BULLETIN OF THE
HELLENIC MATHEMATICAL SOCIETY
Volume 66, 2022 (98–122)

THE ASYMPTOTIC DISTRIBUTION OF PRIME ELEMENTS

ADRIAN FELIX DOMINIK FELLHAUER

ABSTRACT. In this paper, many hitherto only conjecturally known distributions of prime elements are confirmed. More precisely, the author calculates the distribution of prime elements when only prime elements satisfying either of the following conditions are counted: They lie in a common arithmetic progression, they sum to a certain number with another prime, they lie in the image of polynomial tuples of which the other images are also prime or they lie in a comparatively small interval. The former two distributions are calculated in general arithmetic monoids with the according unavoidable large error terms, whereas the second two distributions are calculated using a different method, which yields better error terms if certain structural requirements are met. (There is a certain redundancy, because the second method may be used in place of the first.)

CONTENTS

I. SOME PRINCIPLES OF PRIME ELEMENT DISTRIBUTION

1.1. Sum sifting in arithmetic monoids.

Date: Received 14 July 2021. Accepted 3 December 2022.
2020 *Mathematics Subject Classification.* 11A41, 11B25, 11C08, 11D04, 11N05, 11N13, 11N25, 11N36, 11N69, 11N80, 11P32, 11R44.
Key words and phrases. Prime element distribution, sieve theory, patterns in prime element grids, asymptotic frequency.

Definition I.1. An *arithmetic monoid*[1] is a commutative unique factorisation monoid (cf. Růžička [Rů18]) G with countably many (prime) elements (the set of which shall be denoted by P) and a so-called *norm* $|\cdot| : G \to \mathbb{R}_+$ on G with the properties

(1) $|1| = 1$
(2) $\forall p \in P : |p| > 1$
(3) $\forall a, b \in G : |ab| = |a||b|$
(4) $N_G(x)$ is finite for all $x \geq 0$,

where

$$N_G(x) := \# \{p \in P : |p| \leq x\}.$$

The concept was discovered by John Knopfmacher, who explained it in his monograph Knopfmacher [Kno75]. Familiarity with this book is not required for understanding what follows, even though in fact some of the methods used therein are useful for obtaining precise error estimates for certain asymptotics, as is shown in a subsequent paper of mine. The study of the distribution of prime elements could also be carried out within the setting of Beurling primes (cf. [Beu37]) because of the functor that sends an arithmetic monoid to the Beurling primes $(|p|)_{p\in P}$ (see below for a norm-compatible well-order on P), but then the notation would become more complicated.

Mostly, we shall concern ourselves with arithmetic monoids whose elements are distributed such that $N_G(x)$ is approximately linear with an error term of a power less than one, that is, there is a constant $A > 0$ such that

$$N_G(x) = Ax + O(x^\delta), \ \delta < 1. \tag{1}$$

This requirement is usually denominated *axiom A* in the literature, cf. eg. Knopfmacher [Kno75, p. 75]. The case when

$$N_G(x) = Ax^\mu + O(x^\delta), \ \delta < \mu,$$

for an arbitrary $\mu > 0$ may then also be treated, since it is easily proven that in this case, we may replace $|\cdot|$ by $|\cdot|^\mu$ and then G satisfies Eq. (1), as Knopfmacher explained in his monograph (cf. [Kno75, p. 75]).

This first chapter shall treat the theorem of Rosser–Iwaniec [Iwa80] (namely, Theorem I.5), which is based on previous work by Brun [Bru19], regarding the sifting of sums that will subsequently be necessary. The chapter will more or less follow Terence Tao's lecture notes [Tao15], adapting it to our setting.

Proposition I.2. *Let G be an arithmetic monoid with norm $|\cdot|$ and primes P. Then there exists a strict well-order $\prec$ on P such that for all $p, q \in P$*

$$p \prec q \Rightarrow |p| \leq |q|.$$

Proof. Since P is countable, there exists an arbitrary well-order of P. We now construct the new order as follows: First, we partition P grouping together elements of equal norm, then we order each class (ie. elements of the same norm) according to the previous order, and finally we concatenate the orders of the classes ascendingly. □

In what follows, we shall fix such an order of the prime elements of the arithmetic monoid in question. This order will *not* be unique if (and only if) there are two primes of equal norm.

[1] This concept is usually called *arithmetic semigroup*, but since they are assumed to possess an identity, the author feels that it's better to refer to them as what they really are: Monoids.

Thus, for the positive integers, it is unique, but for the Gaussian integers, it is not unique; in fact, there are uncountably many possible choices.

What we want to do in this chapter is this: Consider the sum

$$\sum_{|g| \leq x} a_g,$$

where for each $g \in G$ an $a_g \in \mathbb{C}$ is given. This is what one would call an unsifted sum, and we are not really interested in its value, which in most cases is easy to evaluate (a typical case, and one that shall be considered below, would be $a_g = 1$ for all $g \in G$). Instead, what we want to evaluate is the *sifted sum*

$$\sum_{\substack{|g| \leq x \\ g \notin \cup_{q \prec p} E_q}} a_g,$$

where for each $p \in P$, the set E_p is an arbitrary subset of G. For example, E_p might be the set of all multiples of p, and a_g might be 1 for all $g \in G$. Then the latter sum becomes the cardinality of the set of elements $g \in G$ with $|g| \leq x$ such that g is not divisible by any q with $q \prec p$.

In many cases, the techniques presented in this chapter yield a decent approximation of such sifted sums when $|q|$ is not too large in comparison to x. If q is so large that $|q| \approx \sqrt{x}$, different techniques are required; these shall be laid out in Section II and Section III.

We shall now endeavour to derive a lower and an upper bound for the sifted sum. Using the Iverson bracket notation (cf. Knuth [Knu97, p. 32, eqn. (16)]), the sifted sum becomes

$$\sum_{|g| \leq x} a_g [g \notin \cup_{p \prec q} E_p].$$

The term $[g \notin \cup_{p \prec q} E_p]$ may now be rewritten. Indeed, the following proposition (when applied to the family of functions dependent on g that arises when multiplicatively continuing the expression $f_g(p) := [g \in E_p]$) makes this possible:

Proposition I.3. *Let G be an arithmetic monoid, and let $f : G \to \mathbb{C}$ be a function. If we define*

$$V(p) := \prod_{q \prec p} (1 - f(q)),$$

then

$$V(p) = 1 - \sum_{q \prec p} f(q) V(q).$$

Proof. We use induction on p, using the trivial equation

$$\prod_{q \prec p} (1 - f(q)) = \prod_{q \prec \operatorname{pre}(p)} (1 - f(q)) (1 - f(p))$$

(where $\operatorname{pre}(p)$ shall denote the predecessor of p regarding $\prec$), expanding the last brackets and applying the induction hypothesis to the first term of the result. □

Proposition I.4. *Let $f : G \to \mathbb{C}$ be multiplicative. Suppose that for each square-free $d \in G$ we are given a truth value $A(d) \in \{TRUE, FALSE\}$ with $A(1) = TRUE$. If we set $B(p_1 \cdots p_r) := \forall k \leq r : A(p_1 \cdots p_k)$ and if $p \in P$, then*

$$V(p) = \sum_{r=0}^{\infty} (-1)^r \sum_{\substack{p \succ p_1 \succ \cdots \succ p_r \\ B(p_1 \cdots p_r)}} f(p_1 \cdots p_r) + \sum_{r=0}^{\infty} (-1)^r \sum_{p_1 \cdots p_r \in \mathcal{E}_r} f(p_1 \cdots p_r) V(p_r)$$

with the convention that one sums the empty tuple also, where

$$\mathcal{E}_r := \left\{ p_1 \cdots p_r \in G : \begin{matrix} p \succ p_1 \succ \cdots \succ p_r \ \& \ \forall k < r : A(p_1 \cdots p_k) \\ \& \ \neg A(p_1 \cdots p_r) \end{matrix} \right\}.$$

Note that the sums over r are in fact finite, because the chains $p_1 \succ \cdots \succ p_r$ cannot grow arbitrarily large when $p \succ p_1$.

