UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

DOTTORATO DI RICERCA IN MATEMATICA

IN CONVENZIONE CON
L'UNIVERSITÀ DEGLI STUDI DI FERRARA
E L'UNIVERSITÀ DEGLI STUDI DI PARMA

CICLO XXXII

DUE PROBLEMI DI TEORIA ANALITICA DEI NUMERI: SOMME ARMONICHE CON I PRIMI E DISTRIBUZIONE DELLE CIFRE DI QUOZIENTI FRA INTERI

TWO PROBLEMS IN ANALYTIC NUMBER THEORY: HARMONIC SUMS WITH PRIMES AND DISTRIBUTION OF THE DIGITS OF RATIOS OF INTEGERS

CANDIDATO: REMIS TONON

RELATORI: DOTT. S. BETTIN
PROF. A. ZACCAGNINI

COORD. DEL CORSO DI DOTTORATO: PROF. C. GIARDINÀ

Abstract

Nella prima parte della tesi si estendono i risultati dell'articolo *Small values of signed harmonic sums* di Bettin, Molteni e Sanna. In esso, gli autori considerano serie armoniche troncate, in cui si ammette per ogni addendo la possibilità del segno positivo o negativo, e studiano la funzione che misura quanto precisamente si possa approssimare un valore reale con tali oggetti. Nello specifico, vengono dimostrate delle limitazioni per tale funzione in alcuni intervalli di validità. Nella tesi si è dimostrato che lo stesso risultato vale non solo per la successione di tutti i naturali, ma anche per ogni sua sottosuccessione che rispetti ragionevoli ipotesi di crescita. Inoltre, nel caso specifico della successione dei numeri che sono il prodotto di k fattori primi distinti, dove k è un numero naturale fissato, è stato possibile migliorare sensibilmente le limitazioni per la funzione approssimante.

Nella seconda parte della tesi si migliora il risultato dell'articolo *Probability of digits by dividing random numbers: a ψ and ζ functions approach* di Gambini, Mingari Scarpello e Ritelli. Gli autori studiano in esso la distribuzione dell'ennesima cifra dopo la virgola (in diverse basi di numerazione) di tutti i possibili quozienti tra i primi N numeri naturali, dimostrando che essa non è uniforme, ma che segue una legge affine alla legge di Benford [1]. Nella tesi, si migliora il termine d'errore ottenuto dai tre autori e si affronta il problema analogo in cui, invece dei numeri naturali, si considerano solamente i numeri primi.

In the first part of this thesis we extend the results of the paper *Small values of signed harmonic sums* by Bettin, Molteni and Sanna. There, the authors consider harmonic truncated series, where the summands can have positive or negative signs; using these objects to approximate any real value, they study the function that measures the precision of this approximation. In particular, they prove some bounds for this function in some specific ranges. In this thesis, we prove that the same result holds not only for the sequence of all natural numbers, but also for any subsequence satisfying reasonable growing hypotheses. Besides, in the case of the sequence of numbers that are the product of k distinct primes, where k is a fixed natural number, we obtain a significant improvement on the bounds for the approximating function.

In the second part of this thesis, we improve the result of the paper *Probability of digits by dividing random numbers: a ψ and ζ functions approach* by Gambini, Mingari Scarpello and Ritelli. In this paper, the authors study the distribution of the n-th digit after the decimal point (in different bases) of all possible ratios between the first N natural numbers: they prove that the distribution is not uniform, but it follows a law analogous to Benford's law [1]. In this thesis, we improve the error term and deal with the analogous problem where, instead of taking natural numbers, we consider just the primes.

Contents

Chapter 1

Preface

In this thesis, we deal with two problems in number theory, both concerning rational numbers. Here we present them and give some context to better understand where their interest lies.

1.1 Signed harmonic sums

We begin with a problem related to harmonic numbers. It is well known that the n-th harmonic number is defined as

$$H_n := \sum_{k=1}^{n} \frac{1}{k}.$$

These objects have been studied for a long time, and some of their properties are still interesting. One of the most prestigious names associated to them is Euler: he made wide use of these numbers and investigated them extensively. It is not surprising, then, that the famous Euler–Mascheroni constant γ, often appearing in analytic number theory, is related to harmonic numbers. Indeed, it is known that, when n diverges,

$$H_n = \log n + \gamma + O(1/n), \tag{1.1}$$

and this can be taken as a definition of γ itself. We will speak about Euler again soon, because part of our work is related to a famous proof by him: the divergence of the series of the reciprocal of primes.

We proceed mentioning some other very interesting classical results, of a more arithmetic nature: the most famous is probably the one by Taesinger, who proved that H_n is never an integer for $n > 1$. After him, Erdős generalized this theorem to the sums of inverses of numbers in arithmetic progressions. Other properties have been discovered since, and some are still object of study: we just recall, as an example, that a currently active field of research is related to the p-adic valuation of H_n.

Another reason that makes harmonic numbers interesting is that they are special cases of Egyptian fractions. These are defined as sums of distinct unit fractions, and it seems that ancient Egyptians used such a method to represent rationals. In fact, Fibonacci proved, with a constructive argument, that any positive rational number can be represented in this way; actually, now we know that any rational number has representations of this kind with arbitrarily many terms and with arbitrarily large denominators. On the other side, many other problems are still open, especially concerning Egyptian fractions with a fixed number of unit fractions: we just mention here the Erdős–Straus conjecture, which states that, for any integer $n \geq 2$, the number $4/n$ can be expressed as the sum of exactly three unit fractions.

We have just given some motivation for the interest in harmonic sums. We now introduce a variant: what if one were allowed to choose between plus and minus sign for any summand? Some cases are quite straightforward from basic analysis. Consider for example the series

$$\sum_{k=1}^{n} \frac{(-1)^{k+1}}{k}.$$

We know that it converges by the Leibniz criterion, and indeed as $n \to \infty$ the limit is $\log 2$. More in general, we can make an observation based on the proof of Riemann's rearrangement theorem. This states that, given any conditionally convergent series and any $\tau \in \mathbb{R} \cup \{ \pm\infty \}$, one could rearrange it and make it converge to τ. The idea of the proof for a finite τ is quite natural: one can take all positive summands until exceeding the target value; at that point, one should start summing negative values to reach a point less than τ; and so on. In the same way, if we fix a target τ, we can apply the same procedure with our signed harmonic series, since, as we saw in (1.1), it diverges absolutely: just pick positive signs for the first summands until the sum exceeds the target, then take negative signs to go below τ, and so on. This means that we have an effective algorithm to pick signs to make our series converge to any $\tau \in \mathbb{R}$. Actually, this algorithmic process was studied in detail by Bettin, Molteni and Sanna [2] in a paper, where they give bounds for the rate of convergence of this approximation, which they call greedy approximation, and make some interesting theoretical properties emerge, like the connection between the series of signs and the Thue–Morse sequence.

