = \int_2^x (\log t)^{-k} dt \quad (x \geq 2),$$

where k is a positive integer (so that $\mathrm{Li}_1(x) = \mathrm{Li}(x)$). This result is of independent interest, and its proof is a good illustration of the method of splitting the range of integration.

Lemma 2.2. *For any fixed positive integer k we have*

$$\mathrm{Li}_k(x) \ll_k \frac{x}{(\log x)^k} \quad (x \geq 2).$$

Proof. First note that the bound holds trivially in any range of the form $2 \leq x \leq x_0$ (with the O-constant depending on x_0). We therefore may assume that $x \geq 4$. In this case we have $2 \leq \sqrt{x} \leq x$, so that we may split the range $2 \leq t \leq x$ into the two subranges $2 \leq t < \sqrt{x}$ and $\sqrt{x} \leq t \leq x$. In the first subrange the integrand is bounded by $1/(\log 2)^k$, so the integral over this range is $\leq (\log 2)^{-k}(\sqrt{x}-2) \ll_k \sqrt{x}$, which is of the desired order of magnitude.

In the remaining range $\sqrt{x} \leq t \leq x$, the integrand is bounded by $\leq (\log\sqrt{x})^{-k} = 2^k(\log x)^{-k}$, so the integral over this range is at most $2^k(\log x)^{-k}(x-\sqrt{x}) \ll_k x(\log x)^{-k}$, which again is of the desired order of magnitude. □

Remark. The choice of $\sqrt{x}$ as the splitting point is sufficient to obtain the asserted upper bound for $\mathrm{Li}_k(x)$, but it is not optimal and choosing a larger division point allows one to derive a more accurate estimate for $\mathrm{Li}_k(x)$ (which, however, is still inferior to what can be obtained with the integration by parts method that we will use to prove Theorem 2.1). For example, splitting the integral at $x(\log x)^{-k-1}$, the contribution of lower subrange is $\ll_k x(\log x)^{-k-1}$, whereas in the upper subrange the integrand can be approximated as follows:

$$(\log t)^{-k} = (\log x + \log(t/x))^{-k} = (\log x)^{-k}\left(1 + O_k\left(\frac{\log\log x}{\log x}\right)\right).$$

This leads to the estimate

$$\mathrm{Li}_k(x) = \frac{x}{(\log x)^k}\left(1 + O_k\left(\frac{\log\log x}{\log x}\right)\right).$$

Proof of Theorem 2.1. Integration by parts shows that, for $i = 1, 2, \ldots$,

$$\begin{aligned}\mathrm{Li}_i(x) &= \frac{x}{(\log x)^i} - \frac{2}{(\log 2)^i} - \int_2^x t\frac{-i}{(\log t)^{i+1}t}dt \\ &= \frac{x}{(\log x)^i} - \frac{2}{(\log 2)^i} + i\,\mathrm{Li}_{i+1}(x).\end{aligned}$$

Applying this identity successively for $i = 1, 2, \ldots, k$ (or, alternatively, using induction on k) gives

$$\mathrm{Li}(x) = \mathrm{Li}_1(x) = O_k(1) + \sum_{i=1}^{k}\frac{(i-1)!x}{(\log x)^i} + k!\,\mathrm{Li}_{k+1}(x).$$

(Here the term $O_k(1)$ absorbs the constant terms $2(\log 2)^{-i}$ that arise when using the above estimate for each $i = 1, 2, \dots, k$.) Since $\mathrm{Li}_{k+1}(x) \ll_k x(\log x)^{-k-1}$ by Lemma 2.2, the asserted estimate follows. □

Remark. Note that, because of the factor $i!$, the series in the main term diverges if one lets $k \to \infty$. The resulting infinite series $\sum_{i=0}^{\infty} i!(\log x)^{-i}$ is an example of a so-called "asymptotic series", a series that diverges everywhere, but which, when truncated at some level k, behaves like an ordinary convergent series, in the sense that the error introduced by truncating the series at the kth term has an order of magnitude equal to that of the next term in the series.

2.2 Sums of smooth functions: Euler's summation formula

2.2.1 Statement of the formula

The simplest types of sums $\sum_{n\le x} f(n)$ are those in which f is a "smooth" function that is defined for real arguments x. Sums of this type are of interest in their own right (for example, Stirling's formula for $n!$ is equivalent to an estimate for a sum of the above type with $f(n) = \log n$—see Theorem 2.6 and Corollary 2.7 below), but they also occur in the process of estimating sums of arithmetic functions like the divisor function or the Euler phi function (see the following sections).

The basic idea for handling such sums is to approximate the sum by a corresponding integral and investigate the error made in the process. The following important result, known as Euler's summation formula, gives an exact formula for the difference between such a sum and the corresponding integral.

Theorem 2.3 (Euler's summation formula)**.** *Let $0 < y \le x$, and suppose $f(t)$ is a function defined on the interval $[y, x]$ and having a continuous derivative there. Then*

$$\sum_{y<n\le x} f(n) = \int_y^x f(t)dt + \int_y^x \{t\}f'(t)dt - \{x\}f(x) + \{y\}f(y),$$

where $\{t\}$ denotes the fractional part of t, i.e., $\{t\} = t - [t]$.

In most applications, one needs to estimate a sum of the form $\sum_{n\le x} f(n)$, taken over all positive integers $n \le x$. In this case, Euler's summation formula reduces to the following result:

Corollary 2.4 (Euler's summation formula, special case)**.** *Let $x \geq 1$ and suppose that $f(t)$ is defined on $[1, x]$ and has a continuous derivative on this interval. Then we have*

$$\sum_{n \leq x} f(n) = \int_1^x f(t)dt + \int_1^x \{t\} f'(t)dt - \{x\} f(x) + f(1).$$

Proof. We apply Euler's summation formula with $y = 1$. The integrals on the right-hand side are then of the desired form, as is the term $\{x\} f(x)$, while the term $\{y\} f(y)$ vanishes. On the other hand, the sum on the left with $y = 1$ is equal to the sum over all positive integers $n \leq x$ minus the term $f(1)$. Adding this term on both sides of Euler's summation formula gives the identity stated in the corollary. □

Proof of Theorem 2.3. Letting $F(x) = [x] = x - \{x\}$, we can write the given sum as a *Stieltjes integral* $\sum_{y<n\leq x} f(n) = \int_y^x f(t)dF(t)dt$. Writing $dF(t) = dt - d\{t\}$, the integral splits into $\int_y^x f(t)dt - \int_y^x f(t)d\{t\}$. The first of these latter two integrals is the desired main term, while the second can be transformed as follows using integration by parts:

$$\int_y^x f(t)d\{t\} = f(x)\{x\} - f(y)\{y\} - \int_y^x f'(t)\{t\}dt.$$

Combining these formulas gives the desired identity. □

Remark. The above proof is quite simple and intuitive, and motivates the particular form of the Euler summation formula. However, it is less elementary in that it depends on a concept beyond the calculus level, namely the Stieltjes integral. In Section 2.3.1 we will give an independent, more elementary, proof; in fact, we will prove a more general result (the partial summation formula), of which Euler's summation formula is a corollary.

2.2.2 Partial sums of the harmonic series

Euler's summation formula has numerous applications in number theory and analysis. We will give here three such applications; the first is to the partial sums of the harmonic series.

Theorem 2.5 (Partial sums of the harmonic series)**.** *We have*

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right) \quad (x \geq 1),$$

where $\gamma = \lim_{x\to\infty}(\sum_{n\le x} 1/n - \log x) = 0.5772\dots$ *is a constant, the so-called Euler constant.*

Remark. The error term $O(1/x)$ here is best-possible, since the left-hand side has a jump of size $1/x$ whenever x crosses an integer, while the main term on the right is continuous in x.

Proof. Let $S(x) = \sum_{n\le x} 1/n$. By Euler's summation formula (in the version given by Corollary 2.4) we have, for $x \ge 1$,

$$\sum_{n\le x} = \int_1^x \frac{1}{t}dt + \int_1^x \{t\}\frac{-1}{t^2}dt + \frac{\{x\}}{x} + 1 = \log x - I(x) + 1 + O\left(\frac{1}{x}\right),$$

where $I(x) = \int_1^x \{t\}t^{-2}dt$. To obtain the desired estimate, it suffices to show that $I(x) = \gamma - 1 + O(1/x)$ for $x \ge 1$. To estimate the integral $I(x)$, we employ the following trick: We extend the integration in the integral to infinity and estimate the tail of the integral. Since the integrand is bounded by $1/t^2$, the integral converges absolutely when extended to infinity, and therefore equals a finite constant, say I, and we have

$$I(x) = I - \int_x^\infty \frac{\{t\}}{t^2}dt = I + O\left(\int_x^\infty \frac{1}{t^2}dt\right) = I + O\left(\frac{1}{x}\right) \quad (x \ge 1).$$

We have thus shown that $S(x) = \log x + 1 - I + O(1/x)$ for $x \ge 1$. In particular, this implies that $S(x) - \log x$ converges to $1 - I$ as $x \to \infty$. Since, by definition, $\gamma = \lim_{x\to\infty}(S(x) - \log x)$, we have $1 - I = \gamma$, and thus obtain the desired formula. □

2.2.3 Partial sums of the logarithmic function and Stirling's formula

Our second application of Euler's summation formula is a proof of the so-called Stirling formula, which gives an asymptotic estimate for $N!$, where N is a (large) integer. This formula will be an easy consequence of the following estimate for the logarithm of $N!$, $\log N! = \sum_{n\le N} \log n$, which is a sum to which Euler's summation formula can be applied.