Proof. This formula follows from taking $m \to \infty$ in the more general formula

$$\begin{aligned} V(p) &= \sum_{r=0}^{m-1} (-1)^r \sum_{\substack{p \succ p_1 \succ \cdots \succ p_r \\ B(p_1 \cdots p_r)}} f(p_1 \cdots p_r) \\ &\quad - (-1)^m \sum_{\substack{p \succ p_1 \succ \cdots \succ p_m \\ B(p_1 \cdots p_m)}} f(p_1 \cdots p_m) V(p_m) \\ &\quad - \sum_{r=0}^{m} (-1)^r \sum_{p_1 \cdots p_r \in \mathcal{E}_r} f(p_1 \cdots p_r) V(p_r) \end{aligned}$$

which is proven by induction on m, where the induction step consists of applying Proposition I.3 to all the V's in the middle line.[2] □

This theorem may now be applied to the functions f_g that are the multiplicative extensions of $f_g(p) := [g \in E_p]$ (note that the dependency of V on f was omitted in the above notation). In this case the formula

$$V(p) = [g \notin \cup_{q \prec p} E_q]$$

holds, so that upon summing the applications of Proposition I.4 to each relevant f_g and changing the order of summation

$$\begin{aligned} \sum_{|g| \le x} a_g [g \notin \cup_{q \prec p} E_q] &= \sum_{r=0}^{\infty} (-1)^r \sum_{\substack{p \succ p_1 \succ \ldots \succ p_r \\ B(p_1 \cdots p_r)}} \sum_{|g| \le x} a_g [g \in E_{p_1 \cdots p_r}] \\ &+ \sum_{r=0}^{\infty} (-1)^r \sum_{p_1 \cdots p_r \in \mathcal{E}_r} \sum_{|g| \le x} a_g [g \in E_{p_1 \cdots p_r}][g \notin \cup_{q \prec p_r} E_q]. \end{aligned} \tag{2}$$

Here,

$$E_d := E_{p_1} \cap \cdots \cap E_{p_r}$$

for a square-free $d = p_1 \cdots p_r$. And now we see the first purpose of our labour: suppose that $a_g \ge 0$ for all $g \in G$. If we choose $A(p_1 \cdots p_r)$ so that it is true whenever r is even, the second summand on the right hand side of Eq. (2) will be non-positive, whereas it will be non-negative if we choose $A(p_1 \cdots p_r)$ to be always true when r is odd. This means that upon simply omitting this summand, we obtain an upper resp. lower bound for the sifted sum.

In order to see the second purpose, we note that for a square-free d, the term

$$X_d := \sum_{|g| \le x} a_g [g \in E_d],$$

[2]One such induction step is sometimes called a *Buchstab iteration*.

which is found in the first line of Eq. (2), is in many cases rather simple. For instance, if $a_g = 1$ for all g and $E_p := pG$, then

$$X_d = \frac{N_G(x)}{|d|} + \text{error},$$

where there is some error term that will later be small enough. But here, the function $d \mapsto \frac{1}{|d|}$ is again multiplicative, so that we may work our way backward: neglecting the error term, we have

$$\sum_{r=0}^{\infty}(-1)^r \sum_{\substack{p \succ p_1 \succ \dots \succ p_r \\ B(p_1 \cdots p_r)}} \sum_{|g| \le x} a_g [g \in E_{p_1 \cdots p_r}] \approx N_G(x) \sum_{r=0}^{\infty}(-1)^r \sum_{\substack{p \succ p_1 \succ \dots \succ p_r \\ B(p_1 \cdots p_r)}} \frac{1}{|p_1 \cdots p_r|},$$

and now we may apply Proposition I.4 to the right hand side, and if the error term (ie. the summand where the $\mathcal{E}_r$'s appear) is small, the result will be approximately equal to $V(p)N_G(x)$, and an appropriate version of Mertens' third theorem may then be used to approximate $V(p)$. Fortunately, there exists a choice of A that makes both error terms that ultimately occur sufficiently small:

Theorem I.5 (Brun–Rosser–Iwaniec)**.** *Suppose that f is such that the axiom*

$$\frac{V(q)}{V(p)} \ll \left(\frac{\ln(|p|)}{\ln(|q|)}\right)^k \quad (p \to \infty)$$

is satisfied uniformly in q; such an estimate follows in many cases of interest from Mertens' third theorem[3]*. Let $p \in P$. Then*

$$\sum_{r=0}^{\infty}(-1)^r \sum_{\substack{p \succ p_1 \succ \dots \succ p_r \\ B_\pm(p_1 \cdots p_r)}} f(p_1 \cdots p_r) = V(p)(1 + O_k(e^{-(1-\varepsilon)s\ln(s)})),$$

where $B_\pm = B_\pm(s)$ comes from an $A_+(s)$ resp. $A_-(s)$ that is suitably chosen among all "boolean-valued" functions on G such that

- *$A_\pm$ makes the error term in Eq.* (2) *non-positive resp. non-negative, and*
- *$B_\pm(d) = \mathit{FALSE}$ whenever $|d| > D(s) := |p|^s$.*

Note that in the equation in the proposition, the sum on the left hand side is the main term in Proposition I.4.

Proof. In fact, we give $A_+(s)$ and $A_-(s)$ explicitly, and then prove that these choices are suitable for our stated purposes. Our choice shall be thus: For $p_1 \succ \cdots \succ p_r$, we shall choose

$$A_\pm(p_1 \cdots p_r) := [|p_1| \cdots |p_{r-1}||p_r|^{1+\beta} \le D],$$

for a $\beta \in [1, s]$ to be chosen later, unless r is even (resp. odd), and then we shall set it to TRUE. The paragraph below Eq. (2) thus explains why the error term therein must be non-positive (resp. non-negative). If now $|d| > D$, we write $d = p_1 \cdots p_r$ with $p_1 \succ \cdots \succ p_r$; if then r is odd (even), the predicate will be false, and because $\beta \ge 1$, also in the other case. All that is now left to do is to prove the error estimate.

To this end, note that by the multiplicativity of f and Proposition I.4, the error term *within* the brackets takes the form

$$\sum_{r=0}^{\infty}(-1)^r \sum_{\substack{p \succ p_1 \succ \dots \succ p_r \\ p_1 \cdots p_r \in \mathcal{E}_r}} f(p_1) \cdots f(p_r) \frac{V(p_r)}{V(p)}.$$

[3]We will see that Mertens' third theorem holds if one imposes a certain condition on G.

We now show that with a suitable choice of β, this error term is as small as asserted in the theorem statement. First, we note that we are restricted to the case $r \le s-\beta$, because otherwise $A_\pm$ is trivially true. Now, we derive an upper bound for the fraction of V's in the error term. First note that due to the theorem's assumption,

$$\frac{V(p_r)}{V(p)} \ll \left(\frac{\ln(|p|)}{\ln(|p_r|)}\right)^k,$$

so that it is sufficient to give an upper bound of the latter expression. First, we derive a *lower* bound for $|p_r|$, because then the denominator is large. Thus, suppose that $p_1 \cdots p_r \in \mathcal{E}_r$. Because of the definition of $\mathcal{E}_r$, if $t < r$ has the same parity as r, then $|p_1| \cdots |p_{t-1}||p_t|^{1+\beta} \le D$, from which follows that $|p_1| \cdots |p_{t-1}||p_t|^{\beta} \le D$ for *all* $t < r$, regardless of the parity. Taking the β-th root, multiplying by D and dividing by $(|p_1| \cdots |p_{t-1}|)^{(\beta-1)/\beta}$, this inequality becomes

$$\frac{D}{|p_1| \cdots |p_t|} \ge \left(\frac{D}{|p_1| \cdots |p_{t-1}|}\right)^{\frac{\beta-1}{\beta}} \overset{\text{iteration}}{\Rightarrow} \frac{D}{|p_1| \cdots |p_{r-1}|} \ge D^{\left(\frac{\beta-1}{\beta}\right)^{r-1}}.$$

But the definition of $\mathcal{E}_r$ also implies that $|p_1| \cdots |p_{r-1}||p_r|^{1+\beta} > D$, so that we obtain

$$|p_r| > D^{\frac{\beta-1}{\beta}/(\beta+1)} > x^{\left(\frac{\beta-1}{\beta}\right)^r} \Rightarrow \frac{\ln(|p|)}{\ln(|p_r|)} < \left(\frac{\beta}{\beta-1}\right)^{kr}, \tag{3}$$

and from this and the triangle inequality we may infer that the error term is bounded by

$$\sum_{r>s-\beta} \left(\frac{\beta}{\beta-1}\right)^{kr} \sum_{\substack{p \succ p_1 \succ \cdots \succ p_r \\ p_1 \cdots p_r \in \mathcal{E}_r}} f(p_1) \cdots f(p_r).$$

The latter sum may now be bounded as thus:

$$\begin{aligned}\sum_{\substack{p \succ p_1 \succ \cdots \succ p_r \\ p_1 \cdots p_r \in \mathcal{E}_r}} f(p_1) \cdots f(p_r) &\le \frac{1}{r!}\left(\sum_{p_r \preceq q \prec p} f(q)\right)^r \le \frac{1}{r!}\left(\ln\left(\prod_{p_r \preceq q \prec p} (1-f(q))^{-1}\right)\right)^r \\ &= \frac{1}{r!}\left(\ln\left(\frac{V(p_r)}{V(p)}\right)\right)^r \ll \frac{1}{r!}\left(kr\ln\left(\frac{\beta}{\beta-1}\right)+O(1)\right)^r;\end{aligned}$$

of course, we don't compute any derivatives in order to obtain the Taylor series for $-\ln(1-x)$, but use the formula $\frac{1}{1-x} = \frac{d}{dx} - \ln(1-x)$, to the left hand side of which we apply the formula for the geometric series and integrate. The last expression in turn may be bounded as follows, using the trivial estimate $r^r/r! \le e^r$ that may be read off the Taylor expansion of e^r:

$$\frac{1}{r!}\left(kr\ln\left(\frac{\beta-1}{\beta}\right)+O(1)\right)^r \le \left(ke\ln\left(\frac{\beta}{\beta-1}\right)+O\left(\frac{1}{r}\right)\right)^r.$$

The integral definition of the logarithm shows that $\ln(\beta/(\beta-1)) = \ln(\beta) - \ln(\beta-1)$ decays like $1/\beta$, whence we may choose β so that

$$\left(ke\ln\left(\frac{\beta}{\beta-1}\right)+O\left(\frac{1}{r}\right)\right)\left(\frac{\beta}{\beta-1}\right)^k \ll \frac{1}{s^{1-\varepsilon'}};$$

here, $\beta \sim s^{1-\varepsilon}$ will do. The total error term then sums to

$$O_k\left(\left(\frac{1}{s^{1-\varepsilon'}}\right)^{s-\beta}\right),$$

which is easily seen to be acceptable upon choosing ε' adequately. □

1.2. **Approximation and error functions for rough numbers.** The prime elements have a cumulative distribution function given by

$$\pi_G(x) := \#\{p \in P : |p| < x\}.$$

Note that in this definition, a strict less-sign has been used. This is because later on, there is a computation where π defined in this way is useful, and there is otherwise no disadvantage in using this convention.