Hence, we are sure that for any τ we can choose at least one sequence of signs to make the signed harmonic series converges to it, but what can we say about the best approximation that we can obtain with the first N terms? Indeed, if the greedy algorithm ensures the existence of a signed harmonic series converging to any target, it does not guarantee that such choice of signs is the one that minimizes the distance to the target at any step. After some computational work, one immediately realizes that the greedy algorithm is hardly the best choice to have a fast convergence, whence the interest in finding some bounds for the rate of convergence of the best approximation given by a signed harmonic sum. This is precisely what Bettin, Molteni and Sanna [3] studied in another recent paper, which we took as a starting point to study an interesting variation of the original problem. Let us introduce some notation to be more precise.

First of all, following [3], we define the set

$$\mathfrak{S}'_N := \left\{ \sum_{n \le N} \frac{s_n}{n} \colon s_n \in \{\pm 1\} \text{ for } n \in \{1, \dots, N\} \right\}, \tag{1.2}$$

whose elements are all the points that can be reached at the N-th step. It is clear that $\mathfrak{S}'_N$ is symmetric with respect to 0 and that

$$\max \mathfrak{S}'_N = H_N \sim \log N,$$

as we saw in (1.1). If we fix $\tau \in \mathbb{R}$, the really interesting quantity is

$$\mathfrak{m}'_N(\tau) := \min\bigl\{ |S_N - \tau| \colon S_N \in \mathfrak{S}'_N \bigr\}.$$

We know that for every $\varepsilon > 0$ and for almost every $\tau \in \mathbb{R}$,

$$\mathfrak{m}'_N(\tau) > \exp(-\varepsilon N + o(N))$$

as $N \to \infty$. This follows from [5], proceeding as for Proposition 2.7 of [3]. In fact, $\mathfrak{m}'_N(\tau)$ can be arbitrary small infinitely often: it is possible to prove that, given $f : \mathbb{N} \in \mathbb{R}_{>0}$, there exist $\tau_f \in \mathbb{R}$ such that $\mathfrak{m}'_N(\tau_f) < f(N)$ for infinitely many N. Now let us state the main results from [3].

Theorem 1.1 (Bettin, Molteni & Sanna, 2018)**.** *For every $\tau \in \mathbb{R}$ and for any positive constant C less than $1/\log 4$, we have that, for all $N > \overline{N}(C, \tau)$,*

$$\mathfrak{m}'_N(\tau) < \exp\left(-C \log^2 N\right).$$

It is really important to notice that a core point of the problem is that at every step all the signs are updated, and thus the sequence of signs does not converge in any sense to a limiting sequence. A computational approach reveals immediately that indeed it is very rare that even the very first signs are maintained for too many steps: this seems to be the reason why the convergence is so fast with respect to the greedy algorithm.

This theorem is in fact a straightforward consequence of another one and its corollary, which are characterized by a probabilistic approach. To state them, we define the random variable

$$X'_N := \sum_{n=1}^{N} \frac{s_n}{n},$$

where the signs s_n are taken uniformly and independently at random in $\{-1, 1\}$. With a slight abuse of notation, we denote by s_n both the signs in the definition (1.2) and the random variables in the definition above. In [3], the authors prove the following results about the small-scale distribution of X'_N.

Theorem 1.2 (Bettin, Molteni & Sanna, 2018). *Let $0 < C < 1/\log 4$ be fixed. Then, for all intervals $I \subseteq \mathbb{R}$ of length $|I| > \exp(-C\log^2 N)$, one has*

$$\mathbb{P}[X'_N \in I] = \int_I g(x)\,\mathrm{d}x + o(|I|),$$

as $N \to \infty$, where

$$g(x) := 2\int_0^\infty \cos(2\pi u x)\prod_{n=1}^{\infty}\cos\left(\frac{2\pi u}{n}\right)\mathrm{d}u.$$

Corollary 1.3 (Bettin, Molteni & Sanna, 2018). *Let $0 < C < 1/\log 4$ be fixed. Then, for all $\tau \in \mathbb{R}$, one has*

$$\left|\left\{ (s_1, \dots, s_N) \in \{\pm 1\}^N : \left|\tau - \sum_{n=1}^{N}\frac{s_n}{n}\right| < \delta \right\}\right| \sim 2^{N+1} g(\tau)\delta(1 + o_{C,\tau}(1))$$

as $N \to \infty$ and $\delta \to 0$, uniformly in $\delta \geq \exp(-C\log^2 N)$. In particular, for all large enough N, one has $\mathfrak{m}'_N(\tau) < \exp(-C\log^2 N)$.

Our idea was to generalize these results to subsequences of integers satisfying suitable growing conditions: in this way, one could apply this theorem to primes or primes in arithmetic progressions, say. In fact we were able to show that the same results hold for more general sequences and that substantial improvements are possible in the case of products of a fixed number of primes. We recall that the problem in the case of primes is related to a very important proof for number theory, due again to Euler: in his famous paper about infinite series [10], he wrote something that could be interpreted as an idea of proof of the fact that

$$\sum_{p\le N}\frac{1}{p}\sim\log\log N,$$

not only providing a new proof of Euclid's result about the existence of infinitely many primes, but also providing a quantitative version of this theorem. Actually the formal proof is due to Mertens, who showed that

$$\sum_{p\le N}\frac{1}{p}=\log\log N+A+O\left(\frac{1}{\log N}\right),$$

where $A\simeq 0.2614972\ldots$ is the Meissel–Mertens constant. Due to the importance of this series, we thought it would be interesting to study the random sign problem in this setting as well, and to also extend the above results to more general sequences. Let us give some new definitions to suit the new setting.

We always denote by $\mathbb{N}$ the set of positive integers. Let $b_n\in\mathbb{N}$ be a strictly increasing sequence and let us assume that

$$\sum_{n\ge1}\frac{1}{b_n}=+\infty \qquad\text{and}\qquad n\le b_n\le nB(n), \tag{1.3}$$

where $B(n)=n^{\beta(n)}$, with β a real-valued decreasing function such that $\beta(n)=o(1)$. We will be studying sums involving $1/b_n$, so we define $a_n:=1/b_n$. In this way we clearly have that

$$\lim_{n\to+\infty}a_n=0 \qquad\text{and}\qquad \sum_{n\ge1}a_n=+\infty.$$

In order to prove Proposition 2.7, we will also assume another condition on β, namely that

$$\beta(n)\le\frac{1}{8\log\log n} \qquad\text{for all sufficiently large } n\in\mathbb{N}. \tag{1.4}$$

Actually, this assumption is not strictly necessary and we will discuss this in Remark 2.1. Nevertheless, since the series $\sum a_n$ must diverge, this condition is not too restrictive, and besides it is satisfied by most interesting sequences, such as arithmetic progressions and the sequence of primes and of primes in arithmetic progressions.