Theorem 2.6 (Partial sums of the logarithmic function)**.** *We have*

$$\sum_{n\le N} \log n = N(\log N - 1) + \frac{1}{2}\log N + c + O\left(\frac{1}{N}\right) \quad (N \in \mathbb{N}),$$

where c is a constant.

Proof. Let $S(N) = \sum_{n\le N}\log n$. Applying Euler's summation formula (again in the special case provided by Corollary 2.4), and noting that $\{N\}=0$ since N is an integer, we obtain

$$S(N) = I_1(N) + I_2(N)$$

with

$$I_1(N) = \int_1^N (\log t)dt = t(\log t - 1)\Big|_1^N = N\log N - N + 1$$

and

$$I_2(N) = \int_1^N \frac{\{t\}}{t}dt = \int_1^N \frac{1/2}{t}dt + \int_1^N \frac{\rho(t)}{t}dt = \frac{1}{2}\log N + I_3(N),$$

where $\rho(t) = \{t\} - 1/2$ is the "row of teeth" function and $I_3(N) = \int_1^N(\rho(t)/t)dt$. Combining these formulas gives

$$S(N) = N\log N - N + \frac{1}{2}\log N + 1 + I_3(N).$$

Thus, to obtain the desired estimate, it suffices to show that $I_3(N) = c' + O(1/N)$, for some constant c'.

We begin with an integration by parts to get

$$I_3(N) = \frac{R(t)}{t}\Big|_1^N + I_4(N),$$

where

$$R(t) = \int_1^t \rho(t)dt, \quad I_4(x) = \int_1^x \frac{R(t)}{t^2}dt.$$

Since $\rho(t)$ is periodic with period 1, $|\rho(t)|\le 1/2$ for all t, and $\int_k^{k+1}\rho(t) = 0$ for any integer k, we have $R(t) = 0$ whenever t is an integer, and $|R(t)|\le 1/2$ for all t. Hence the terms $R(t)/t$ vanish at $t=1$ and $t=N$, so we have $I_3(N) = I_4(N)$. Now, the integral $I_4(N)$ converges as $x\to\infty$, since its integrand is bounded by $|R(t)|t^{-2}\le (1/2)t^{-2}$, and we therefore thave

$$I_4(N) = I - \int_N^\infty \frac{R(t)}{t^2}dt = I - O\left(\int_N^\infty \frac{1}{t^2}dt\right) = I + O\left(\frac{1}{N}\right),$$

where

$$I = \int_1^\infty \frac{R(t)}{t^2}dt.$$

(Note here again the "trick" of extending a convergent integral to infinity and estimating the tail.) Hence we have $I_3(N) = I_4(N) = I + O(1/N)$, as we wanted to show. □

We now use this estimate to prove (modulo the evaluation of a constant) Stirling's formula for $n!$.

Corollary 2.7 (Stirling's formula)**.** *If n is a positive integer, then*

$$n! = C\sqrt{n}n^n e^{-n}\left(1 + O\left(\frac{1}{n}\right)\right),$$

where C is a constant.

Remark. One can show that the constant C is equal to $\sqrt{2\pi}$, and Stirling's formula is usually stated with this explicit value of the constant. However, proving this is far from easy, and since the value of the constant is not important for our applications, we will not pursue this here. The argument roughly goes as follows: Consider the sum $\sum_{k=0}^{n}\binom{n}{k}$. On the one hand, this sum is exactly equal to 2^n. On the other hand, by expressing the binomial coefficients in terms of factorials and estimating the factorials by Stirling's formula one can obtain an estimate for this sum involving the Stirling constant C. Comparing the two evaluations one obtains $C = \sqrt{2\pi}$.

Proof. Since $n! = \exp\{\sum_{k\leq n}\log k\}$, we have, by the theorem,

$$n! = \exp\left\{n\log n - n + \frac{1}{2}\log n + c + O\left(\frac{1}{n}\right)\right\},$$

which reduces to the right-hand side in the estimate of the corollary, with constant $C = e^c$. □

The estimate of Theorem 2.6 applies only to sums $\sum_{n\leq x}\log n$ when x is a positive integer; this is the case of interest in the application to Stirling's formula. In numbertheoretic applications one needs estimates for these sums that are valid for all (large) *real* x. The following corollary provides such an estimate, at the cost of a weaker error term.

Corollary 2.8. *We have*

$$\sum_{n\leq x}\log n = x(\log x - 1) + O(\log x) \quad (x \geq 2).$$

Proof. We apply the estimate of the theorem with $N = [x]$. The left-hand side remains unchanged when replacing x by N. On the other hand, the main term on the right, $x(\log x - 1)$, has derivative $\log x$. so it changes by an amount of order at most $O(\log x)$ if x is replaced by $[x]$. Since the error term $O(1/x)$ on the right is also of this order of magnitude, the asserted estimate follows. (Note here the restriction $x \geq 2$; in the larger range $x \geq 1$ this estimate would not be valid, since the error term $O(\log x)$ is 0 at $x = 1$, whereas the main terms on the left and right are clearly not equal when $x = 1$.) □

2.2.4 Integral representation of the Riemann zeta function

The Riemann zeta function is defined for complex arguments s with $\operatorname{Re} s > 1$ by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

As an application of Euler's summation formula, we now derive an integral represention for this function. This representation will be crucial in deriving deeper analytic properties of the zeta function.

Theorem 2.9 (Integral representation of the zeta function)**.** *For* $\operatorname{Re} s > 1$ *we have*

$$\zeta(s) = \frac{s}{s-1} - s \int_1^{\infty} \{x\} x^{-s-1} dx. \tag{2.1}$$

Proof. Fix s with $\operatorname{Re} s > 1$, and let $S(x) = \sum_{n \leq x} n^{-s}$. Applying Euler's summation formula in the form of Corollary 2.4 with $f(x) = x^{-s}$, we get, for any $x \geq 1$,

$$S(x) = I_1(x) + I_2(x) - \{x\} x^{-s} + 1,$$

where

$$I_1(x) = \int_1^x y^{-s} dy = \frac{1 - x^{1-s}}{s-1} = \frac{1}{s-1} + O_s(x^{1-\operatorname{Re} s})$$

and

$$\begin{aligned} I_2(x) &= \int_1^x \{y\}(-s) y^{-s-1} dy \\ &= -s \int_1^{\infty} \{y\} y^{-s-1} dy + O_s\left(\int_x^{\infty} y^{-\operatorname{Re} s - 1} dy \right) \\ &= -s \int_1^{\infty} \{y\} y^{-s-1} dy + O_s(x^{-\operatorname{Re} s}). \end{aligned}$$

Letting $x \to \infty$, the O-terms in the estimates for $I_1(x)$ and $I_2(x)$, as well as the term $\{x\}x^{-s}$, tend to zero, and we conclude

$$\begin{aligned}\zeta(s) = \lim_{x\to\infty} S(x) &= \frac{1}{s-1} + 1 - s\int_1^\infty \{y\}y^{-s-1}dy \\ &= \frac{s}{s-1} - s\int_1^\infty \{y\}y^{-s-1}dy,\end{aligned}$$

which is the asserted identity. □

2.3 Removing a smooth weight function from a sum: Summation by parts

2.3.1 The summation by parts formula

Summation by parts (also called partial summation or Abel summation) is the analogue for sums of integration by parts. Given a sum of the form $\sum_{n\le x} a(n)f(n)$, where $a(n)$ is an arithmetic function with summatory function $A(x) = \sum_{n\le x} a(n)$ and $f(n)$ is a "smooth" weight, the summation by parts formula allows one to "remove" the weight $f(n)$ from the above sum and reduce the evaluation or estimation of the sum to that of an integral over $A(t)$. The general formula is as follows:

Theorem 2.10 (Summation by parts formula)**.** *Let $a : \mathbb{N} \to \mathbb{C}$ be an arithmetic function, let $0 < y < x$ be real numbers and $f : [y,x] \to \mathbb{C}$ be a function with continuous derivative on $[y,x]$. Then we have*

$$\sum_{y<n\le x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t)dt, \tag{2.2}$$

where $A(t) = \sum_{n\le t} a(n)$.