There is a related function, given by

$$\pi_G(x,y) := \#\{|g| < x : p|x \Rightarrow |p| \geq y\}, \tag{4}$$

where p denotes a prime element of G. This function equals the cardinality of rough numbers. The equation (which certainly holds under the assumption of Eq. (1))

$$\pi_G(x, \sqrt{x}) = \pi_G(x) + O(\sqrt{x}) \tag{5}$$

signifies that once one has sieved all primes up to $\sqrt{x}$ from the elements of norm $< x$, one is only left with primes.

Another fundamental identity concerning the prime elements was found by Buchstab [Buc37] in or before 1937:

Theorem I.6 (Sieve equation for $\mathbb{N}$)**.** *For all real numbers $\alpha > \beta > 0$,*

$$\pi(x, x^{1/\beta}) = \pi(x, x^{1/\alpha}) - \sum_{x^{1/\alpha} \leq p < x^{1/\beta}} \pi\left(\frac{x}{p}, p\right).$$

Proof. When we sift the prime p, we remove precisely the elements gp where g is not divisible by any prime smaller than p, such that $|gp| < x$. □

In this section, we are going to study relatively simple functions by which one may approximate the bivariate version of π_G (and related functions which describe, for instance, the number of prime tuples). It is not surprising that we shall see that they are approximate solutions to appropriate versions of Eq. (5) and Theorem I.6. Each has its own advantages and disadvantages:

(1) Buchstab's approximation is separated in the two variables, but not very precise,
(2) the approximation presented in this paper (which has not been spotted by the author in the literature and so is presumed new, even though its definition is very natural) is a bit more precise, continuously differentiable almost everywhere and defined in an obvious, natural way, but not separated in the two variables, and
(3) de Bruijn's approximation is by far the most precise and smooth in all variables, but in the author's opinion, computing with it is somewhat more cumbersome than with the one introduced in this paper.

1.2.1. *Buchstab's approximation.* Buchstab [Buc37] computed the following approximation to the cumulative distribution function for numbers of a certain roughness, given by Eq. (4):

$$\pi(x, x^{1/\alpha}) = \varphi(\alpha)\frac{x}{\ln(x)} + o_\alpha\left(\frac{x}{\ln(x)}\right), \ (\alpha > 1),$$ [4]

[4] In fact, this error term may be improved to $O_\alpha\left(\frac{x}{\ln(x)^2}\right)$, as will be demonstrated further on.

where φ is defined inductively by the equation

$$\varphi(\alpha) = \varphi(n) + \int_{n-1}^{\alpha-1} \frac{\varphi(z)}{z} dz$$

with the induction base being given by $\varphi(\alpha) = 1$ for $\alpha \in [1, 2]$. Since then, it seems to have become customary (cf. eg. [dB51a][p. 803]) to write the function φ as

$$\varphi(\alpha) = \alpha\omega(\alpha),$$

and under this equation the integral definition of φ may be rewritten as the following definition:

Definition I.7 (Buchstab's function)**.**

$$\begin{cases} \omega(\alpha) = \frac{1}{\alpha} & (\alpha \in [1, 2]) \\ \frac{d}{d\alpha}\left(\alpha\omega(\alpha)\right) = \omega(\alpha - 1) & (\alpha > 2). \end{cases} \tag{6}$$

It is the function ω and not the function φ which is now known as *Buchstab's function.* It rapidly converges to $e^{-\gamma}$ as $\alpha \to \infty$. First, we prove that Buchstab's function converges very quickly; the limit will then be computed in the subsequent proposition.

Proposition I.8. *There exists a constant $A \in \mathbb{R}$ such that*

$$\left|\omega(\alpha)^k - A^k\right| = O\left(\frac{1}{\Gamma(\alpha + 1)}\right) \ (\alpha \to \infty),$$

where there is no implied dependence on k in the Bachmann symbol.

Proof. For $\alpha > 2$ define

$$M(\alpha) := \sup_{\alpha \le \beta < \infty} \omega'(\beta).$$

From the definition of Buchstab's function (Definition I.7), we infer that for $t > 2$,

$$\omega'(t) = \frac{\omega(t-1) - \omega(t)}{t}, \tag{7}$$

whence the mean value theorem implies that $M(t) \le \frac{1}{t} M(t-1)$ and therefore $M(t) \le C/\Gamma(t+1)$ by induction.

Using the triangle inequality for the integral and the boundedness of ω on the interval $[2, \infty)$, we obtain for $\alpha > 2$

$$\left|\omega(\alpha)^k - \omega(\beta)^k\right| = \left|\int_\alpha^\beta k\omega(t)^{k-1}\omega'(t)dt\right| \le \frac{C}{\Gamma(\alpha + 1)} \int_\alpha^\beta \frac{\Gamma(\alpha + 1)}{\Gamma(t + 1)} dt,$$

where

$$\begin{aligned} \int_\alpha^\beta \frac{\Gamma(\alpha + 1)}{\Gamma(t + 1)} dt &\le 1 + \int_{\alpha+1}^{\alpha+2} \frac{1}{t} dt + \int_{\alpha+2}^\infty \frac{1}{t^2} dt \\ &\le 2 + \int_1^\infty \frac{1}{t^2} dt < \infty, \end{aligned}$$

whence $(\omega(\alpha))_{\alpha \ge 2}$ converges[5]. The constant C does not depend on k because ω is bounded away from 1 on any interval of the form $[a, \infty)$ with $a > 1$; indeed, this follows from ω's convergence

[5]It follows from what has been proved that $(\omega(\alpha))_{\alpha>0}$ is a Cauchy net, and upon taking the limit $\beta \to \infty$ and using the triangle inequality, we see the convergence.

and the mean value theorem, which, in conjunction with Eq. (7), yields ever smaller roots of ω'. ☐

Proposition I.9.

$$\lim_{\alpha\to\infty}\omega(\alpha)=e^{-\gamma}.$$

Proof. For this proof, we will use the integral definition of the logarithm. It is little known though that the n-th harmonic number too admits an integral representation. Indeed,

$$H_n=\int_0^n \frac{1-\left(1-\frac{t}{n}\right)^n}{t}dt.$$

This equation is proved as follows: First, note that

$$\frac{1}{k}=\int_0^1 x^{k-1}dx,$$

so that

$$\sum_{k=1}^n \frac{1}{k}=\int_0^1\sum_{k=1}^n x^{k-1}dx \overset{\text{geom. sum}}{=} \int_0^1 \frac{1-x^n}{1-x}dx.$$

Now all that is left to do is to transform variables according to the equation $x=1-t/n$. From this, we now deduce the following limit, which is essential to this proposition's proof:

$$\lim_{x\to\infty}\int_0^x \frac{1-e^{-t}}{t}dt-\ln(x)=\gamma. \tag{8}$$

Indeed, the function $x\mapsto(1-t/x)^x$ is non-increasing (which may be easily seen by computing the derivative), so that by the monotone convergence theorem

$$\begin{aligned}\lim_{x\to\infty}\int_0^x \frac{1-e^{-t}}{t}dt-\ln(x)&=\int_0^\infty \frac{[t\le 1]-e^{-t}}{t}dt\\&=\lim_{n\to\infty}\int_0^n \frac{[t\le 1]-\left(1-\frac{t}{n}\right)^n}{t}dt\\&=\lim_{n\to\infty}H_n-\ln(n)=\gamma.\end{aligned}$$