Let us introduce some more notation. We now consider the set

$$\mathfrak{S}_N:=\left\{\sum_{n\le N}s_na_n\colon s_n\in\{\pm1\}\text{ for } n\in\{1,\ldots,N\}\right\},$$

and, for a given $\tau \in \mathbb{R}$, we set

$$\mathfrak{m}_N(\tau) := \min\big\{|S_N - \tau| \colon S_N \in \mathfrak{S}_N\big\}.$$

In other words, for a given $N \in \mathbb{N}$, the goal is to find the choice of signs such that $|S_N - \tau|$ attains its minimum value. Finally, we define the random variable

$$X_N := \sum_{n=1}^{N} s_n a_n,$$

where the signs s_n are taken uniformly and independently at random in $\{-1, 1\}$.

Now we are ready to state our findings. We remark here that we obtained these results concerning an upper bound for $\mathfrak{m}_N(\tau)$; we did not have the time to study possible results in the other direction, that is finding a nontrivial lower bound, for the more general case or for products of k primes. For ease of comparison with Theorem 1.1, we now state our main results in the following simplified form. We remark that the previous definitions as well as the following theorems depend on the sequence a_n.

Theorem 1.4. *Let β satisfy* (1.4). *Then there exists $C > 0$ such that for every $\tau \in \mathbb{R}$ we have*

$$\mathfrak{m}_N(\tau) < \exp(-C \log^2 N)$$

for all sufficiently large N depending on τ and on the sequence a_n.

Theorem 1.5. *Let $k \in \mathbb{N}$, $k \geq 1$ be fixed. Let us consider $b_n = b_n^{(k)}$, where $\big(b_n^{(k)}\big)_{n\in\mathbb{N}}$ denotes the ordered sequence of the products of k distinct positive primes. Then, for every $\tau \in \mathbb{R}$ and any $\varepsilon > 0$, we have*

$$\mathfrak{m}_N(\tau) < \exp\left(-N^{1/(2k+1)-\varepsilon}\right),$$

for all sufficiently large N depending on τ, the sequence a_n and ε.

We collect some numerical results for $k = 1$ in Tables 4.1, 4.2, 4.3 and 4.4.

As before, each of these theorems is the consequence of a corresponding theorem and corollary in the probabilistic setting. We report them as well in a simplified form, avoiding some technicalities. The precise statements are to be found in the next chapter. Let us begin with the general case.

Theorem 1.6. *Let β satisfy* (1.4)*. Then there exist $C' > 0$ such that, for all intervals $I \subseteq \mathbb{R}$ of length $|I| > \exp(-C'(\log N)^2)$, one has*

$$\mathbb{P}[X_N \in I] = \int_I g(x)\,\mathrm{d}x + o(|I|)$$

as $N \to \infty$, where

$$g(x) := 2\int_0^\infty \cos(2\pi ux)\prod_{n=1}^\infty \cos\left(\frac{2\pi u}{b_n}\right)\mathrm{d}u.$$

Corollary 1.7. *Let β satisfy* (1.4)*. Then there exist $C' > 0$ such that for all $\tau \in \mathbb{R}$ we have*

$$\left|\left\{ (s_1,\dots,s_N) \in \{\pm 1\}^N : \left|\tau - \sum_{n=1}^N \frac{s_n}{b_n}\right| < \delta \right\}\right| \sim 2^{N+1} g(\tau)\delta(1 + o_{C',\tau}(1))$$

as $N \to \infty$ and $\delta \to 0$, uniformly in $\delta \geq \exp(-C'(\log N)^2)$. In particular, for large enough N, one has $\mathfrak{m}_N(\tau) < \exp(-C'(\log N)^2)$.

Instead, for products of k primes, we have the following ones.

Theorem 1.8. *Let $k \in \mathbb{N}$, $k \geq 1$ be fixed. Let us consider $b_n = b_n^{(k)}$, where $\big(b_n^{(k)}\big)_{n\in\mathbb{N}}$ denotes the ordered sequence of the products of k distinct positive primes. Then for all $\varepsilon > 0$ and for all intervals $I \subseteq \mathbb{R}$ of length $|I| > \exp(-x^{1/(2k+1)-\varepsilon})$, one has*

$$\mathbb{P}[X_N \in I] = \int_I g(x)\,\mathrm{d}x + o(|I|),$$

as $N \to \infty$, where

$$g(x) := 2\int_0^\infty \cos(2\pi ux)\prod_{n=1}^\infty \cos\left(\frac{2\pi u}{b_n^{(k)}}\right)\mathrm{d}u.$$

Corollary 1.9. *Let k and $(b_n^{(k)})_{n\in\mathbb{N}}$ be as in Theorem 1.8. Then for any $\varepsilon > 0$ and for all $\tau \in \mathbb{R}$ we have*

$$\left|\left\{ (s_1,\dots,s_N) \in \{\pm 1\}^N : \left|\tau - \sum_{n=1}^N \frac{s_n}{b_n}\right| < \delta \right\}\right| \sim 2^{N+1} g(\tau)\delta(1 + o_{\tau}(1))$$

as $N \to \infty$ and $\delta \to 0$, uniformly in $\delta \geq \exp(-x^{1/(2k+1)-\varepsilon})$. In particular, for N large enough, one has $\mathfrak{m}_N(\tau) < \exp\big(-x^{1/(2k+1)-\varepsilon}\big)$.

1.2 The digits of quotients of integers

The first attempts to study number theoretical problems by means of probabilistic methods dates back to the 19th century and involve mainly two mathematicians: first Gauss, who was interested in the number of products of exactly k distinct primes below a certain threshold (and the solution of this problem for $k = 1$ is the famous prime number theorem, proved by Hadamard and de la Vallée Poussin, building on the ideas of Riemann, only in 1896) and then Cesàro, who showed in 1881 that the probability that two randomly chosen integers are coprime is $6/\pi^2$. In 1885 he then published a book [8] collecting some interesting problems from some articles he published in *Annali di matematica pura ed applicata*: the article we are interested in is *Eventualités de la division arithmétique*, which originally appeared as [7]. There he states that, dividing two random integers, the probability that the i-th digit after the decimal point is r is given by

$$\frac{1}{20} + \frac{10^i}{2} \int_0^1 \frac{1-\varphi}{1-\varphi^{10}} \varphi^{10^i-1+r} \, \mathrm{d}\varphi.$$

Taking this as a starting point, Gambini, Mingari Scarpello and Ritelli [11] studied the problem in more detail, and recognised that the integral can be expressed in terms of the digamma function

$$\psi(x) := \frac{\Gamma'(x)}{\Gamma(x)}.$$

We collect some of its properties in the first section of the appendix. The representation

$$\psi(z) = -\gamma - \frac{1}{z} + \sum_{k\geq 1} \frac{z}{k(z+k)} = -\gamma + \sum_{k\geq 0} \Big(\frac{1}{k+1} - \frac{1}{k+z} \Big),$$

led the authors to a different form for the integral, which made them suspect that an elementary proof of the result was possible. Indeed, they were able to find it and this is the starting point for our study.