This formula is easy to remember since it has the same form as the formula for integration by parts, if one thinks of $A(x)$ as the "integral" of $a(n)$.

In nearly all applications, the sums to be estimated are sums of the form $\sum_{n\le x} a(n)f(n)$, where n ranges over all positive integers $\le x$. We record the formula in this special case separately.

Corollary 2.11 (Summation by parts formula, special case)**.** *Let* $a : \mathbb{N} \to \mathbb{C}$ *be an arithmetic function, let* $x \ge 1$ *be a real number and* $f : [1, x] \to \mathbb{C}$ *a function with continuous derivative on* $[1, x]$*. Then we have*

$$\sum_{n \le x} a(n)f(n) = A(x)f(x) - \int_1^x A(t)f'(t)dt. \tag{2.3}$$

Proof. Applying the theorem with $y = 1$ and adding the term $a(1)f(1) = A(1)f(1)$ on both sides of (2.2) gives (2.3). □

In the case when $a(n) \equiv 1$ the sum on the left of (2.2) is of the same form as the sum estimated by Euler's summation formula (Theorem 2.3), which states that, under the same conditions on f, one has

$$\sum_{y<n\le x} f(n) = \int_y^x f(t)dt + \int_y^x \{t\}f'(t)dt - \{x\}f(x) + \{y\}f(y).$$

In fact, as we now show, this formula can be derived from the partial summation formula.

Alternate proof of Euler's summation formula (Theorem 2.3). Applying the partial summation formula with $a(n) \equiv 1$ and $A(t) = \sum_{n\le t} 1 = [t]$, we obtain

$$\begin{aligned}\sum_{y<n\le x} f(n) &= [x]f(x) - [y]f(y) - \int_y^x [t]f'(t)dt \\ &= xf(x) - yf(y) - \int_y^x tf'(t)dt \\ &\qquad - \{x\}f(x) + \{y\}f(y) + \int_y^x \{t\}f'(t)dt.\end{aligned}$$

By an integration by parts, the first integral on the right-hand side equals $xf(x) - yf(y) - \int_y^x f(t)dt$, so the above reduces to

$$\sum_{y<n\le x} f(n) = \int_y^x f(t)dt - \{x\}f(x) + \{y\}f(y) + \int_y^x \{t\}f'(t)dt,$$

which is the desired formula. □

Proof of Theorem 2.10. Let $0 < y < x$, a, and f be given as in the theorem. and let I denote the integral on the right of (2.2). Setting $\chi(n,t) = 1$ if $n \le t$ and $\chi(n,t) = 0$ otherwise, we have

$$\begin{aligned} I &= \int_y^x \sum_{n\le x} a(n)\chi(n,t)f'(t)dt = \sum_{n\le x} a(n) \int_y^x \chi(n,t)f'(t)dt \\ &= \sum_{n\le x} a(n) \int_{\max(n,y)}^x f'(t)dt, \end{aligned}$$

where the interchanging of integration and summation is justified since the sum involves only finitely many terms. Since f' is continuous on $[y,x]$, the inner integrals can be evaluated by the fundamental theorem of calculus, and we obtain

$$\begin{aligned} I &= \sum_{n\le x} a(n)(f(x) - f(\max(n,y)) \\ &= \sum_{n\le x} a(n)f(x) - \sum_{n\le y} a(n)f(y) - \sum_{y<n\le x} a(n)f(n) \\ &= A(x)f(x) - A(y)f(y) - \sum_{y<n\le x} a(n)f(n). \end{aligned}$$

Hence,

$$\sum_{y<n\le x} a(n)f(n) = A(x)f(x) - A(y)f(y) - I,$$

which is the desired formula. □

Partial summation is an extremely useful tool that has numerous applications in number theory and analysis. In the following subsections we give three such applications. We will encounter a number of other applications in later chapters.

2.3.2 Kronecker's Lemma

As a first, and simple, illustration of the use of the partial summation formula we prove the following result, known as "Kronecker's Lemma", which is of independent interest and has a number of applications in its own right, in particular, in probability theory and analysis.

Theorem 2.12 (Kronecker's Lemma). *Let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function. If s is a complex number with* $\operatorname{Re} s > 0$ *such that*

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s} \quad \textit{converges,} \tag{2.4}$$

then

$$\lim_{x\to\infty} \frac{1}{x^s} \sum_{n\le x} f(n) = 0. \tag{2.5}$$

In particular, the convergence of $\sum_{n=1}^{\infty} f(n)/n$ *implies that f has mean value zero in the sense that* $\lim_{x\to\infty}(1/x)\sum_{n\le x} f(n) = 0$.

Remarks. Kronecker's lemma is often stated only in the special case mentioned at the end of the above theorem (i.e., the case $s = 1$), and for most applications it is used in this form. We have stated a slightly more general version involving "weights" n^{-s} instead of n^{-1}, as we will need this version later. In fact, the result holds in greater generality, with the function x^{-s} replaced by a general "weight function" $w(x)$ in both (2.4) and (2.5).

Proof. Fix a function $f(n)$ and $s \in \mathbb{C}$ with $\operatorname{Re} s > 0$ as in the theorem, and set

$$S(x) = \sum_{n\le x} f(n), \quad T(x) = \sum_{n\le x} \frac{f(n)}{n^s}.$$

The hypothesis (2.4) means that $T(x)$ converges to a finite limit T as $x \to \infty$, and the desired conclusion (2.5) is equivalent to $\lim_{x\to\infty} S(x)/x^s = 0$.

Let $\epsilon > 0$ be given. Since $\lim_{x\to\infty} T(x) = T$, there exists $x_0 = x_0(\epsilon) \ge 1$ such that

$$|T(x) - T| \le \epsilon \quad (x \ge x_0). \tag{2.6}$$

Let $x \ge x_0$. Applying the summation by parts formula with $f(n)/n^s$ and n^s in place of $a(n)$ and $f(n)$, respectively, we obtain

$$\begin{aligned} S(x) &= \sum_{n\le x} \frac{f(n)}{n^s} \cdot n^s = T(x)x^s - \int_1^x T(t)st^{s-1}dt \\ &= \int_0^x T(x)st^{s-1}dt - \int_1^x T(t)st^{s-1}dt. \end{aligned}$$

Defining $T(t)$ to be 0 if $t \le 1$, we can combine the last two integrals to a single integral over the interval $[0, x]$ and obtain

$$\begin{aligned} |S(x)| &= \left| \int_0^x (T(x) - T(t)) s t^{s-1} dt \right| \\ &\le \int_0^x |T(x) - T(t)| |s| t^{\operatorname{Re} s - 1} dt. \end{aligned}$$

To estimate the latter intergral, we split the interval of integration into the two subintervals $[0, x_0)$ and $[x_0, x]$, and bound the integrand separately in these two intervals. For $x_0 \le t \le x$ we have, by (2.6),

$$|T(t) - T(x)| \le |T(t) - T| + |T - T(x)| \le 2\epsilon,$$

while for $0 \le t \le x_0$ we use the trivial bound

$$\begin{aligned} |T(t) - T(x)| &\le |T(t)| + |T| + |T - T(x)| \\ &\le \sum_{n \le x_0} \left| \frac{f(n)}{n^s} \right| + |T| + \epsilon = M, \end{aligned}$$

say, with $M = M(\epsilon)$ a constant depending on ϵ, but not on x. It follows that

$$\begin{aligned} |S(x)| &\le \epsilon \int_{x_0}^x |s| t^{\operatorname{Re} s - 1} dt + M \int_0^{x_0} |s| t^{\operatorname{Re} s - 1} dt \\ &\le \frac{|s|}{\operatorname{Re} s} \left(\epsilon \left(x^{\operatorname{Re} s} - {x_0}^{\operatorname{Re} s} \right) + M x_0^{\operatorname{Re} s} \right) \end{aligned}$$

and hence

$$\left| \frac{S(x)}{x^s} \right| \le \frac{|s|}{\operatorname{Re} s} \left(\epsilon + \frac{M x_0^{\operatorname{Re} s}}{x^{\operatorname{Re} s}} \right).$$

Since, by hypothesis, $\operatorname{Re} s > 0$, the last term on the right tends to zero as $x \to \infty$, so we obtain $\limsup_{x \to \infty} |S(x)/x^s| \le \epsilon |s| / \operatorname{Re} s$. Since $\epsilon > 0$ was arbitrary, we conclude that $\lim_{x \to \infty} S(x)/x^s = 0$, as desired. □

2.3.3 Relation between different notions of mean values of arithmetic functions

We next use partial summation to study the relation between two different types of "mean values", or averages, of an arithmetic function f: the *ordinary (or asymptotic) mean value*

$$M(f) = \lim_{x \to \infty} \frac{1}{x} \sum_{n \le x} f(n), \tag{2.7}$$

and the *logarithmic mean value*

$$L(f) = \lim_{x\to\infty} \frac{1}{\log x} \sum_{n\le x} \frac{f(n)}{n}, \tag{2.8}$$

The asymptotic mean value is (a limit of) an ordinary average, or arithmetic mean, of the values $f(n)$, while the logarithmic mean value can be regarded as a weighted average of these values, with the weights being $1/n$. Thus, to convert between these two mean values it is natural to use partial summation to remove or re-instate the weights $1/n$. The application of partial summation in this way is very common, and it is also quite instructive as it illustrates both a siutation in which this approach is successful, and a situation in which the method fails.