Now for $a\ge 2$ we define the scalar product

$$\langle f,g\rangle_a:=\int_{a-1}^a f(t)g(t)dt+ag(a)f(a-1)$$

on a suitable function space, for instance $C(\mathbb{R})$, the continuous functions on $\mathbb{R}$. Suppose that the function χ is a solution of the equation

$$a\chi'(a-1)+\chi(a)=0. \tag{9}$$

Then from the definition of Buchstab's function (Definition I.7) we may deduce without effort that $\langle\chi,\omega\rangle_a$ does not depend on a by differentiating in a. But Eq. (9) admits the explicit solution

$$\chi(a):=\int_0^\infty \exp\left(-ax-x+\int_0^x\frac{e^{-t}-1}{t}dt\right)$$

on all of $\mathbb{R}$; this may be seen by conducting an integration by parts, applying the product rule to the function

$$x \mapsto x \exp\left(\int_0^x \frac{e^{-t}-1}{t} dt\right).$$

Now it is easily seen that $\chi(a)$ grows like $1/a$ as $a \to \infty$; this is because disregarding parts of the argument of exp in the definition of χ, we may easily achieve upper and lower bounds that are simple shifts of the reciprocal function. Therefore, $\langle \chi, \omega \rangle_a$ equals the limit of ω as $a \to \infty$. On the other hand, applying Eq. (9),

$$\begin{aligned}\langle \chi, \omega \rangle_2 &= \int_1^2 \frac{\chi(t)}{t} dt + \chi(1) = \lim_{r \downarrow 0} \int_r^2 -\chi'(t-1) dt + \chi(1) \\ &= \lim_{r \downarrow 0} -\chi(r) = \lim_{r \downarrow 0} r\chi'(r-1) \\ &= \lim_{r \downarrow 0} -r \int_0^\infty \exp\left(-ux + \int_0^x \frac{e^{-t}-1}{t} dt + \ln(x)\right).\end{aligned}$$

But the latter limit equals $e^{-\gamma}$, because the outer integral may be split in two, where the former part is arbitrarily small, and the latter arbitrarily close to $e^{-\gamma}$ because of Eq. (8). □

The upper bound on the convergence rate given in Proposition I.8 is not optimal. A theorem of Jurkat–Richert [JR65][p. 227, 5.13] states that

$$|\omega(\alpha) - e^{-\gamma}| \leq \frac{\rho(\alpha-1)}{\alpha},$$

where ρ is the Ramanujan function, given as the continuous solution to the delay differential equation

$$\begin{cases} u\rho'(u-1) + \rho(u) = 0 & (u > 1) \\ \rho(u) = 1 & (u \in [0,1]) \end{cases},$$

which was introduced by Ramanujan [Ram20][p. 78] (cf. also one of the books of Andrews and Berndt [AB13, p. 185, Entry 8.2.1]) as the function that satisfies

$$\psi(x, x^{1/\alpha}) = \rho(\alpha)x + o(x)^6,$$

where

$$\psi(x,y) := \left|[1,x] \setminus \bigcup_{p>y} p\mathbb{N}\right|.$$

de Bruijn [dB51b][p. 26, 1.8] proved that

$$\rho(\alpha) \sim \frac{e^\gamma}{\sqrt{2\pi\alpha}} \exp\left(\int_0^{\xi(\alpha)} \frac{(s-1)e^s+1}{s} ds\right),$$

where $\xi(\alpha)$ is the positive solution to

$$e^\xi - 1 = \alpha\xi.$$

Combining this with the aforementioned result of Jurkat and Richert yields a sharper bound on the convergence rate of ω. Given the eminent position that Buchstab's function occupies in the theory of primes, it would be highly desirable to know its precise asymptotic behaviour.

[6] The first rigorous proof of this asymptotic is due to Ramaswami [Ram49].

There is an analogue of Buchstab's function for prime tuples. The function φ may be modified to yield a function φ_k $(k > 2)$ which satisfies

$$\varphi_k(\alpha) = \varphi_k(n) + k \int_{n-1}^{\alpha-1} \left(\frac{z+1}{z}\right)^n \varphi_k(z) dz.$$

This function will be used to approximate the number of numbers of a certain roughness less than a certain size that satisfy a k-dimensional sum sifting problem as follows:

$$\pi_k(x, x^{1/\alpha}) \sim C \frac{x}{\ln(x)^k} \varphi_k(\alpha),$$

where C is a constant that depends on the problem at hand and may be determined by carefully considering the respective sum sifting problem which is inserted into the Brun–Rosser–Iwaniec theorem.

Asymptotically, $\varphi_k(\alpha) \sim \varphi(\alpha)^k$; indeed, first we show by induction and the approximate identity $\varphi(\alpha) \approx \varphi(\alpha - 1)$ that the difference of the derivatives of the two functions converges so rapidly to zero that the difference of the two functions itself converges to a constant, and then we show that the constant must be zero.

It is prudent to also define

$$\omega_k(\alpha) := \frac{\varphi_k(\alpha)}{\alpha}.$$

That the aforementioned expressions really do constitute approximations to the respective prime counting functions will be proven in the next section for the one-dimensional case, and later for dimension ≥ 2. This is because the approximation functions introduced in this paper provide an easy-to-use tool for doing so; indeed, one may show the accuracy of the approximations introduced in this paper, and then that Buchstab's approximations approximate these in turn.

1.2.2. *The novel approximation.* From the sieve equation (Theorem I.6) and the prime number theorem, which implies that the density of the prime numbers around any $t > 2$ is approximately $1/\ln(t)$, we may guess that a better approximation to the number of numbers of a certain roughness below x may be given by

$$\pi(x, y) \approx h_1(x, y),$$

where h_1 is the bivariate function which is defined inductively by

$$h_1(x, y) := \int_y^x \frac{1}{\ln(t)} dt \quad (y \geq \sqrt{x})$$

and

$$h_1(x, x^{1/\alpha}) := h_1(x, x^{1/n}) + \int_{x^{1/\alpha}}^{x^{1/n}} h_1\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt$$

for $\alpha \in (n, n+1]$. For the case of polynomial tuples, there is a similar sieve equation, which shall be derived in Section 2.2. From this equation, it is seen in the same way that a good approximation to the rough number k-tuples is given by

$$h_k(x, y) := \int_y^x \frac{1}{\ln(t)^k} dt \quad (y \geq \sqrt{x}, l \in \mathbb{N}),$$

and

$$h_k(x, x^{1/\alpha}) := h_k(x, x^{1/n}) + k \int_{x^{1/\alpha}}^{x^{1/n}} h_k\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt$$

for $\alpha \in (n, n+1]$.

The best approximation to the number of rough numbers was given by de Bruijn [dB51a], who had had the brilliant insight to separate the product occurring in Mertens' third theorem from π: He wrote

$$\pi(x, y) = x \prod_{k=1}^{n} \left(1 - \frac{1}{|p_k|}\right) \psi(x, y),$$

where n is such that p_n is the largest prime less than y, and then he gave

$$\psi(y^u, y) \approx \ln(y) e^\gamma \int_1^u y^{t-u} \omega(t) dt.$$

In this chapter, we shall nevertheless investigate the second type of approximation function, because they are well-suited for a kind of algebraic manipulation that we will make use of. Yet, since de Bruijn's functions will be strictly necessary in the third chapter, the author believes that the textbook proof of the propositions put forth within this paper will only use de Bruijn's function, because it minimises the total workload and yet proves all propositions of interest to most readers. A secondary purpose of this paper, however, shall be the investigation of the functions h_k, which the author feels to be of interest in itself. Moreover, the novel approximation functions have close relatives which are well-suited as error functions and therefore yield a proof of certain asymptotics. These relatives we shall define by the following two equations:

$$h_{k,l}(x, y) := \int_y^x \frac{1}{\ln(t)^l} dt \quad (y \geq \sqrt{x}, l \in \mathbb{N})$$

and

$$h_{k,l}(x, x^{1/\alpha}) := h_{k,l}(x, x^{1/n}) + k \int_{x^{1/\alpha}}^{x^{1/n}} h_{k,l}\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt$$

for $\alpha \in (n, n+1]$. With these definitions, we have $h_k = h_{k,k}$, but we may choose $l > k$, and we shall see that $h_{k,l}$ is then well-suited as an upper bound for the error term, indeed for arbitrarily large l.

Using induction, one may prove that

$$h_{k,l}(x, z) = h_{k,l}(x, y) - k \int_y^z h_{k,l}\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt$$

whenever $\sqrt{x} \geq z \geq y$. In particular, setting $z = \sqrt{x}$, we obtain

$$h_{k,l}(x, y) = h_{k,l}(x, \sqrt{x}) + k \int_y^{\sqrt{x}} h_{k,l}\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt \tag{10}$$

for $y \leq \sqrt{x}$.

Proposition I.10.

$$\frac{\partial h_{k,l}}{\partial y}(x, y) = \begin{cases} -\frac{1}{\ln(y)^l} & (y > \sqrt{x}) \\ -k h_{k,l}\left(\frac{x}{y}, y\right) \frac{1}{\ln(y)} & (y < \sqrt{x}) \end{cases}$$

and

$$\frac{\partial h_{k,l}}{\partial x}(x,y) = \begin{cases} \frac{1}{\ln(x)^l} & (y > \sqrt{x}) \\ \frac{1}{\ln(x)^l} + \frac{2^{l-1}}{\ln(x)^l\sqrt{x}} + k\int_y^{\sqrt{x}} \frac{1}{t}\frac{\partial h_{k,l}}{\partial x}\left(\frac{x}{t},t\right)\frac{1}{\ln(t)}dt & \textit{otherwise} \end{cases}$$

Proof. The formulae follow from the fundamental theorem of calculus and the Leibniz integral rule respectively, when applied to equation (10). □

The following proposition describes the asymptotic behaviour of the functions h_l.