In order to be more precise, we start giving some definitions. We consider a number basis $b \geq 2$ and the set of digits $S_b = \{0, \ldots, b-1\}$. Given a positive real number x and a positive integer i, we are concerned with the i-th digit to the right of the point of the representation in base b of x: we will call it $\phi(x; b; i)$. With b and i as above and a digit $r \in S_b$, we also define $\Phi(T; b, r; i) := |\mathcal{A}(T; b, r; i)|$, where

$$\mathcal{A}(T; b, r; i) := \{(n, m) \in \mathbb{N}^2 \cap [1, T]^2 \colon \phi(n/m; b; i) = r\}.$$

Throughout the thesis, for brevity we often drop the dependency on b, r and i of our functions, whenever there is no possibility of misunderstanding. We recall that Gambini, Mingari Scarpello and Ritelli [11] implicitly obtained the asymptotic formula

$$\Phi(T; b, r; i) = c(b, r; i)T^2 + O\left(T^{3/2}\right),$$

as $T \to +\infty$. The constant $c(b, r; i)$ is defined as an infinite series and can be expressed by means of the digamma function, as follows:

$$c(b, r; i) = \frac{1}{2b} + \frac{1}{2}b^{i-1}\Big(\psi\Big(\frac{b^i + r + 1}{b}\Big) - \psi\Big(\frac{b^i + r}{b}\Big)\Big). \tag{1.5}$$

We remark here that this problem is appropriately situated among the classical problems of counting lattice points that belong to some precise region of the plane. One of the most famous of these results is Minkowski's theorem, which states that every convex set in $\mathbb{R}^n$ that is symmetric with respect to the origin and with volume greater than 2^n contains at least one point with integer coordinates distinct from the origin. This theorem, proved in 1889, gave birth to a new branch of number theory: the geometry of numbers. Another famous result in this direction is Pick's theorem: proved in 1899, it relates the area of a simple polygon with integer coordinates with the number of lattice points in its interior and the number of lattice points on its boundary. Finally, it is mandatory to refer to two of the most famous problems in analytic number theory, which are related to ours: Gauss' circle problem and Dirichlet's divisor problem. Both of them deal with counting points with integer coordinates belonging to some region delimited by conics: a circumference and a hyperbola. The two mathematicians were able to provide a formula with the area of the figure as a main term plus some error due to the integer points near the border. For example, Gauss proved that in the circle of radius r there are

$$\pi r^2 + E(r)$$

points with integer coordinates, where $|E(r)| \leq 2\sqrt{2}\pi r$. By this simple formulation, one could think that guessing the right order of magnitude of the error term should not be a hard problem; actually, although many (slow) improvements have been done, both problems remain open.

After this excursus, which gives some motivation to our research for a better error term, we come back to our results. In the last chapter of this thesis, we introduce some number-theoretic devices which allow us to improve upon the result by Gambini, Mingari Scarpello and Ritelli, and specifically to obtain a better error term, but just for the case $i = 1$, as stated here.

Theorem 1.10. *We have*

$$\Phi(T; b, r; 1) = c(b, r; i)T^2 + O\left(T^{4/3}\right),$$

as $T \to +\infty$*, where* $c(b, r; 1)$ *is the constant defined in* (1.5)*. The implicit constant may depend on* b*.*

Our improvement stems largely from the fact that we evaluate more carefully the error terms arising from computing ratios of integers with the desired digit and that we introduce a variable threshold, to be chosen at the end of the proof, which allows us to ignore some points in $\mathcal{A}(T; b, r; i)$.

In the second part of the last chapter, we deal with a variation of the same problem: we consider the case of primes. So we actually consider the function $\Phi'(T; b, r; i) := |\mathcal{A}'(T; b, r; i)|$, where

$$\mathcal{A}'(T; b, r; i) := \{(p, q) \in \mathbb{P}^2 \cap [1, T]^2 \colon \phi(p/q; b; i) = r\}$$

and $\mathbb{P}$ is the set of positive prime integers. All the other definitions are to be adjusted in the same straightforward way. We are able to obtain an error term which is just slightly smaller than the main one. The result that we have reached is the following.

Theorem 1.11. *In the case of primes, we have*

$$\Phi'(T; b, r; 1) = c(b, r; 1)\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right),$$

as $T \to +\infty$*, where* $c(b, r; 1)$ *is the constant defined in* (1.5)*. The implicit constant may depend on* b*.*

1.3 Papers and possible developments

The results about signed harmonic sums are part of a paper [12] which has been accepted and will be published in *Rendiconti del Seminario Matematico di Torino*, while the contents of the last chapter are the core matter for a paper [6] in preparation.

Indeed, we have many questions and variants that we would like to study in more detail. First of all, Sandro Bettin has proposed a different approach, which could further improve the error term and maybe provide an omega result for it: in particular, it seems that in Theorem 1.10, instead of 4/3 as an exponent, we could obtain around 13/10 and, assuming some of the classical conjectures, reach 5/4. Besides, if the heuristics are correct, one should not be able to go below this value. In addition to this, instead of dealing just with the first digit after the decimal point, we would like to complete the study for any i: this should be possible, but it will probably present some slight complications in the details. Other variants, inspired by some computations, concern coprimality conditions between numerators and denominators (with connections with the Farey sequence), and with the basis. For more heuristics about this, see the plots in the appendix. Besides, we would like to examine: the uniformity in the variables of the error term; the behaviour of the same problem if the two bounds for the numerator and the denominator go to infinity at different rates; the behaviour of digits on the left of the point; a possibly different approach with primes to employ the cancellation that should occur in some sums. Last, but not least, it might be possible to generalize these results with an axiomatic approach from lattice points to more general points in the plane, through discrepancy theory.

1.4 Acknowledgements

First of all, I would like to thank my advisors Dr. Sandro Bettin and Prof. Alessandro Zaccagnini for their helpfulness, patience, knowledge and insight, which were a real source of inspiration for me along all my PhD study. Especially in the last period, their guidance and support have been essential to complete this thesis, and I feel sincerely grateful to them.

Besides my advisors, I would like to thank Mattia Cafferata, with whom I worked for much time on the results that are presented in this thesis, and especially on the computational side of them, and Alessandro Gambini for his great motivation and availability to talk and work on these topics. My sincere thanks also go to Prof. Giuseppe Molteni, for sharing useful intuitions about the research that we were pursuing, and to him again and to Dr. Timothy Trudgian for their many precious comments as referees.