In one direction (namely, going from $M(f)$ to $L(f)$), the method works well, and we have the following result.

Theorem 2.13. *Let f be an arithmetic function, and suppose that the ordinary mean value $M(f)$ exists. Then the logarithmic mean value $L(f)$ exists as well, and is equal to $M(f)$.*

Proof. Suppose f has mean value $M(f) = A$. Let $S(x) = \sum_{n\le x} f(n)$ and $T(x) = \sum_{n\le x} f(n)/n$. By the assumption $M(f) = A$, we have $\lim_{x\to\infty} S(x)/x = A$, and we need to show that $\lim_{x\to\infty} T(x)/\log x = A$.

To obtain an estimate for $T(x)$, we apply the partial summation formula with $a(n) = f(n)$, $A(x) = S(x)$, and with the function $f(x) = 1/x$ as the weight function to be removed from the sum. We obtain

$$T(x) = \frac{S(x)}{x} + \int_1^x \frac{S(t)}{t^2} dt = \frac{S(x)}{x} + I(x), \tag{2.9}$$

say. Upon dividing by $\log x$, the first term, $S(x)/(x\log x)$, tends to zero, since, by hypothesis, $S(x)/x$ converges, and hence is bounded. To show that the limit $L(f) = \lim_{x\to\infty} T(x)/\log x$ exists and is equal to A, it remains therefore to show that the integral $I(x)$ satisfies

$$\lim_{x\to\infty} \frac{I(x)}{\log x} = A. \tag{2.10}$$

Let $\epsilon > 0$ be given. By our assumption $\lim_{t\to\infty} S(t)/t = A$, there exists $t_0 = t_0(\epsilon) \ge 1$ such that $|S(t)/t - A| \le \epsilon$ for $t \ge t_0$. Moreover, for $1 \le t \le t_0$ we have

$$\left| \frac{S(t)}{t} - A \right| \le \frac{1}{t} \sum_{n\le t_0} |f(n)| + |A| \le \sum_{n\le t_0} |f(n)| + |A| = K_0,$$

say, where $K_0 = K_0(\epsilon)$ is a constant depending on ϵ. Hence, for $x \geq t_0$ we have

$$\begin{aligned} |I(x) - A\log x| &= \left|\int_1^x \frac{S(t)/t - A}{t} dt\right| \\ &\leq \int_1^{t_0} \frac{K_0}{t} dt + \int_{t_0}^x \frac{\epsilon}{t} dt \\ &\leq K_0 \log t_0 + \epsilon \log(x/t_0) \\ &\leq K_0 \log t_0 + \epsilon \log x. \end{aligned}$$

Since ϵ was arbitrary, it follows that

$$\limsup_{x\to\infty} \frac{|I(x) - A\log x|}{\log x} = 0,$$

which is equivalent to (2.10). □

In the other direction (going from $L(f)$ to $M(f)$) the method fails; indeed, the converse of Theorem 2.13 is false:

Theorem 2.14. *There exist arithmetic functions f such that $L(f)$ exists, but $M(f)$ does not exist.*

Proof. Define a function f by $f(n) = n$ if $n = 2^k$ for some nonnegative integer k, and $f(n) = 0$ otherwise. This function does not have an ordinary mean value since the average $(1/x)\sum_{n\leq x} f(n)$, as a function of x, has a jump of size 1 at all powers of 2, and hence does not converge as $x \to \infty$. However, f has a logarithmic mean value (namely $1/\log 2$), since

$$\begin{aligned} \frac{1}{\log x}\sum_{n\leq x}\frac{f(n)}{n} &= \frac{1}{\log x}\sum_{2^k\leq x}\frac{2^k}{2^k} \\ &= \frac{1}{\log x}\left(\left[\frac{\log x}{\log 2}\right] + 1\right) = \frac{1}{\log 2} + O\left(\frac{1}{\log x}\right). \end{aligned}$$ □

For an arithmetic function to have an asymptotic mean value is therefore a stronger condition than having a logarithmic mean value, and the existence of an asymptotic mean value is usually much harder to prove than the existence of a logarithmic mean value. For example, it is relatively easy to prove (as we will see in the next chapter) that the von Mangoldt function $\Lambda(n)$ has logarithmic mean value 1, and the Moebius function $\mu(n)$ has logarithmic mean value 0. By contrast, the existence of an ordinary asymptotic mean value for Λ or μ is equivalent to the prime number theorem and much more difficult to establish.

Failure of partial summation. It is tempting to try to use partial summation in an attempt to show that the existence of $L(f)$ implies that of $M(f)$. Of course, since this implication is not true, such an approach is bound to fail, but it is instructive to see what exactly goes wrong if one tries to apply partial summation in the "converse" direction. Thus, assume that $L(f)$ exists and is equal to A. In an attempt to show that $M(f)$ exists as well and is equal to A, one would start with the sum $S(x) = \sum_{n\le x}(f(n)/n)n$, and then "remove" the factor n by partial summation. Applying partial summation as in (2.9), but with the roles of $S(x)$ and $T(x)$ interchanged, gives the identity

$$S(x) = xT(x) - \int_1^x T(t)dt,$$

so to show that $M(f) = A$ we would need to show

$$\frac{S(x)}{x} = T(x) - (1/x)\int_1^x T(t)dt \to A \quad (x \to \infty). \tag{2.11}$$

However, the assumption that f has logarithmic mean value A is equivalent to the estimate $T(x) = A\log x + o(\log x)$, and substituting this estimate into (2.11) introduces an error term $o(\log x)$ that prevents one from drawing any conclusions about the convergence of $S(x)/x$ in (2.11). To obtain (2.11) would require a much stronger estimate for $T(x)$, in which the error term is $o(1)$ instead of $o(\log x)$.

The reason why (2.11) is so ineffective is because the right-hand side is a difference of two large terms of nearly the same size, both of which are much larger than the left-hand side. By contrast, the right-hand side of (2.9) is a sum of two expressions, each of the same (or smaller) order of magnitude than the function on the left.

The logarithmic mean value as an average version of the asymptotic mean value. Further insight into the relation between the asymptotic and logarithmic mean values is provided by rewriting the identity (2.9) in terms of the functions

$$\mu(t) = m(e^t), \quad \lambda(t) = l(e^t),$$

where

$$m(x) = \frac{1}{x}\sum_{n\le x} f(n), \quad \lambda(x) = \frac{1}{\log x}\sum_{n\le x}\frac{f(n)}{n}$$

are the finite asymptotic, resp. logarithmic, mean values. Assuming that $m(x) = o(\log x)$ (a very mild assumption that holds, for example, if the function f is bounded), (2.9) becomes

$$\lambda(t) = o(1) + \frac{1}{t}\int_0^t \mu(s)ds.$$

Thus, the convergence of $\lambda(t)$ (i.e., the existence of a logarithmic mean value) is equivalent to the convergence of $\overline{\mu}(t) = (1/t)\int_0^t \mu(s)ds$, i.e., the convergence of a certain average of $\mu(s)$, the ordinary (finite) mean value. It is obvious that if a function $\mu(t)$ converges, then so does its average $\overline{\mu}(t)$, and it is also easy to construct functions $\mu(t)$ for which the converse does not hold. Interpreting $M(f)$ as the limit of a function $\mu(t)$, and $L(f)$ as the limit of the corresponding average function $\overline{\mu}(t)$, it is then clear that the existence of $M(f)$ implies that of $L(f)$, but not vice versa.

2.3.4 Dirichlet series and summatory functions

As a final illustration of the use of partial summation, we prove an integral representation for the so-called Dirichlet series of an arithmetic function.

Given an arithmetic function f, the *Dirichlet series of f* is the (formal) infinite series

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s},$$

where s is any complex number. The following result gives a representation of this series as a certain integral involving the partial sums $S(x) = \sum_{n\le x} f(n)$.

Theorem 2.15 (Mellin transform representation of Dirichlet series)**.** *Let f be an arithmetic function, let $S_f(x) = \sum_{n\le x} f(n)$ be the associated summatory function, and let $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s}$ be the "Dirichlet series" associated with f, whenever the series converges.*

(i) For any complex number s with $\operatorname{Re} s > 0$ such that $F(s)$ converges we have

$$F(s) = s\int_1^{\infty} \frac{S_f(x)}{x^{s+1}}dx, \tag{2.12}$$

(ii) If $S_f(x) = O(x^{\alpha})$ for some $\alpha \ge 0$, then $F(s)$ converges for all complex numbers s with $\operatorname{Re} s > \alpha$, and (2.12) holds for all such s.