Proposition I.11. *Let $\alpha \in (n, n+1])$. There is a constant $C > 0$ such that*

$$\left| h_l(x, x^{1/\alpha}) - \varphi_{1,l}(\alpha)\frac{x}{\ln(x)^l} \right| \leq C\ln(n)\frac{x}{\ln(x)^{l+1}},$$

where $\varphi_{1,l}(\alpha) = 1$ for $\alpha \in [1,2]$ and

$$\varphi_{1,l}(\alpha) = \varphi_l(n) + \int_{n-1}^{\alpha-1} \frac{\varphi_{1,l}(z)}{z}\left(1+\frac{1}{z}\right)^{l-1} dz.$$

Proof. For $\alpha \leq 2$, we first observe that integration by parts yields

$$\int_{x^{1/\alpha}}^{x} \frac{1}{\ln(t)^l}dt = \left[\frac{t}{\ln(t)^l}\right]_{t=x^{1/\alpha}}^{t=x} + l\int_{x^{1/\alpha}}^{x} \frac{1}{\ln(t)^{l+1}}dt,$$

which proves the claim for $\alpha \leq 2$. For larger α, we set $n = \lceil \alpha - 1 \rceil$ and use the inductive definition of h:

$$\begin{aligned} h(x, x^{1/\alpha}) &= \varphi(n)\frac{x}{\ln(x)} + R(x,n,n-1) \\ &+ \int_{x^{1/\alpha}}^{x^{1/n}} \varphi\left(\log_t\left(\frac{x}{t}\right)\right)\frac{x}{t\ln\left(\frac{x}{t}\right)^l}\frac{1}{\ln(t)}dt \\ &+ \int_{x^{1/\alpha}}^{x^{1/n}} R\left(\frac{x}{t}, \log_t\left(\frac{x}{t}\right), n-1\right)\frac{1}{\ln(t)}dt, \end{aligned}$$

where R is the remainder that by the induction hypothesis satisfies

$$R(x,\alpha,n) \leq C\ln(n)\frac{x}{\ln(x)^{l+1}}.$$

Using the equation

$$\int_{x^{1/\alpha}}^{x^{1/n}} \frac{1}{t\ln(t)} = [\ln(\ln(t))]_{x^{1/\alpha}}^{x^{1/n}} = \ln(n) - \ln(\alpha) \leq \frac{1}{n-1}$$

(the inequality using the integral definition of the logarithm) and the asymptotics for the harmonic series, we see that $R(x,\alpha,n+1)$ satisfies the same constraints, and noting that

$$\log_t\left(\frac{x}{t}\right) = \frac{\ln(x)}{\ln(t)} - 1,$$

we may use the substitution

$$z = \frac{\ln(x)}{\ln(t)} - 1$$

to transform the former integral into the integral

$$\int_{n-1}^{\alpha-1} \frac{\varphi_l(z)}{z} \left(1+\frac{1}{z}\right)^{l-1} \frac{x}{\ln(x)^l} dz,$$

which is precisely what we wanted to obtain. □

An alternative formula for the derivative of h with respect to its first variable is given by

$$\frac{\partial h_l}{\partial x}(x, x^{1/\alpha}) = \frac{\partial h_l}{\partial x}(x, x^{1/n}) + \int_{x^{1/\alpha}}^{x^{1/n}} \frac{1}{t}\frac{\partial h_l}{\partial x}\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt \quad (\alpha \in (n, n+1]),$$

as one obtains by applying the Leibniz integral rule as well as Proposition I.10 to the inductive definition of h. From this formula one inductively deduces that *this partial derivative is positive*, and as in the proof of Proposition I.11 one is led to the bound

$$\frac{\partial h_l}{\partial x}(x, x^{1/\alpha}) \leq C \ln(n) \frac{1}{\ln(x)^l}, \tag{11}$$

where $\alpha \in (n, n+1]$.

Up to error terms which are sufficiently small for our purposes, the functions h_l satisfy the two equations for π mentioned in the introduction (Eq. (5) and Theorem I.6). In proving that it is so, we shall assume the prime number theorem.

Proposition I.12. *Let $l \in \mathbb{N}$. For all $y \geq \sqrt{x}$,*

$$h_l(x, y) = h_l(x, \sqrt{x}) - h_l(y, \sqrt{y}) + O\left(\frac{\sqrt{x}}{\ln(x)^l}\right)$$

and if $y, z \in [x^{1/2}, x^{1/\lambda}]$ $(\lambda > 2)$ st. $y > z$ and $r \in \mathbb{N}$ is arbitrary,

$$\sum_{k=n(z)}^{n(y)-1} h_l\left(\frac{x}{|p_k|}, |p_k|\right) = h_l(x, z) - h_l(x, y) + O\left(\frac{\lambda^{r+1} x}{\ln(x)^r}\right),$$

where

$$n(z) := \min\{n \in \mathbb{N} | |p_n| \geq z\}.$$

Proof. The first identity is a consequence of the additive properties (both set-theoretical and regarding functions) of the integral.

In order to prove the second identity, we write

$$\sum_{k=n(z)}^{n(y)-1} h_l\left(\frac{x}{|p_k|}, |p_k|\right) = \int_z^y h\left(\frac{x}{t}, t\right) d\pi(t).$$

It is evident that

$$\int_z^y h\left(\frac{x}{t}, t\right) d\pi(t) - \int_z^y h\left(\frac{x}{t}, t\right) \frac{1}{\ln(t)} dt = \int_z^y h\left(\frac{x}{t}, t\right) d(\pi - \mathrm{Li})(t).$$

We apply integration by parts:

$$\int_z^y h\left(\frac{x}{t}, t\right) d(\pi - \mathrm{Li})(t) = \left[h\left(\frac{x}{t}, t\right)(\pi(t) - \mathrm{Li}(t))\right]_{t=z}^{t=y} - \int_z^y \frac{d}{dt} h\left(\frac{x}{t}, t\right)(\pi(t) - \mathrm{Li}(t)) dt.$$

The absolute value of

$$\frac{d}{dt}h\left(\frac{x}{t},t\right)$$

is proven by the chain rule, Proposition I.10 and Eq. (11) to be

$$\leq C\ln(\lambda)\frac{x}{t^2\ln(x/t^2)^l} \leq C\lambda\frac{x}{t^2}.$$

The claim hence follows from any version of the prime number theorem that gives a sufficient number of logarithms. □

1.3. **Mertens' theorems in arithmetic monoids.** In this section, we shall prove that Mertens' theorems still hold in arithmetic monoids for which $N_G(x)$ satisfies *axiom A*, ie. is approximately linear in the sense of Eq. (1); only the constant in the third theorem is not explicitly derived. Fortunately, the proofs remain valid rather verbatim, with only few modifications being necessary in order to account for the fact that we are working with an arithmetic monoid.

Proposition I.13 (Mertens' first theorem)**.** *Suppose G satisfies* axiom A *(Eq.* (1)*). Then*

$$\sum_{\substack{p\in P\\|p|\leq x}}\frac{\ln(|p|)}{|p|} = \ln(x)+O(1).$$

Proof. In order to adapt the usual proof to our setting, we need a generalisation of the von Mangoldt function Λ, because we shall make use of the equation

$$\Lambda * \mathbf{1} = \ln(|\cdot|).^7$$

Since the logarithm function is defined on the real numbers and not on G, we define for $g \in G$

$$\Lambda(g) := \begin{cases}\ln(|p|) & \text{if } g = p^k\\ 0 & \text{otherwise}\end{cases}$$

as did Knopfmacher [Kno75, p. 40, (viii)].[8] The function $N_G(x)$ is right-continuous and non-decreasing, whence we may Stieltjes integrate with respect to it, and if we apply integration by parts, we obtain

$$\begin{aligned}\sum_{\substack{g\in G\\|g|\leq x}}\ln(|g|) &= \int_{1/2}^x \ln(t)dN_G(t) = \ln(x)\overbrace{N_G(x)}^{=Ax+O(x^\delta)} - \overbrace{\int_{1/2}^x\frac{N_G(t)}{t}dt}^{=O(x)}\\ &= Ax\ln(x)+O(x).\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{\substack{g\in G\\|g|\leq x}}\ln(|g|) &= \sum_{\substack{g\in G\\|g|\leq x}}\Lambda*\mathbf{1}(g) = \sum_{\substack{g\in G\\|g|\leq x}}\sum_{d|g}\Lambda(d) = \sum_{\substack{g\in G\\|g|\leq x}}\sum_{d\in G}[d|g]\Lambda(d)\\ &= \sum_{d\in G}\sum_{\substack{g\in G\\|g|\leq x}}[d|g]\Lambda(d) = \sum_{d\in G}\Lambda(d)\sum_{\substack{g\in G\\|g|\leq x}}[d|g].\end{aligned}$$

---

[7]The absolute value is necessary in our setting because the logarithm is defined on $\mathbb{R}_+$.

[8]Note that we don't need another notation for this new Λ because the old one is the case $G = \mathbb{N}$.