Chapter 2

Signed harmonic sums of integers with k distinct prime factors

2.1 Introduction and general setting

As explained in the preface, it is well known that the harmonic series restricted to prime numbers diverges, as the harmonic series itself. This was first proved by Leonhard Euler in 1737 [10], and it is considered as a landmark in number theory. The proof relies on the fact that

$$\sum_{n=1}^{N} \frac{1}{n} = \log N + \gamma + O\left(1/N\right),$$

where $\gamma \simeq 0.577215\ldots$ is the Euler–Mascheroni constant. The corresponding result for primes is one of the formulae proved by Mertens, namely

$$\sum_{p \leq N} \frac{1}{p} = \log\log N + A + O\left(\frac{1}{\log N}\right),$$

where $A \simeq 0.2614972\ldots$ is the Meissel–Mertens constant.

Recently, Bettin, Molteni and Sanna [3] studied the random harmonic series

$$X' := \sum_{n=1}^{\infty} \frac{s_n}{n}, \tag{2.1}$$

where $s_1, s_2, \ldots$ are independent uniformly distributed random variables in $\{-1, +1\}$. Based on the previous work by Morrison [14, 15] and Schmuland [18], they proved the almost sure convergence of (2.1) to a density function g, getting lower and upper bounds of the minimum of the distance of a number $\tau \in \mathbb{R}$ to a partial sum

$\sum_{n=1}^{N} s_n/n$. For further references, see also Bleicher and Erdős [5, 4], where the authors treated the number of distinct subsums of $\sum_1^N 1/n$, which corresponds to taking s_i independent uniformly distributed random variables in $\{0,1\}$. A more complete list of references can be found in [3].

The purpose of this chapter is firstly to show that basically the same results hold for a general sequence of integers under some suitable, and not too restrictive, conditions; moreover, that a stronger result can be reached if we restrict to integers with exactly k distinct prime factors.

Although Bettin, Molteni and Sanna [3] treat both the lower bound and the upper bound, we are mainly interested in the upper bound using a probabilistic approach. As we will see, in the cases that we treat, we will not be able to say anything about the lower bound, except in terms of numerical computations.

We will use a consistent notation with the previous works by Bettin, Molteni and Sanna [2], [3], Crandall [9] and Schmuland [18].

2.1.1 General setting of the problem

We here recall and make precise some notations and hypotheses that we introduced in the preface. We denote by $\mathbb{N}$ the set of positive integers. Let $(a_n)_{n\in\mathbb{N}}$ be a strictly decreasing sequence of positive real numbers such that

$$\lim_{n\to+\infty} a_n = 0 \qquad \text{and} \qquad \sum_{n\geq 1} a_n = +\infty. \tag{2.2}$$

As we said before, we are mainly interested in integers satisfying some multiplicative constraints, so we introduce some further reasonable hypotheses on the sequence a_n: we assume that $b_n = a_n^{-1} \in \mathbb{N}$, so that b_n is strictly increasing, and that

$$n \leq b_n \leq nB(n), \tag{2.3}$$

where $B(n) = n^{\beta(n)}$, with β a real-valued decreasing function such that $\beta(n) = o(1)$. In order to prove Proposition 2.7 below, we will assume also the condition

$$\beta(n) \leq \frac{1}{8\log\log n} \qquad \text{for all sufficiently large } n \in \mathbb{N}. \tag{2.4}$$

Actually, this assumption is not strictly necessary and we will discuss this in Remark 2.1. Nevertheless, since the series $\sum a_n$ must diverge, this condition is not too restrictive, and besides it is satisfied by most interesting sequences, such as arithmetic progressions and the sequence of primes and of primes in arithmetic progressions.

Let us recall that, by definition,

$$\mathfrak{S}_N := \left\{ \sum_{n\le N} s_n a_n \colon s_n \in \{\pm 1\} \text{ for } n \in \{1,\dots,N\} \right\}, \tag{2.5}$$

and that, for a given $\tau \in \mathbb{R}$,

$$\mathfrak{m}_N(\tau) := \min\bigl\{|S_N - \tau| \colon S_N \in \mathfrak{S}_N\bigr\}.$$

For a given $N \in \mathbb{N}$, the goal is to find the choice of signs such that $|S_N - \tau|$ attains its minimum value. Finally, we recall the definition of the random variable

$$X_N := \sum_{n=1}^{N} s_n a_n,$$

where the signs s_n are taken uniformly and independently at random in $\{-1, 1\}$. We will study its small scale distribution. With a slight abuse of notation, we denote by s_n both the signs in the definition (2.5) and the random variables in the definition above.

2.1.2 Results

For ease of comparison with the results in Bettin, Molteni and Sanna [3], we now state our main results in the following form, even though more precise versions of them are to be found in the next section.

Theorem 2.1. *Let β satisfy* (2.4)*. Then there exists $C > 0$ such that for every $\tau \in \mathbb{R}$ we have*

$$\mathfrak{m}_N(\tau) < \exp(-C\log^2 N)$$

for all sufficiently large N depending on τ and on the sequence a_n.

Theorem 2.2. *Let $k \in \mathbb{N}$, $k \ge 1$ be fixed. Let us consider $b_n = b_n^{(k)}$, where $\bigl(b_n^{(k)}\bigr)_{n\in\mathbb{N}}$ denotes the ordered sequence of the products of k distinct positive primes. Then, for every $\tau \in \mathbb{R}$ and for all sufficiently large N depending on τ and k, we have*

$$\mathfrak{m}_N(\tau) < \exp(-f(N)),$$

where f is any function satisfying

$$f(N) = o\left(N^{1/(2k+1)-\varepsilon}\right)$$

for some $\varepsilon > 0$.

We collect some numerical results for $k = 1$ in Tables 4.1, 4.2, 4.3 and 4.4.

2.2 The case of natural numbers

2.2.1 Lemmas

In this section we study some properties of the general sequence defined in (2.2), using the classical notation: $\mathbb{E}[X]$ denotes the expected value of a random variable X, $\mathbb{P}(E)$ the probability of an event E. For each continuous function with compact support $\Phi \in \mathbb{C}_c(\mathbb{R})$ we denote by $\widehat{\Phi}$ its Fourier transform defined as follows:

$$\widehat{\Phi}(x) := \int_{\mathbb{R}} \Phi(y)\, \mathrm{e}^{-2\pi \mathrm{i} xy} \mathrm{d}y.$$

We are actually interested in smooth functions, because the smoothness of the density of any random variable X is related to the decay at infinity of its characteristic function, defined precisely by its Fourier transform.

For each $N \in \mathbb{N} \cup \{\infty\}$, for any $x \in \mathbb{R}$ and for any sequence satisfying (2.2), we also define the product

$$\varrho_N(x) := \prod_{n=1}^{N} \cos(\pi x a_n) \qquad \text{and} \qquad \varrho(x) := \varrho_\infty(x).$$

We begin with the following lemma, which is a more general version of Lemma 2.4 from [3].

Lemma 2.3. *We have*

$$\mathbb{E}[\Phi(X_N)] = \int_{\mathbb{R}} \widehat{\Phi}(x) \varrho_N(2x) \mathrm{d}x$$

for all $\Phi \in \mathbb{C}_c^1(\mathbb{R})$.