The identity (2.12) can be interpreted in terms of so-called Mellin transforms. Given a function $\phi(x)$ defined on the positive real axis, the *Mellin transform* of ϕ is the function $\hat{\phi}$ defined by

$$\hat{\phi}(s) = \int_0^\infty \phi(s) x^{-s} dx,$$

provided the integral exists. In this terminology (2.12) says that $F(s)/s$ is the Mellin transform of the function $S_f(x)/x$ (with the convention that $S_f(x) = 0$ if $x < 1$).

Proof. (i) Suppose that $F(s)$ converges for some s with $\operatorname{Re} s > 0$. We want to apply partial summation to remove the factor n^{-s} in the summands of $F(s)$ in order express $F(s)$ in terms of the partial sums $S_f(x)$. Since $F(s)$ is an infinite series, we cannot apply the partial summation formula directly to $F(s)$. However, we can apply it to the partial sums $F_N(s) = \sum_{n=1}^N f(n) n^{-s}$ and obtain, for any positive integer N,

$$F_N(s) = \frac{S_f(N)}{N^s} + s \int_1^N \frac{S_f(x)}{x^{s+1}} dx. \tag{2.13}$$

Now let $N \to \infty$ on both sides of this identity. Since, by assumption, the series $F(s)$ converges, the partial sums $F_N(s)$ on the left tend to $F(s)$. Also, by Kronecker's Lemma (Theorem 2.12), the convergence of $F(s)$, along with the hypothesis $\operatorname{Re} s > 0$, implies that the first term on the right, $S_f(N)/N^s$, tends to zero. Hence the integral on the right-hand side converges as N tends to infinity, and we obtain (2.12).

(ii) Suppose that $S_f(x) = O(x^\alpha)$ for some $\alpha > 0$, let s be a complex number with $\operatorname{Re} s > \alpha$, and set $\delta = \operatorname{Re} s - \alpha > 0$. We again apply (2.13), first for fixed finite $N \in \mathbb{N}$, and then let $N \to \infty$. First note that, by our assumptions on $S_f(x)$ and s, the term $S_f(N)N^{-s}$ is of order $O(N^{\alpha - \operatorname{Re} s}) = O(N^{-\delta})$ and hence tends to zero as $N \to \infty$. Also, the integrand $S_f(x)x^{-s-1}$ in the integral on the right of (2.13) is of order $O(x^{-1-\delta})$, so this integral is absolutely convergent when extended to infinity. Letting $N \to \infty$, we conclude that the limit $\lim_{N\to\infty} F_N(s)$ exists and is equal to $s \int_1^\infty S_f(x) x^{-s-1} dx$. But this means that $F(s)$ converges and the identity (2.12) holds. □

2.4 Approximating an arithmetic function by a simpler arithmetic function: The convolution method

2.4.1 Description of the method

Among the various methods for estimating sums of arithmetic functions, one of the most widely applicable is the "convolution method" presented in this section. The basic idea of is as follows. Given an arithmetic function f whose partial sums $F(x) = \sum_{n\le x} f(n)$ we want to estimate, we try to express f as a convolution $f = f_0 * g$, where f_0 is a function that approximates f (in a suitable sense) and which is well-behaved in the sense that good estimates for the partial sums $F_0(x) = \sum_{n\le x} f_0(n)$ are available, and where g is a "perturbation" that is small (again in a suitable sense). Writing $f(n) = \sum_{d|n} g(d)f_0(n/d)$, we have

$$
\begin{aligned}
(2.14)\qquad F(x) = \sum_{n\le x} f(n) &= \sum_{n\le x}\sum_{d|n} g(d)f_0(n/d) = \sum_{d\le x} g(d) \sum_{\substack{n\le x\\ d|n}} f_0(n/d)\\
&= \sum_{d\le x} g(d) \sum_{n'\le x/d} f_0(n') = \sum_{d\le x} g(d)F_0(x/d).
\end{aligned}
$$

Substituting known estimates for $F_0(y)$ then yields an estimate for $F(x) = \sum_{n\le x} f(n)$.

We call this method the convolution method, since the idea of writing an unknown function as a convolution of a known function with a perturbation factor is key to the method.

In practice, the approximating function f_0 is usually a very simple and well-behaved function, such as the function 1, or the identity function $f(n) = n$, though other choices are possible, too. In most applications the function f is multiplicative, and an appropriate approximation is usually easily obtained by taking for f_0 a simple multiplicative function whose values on primes are similar (or equal) to the corresponding values of f.

The following examples illustrate typical situations in which the method can be successfully applied, along with appropriate choices of the approximating function. We will carry out the argument in detail for two of these cases.

Examples

(1) $f(n) = \phi(n)$: f is multiplicative with $f(p) = p - 1$ for all primes p. Thus, a natural approximation to f is provided by the identity function id, which at a prime p has value p. The identity $\phi * 1 = \mathrm{id}$ proved earlier implies $\phi = \mathrm{id} * \mu$, so we have $\phi = f_0 * g$ with $f_0 = \mathrm{id}$ and $g = \mu$. The estimation of $\sum_{n \le x} \phi(n)$ will be carried out in detail in Theorem 2.16 below.

(2) $f(n) = \sigma(n)$: This case is very similar to the previous example. The function $\sigma(n)$ is multiplicative with values $\sigma(p) = p + 1$ at primes, and choosing id as the approximating function f_0 leads to an estimate for $\sum_{n \le x} \sigma(n)$ of the same quality as the estimate for $\sum_{n \le x} \phi(n)$ given in Theorem 2.16.

(3) $f(n) = \phi(n)/n$: f is multiplicative with $f(p) = 1 - 1/p$ for all primes p, so $f_0 = 1$ is a natural choice for an approximating function. The corresponding perturbation factor is $g = \mu/\mathrm{id}$ which can be seen as follows: Starting from the identity $\phi = (\mathrm{id} * \mu)$, we obtain $f = \phi/\mathrm{id} = (\mathrm{id} * \mu)/\mathrm{id}$. Since the function $1/\mathrm{id}$ is completely multiplicative, it "distributes" over the Dirichlet product (see Theorem 1.10), so $f = (\mathrm{id} * \mu)/\mathrm{id} = 1 * \mu/\mathrm{id}$.

(4) $f(n) = \mu^2(n)$: f is multiplicative with values 1 at all primes p, so $f_0 = 1$ serves as the obvious approximating function. See Theorem 2.18 below for a detailed argument in this case.

(5) $f(n) = \lambda(n)$: Suppose we have information on the behavior of $M(x) = \sum_{n \le x} \mu(n)$, such as the relation $M(x) = o(x)$ (a result which, as we will see in the next chapter, is equivalent to the prime number theorem), or the relation $M(x) = O_\epsilon(x^{1/2+\epsilon})$ for $\epsilon > 0$ (which is equivalent to the Riemann Hypothesis). Applying the convolution method with $f = \lambda$ and $f_0 = \mu$ then allows one to show that estimates of the same type hold for $\lambda(n)$.

(6) $f(n) = 2^{\omega(n)}$: Since $\omega(n)$, the number of distinct prime divisors of n, is an additive function, the function $f = 2^\omega$ is multiplicative. At primes f has the same values as the divisor function. This suggests to apply the convolution method with the divisor function as the approximating function, and to try to derive estimates for the partial sums of f from estimates for the partial sums of the divisor function provided by Dirichlet's theorem (see Theorem 2.20 in the following section). This

approach works, and it yields an estimate for $\sum_{n\le x} 2^{\omega(n)}$ of nearly the same quality as Dirichlet's estimate for $\sum_{n\le x} d(n)$.

2.4.2 Partial sums of the Euler phi function

We will prove the following estimate.

Theorem 2.16. *We have*

$$\sum_{n\le x}\phi(n) = \frac{3}{\pi^2}x^2 + O(x\log x) \quad (x\ge 2). \tag{2.15}$$

Before proving this result, we present some interesting applications and interpretations of the result.

Number of Farey fractions of given order. Let Q be a positive integer. The **Farey fractions of order** Q are the rational numbers in the interval $(0,1]$ with denominator (in reduced form) at most Q. From the definition of $\phi(n)$ it is clear that $\phi(n)$ represents the number of rational numbers in the interval $(0,1]$ that in reduced form have denominator n. The sum $\sum_{n\le Q}\phi(n)$ is therefore equal to the number of rationals in the interval $(0,1]$ with denominators $\le Q$, i.e., the number of Farey fractions of order Q. The theorem shows that this number is equal to $(3/\pi^2)Q^2 + O(Q\log Q)$.