But the multiples of d of norm less than or equal to x (which is precisely what the very last sum represents) are of the form dh ($h \in G$), where h must satisfy

$$|dh| \leq x \Leftrightarrow |h| \leq \frac{x}{|d|},$$

whence

$$\sum_{\substack{g\in G\\|g|\leq x}} [d|g] = N_G\left(\frac{x}{|d|}\right) = A\frac{x}{|d|} + O\left(\left(\frac{x}{|d|}\right)^{\delta}\right).$$

Whenever $|d| > x$, the number of multiples of d that is less than or equal to x is zero, because it is easily proved that the axioms required for an arithmetic monoid imply that the norm of every element of G is ≥ 1.

Now, using integration by parts as before it is not difficult to verify that for $0 \leq \beta < 1$

$$\sum_{\substack{d\in G\\|d|\leq x}} \frac{1}{|d|^{\beta}} = O(x^{1-\beta}). \tag{12}$$

In the computation one has to use at one point that there are only finitely many elements of G of norm less than a given number in order to bound the remainder terms.

Therefore,

$$Ax\sum_{\substack{d\in G\\|d|\leq x}} \frac{\Lambda(d)}{|d|} = O(\overbrace{x^{\delta}x^{1-\delta}}^{=x}) + \sum_{d\in G} \Lambda(d) \sum_{\substack{g\in G\\|g|\leq x}} [d|g] = Ax\ln(x) + O(x).$$

Cancelling x, we get

$$\sum_{\substack{d\in G\\|d|\leq x}} \frac{\Lambda(d)}{|d|} = \ln(x) + O(1).$$

But

$$\sum_{\substack{d\in G\\|d|\leq x}} \frac{\Lambda(d)}{|d|} = \sum_{|p|\leq x} \frac{\ln(p)}{|p|} + O(1),$$

since a complement-finite subset of the summands appearing in the remainder term is bounded from above by a multiple of

$$\sum_{d\in G} \frac{1}{|d|^{1.5}},$$

which integration by parts demonstrates to be finite. □

Exercise I.14. *Prove that if* axiom A *(Eq.* (1)*) holds, then*

- *There exists* $\gamma_G \in \mathbb{R}$ *such that* $\sum_{\substack{d\in G\\|d|\leq x}} \frac{1}{|d|} = A\ln(x) + \gamma_G + o(1)$, *and for* $\beta \neq 1$
- $\sum_{\substack{d\in G\\|d|\leq x}} |d|^{\beta} = \frac{A}{\beta+1}x^{\beta+1} + O(x^{\beta+\delta})$.

Theorem I.15 (Mertens' second theorem)**.** *Suppose G satisfies* axiom A *(Eq.* (1)*). Then*

$$\sum_{|p|\leq x} \frac{1}{|p|} = \ln(\ln(x)) + O(1).$$

Proof. The usual proof carries over without modification: Set

$$S(x) := \sum_{\substack{p \in P \\ |p| \leq x}} \frac{\ln(|p|)}{|p|}$$

and calculate (using the previous theorem)

$$\sum_{|p| \leq x} \frac{1}{|p|} = \int_{1/2}^{x} \frac{1}{\ln(t)} dS(t) = \frac{S(x)}{\ln(x)} + \int_{1/2}^{x} \frac{S(t)}{t \ln(t)^2} dt$$
$$= \ln(\ln(x)) + O(1).$$

□

Theorem I.16 (Weak Mertens' third theorem)**.** *Suppose G satisfies* axiom A *(Eq. (1)). Then there exists a constant $C > 0$ such that*

$$\prod_{|p| \leq x} \left(1 - \frac{1}{|p|}\right) = \frac{C}{\ln(x)} + o(1).$$

Proof. As in the proof for $\mathbb{N}$, we use

$$\ln\left(\prod_{|p| \leq x} \left(1 - \frac{1}{|p|}\right)\right) = \sum_{|p| \leq x} \ln\left(1 - \frac{1}{|p|}\right);$$

from the Taylor series of the logarithm we get that

$$\sum_{|p| \leq x} \ln\left(1 - \frac{1}{|p|}\right) = -\sum_{|p| \leq x} \frac{1}{|p|} + O(1),$$

because

$$\sum_{k=2}^{\infty} (-1)^k \frac{1}{k} \sum_{p \in P} \frac{1}{|p|^k}$$

converges as an alternating series of terms that converge to zero. Now we can apply Theorem I.15. □

II. Prime element distribution in general arithmetic monoids

2.1. **Prime elements within congruence classes.** In this chapter, we shall demonstrate the usefulness of the analytical bivariate approximation function by proving an equidistribution result for primes among so-called *arithmetic classes*[9].

Definition II.1. A partition of G into *arithmetic classes* is a partition of G into finitely many classes $C_1, \ldots, C_H$ such that the operation

$$[a][b] = [ab]$$

($[a]$ denoting the class containing a) is well-defined and turns the set $\{C_1, \ldots, C_H\}$ into a group. The cardinality $H \in \mathbb{N}$ of this group is called the *class number*.[10]

[9]Knopfmacher [Kno75, p. 251] speaks of "arithmetical formations", but since the term *formation* usually includes every kind of regular shape, it seems inadequate to the author.

[10]For instance, G could be the ideals of an algebraic number field, which are partitioned into the ideal class group.

We define

$$\pi(x,y;C) := \left|\{g \in C : |g| < x \wedge (|p| < y \Rightarrow p \nmid g)\}\right|,$$

and it is not difficult to prove (by considering which numbers precisely are sifted when multiples of p are removed) that for $y > z > 0$

$$\pi(x,y;C) = \pi(x,z;C) - \sum_{z \le |p| < y} \pi\left(\frac{x}{|p|}, k, [p]^{-1}C\right).$$

Consequently, if we define the maximal discrepancy

$$\Delta(x,y) := \max_{C,C'} |\pi(x,y;C) - \pi(x,y;C')|,$$

the triangle inequality implies that

$$\Delta(x,y) \le \Delta(x,z) + \sum_{z \le |p| < y} \Delta\left(\frac{x}{|p|}, |p|\right).$$

In addition,

$$\pi(x,y);C) = \pi(x,\sqrt{x};C) - \pi(y,\sqrt{y};C) + O(\sqrt{x}).$$

Therefore,

$$\Delta(x,y) \le \Delta(x,\sqrt{x}) + \Delta(y,\sqrt{y}) + O(\sqrt{x}). \tag{13}$$

In what follows, we shall assume **axiom A***, which states that if

$$N_C(x) := \{g \in G||g| \le x \wedge x \in C\},$$

then

$$N_C(x) = \frac{N_G(x)}{H} + O(x^\delta) = A\frac{x}{H} + O(x^\delta),$$

where $A > 0$ is the constant from axiom A and H is the class number. Further, when considering several partitions of G into arithmetic classes, we shall assume that axiom A* holds uniformly, and further that

$$H \ll x^{1-\delta'},$$

where $\delta' > \delta$ is fixed. We now want to apply the Rosser–Iwaniec sum sifting theorem. We choose

$$a_g := \frac{H}{A}\mathbb{1}_C(g)$$

as the summands. The first error term (the one coming from approximating the X_d's by a simple function) is then bounded as follows:

$$\sum_{d \le D} r_d \le Hx^\delta \sum_{d \le D} \frac{1}{|d|^\delta} \ll Hx^\delta D^{1-\delta} \ll x^{1-\delta'+\delta} D^{1-\delta}.$$

The Rosser–Iwaniec theorem with $s = \ln(\ln(x))/m$ for a sufficiently large power of m then yields a $B > 0$ with

$$\Delta\left(x, x^{1/(\ln(\ln(x))+\lambda)}\right) \le B/He^{-\sigma(\ln(\ln(x))+\lambda)} h_l\left(x, x^{1/(\ln(\ln(x))+\lambda)}\right)$$

for sufficiently large x and arbitrary $l \in \mathbb{N}$, $\lambda \in [0,1]$ and $\sigma > 0$. (The expression on the right basically equals x by a large power of $\ln(x)$, but the reason for this complicated expression will become apparent shortly).

Proposition II.2. *For all $\alpha \leq \ln(\ln(x)) + 1$ and sufficiently large x,*

$$\Delta(x, x^{1/\alpha}) \leq \begin{cases} 3B\frac{A}{H}h_l(x, x^{1/2}) & (\alpha \in [1,2]) \\ B\frac{A}{H}e^{-\sigma(\alpha-1)}h_l(x, x^{1/\alpha}) \end{cases},$$

where σ is to be chosen later.

For all α in the given range we obtain a uniform bound on Δ provided that the class number is bounded by x to a power $< 1-\delta'$. Additional strength may be gained by assuming a stronger version of axiom A*, namely

$$N_C(x) = \frac{N_G(x)}{H} + O(x^\varepsilon),$$

where $\varepsilon < \delta$, because in this situation the sum sifting problem may be carried out with $N_G(x)$ in the place of x. This means that we only have to assume that $\delta' > \varepsilon$ in order to get a uniform bound.