Proof. By the definition of expected value we have

$$\mathbb{E}[\Phi(X_N)] = \frac{1}{2^N} \sum_{s_1,\dots,s_N \in \{-1,1\}} \Phi\left(\sum_{n=1}^{N} s_n a_n\right).$$

Using the inverse Fourier transform, we get

$$\begin{aligned}
\mathbb{E}[\Phi(X_N)] &= \frac{1}{2^N} \sum_{s_1,\dots,s_N \in \{-1,1\}} \int_{\mathbb{R}} \widehat{\Phi}(x) \exp\left(2\pi \mathrm{i} x \sum_{n=1}^{N} s_n a_n\right) \mathrm{d}x \\
&= \int_{\mathbb{R}} \widehat{\Phi}(x) \frac{1}{2^N} \sum_{s_1,\dots,s_N \in \{-1,1\}} \exp\left(2\pi \mathrm{i} x \sum_{n=1}^{N} s_n a_n\right) \mathrm{d}x.
\end{aligned}$$

Exploiting the fact that $\mathrm{e}^{\mathrm{i}\alpha} + \mathrm{e}^{-\mathrm{i}\alpha} = 2\cos(\alpha)$, we have

$$\sum_{s_1,\dots,s_N\in\{-1,1\}} \exp\left(2\pi ix\sum_{n=1}^{N} s_n a_n\right) = \frac{1}{2}\sum_{s_1,\dots,s_N\in\{-1,1\}} 2\cos\left(2\pi x\sum_{n=1}^{N} s_n a_n\right).$$

We now extend Werner's trigonometric identity

$$\cos\alpha_1\cdot\cos\alpha_2 = \frac{1}{2}(\cos(\alpha_1+\alpha_2)+\cos(\alpha_1-\alpha_2))$$

to the case of N factors. It easy to see that

$$\prod_{n=1}^{N}\cos\alpha_i = \frac{1}{2^{N-1}}\sum_{s_2,\dots,s_N\in\{-1,1\}}\cos\left(\alpha_1+\sum_{n=2}^{N} s_n\alpha_n\right);$$

since

$$\sum_{s_2,\dots,s_N\in\{-1,1\}}\cos\left(-\alpha_1+\sum_{n=2}^{N} s_n\alpha_n\right) = \sum_{s_2,\dots,s_N\in\{-1,1\}}\cos\left(-\alpha_1+\sum_{n=2}^{N}(-s_n)\alpha_n\right)$$

because we just rearranged the summands, and since the cosine is an even function, we have that

$$\prod_{n=1}^{N}\cos\alpha_n = \frac{1}{2^N}\sum_{s_1,\dots,s_N\in\{-1,1\}}\cos\left(\sum_{n=1}^{N} s_n\alpha_n\right).$$

So this implies that

$$\varrho_N(x) = \frac{1}{2^N}\sum_{s_1,\dots,s_N\in\{-1,1\}}\cos\left(\pi x\sum_{n=1}^{N} s_n a_n\right),$$

and we finally obtain

$$\mathbb{E}[\Phi(X_N)] = \int_{\mathbb{R}}\widehat{\Phi}(x)\varrho_N(2x)\,\mathrm{d}x. \qquad \square$$

We will need also a generalization of Lemma 2.5 from [3], which is the following

Lemma 2.4. *For all $N\in\mathbb{N}$ and $x\in[0,\sqrt{N}]$, we have*

$$\varrho_N(x) = \varrho(x)\left(1+O\left(x^2/N\right)\right). \tag{2.6}$$

Proof. We recall that a_n is defined as in (2.2) and that $b_n = 1/a_n$ satisfies (2.3); in particular $a_n = O(1/n)$. Following the proof of Lemma 2.5 of [3], we have

$$\prod_{n=N+1}^{\infty}\cos(\pi x/n) = \prod_{n=N+1}^{\infty}\left(1+O\big((x/n)^2\big)\right) = \exp\left(O\big(x^2/N\big)\right) = 1+O\big(x^2/N\big).$$

Hence

$$\varrho_N(x) = \frac{\varrho(x)}{1+O\big(x^2/N\big)} = \varrho(x)\left(1+O\left(x^2/N\right)\right). \qquad \square$$

Let us now define, for every positive integer N and any real δ and x, the set

$$\mathcal{S}\big(N,\delta,x,(a_n)_{n\geq 1}\big) := \{\, n \in \{1,\dots,N\} \colon \|x\,a_n\| \geq \delta \,\},$$

where $\|\cdot\|$ denotes the distance from the nearest integer. For brevity, we will not indicate the dependency on the sequence $(a_n)_{n\geq 1}$ when it is not necessary.

Lemma 2.5. *For all $N \in \mathbb{N}$ and for all $x,\delta \geq 0$ we have*

$$|\varrho_N(x)| \leq \exp\left(-\frac{\pi^2\delta^2}{2}\cdot|\mathcal{S}(N,\delta,x)|\right).$$

Proof. Using the inequality

$$|\cos(\pi x)| \leq \exp\left(-\frac{\pi^2\|x\|^2}{2}\right),$$

we have

$$\begin{aligned}|\varrho_N(x)| = \prod_{n=1}^{N}|\cos(\pi x a_n)| &\leq \exp\left(-\frac{\pi^2}{2}\sum_{n=1}^{N}\|xa_n\|^2\right)\\ &\leq \exp\left(-\frac{\pi^2}{2}\sum_{\substack{n\leq N\\ \|xa_n\|\geq\delta}}\|xa_n\|^2\right) \leq \exp\left(-\frac{\pi^2\delta^2}{2}\cdot|\mathcal{S}(N,\delta,x)|\right).\end{aligned}$$

□

Lemma 2.6. *For any $N \in \mathbb{N}$, $x \in \mathbb{R}$ and $0 < \delta < 1/2$ we have*

$$\frac{N}{2} - D(N,y(\delta),x) < |\mathcal{S}(N,\delta,x)| < N - D(N,y(\delta)/2,x),$$

where

$$D(N,y,x) = D\left(N,y,x,(b_n)_{n\geq 1}\right) := \sum_{x-y<m<x+y}\ \sum_{\substack{b_n\mid m\\ N/2\leq n\leq N}} 1$$

and $y(\delta) := \delta N B(N)$.