Lattice points visible from the origin. A second application of the theorem is obtained by interpreting the pairs (n,m), with $1\le m\le n$ and $(m,n)=1$ as lattice points in the plane. The number of such pairs is equal to the sum $\sum_{n\le x}\phi(n)$ estimated in the theorem. It is easy to see that the condition $(m,n)=1$ holds if and only if the point (m,n) is visible from the origin, in the sense that the line segment joining this point with the origin does not pass through another lattice point. The theorem therefore gives an estimate for the number of lattice points in the triangular region $0<n\le x$, $0<m\le n$, that are visible from the origin. By a simple symmetry argument, it follows that the total number of lattice points in the first quadrant that are visible from the origin and have coordinates at most x is $(6/\pi^2)x^2 + O(x\log x)$.

Probability that two random integers are coprime. Defining this probability as the limit

$$\lim_{N\to\infty} \frac{1}{N^2} \#\{(n,m) : 1 \le n, m \le N, (n,m) = 1\},$$

we see from the previous application that this limit exists and is equal to $6/\pi^2$.

Proof of Theorem 2.16. We apply the identity (2.14) with $f = \phi$ and $f_0 = \mathrm{id}$ as the approximating function. As noted above, the identity $\mathrm{id} = \phi * 1$ implies $\phi = \mathrm{id} * \mu$, so we have $g = \mu$. Moreover, the summatory function of $f_0 (= \mathrm{id})$ equals

$$F_0(x) = \sum_{n\le x} n = \frac{1}{2}[x]([x]+1) = \frac{1}{2}x^2 + O(x).$$

Substituting this estimate into (2.14) gives

$$\begin{aligned}
\sum_{n\le x} \phi(n) &= \sum_{d\le x} \mu(d) \left(\frac{1}{2} \left(\frac{x}{d}\right)^2 + O\left(\frac{x}{d}\right) \right) \\
&= \frac{1}{2}x^2 \sum_{d\le x} \frac{\mu(d)}{d^2} + O\left(x \sum_{d\le x} \frac{1}{d} \right) \\
&= \frac{1}{2}x^2 \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} + O\left(x^2 \sum_{d>x} \frac{1}{d^2} \right) + O\left(x \sum_{d\le x} \frac{1}{d} \right).
\end{aligned}$$

(Note here that for the estimation of $\sum_{d\le x} \mu(d)d^{-2}$ in the last step we used the "trick" of extending the sum to infinity and estimating the tail of the infinite series. This is a very useful device that can be applied to any finite sum that becomes convergent, and hence equal to a constant, when the summation is extended to infinity.) Since $\sum_{d>x} d^{-2} \ll 1/x$ (e.g., by Euler's summation formula, or, simpler, by noting the sum is $\le \int_{x-1}^{\infty} t^{-2}dt = (x-1)^{-1}$) and $\sum_{d\le x} 1/d \ll \log x$ for $x \ge 2$, the two error terms are of order $O(x)$ and $O(x\log x)$, respectively, while the main term is Cx^2, with $C = (1/2)\sum_{d=1}^{\infty} \mu(d)/d^2$.

To complete the proof, it remains to show that the constant C is equal to $3/\pi^2$. This follows from the following lemma. □

Lemma 2.17. *We have* $\sum_{n=1}^{\infty} \mu(n)n^{-2} = 6/\pi^2$.

Proof. By the Moebius identity $e(n) = \sum_{d|n} \mu(d)$ we have

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} \frac{e(n)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{d|n} \mu(d) \\ &= \sum_{d=1}^{\infty} \sum_{m=1}^{\infty} \frac{\mu(d)}{(dm)^2} \\ &= \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} \sum_{m=1}^{\infty} \frac{1}{m^2}. \end{aligned}$$

Hence

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^2} = \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \right)^{-1}.$$

By Theorem A.1 of the Appendix, the sum $\sum_{n=1}^{\infty} n^{-2}$ is equal to $\pi^2/6$. The result now follows. □

2.4.3 The number of squarefree integers below x

Since $\mu^2(n)$ is the characteristic function of the squarefree integers, the summatory function of μ^2 is the counting function for the squarefree integers. The following theorem gives an estimate for this function.

Theorem 2.18. *We have*

$$\sum_{n \le x} \mu^2(n) = \frac{6}{\pi^2} x + O(\sqrt{x}) \quad (x \ge 1). \tag{2.16}$$

Thus, the "probability" that a random integer is squarefree is $6/\pi^2 = 0.6079\ldots$.

Proof. Since the function μ^2 is multiplicative and equal to 1 at primes, it is natural to apply the convolution method with $f_0 = 1$ as approximating function. Writing $\mu^2 = f_0 * g = 1 * g$, we have $g = \mu^2 * \mu$ by Moebius inversion.

We begin by explicitly evaluating the function g. Since μ^2 and μ are multiplicative functions, so is the function g, and its value at a prime power p^m is given by

$$g(p^m) = \sum_{k=0}^{m} \mu^2(p^k)\mu(p^{m-k}) = \mu(p^m) + \mu(p^{m-1}) = \begin{cases} 0 & \text{if } m = 1, \\ -1 & \text{if } m = 2, \\ 0 & \text{if } m \ge 3. \end{cases}$$

It follows that $g(n) = 0$ unless $n = m^2$ where m is squarefree, and in this case $g(n) = \mu(m)$. In fact, since $\mu(m) = 0$ if m is not squarefree, we have $g(m^2) = \mu(m)$ for all positive integers m, and $g(n) = 0$ if n is not a square.

The identity (2.14) with g defined as above and $F_0(x) = \sum_{n\le x} 1 = [x]$ then gives

$$\begin{aligned}\sum_{n\le x}\mu^2(n) &= \sum_{d\le x} g(d)[x/d] = \sum_{m\le\sqrt{x}}\mu(m)\left(\frac{x}{m^2}+O(1)\right)\\ &= x\sum_{m\le\sqrt{x}}\frac{\mu(m)}{m^2} + O\left(\sum_{m\le\sqrt{x}}|\mu(m)|\right)\\ &= x\sum_{m=1}^{\infty}\frac{\mu(m)}{m^2} + O\left(x\sum_{m>\sqrt{x}}\frac{1}{m^2}\right) + O(\sqrt{x}).\end{aligned}$$

(Note again the trick of extending a convergent sum to an infinite series and estimating the tail.) The coefficient of x in the main term is $\sum_{m=1}^{\infty}\mu(m)/m^2 = 6/\pi^2$ by Lemma 2.17, the second of the two error terms is of the desired order of magnitude $O(\sqrt{x})$, and in view of the estimate $\sum_{n>y} 1/n^2 \ll 1/y$ the same holds for the first error term. The asserted estimate therefore follows. □

2.4.4 Wintner's mean value theorem

Given an arithmetic function f, we say that f has a **mean value** if the limit

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n\le x} f(n)$$

exists (and is finite), and we denote the limit by $M(f)$, if it exists. The concept of a mean value is a useful one, as many results in number theory can be phrased in terms of existence of a mean value. For example, as we will show in the next chapter, the prime number theorem is equivalent to the assertions $M(\Lambda) = 1$ and $M(\mu) = 0$; Theorem 2.18 above implies $M(\mu^2) = 6/\pi^2$; and a similar argument shows that $M(\phi/\mathrm{id}) = 6/\pi^2$.

As a first illustration of the convolution method, we prove a result due to A. Wintner, that gives a general sufficient condition for the existence of a mean value. Note that this theorem does not require the function f to be multiplicative.

Theorem 2.19 (Wintner's mean value theorem)**.** *Suppose $f = 1 * g$, where $\sum_{n=1}^{\infty} |g(n)|/n < \infty$. Then f has a mean value given by $M(f) = \sum_{n=1}^{\infty} g(n)/n$.*

As an illustration of this result, we consider again the function $f = \mu^2$. We have $f = 1 * g$, where the function $g = \mu^2 * \mu$ is given by $g(n) = \mu(m)$ if $n = m^2$ and $g(n) = 0$ if n is not a square, as shown in the proof of Theorem 2.18. Hence the series $\sum_{n=1}^{\infty} g(n)/n$ equals $\sum_{m=1}^{\infty} \mu(m)/m^2$, which converges absolutely, with sum $6/\pi^2$. Wintner's theorem therefore applies and shows that μ^2 has mean value $6/\pi^2$, as we had obtained in Theorem 2.18. (Of course, a direct application of the convolution method, as in the proof of Theorem 2.18, may yield more precise estimates with explicit error terms in any given case. The main advantage of Wintner's mean value theorem lies in its generality.)