Proof. We have already proven this when $\alpha \in [\ln(\ln(x)), \ln(\ln(x))+1]$ and x is sufficiently large. For $\alpha \in [3, \ln(\ln(x))]$, we proceed as follows: We have

$$\Delta(x, x^{1/\alpha}) \leq \Delta(x, x^{1/(\alpha+\tau)}) + \sum_{x^{1/(\alpha+\tau)} \leq |p| < x^{1/\alpha}} \Delta\left(\frac{x}{|p|}, |p|\right),$$

where $\tau \in [0,1]$ is a parameter. We induct along the real line from $\ln(\ln(x))$ downwards. We know already that

$$\Delta(x, x^{1/(\alpha+\tau)}) \leq e^{-\sigma(\alpha+\tau-1)}h_l(x, x^{1/(\alpha+\tau)}),$$

from which we gain a factor of $e^{-\sigma\tau}$ (which is small enough if σ is big enough, dependent on τ to be chosen below), and by induction on x and the asymptotics in Proposition I.11,

$$\sum_{x^{1/(\alpha+\tau)} \leq |p| < x^{1/\alpha}} \Delta\left(\frac{x}{|p|}, |p|\right) \sim (\varphi_l(\alpha+\tau) - \varphi_l(\alpha))\frac{x}{\ln(x)^l},$$

where by the integral definition of φ_l the difference $\varphi_l(\alpha+\tau) - \varphi_l(\alpha)$ is $\leq 2^l\tau\varphi_l(\alpha)$, so that if τ is sufficiently small, here too a sufficient factor is gained. Given this τ, we are now able to choose σ sufficiently large; eg. such that $e^{-\sigma\tau} \leq \frac{1}{2}\frac{\varphi_l(\alpha)}{\varphi_l(\alpha+\tau)}$, if τ is such that the sum of the latter term and $2^l\tau\varphi_l(\alpha)$ is $\ll 1$. [11]

If now $\alpha < 3$, it may happen that

$$\frac{x}{|p|}$$

is a power of $|p|$ that is less than 2. In this case, we don't have the nice bound

$$Bh_l\left(\frac{x}{|p|}, |p|\right),$$

but instead the worse bound

$$Bh_l\left(\frac{x}{|p|}, \sqrt{\frac{x}{|p|}}\right).$$

Thus, we have to use Mertens' second theorem in order to bound the difference between the bound induced by these two bounds, and only then apply the argument that we used in $[3, \ln(\ln(x))]$.

[11]Note that a larger σ will lead to a larger constant B, since the Rosser–Iwaniec theorem is only valid for sufficiently large x, increasingly depending on σ.

Finally, for $\alpha \in [1, 2]$, we simply use Eq. (13).

In order to show that the induction on x does not terminate, we note that for a given x, we need information up to

$$x^{1-\frac{1}{\ln(\ln(x))}} =: y,$$

whence from y we gain information at least up to

$$y^{\frac{\ln(\ln(y))}{\ln(\ln(y))-1}}.$$

Since the equation

$$y = y^{\frac{\ln(\ln(y))}{\ln(\ln(y))-1}}$$

is not solvable for $y > 1$, the induction does not terminate. □

2.2. **A general linear sieve equation.** In this chapter, we assume that G is a semiring. Suppose then that $a_1, \dots, a_n$ are elements of G. We define

$$\pi(x, y, a_1, \dots, a_n, w) := \# \left\{ g \in G : \begin{matrix} |g| < x - w \ \& \ \forall p \in P, |p| < y : \\ \forall k \in \{1, \dots, n\} : a_k \not\equiv g \mod p \end{matrix} \right\}.$$

This quantity occurs naturally in the sieve equations for several quantities. For example, the expression

$$\pi(x, \sqrt{x}, 0, -2, 0)$$

equals the number of twin prime pairs in the interval $[\sqrt{x}, x+2)$, and the corresponding sieve equation is given by

$$\pi(x, \sqrt{x}, 0, -2, 0) = \pi(x, y, 0, -2, 0) - \sum_{y \le |p| < x} \pi\left(\frac{x}{p}, p, \right).$$

Hardy and Littlewood [HL23] and Leonard Dickson [Dic04] conjectured asymptotic formulas for π in certain special cases, and it turns out that they are all correct. Some formulas of Hardy and Littlewood concern non-linear polynomials, but to the best knowledge of the author, they only yield to the techniques of the next chapter.

This quantity has many important special cases. For instance,

$$\pi(x, \sqrt{x}, (0, -2)_p)$$

approximately represents the number of twin primes up to x, whereas $\pi(x, \sqrt{x}, (0, x)_p)$ represents the approximate number of representations of x as the sum of two prime numbers.

For these functions, the following two equations hold in analogy to the simple rough numbers:

$$\begin{aligned}\pi(x, y, a_1, \dots, a_n, w) =& \pi(x, z, a_1, \dots, a_n) \\ & - \sum_{z \le p < y} \sum_{k=1}^{n} \pi\left(\frac{x}{p}, p, p^{-1}a_1, \dots, \widehat{p^{-1}}a_k, \dots, a_n, w + \frac{a_k}{p}\right),\end{aligned}$$

where the hat denotes omission and the $^{-1}$ sign is understood modulo every prime occuring in the definition of $\pi(\cdot, \cdot, \cdot)$ separately.

We may apply the exact same proof as in the last chapter in order to show that for w sufficiently small compared to x

$$\pi(x, x^{1/\alpha}, a_1, \dots, a_n, w) \sim \varphi(\alpha)^n \frac{x}{\ln(x)^n},$$

only replacing the discrepancy Δ by the difference between π and the generalised de Bruijn function (see also appendix A)

$$\theta_k(y^u, y) := e^\gamma \ln(y) \int_1^u y^{-t} \omega_k(t) dt.$$

This is because θ_k is an approximate solution to the idealised equation

$$\theta_k(y^u, y) = \int_1^h \theta_k(y^{u-\sigma}, y^\sigma) dW_k(\sigma) + \theta_k(y^u, y^h)(1 - W_k(h))$$

with

$$W_k(\sigma) = 1 - \frac{P_k(y^\sigma)}{P_k(y)},\ P_k(y) = \prod_{|p|<y} \left(1 - \frac{k}{|p|}\right)$$

(cf. de Bruijn [dB51a, p. 807f.]), as is shown by computations analogous to those of de Bruijn.

III. Prime elements in arithmetic monoids with additional regularity

It may be prudent to introduce here a second equation concerning $\pi(x, y)$. It may be stated as

$$\forall y \geq \sqrt{x}:\ \pi(x, y) = \pi(x, \sqrt{x}) - \pi(y, \sqrt{y}) + O(\sqrt{x})$$

or more precisely as

$$\pi(x, y) = \pi(x, \sqrt{x}) - \pi(y, \sqrt{y}) + \pi(\sqrt{x}, x^{1/4}) - \pi(\sqrt{y}, y^{1/4}) \pm \text{etc.} \tag{14}$$

for $y \geq \sqrt{x}$. This equation expresses the facts that for $y \geq \sqrt{x}$, $\pi(x, y)$ represents the prime numbers between y and x, so that we may obtain $\pi(x, y)$ by taking the prime numbers between $\sqrt{x}$ and x, then removing the prime numbers between $\sqrt{y}$ and y and then adding what was removed in excess (and so on). Please note that this sum is finite, because if

$$x^{1/2^n} < 2 \Leftrightarrow n > \log_2(\log_2(x)), \tag{15}$$

the $(n+1)$st term will be $1 - 1 = 0$. (The reason why we don't use the simple prime counting function here is that the bi-variate function we use above may be directly approximated by sum sifting when the second argument is sufficiently small.)

Subsequently, it shall be necessary (or if not necessary, then at least convenient for the computations) to generalise these two equations, using a generalised version of the function $\pi(x, y)$. To this end, we consider for a fixed $n \in \mathbb{N}$ the set

$$S_n := \{(c_1, x_1, \dots, c_n, x_n) | \forall k : c_k \in \mathbb{R}, x_k > 0\}$$

and also their union

$$T := \bigcup_{n \in \mathbb{N}} S_n.$$

It is obvious that these form topological spaces (the latter with the final topology), are closed under addition and are subspaces of a suitable $\mathbb{R}^{2n}$ or $\mathbb{R}^\infty$ resp., but we shan't be concerned with that at all in what follows. For us, the following definitions shall be of importance:

Definition III.1. Let

$$v = (c_1, x_1, \dots, c_n, x_n), w = (c_1, y_1, \dots, c_n, y_n) \in S_n \subset T;$$

it is of great importance here that the oddly indexed places coincide, and that further

$$\forall k \in \{1, \dots, n\} : x_n \leq y_n.$$

Then we define

$$\int_v^w a(t)db(t) := \sum_{k=1}^n c_k \int_{x_k}^{y_k} a(t)db(t),$$

where a and b are functions such that the Stieltjes integrals on the right hand side are all defined.