Proof. As in Lemma 3.3 of [3], we observe that

$$\frac{N}{2} - T(N,\delta,x) < \left[\frac{N}{2}\right] + 1 - T(N,\delta,x) \leq |\mathcal{S}(N,\delta,x)| < N - T(N,\delta,x),$$

where

$$T(N,\delta,x) := \left|\{n \in \mathbb{N}\cap[N/2,N] \colon \|xa_n\| < \delta\}\right|.$$

Now, recalling that $a_n = 1/b_n$, we have

$$\begin{aligned} T(N,\delta,x) &= |\{n \in \mathbb{N} \cap [N/2,N] : \exists \ell \in \mathbb{N},\ \ell - \delta < xa_n < \ell + \delta\}| \\ &= |\{n \in \mathbb{N} \cap [N/2,N] : \exists \ell \in \mathbb{N},\ x - \delta b_n < \ell b_n < x + \delta b_n\}| \,. \end{aligned}$$

From our hypothesis (2.3) we know that $b_n \leq NB(N)$; then

$$\begin{aligned} T(N,\delta,x) &< |\{n \in \mathbb{N} \cap [N/2,N] \colon \exists \ell \in \mathbb{Z},\ x - y(\delta) < \ell b_n < x + y(\delta)\}| \\ &= D(N,y(\delta),x). \end{aligned}$$

This proves the lower bound; the upper bound follows with the same argument. □

2.2.2 The main results

Proposition 2.7. *Let A be a fixed positive constant and, for N sufficiently large,*

$$\beta(N) \leq \frac{1}{8\log\log N}.$$

Then there exists $C' > 0$ such that $|\varrho_N(x)| < x^{-A}$ for all sufficiently large positive integers N and for all $x \in [N, \exp(C'(\log N)^2)]$.

Proof. The proof follows along the same lines as Proposition 3.2 of [3]. We take

$$\overline{\delta} = \frac{2\sqrt{2A\log x}}{\pi}N^{-1/2} \quad \text{and} \quad x \in \left[N, \exp\left(\frac{\pi^2 N}{32A}\right)\right),$$

so that, for sufficiently large N, it holds that $0 < \overline{\delta} < 1/2$ and $y(\overline{\delta}) = \overline{\delta}NB(N) < x$. By Lemmas 2.5 and 2.6, if we show that $D(N, y(\overline{\delta}), x) < N/4$, then we get $|\varrho_N(x)| < 1/x^A$. Considering that b_n is a sequence of positive integers, we use Rankin's trick with $w \in (1/4, 1/2)$ and Ramanujan's result on $\sigma_{-s}(n)$ [17] (see the Appendix) to obtain

$$\begin{aligned}
D(N, y(\overline{\delta}), x) &< \frac{4}{\pi}\sqrt{2AN\log x}\, B(N) \cdot \max_{m\le 2x} \sum_{\substack{b_n \mid m \\ N/2\le n\le N}} 1 \\
&< \frac{4}{\pi}\sqrt{2AN\log x}\, B(N) \cdot \max_{m\le 2x} \sum_{\substack{k \mid m \\ N/2\le k\le NB(N)}} 1 \\
&\le \frac{4}{\pi}\sqrt{2AN\log x}\, B(N) \cdot \max_{m\le 2x} \sum_{\substack{k \mid m \\ N/2\le k\le NB(N)}} \left(\frac{NB(N)}{k}\right)^{w} \\
&= \frac{4}{\pi} N^{\frac{1}{2}+w} B(N)^{1+w}\sqrt{2A\log x} \cdot \max_{m\le 2x} \sum_{\substack{k \mid m \\ N/2\le k\le NB(N)}} k^{-w} \\
&\le \frac{4}{\pi} N^{\frac{1}{2}+w} B(N)^{1+w}\sqrt{2A\log x} \cdot \max_{m\le 2x} \sigma_{-w}(m) \\
&< \frac{4}{\pi} N^{\frac{1}{2}+w} B(N)^{1+w}\sqrt{2A\log x} \cdot \exp\left(C_1 \frac{(\log 2x)^{1-w}}{\log\log 2x}\right),
\end{aligned}$$

where C_1 is the constant of Ramanujan's theorem, as it is stated in Theorem 4.1.

Let $w = w(x) := 1/2 - \varphi(x)$, where φ is a positive decreasing function that we will choose later, with the constraint that $\varphi(x) < 1/2$. We will eventually choose $\varphi(x) \to 0$ as $x \to +\infty$. Then we have

$$B(N)^{1+w} = \exp\left(\left(\frac{3}{2} - \varphi(x)\right)\beta(N)\log N\right),$$

and so we would be done if we showed that

$$N^{1-\varphi(x)+(3/2-\varphi(x))\beta(N)}\sqrt{\log x} \cdot \exp\left(C_1 \frac{(\log 2x)^{1/2+\varphi(x)}}{\log\log 2x}\right) = o(N),$$

that is

$$\sqrt{\log x} \cdot \exp\left(C_1 \frac{(\log 2x)^{1/2+\varphi(x)}}{\log\log 2x}\right) = o(N^{\varphi(x)+(\varphi(x)-3/2)\beta(N)}). \tag{2.7}$$

Hence we must have

$$\varphi(x) + (\varphi(x) - 3/2)\beta(N) > 0,$$

that is

$$\beta(N) < \frac{\varphi(x)}{3/2 - \varphi(x)} \approx \frac{2}{3}\varphi(x). \tag{2.8}$$

Since φ is decreasing and we want to maintain the same range for x as in [3], that is $x \in [N, \exp(C'(\log N)^2)]$, it must hold that

$$\beta(N) \lesssim \frac{2}{3}\,\varphi\Big(\exp\left(C'(\log N)^2\right)\Big).$$

Let us take $\varphi(x) = (\log\log 2x)^{-1}$ and $\beta(N)$ such that, for $x \in [N, \exp(C'(\log N)^2)]$ and sufficiently large N, it holds

$$\beta(N) \leq \frac{2}{3J}\varphi(x) = \frac{2}{3J}\frac{1}{\log\log 2x}, \tag{2.9}$$

where $J \in \mathbb{R}$, $J > 1$. Then we would achieve our goal if we showed that

$$\sqrt{\log x} \cdot \exp\left(C_1 \mathrm{e} \frac{(\log 2x)^{1/2}}{\log\log 2x}\right) = o\left(\exp\left(\left(1 - \frac{1}{J} + o(1)\right)\frac{\log N}{\log\log 2x}\right)\right),$$

that is

$$\exp\left(C_1 \mathrm{e} \frac{(\log 2x)^{1/2}}{\log\log 2x} - \left(1 - \frac{1}{J} + o(1)\right)\frac{\log N}{\log\log 2x} + \frac{1}{2}\log\log x\right) = o(1).$$

This condition is equivalent to

$$C_1 \mathrm{e} \frac{(\log 2x)^{1/2}}{\log\log 2x} - \left(1 - \frac{1}{J} + o(1)\right)\frac{\log N}{\log\log 2x} + \frac{1}{2}\log\log x \to -\infty.$$

Taking into account the ranges for x, we see that it is sufficient to have

$$\frac{1}{\log\log N}\left[C_1\sqrt{C'}\,\mathrm{e}\log N(1 + o(1)) - \left(1 - \frac{1}{J}\right)\log N + O\left((\log\log N)^2\right)\right] \to -\infty.$$

We recall that, by our choice of x and N, we have $\log\log x \asymp \log\log N$. Hence, we just need to take C' sufficiently small, in a way that

$$C' < \left(\frac{J-1}{C_1 \mathrm{e} J}\right)^2, \tag{2.10}$$

to guarantee that $D(N, y(\overline{\delta}), x) < N/4$ for large N. For the sake of simplicity, we take $J = 2$ and the proposition is proved as stated. □

We remark here that the condition (1.4) on β, which we assumed to prove the proposition, was necessary to ensure the existence of the function φ satisfying all the properties we needed, and in particular (2.9).