Proof. Applying again the identity (2.14) with $f_0 = 1$, $F_0(x) = [x]$, we obtain

$$\begin{aligned}\frac{1}{x}\sum_{n\le x} f(n) &= \frac{1}{x}\sum_{d\le x} g(d)[x/d] \\ &= \sum_{d=1}^{\infty} \frac{g(d)}{d} + O\left(\sum_{d>x} \frac{|g(d)|}{d}\right) + O\left(\frac{1}{x}\sum_{d\le x} |g(d)|\right).\end{aligned}$$

As $x \to \infty$, the first of the two error terms tends to zero, by convergence of the series $\sum_{d=1}^{\infty} |g(d)|/d$. The same is true for the second error term, in view of Kronecker's Lemma (Theorem 2.12) and the hypothesis that $\sum_{d=1}^{\infty} |g(d)|/d$ converges. Hence, as $x \to \infty$, the left-hand side converges to the sum $\sum_{d=1}^{\infty} g(d)/d$, i.e., $M(f)$ exists and is equal to the value of this sum. □

2.5 A special technique: The Dirichlet hyperbola method

2.5.1 Sums of the divisor function

In this section we consider a rather special technique, the "Dirichlet hyperbola method," invented by Dirichlet to estimate the partial sums of the divisor function. Dirichlet's result is as follows:

Theorem 2.20 (Dirichlet). *We have*

$$\sum_{n\le x} d(n) = x\log x + (2\gamma-1)x + O(\sqrt{x}) \quad (x\ge 1),$$

where γ is Euler's constant (see Theorem 2.5).

Proof. Let $D(x)=\sum_{n\le x} d(n)$. Writing $d(n)=\sum_{ab=n} 1$, where a and b run over positive integers with product n, we obtain

$$D(x)=\sum_{n\le x}\sum_{ab=n} 1 = \sum_{\substack{a,b\le x\\ ab\le x}} 1.$$

Note that, in the latter sum, the condition $ab\le x$ forces at least one of a and b to be $\le\sqrt{x}$. The key idea now is to split this sum into $\sum_1+\sum_2-\sum_3$, where

$$\sum\nolimits_1=\sum_{a\le\sqrt{x}}\sum_{b\le x/a}, \quad \sum\nolimits_2=\sum_{b\le\sqrt{x}}\sum_{a\le x/b}, \quad \sum\nolimits_3=\sum_{a\le\sqrt{x}}\sum_{b\le\sqrt{x}}.$$

The last sum, $\sum_3$, here compensates for the overlap, i.e., those terms (a,b) that are counted in both $\sum_1$ and $\sum_2$.

The last sum is trivial to estimate. We have

$$\sum\nolimits_3=\left(\sum_{a\le\sqrt{x}} 1\right)\left(\sum_{b\le\sqrt{x}} 1\right)=[\sqrt{x}]^2=(\sqrt{x}+O(1))^2=x+O(\sqrt{x}).$$

Also, $\sum_2=\sum_1$, so it remains to estimate $\sum_1$.

This is rather straightforward, using the estimate for the partial sums of the harmonic series (Theorem 2.5). We have

$$\begin{aligned}\sum\nolimits_1&=\sum_{a\le\sqrt{x}}\left[\frac{x}{a}\right]=x\sum_{a\le\sqrt{x}}\frac{1}{a}+O\left(\sum_{a\le\sqrt{x}} 1\right)\\&=x\left(\log\sqrt{x}+\gamma+O\left(\frac{1}{\sqrt{x}}\right)\right)+O(\sqrt{x})\\&=\frac{1}{2}x\log x+\gamma x+O(\sqrt{x}).\end{aligned}$$

Hence

$$\sum\nolimits_1+\sum\nolimits_2-\sum\nolimits_3=x\log x+(2\gamma-1)x+O(\sqrt{x}),$$

which is the desired estimate. □

2.5.2 Extensions and remarks

Geometric interpretation. The argument in this proof has the following simple geometric interpretation, which explains why it is called the "hyperbola method." The sum $D(x)$ is equal to the number of pairs (a,b) of positive integers with $ab \le x$, i.e., the number of lattice points in the first quadrant (not counting points on the coordinate axes) that are to the left of the hyperbola $ab = x$. The sums $\sum_1$ and $\sum_2$ count those points which, in addition, fall into the infinite strips defined by $0 < a \le \sqrt{x}$, and $0 < b \le \sqrt{y}$, respectively, whereas $\sum_3$ counts points that fall into the intersection of these two strips. It is geometrically obvious that $\sum_1 + \sum_2 - \sum_3$ is equal to $D(x)$, the total number of lattice points located in the first quadrant and to the left of the hyperbola $ab = x$.

The hyperbola method for general functions. The method underlying the proof of Dirichlet's theorem can be generalized as follows. Consider a sum $F(x) = \sum_{n\le x} f(n)$, and suppose f can be represented as a convolution $f = g * h$. Letting $G(x)$ and $H(x)$ denote the partial sums of the functions $g(n)$ and $h(n)$, respectively, we can try to obtain a good estimate for $F(x)$ by writing $F(x) = \sum_1 + \sum_2 - \sum_3$ with

$$\sum\nolimits_1 = \sum_{a\le\sqrt{x}} g(a)H(x/a), \quad \sum\nolimits_2 = \sum_{b\le\sqrt{x}} h(a)G(x/b),$$
$$\sum\nolimits_3 = G(\sqrt{x})H(\sqrt{x}),$$

and estimating each of these sums individually. This can lead to better estimates than more straightforward approaches (such as writing $F(x) = \sum_{a\le x} g(a)H(x/a)$), provided good estimates for the functions $H(x)$ and $G(x)$ are available. The Dirichlet divisor problem is an ideal case, since here the functions g and h are identically 1, and $H(x) = G(x) = [x]$ for all $x \ge 1$. In practice, the usefulness of this method is limited to a few very special situations, which are similar to that of the Dirichlet divisor problem, and in most cases the method does not provide any advantage over simpler approaches. In particular, the convolution method discussed in the previous section has a much wider range of applicability, and for most problems this should be the first method to try.

Maximal order of the divisor function. Dirichlet's theorem gives an estimate for the "average order" of the divisor function, but the divisor function can take on values that are significantly smaller or significantly larger than this average. Regarding lower bounds, we have $d(n) = 2$ whenever n is prime, and this bound is obviously best possible. The problem of obtaining a similarly optimal upper bound is harder. It is easy to prove that $d(n)$ grows at a rate slower than any power of n, in the sense that, for any given $\epsilon > 0$ and all sufficiently large n, we have $d(n) \leq n^\epsilon$. This can be improved to $d(n) \leq \exp\{(1+\epsilon)(\log 2)(\log n)/(\log\log n)\}$, for any $\epsilon > 0$ and $n \geq n_0(\epsilon)$, a bound that is best-possible, in the sense that, if $1+\epsilon$ is replaced by $1-\epsilon$, it becomes false.

The Dirichlet divisor problem. Let $\Delta(x)$ denote the error term in Dirichlet's theorem, i.e., $\Delta(x) = \sum_{n\leq x} d(n) - x\log x - (2\gamma - 1)x$. Thus $\Delta(x) = O(\sqrt{x})$ by Dirichlet's theorem. The problem of estimating $\Delta(x)$ is known as the Dirichlet divisor problem and attained considerable notoriety. The problem is of interest, partly because it is a difficult problem that is still largely unsolved, but mainly because in trying to approach this problem one is led to other deep problems (involving so-called "exponential sums") which have connections with other problems in number theory, including the Riemann Hypothesis. Thus, significant progress on this problem will likely have ramifications on a host of other problems. Most of the known results are estimates of the form $(*)$ $\Delta(x) = O(x^\theta)$ with a certain constant θ. Dirichlet's theorem shows that one can take $\theta = 1/2$. In the other direction, G.H. Hardy proved in the early part of the 20th century that the estimate does not hold with a value of θ that is less than $1/4$. It is conjectured that $1/4$ is, in fact, the "correct" exponent, but this is still open. Nearly 100 years ago, G.F. Voronoii proved that one can take $\theta = 1/3 = 0.333\ldots$, but despite enormous efforts by many authors not much progress has been made: the current record for θ is near 0.31.

2.6 Exercises

2.1 For $x \geq e$ define $I(x) = \int_e^x \log\log t \, dt$. Obtain an estimate for $I(x)$ to within an error term $O(x/\log^2 x)$.

2.2 Let $f(x)$ and $g(x)$ be positive, continuous functions on $[0, \infty)$, and set $F(x) = \int_0^x f(y)dy$, $G(x) = \int_0^x g(y)dy$.

(i) Show (by a counterexample) that the relation

$$f(x) = o(g(x)) \quad (x \to \infty) \tag{1}$$

does *not* imply

$$F(x) = o(G(x)) \quad (x \to \infty). \tag{2}$$

(ii) Find an appropriate *general* condition on $g(x)$ under which the implication (1) $\Rightarrow$ (2) becomes true.

Remark: It is trivial to show that, if "o" is replaced by "O" in (1) and (2), then the implication holds. In other words, one can "pull out" a O-sign from an integral (provided the integrand is positive).

2.3 Show that if $f(x)$ satisfies $f(x) = x^2 + O(x)$, and f is differentiable *with nondecreasing derivative* $f'(x)$ for sufficiently large x, then $f'(x) = 2x + O(\sqrt{x})$.