Definition III.2. Let

$$v = (c_1, x_1, \dots, c_n, x_n), w = (c_1, y_1, \dots, c_n, y_n) \in S_n \subset T$$

and this time assume the reverse inequality

$$\forall k \in \{1, \dots, n\} : x_n \geq y_n.$$

Then we define

$$\pi(v, w) := \sum_{k=1}^n c_k \pi(x_k, y_k).$$

Further, for v and w as in the previous two definitions, we define the functions

$$\begin{aligned} r(v) &:= (c_1, \sqrt{x_1}, \dots, c_n, \sqrt{x_n}), \\ s(v) &:= (c_1, x_1^{1/3}, \dots, c_n, x_n^{1/3}), \\ f(v, w) &:= (f_1(c_1, x_1, y_1) \times \cdots \times f_1(c_n, x_n, y_n), \\ &\qquad f_2(c_1, x_1, y_1) \times \cdots \times f_2(c_n, x_n, y_n)) \\ u(v, z) &:= (c_1, z, \dots, c_n, z) \ (z > 0) \text{ and} \\ \text{mi2}(v, w) &:= (c_1, \min\{x_1, y_1\}, \dots, c_n, \min\{x_n, y_n\}), \end{aligned}$$

where $\times$ denotes tuple juxtaposition and

$$\begin{aligned} f_1(c, x, y) &:= (c, x, c, y, \dots, c, x^{1/2^m}, y^{1/2^m}) \\ f_2(c, x, y) &:= (c, \sqrt{x}, c, \sqrt{y}, \dots, c, x^{1/2^{m+1}}, y^{1/2^{m+1}}) \end{aligned}$$

where in turn

$$m := \lfloor \log_2(\log_2(x)) \rfloor.$$

With these definitions, we may formulate the equation upon which this entire writing is based:

Theorem III.3 (Iterated augmented sieve equation)**.**

$$\begin{aligned} \pi(v, w) &= \pi(v, u(v, z)) - \int_{u(v,z)}^{\text{mi2}(w, r(v))} \pi\left(\frac{1}{t} v, u(v, t)\right) d\pi(t) \\ &\quad - \int_{\text{mi2}(w, r(v))}^{\text{mi2}(w, s(v))} \pi\left(f\left(\frac{1}{t} v, u(v, t)\right)\right) d\pi(t). \end{aligned}$$

Proof. Equation 14 shows that

$$\pi((c, x, y)) = \pi(f((c, x, y)))$$

whenever $y \geq \sqrt{x}$. Thus, the claim follows from splitting the sum of the basic sieve equation in two at the splitting point $x^{1/3}$. □

3.1. **Integrated sum sifting.** The equation formulated in the previous section allows for the fruitful application of sum sifting. To this end, we consider any integrands

$$\pi\left(f^k\left(\frac{1}{t_1\cdots t_n}v, u(v, t_1\cdots t_n)\right)\right)$$

that arise from the repeated application of said equation. To these integrands we may associate a sifting problem dependent on the t_k's, and importantly, the error terms may be integrated over the t_k's, whence they easily cancel to yield an acceptable error term.

The number z has to be chosen sufficiently small (ie. $z = x^{1/\alpha}$ with α a multiple of $\ln(\ln(x))$ as in the previous paper) in order to gain the usual logarithmic error term. One may even be so bold as to choose $\alpha \sim \ln(x)$, but obviously one has to divide by an appropriate large (but fixed) number in order to keep the tree size at an acceptable x^ε for the desired $\varepsilon > 0$. The approximants for the "blocks" of sieve problems (whose error terms are evaluated cumulatively in order to achieve cancellation) are given by the respective differences between the respective de Bruijn approximations.

This method may also be used for number fields, utilising the formula for ideals below a given norm given by the geometry of numbers.

3.2. **Multi-dimensional problems.** In order to attack problems such as the one considered in Bateman–Horn [BH62], one must realise that the classical sum sifting methods will yield the usual bounds in this case as well, so that we may speak of the applicability of the sieve. But in this situation, it is necessary to sift the interval $[1, x]$ for the numbers n such that a given small prime p is a divisor of $f_j(n)$ for some $j \leq l$ where l is the number of polynomials. Therefore, we must stray from the classical path of letting E_p be a union of residue classes mod p. Instead,

$$E_p := \{n|\exists j \leq l : p|f_j(n)\}.$$

Furthermore, a multi-dimensional analogue of the de Bruijn approximation is to be used, with ω being replaced by ω_k.

IV. Acknowledgements

I'd like to thank the Bulletin of the Hellenic Mathematical Society for an incorruptible editing process and Prof. Beligiannis for much encouragement. Furthermore, I'd like to thank Prof. Blomer of Bonn University for excellent advice on time management.

Appendix A. Properties of de Bruijn's function

In this appendix, we shall consider an approximation to de Bruijn's approximation function given by

$$T_k(x, y) := \frac{1}{\ln(y)^{k-1}} \int_1^\alpha y^t \omega(t)^k dt;$$

here and in what follows, we shall utilise the notation $x = y^\alpha$, which makes it unnecessary to substitute $\alpha = \log_y(x)$ each time.

The transformation of variables $t = \log_y(s)$ shows that for $1 \leq \alpha \leq 2$

$$T_k(x, y) = \int_y^x \frac{1}{\ln(s)^k} ds.$$

This means that the functions T_k equal the functions h_k whenever $x \geq y \geq \sqrt{x}$.

When $\alpha \to \infty$, then $T_k(x, y)$ converges rapidly to $e^{-k\gamma}$:

Proposition A.1. *For every $c \in (0,1)$*

$$\left| T_k(x,y) - e^{-k\gamma}\frac{x}{\ln(y)} \right| \ll_k x^c + \frac{x}{\Gamma(cv+1)}.$$

Proof. We split the integral that defines T_k in two:

$$T_k(x,y) = \frac{1}{\ln(y)^{k-1}} \left(\int_1^{c\alpha} y^t \omega(t)^k dt + \int_{c\alpha}^{\alpha} y^t \omega(t)^k dt \right).$$

In order to make any sense of both parts, we first note that

$$\int_a^b y^t dt = \left[\frac{y^t}{\ln(y)} \right]_a^b.$$

Thus, the boundedness of Buchstab's function (indeed, it converges by Proposition I.8 and is continuous on $[1,\infty)$, whence bounded) implies that the first part of the integral is bounded by $O(x^c)$. For the second part, we note that Proposition I.8 and Proposition I.9 imply that

$$\left|\omega(t)^k - e^{-k\gamma}\right| = O\left(\frac{1}{\Gamma(c\alpha+1)}\right)$$

whenever $t \geq c\alpha$, so that the second part of the integral is bounded as desired, where each of the terms resulting from the fundamental theorem of calculus is absorbed into one of the summands of the claimed bound. □

References

[AB13] George Eyre Andrews and Bruce Carl Berndt. *Ramanujan's lost notebook*, volume IV. Springer Science+Business Media, 2013.

[Beu37] Arne Beurling. Analyse de la loi asymptotique de la distribution des nombres premiers généralisés. *Acta Mathematica*, 68:255–291, 1937.

[BH62] Paul Trevier Bateman and Roger Alan Horn. A Heuristic Asyptotic Formula Concerning the Distribution of Prime Numbers. *Mathematics of Computation*, 16(79):363–367, 1962.

[Bru19] Viggo Brun. La série $\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\frac{1}{13}+\frac{1}{17}+\frac{1}{19}+\frac{1}{29}+\frac{1}{31}+\frac{1}{41}+\frac{1}{43}+\frac{1}{59}+\frac{1}{61}$ où les denominateurs sont «nombres premiers jumaux» est convergente ou finie. *Bulletin des Sciences Mathématiques*, 43:100–104 & 124–128, April 1919.

[Buc37] Alexander Buchstab. "Асимптотическая оценка одной общей теоретикочисловой функции (Asymptotic Analysis of a General Number-Theoretical Function)". *Математический сборник*, (2)44(6):1239–1246, 1937.

[dB51a] Nicolaas Govert de Bruijn. On the number of uncancelled elements in the sieve of Eratosthenes. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, 53:803–812, 1951.

[dB51b] Nicolaas Govert de Bruijn. The asymptotic behaviour of a function occuring in the theory of primes. *Journal of the Indian Mathematical Society*, 15:25–32, 1951.

[Dic04] Leonard Eugene Dickson. A new extension of Dirichlet's theorem on prime numbers. *The Messenger of Mathematics*, 33:155–161, 1904.

[HL23] Godfrey Harold Hardy and John Edensor Littlewood. Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes. *Acta Mathematica*, 44:1–70, 1923.

[Iwa80] Henryk Iwaniec. Rosser's sieve. *Acta Arithmetica*, 36(2):171–203, 1980.

[JR65] Wolfgang Bernhard Jurkat and Hans-Egon Richert. An improvement of Selberg's sieve method. *Acta Arithmetica*, 16(2):217–240, 1965.

[Kno75] John Knopfmacher. *Abstract Analytic Number Theory*, volume 12 of *North Holland Mathematical Library*. Elsevier, 1st edition, 1975.

[Knu97] Donald Knuth. *The Art of Computer Programming*, volume 1. Addison–Wesley, 3rd edition, 1997.

[Ram20] Srinivasa Ramanujan. The lost notebook. Posthumously published private notebook, 1920.

[Ram49] Vommi Ramaswami. On the number of positive integers less than x and free of prime divisors greater than x^c. *Bulletin of the American Mathematical Society*, 55(1):1122–1127, 1949.

[Rů18] Pavel Růžička. Unique factorization domains. Web server, 2018.
[Tao15] Terence Tao. Chapter 4 of Terence Tao's lecture notes on analytic number theory. Blog post, January 2015.

Email address: fellhauer@uni-bonn.de

Albtalstrasse 14, 76359 Marxzell,Germany