Remark. While O-estimates can be integrated provided the range of integration is contained in the range of validity of the estimate, in general such estimates cannot be differentiated. The above problem illustrates a situation where, under certain additional conditions (namely, the monoticity of the derivative), differentiation of a O-estimate is allowed.

2.4 Let n be an integer ≥ 2 and p a positive real number. A useful estimate is

$$\left(\sum_{i=1}^n a_i\right)^p \asymp_{n,p} \sum_{i=1}^n a_i^p \quad (a_1, a_2, \ldots, a_n > 0).$$

Prove this estimate with explicit and *best-possible* values for the implied constants. In other words, determine the largest value of $c_1 = c_1(n,p)$ and the smallest value of $c_2 = c_2(n,p)$ such that

$$c_1 \sum_{i=1}^n a_i^p \leq \left(\sum_{i=1}^n a_i\right)^p \leq c_2 \sum_{i=1}^n a_i^p \quad (a_1, a_2, \ldots, a_n > 0).$$

2.5 Obtain an estimate for the sum $\sum_{n\le x}(\log n)/n$ with error term $O((\log x)/x)$.

2.6 Given a positive integer k, let $S_k(x) = \sum_{n\le x}(\log n - \log x)^k$. Estimate $S_k(x)$ to within an error $O_k((\log x)^k)$. Deduce that, for each k, the limit $\lambda_k = \lim_{x\to\infty}(1/x)S_k(x)$ exists (as a finite number), and obtain an explicit evaluation of the constant λ_k.

2.7 Obtain an estimate for the sum

$$S(x) = \sum_{2\le n\le x} \frac{1}{n\log n}$$

with error term $O(1/(x\log x)$.

2.8 Given an arithmetic function f define a mean value $H(f)$ by

$$H(f) = \lim_{x\to\infty} \frac{1}{x\log x}\sum_{n\le x} f(n)\log n,$$

if the limit exists. Show that $H(f)$ exists if and only if the ordinary mean value $M(f) = \lim_{x\to\infty}(1/x)\sum_{n\le x} f(n)$ exists.

2.9 Given an arithmetic function $a(n)$, $n = 1, 2, \dots$, and a real number $\alpha > -1$ define a mean value $M_\alpha(a)$ by

$$M_\alpha(a) = \lim_{x\to\infty} \frac{1+\alpha}{x^{1+\alpha}}\sum_{n\le x} n^\alpha a(n),$$

provided the limit exist. (In particular, $M_0(a) = M(a)$ is the usual asymptotic mean value of a.) Prove, using a rigorous $\epsilon - x_0$ argument, that the mean value $M_\alpha(a)$ exists if and only if the ordinary mean value $M(a) = M_0(a)$ exists. (As a consequence, if one of the mean values $M_\alpha(a)$, $\alpha > -1$, exists, then all of these mean values exist.)

2.10 Say that an arithmetic function f has an *analytic mean value A* if the Dirichlet series $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s}$ satisfies

$$(0) \qquad F(s) = \frac{A}{s-1} + o\left(\frac{1}{s-1}\right) \quad (s\to 1+).$$

Show that if f has a logarithmic mean value $L(f) = A$, then f also has an analytic mean value, and the two mean values are equal.

2.11 Let f be an arithmetic function having a non-zero mean value $M(f) = A$, and let α be a fixed real number. Obtain an asymptotic formula for the sums $\sum_{n\le x} f(n)n^{i\alpha}$.

2.12 Say that an arithmetic function f has a *strong logarithmic mean value* A, and write $L^*(f) = A$, if f satisfies an estimate of the form

$$\sum_{n\le x} \frac{f(n)}{n} = A\log x + B + o(1) \quad (x \to \infty)$$

for some constants A and B. This is obviously a stronger condition than the existence of a logarithmic mean value which would correspond to an estimate of the above form with $o(\log x)$ in place of $B + o(1)$.

(i) Show that, in contrast to the (ordinary) logarithmic mean value, this stronger condition is sufficient to imply the existence of the asymptotic mean value. That is, show that if f has a strong logarithmic mean value A in the above sense, then the ordinary mean value $M(f)$ also exists and is equal to A.

(ii) Is the converse true, i.e., does the existence of $M(f)$ imply that of a strong logarithmic mean value?

2.13 Let $\lambda > 1$ and $t \ne 0$ be fixed real numbers, and $S_{t,\lambda}(x) = \sum_{x<n\le\lambda x} n^{-1-it}$. Obtain an estimate for $S_{t,\lambda}(x)$ as $x \to \infty$ with error term $O_{t,\lambda}(1/x)$. Deduce from this estimate that, for any non-zero t and any $\lambda > 1$, the limit $\lim_{x\to\infty} |S_{t,\lambda}(x)|$ exists, and that, for given $t \ne 0$ and *suitable* choices of λ, this limit is non-zero. (Thus, by Cauchy's criterion, the series $\sum_{n=1}^{\infty} n^{-1-it}$ diverges for every real $t \ne 0$.)

2.14 Obtain an asymptotic estimate with error term $O(x^{1/3})$ for the number of squarefull integers $\le x$, i.e., for the quantity

$$S(x) = \#\{n \le x : p|n \Rightarrow p^2|n\}.$$

2.15 Let $f(n) = \sum_{d^3|n} \mu(d)$, where the sum runs over all cubes of positive integers dividing n. Estimate $\sum_{n\le x} f(n)$ with as good an error as you can get.

2.16 For any positive integer n define its squarefree kernel $k(n)$ by $k(n) = \prod_{p|n} p$. Obtain an estimate for $\sum_{n\le x} k(n)/n$ with error term $O(\sqrt{x})$.

2.17 (i) Obtain an estimate for the sum

$$S(x) = \sum_{\substack{n \le x \\ n \text{ odd}}} \frac{1}{n}, \quad x \ge 1,$$

with error term $O(1/x)$. (The estimate should not involve any unspecified constants.)

(ii) Let

$$D(x) = \sum_{\substack{n \le x \\ n \text{ odd}}} d(n),$$

where $d(n)$ is the divisor function. Give an estimate for $D(x)$ with error term $O(\sqrt{x})$. As in (i), any constants arising in this estimate should be worked out explicitly. (Hint: Use Dirichlet's hyperbola method and the result of part (i).)

2.18 Obtain an estimate, of similar quality as Dirichlet's estimate for $\sum_{n \le x} d(n)$, for the sum $\sum_{n \le x} 2^{\omega(n)}$.

2.19 Obtain an estimate, similar to the estimate for $\sum_{n \le x} 1/n$ proved in Theorem 2.5, for the sum $\sum_{n \le x} 1/\phi(n)$. (Hint: Convolution method.)

2.20 Let $q_1 = 1, q_2 = 2, q_3 = 3, q_4 = 5 \ldots$ denote the sequence of squarefree numbers.

(i) Obtain an asymptotic estimate with error term $O(\sqrt{n})$ for q_n.

(ii) Show that there are arbitrarily large gaps in the sequence $\{q_n\}$, i.e., $\limsup_{n\to\infty}(q_{n+1} - q_n) = \infty$. (Hint: Chinese Remainder Theorem.)

(iii) Prove the stronger bound

$$\limsup_{n\to\infty} \frac{q_{n+1} - q_n}{\log n / \log\log n} \ge \frac{1}{2}.$$

(iv)* (Harder) Prove that (iii) holds with $1/2$ replaced by the constant $\pi^2/12$, i.e., that the limsup above is at least $\pi^2/12$.

2.21 Show that the inequality $\phi(n) \ge n/4$ holds for at least $1/3$ of all positive integers n (in the sense that if A is the set of such n, then $\liminf_{x\to\infty}(1/x)\#\{n \le x : n \in A\} \ge 1/3$). (Hint: use the fact that (1) $\sum_{n \le x} \phi(n) \sim (3/\pi^2)x^2$ (which was proved in Theorem 2.16) or (2) $\sum_{n \le x} \phi(n)/n \sim (6/\pi^2)x$ (an easy consequence of Wintner's theorem, or of (1), by partial summation).)

2.22 Let $f = 1 * g$. Wintner's theorem (Theorem 2.19) shows that if the series

$$\sum_{n=1}^{\infty} \frac{g(n)}{n} \tag{1}$$

converges *absolutely*, then the mean value $M(f)$ of f exists and is equal to the sum of the series (1).

(i) Show that the conclusion of Wintner's theorem remains valid if the series (1) converges only conditionally and if, in addition,

$$\limsup_{x\to\infty} \frac{1}{x} \sum_{n\le x} |g(n)| < \infty. \tag{2}$$

(ii) Show that condition (2) cannot be dropped; i.e., construct an example of a function g for which the series (1) converges, but the function $f = 1 * g$ does not have a mean value.

2.23 Using the Dirichlet hyperbola method (or some other method), obtain an estimate for the sum $\sum_{n\le x} d(n)/n$ with an error term $O((\log x)/\sqrt{x})$.