Corollary 2.8. *Let A be a fixed positive constant and β satisfy* (2.4)*. Then, for all sufficiently large $x \in \mathbb{R}$, we have $|\varrho(x)| < x^{-A}$.*

Proof. We have

$$|\varrho(x)| = \left|\varrho_{\lfloor x\rfloor}(x)\prod_{n=\lfloor x\rfloor+1}^{\infty}\cos(\pi x a_n)\right| < x^{-A},$$

where to conclude we used Proposition 2.7. □

Theorem 2.9. *Let $C' > 0$ satisfy* (2.10) *and β satisfy* (2.4)*. Then for all intervals $I \subseteq \mathbb{R}$ of length $|I| > \exp(-C'(\log N)^2)$ one has*

$$\mathbb{P}[X_N \in I] = \int_I g(x)\,\mathrm{d}x + o(|I|),$$

as $N \to \infty$, where

$$g(x) := 2\int_0^\infty \cos(2\pi ux)\prod_{n=1}^\infty \cos\left(\frac{2\pi u}{b_n}\right)\mathrm{d}u = 2\int_0^\infty \cos(2\pi ux)\varrho(2u)\,\mathrm{d}u.$$

The proof follows along the same lines as Theorem 2.1 in [3] and we omit the details for brevity.

Corollary 2.10. *Let β satisfy* (2.4)*. For all $\tau \in \mathbb{R}$ and $C' > 0$ satisfying* (2.10)*, we have*

$$\left|\left\{ (s_1,\dots,s_N) \in \{\pm 1\}^N : \left|\tau - \sum_{n=1}^N \frac{s_n}{b_n}\right| < \delta \right\}\right| \sim 2^{N+1} g(\tau)\delta(1 + o_{C',\tau}(1))$$

as $N \to \infty$ and $\delta \to 0$, uniformly in $\delta \geq \exp(-C'(\log N)^2)$. In particular, for large enough N, one has $\mathfrak{m}_N(\tau) < \exp(-C'(\log N)^2)$.

Remark 2.1. We have imposed condition (2.4) for β to keep the same range of validity for x as in [3]. We remark that the hypotheses on β could be relaxed at the price of restricting this range: for example, we could take

$$\beta(N) = \frac{\log\log\log N}{\log\log N},$$

and obtain the result of Proposition 2.7 for $x \in [N, \exp(\log^a N)]$, where $a \in (1,2)$ is a suitable constant. In fact, this would weaken directly the estimates that we have just found in Theorem 2.9 and Corollary 2.10, where $\exp(-C'(\log N)^2)$ would be replaced by $\exp(-\log^a N)$.

2.3 Products of k primes

We now leave the general case and concentrate on primes and products of k distinct primes. Hence, we define

$$\mathcal{P}_k := \{\, n \in \mathbb{N} \mid n \text{ is the product of } k \text{ distinct primes} \,\};$$

we will denote by $b_n^{(k)}$ the n-th element of the ordered set $\mathcal{P}_k$. Let us recall the definition of $\mathcal{S}(N,\delta,x)$ in the case $a_n = 1/b_n^{(k)}$:

$$\mathcal{S}(N,\delta,x) := \{\, n \in \{1,\dots,N\} : \|x/b_n^{(k)}\| \geq \delta \,\}\,.$$

We remark that, since we left the general case, we can now take $B(n) = b_n^{(k)}/n$, and denote it by $B_k(n)$. In 1900, Landau [13] proved that

$$\pi_k(t) := |\mathcal{P}_k \cap \{\, n \in \mathbb{N} \mid n \leq t \,\}\,| = \frac{t}{\log t}\frac{(\log\log t)^{k-1}}{(k-1)!} + O\left(\frac{t(\log\log t)^{k-2}}{\log t}\right),$$

which implies that

$$B_k(n) \sim \log n \frac{(k-1)!}{(\log\log n)^{k-1}}. \tag{2.11}$$

We can now start with a refinement of Proposition 2.7, where we extend the interval of validity for x in the case $b_n = b_n^{(k)}$.

Proposition 2.11. *Let A be a fixed positive constant, $k \in \mathbb{N}$ be fixed and $a_n = 1/b_n^{(k)}$, where $b_n^{(k)}$ is the n-th element of the ordered set $\mathcal{P}_k$. Then $|\varrho_N(x)| < x^{-A}$ for all sufficiently large positive integers N and for all $x \in [U, \exp(f(N))]$, where $\log N = o(f(N))$ and*

$$f(N) = o\left(\left(\frac{N}{B_k^2(N)}\right)^{1/(2k+1)}\right),$$

and $U > 1$ is a constant depending on f.

Proof. Let $x \in [N, \exp(f(N))]$. As in the proof of Proposition 2.7, we need to show that $D(N, y(\overline{\delta}), x) < N/4$, where $\overline{\delta}$ is chosen in the same way and $y(\overline{\delta}) = \overline{\delta} N B_k(N)$. Since now we are considering $x \geq N$, it is easy to see that for sufficiently large N we have $y(\overline{\delta}) \leq x$. We recall here that the prime omega function $\omega(n)$ is defined as the number of different prime factors of n, and that

$$\omega(n) \ll \frac{\log n}{\log\log n},$$

as a consequence of the prime number theorem. Hence, we have

$$\begin{aligned}
D(N, y(\overline{\delta}), x) := & \sum_{x-y(\overline{\delta})<m<x+y(\overline{\delta})} \sum_{\substack{b_n^{(k)}|m \\ N/2\leq n\leq N}} 1 \leq \sum_{x-y(\overline{\delta})<m<x+y(\overline{\delta})} \sum_{\substack{p_{j_1}\cdots p_{j_k}|m \\ p_{j_i}\text{ distinct primes}}} 1 \\
\leq & \sum_{x-y(\overline{\delta})<m<x+y(\overline{\delta})} \omega(m)^k \leq (2y(\overline{\delta})+1) \max_{m<x+y(\overline{\delta})} \omega(m)^k \\
\ll & (N\log x)^{1/2} B_k(N) \left(\frac{\log 2x}{\log\log 2x}\right)^k \ll N^{1/2} B_k(N)\,(\log x)^{k+1/2},
\end{aligned}$$

where we used the trivial bound for the prime omega function. If we show that this quantity is $o(N)$, we are done. So we need

$$\log x = o\left(\left(\frac{N}{B_k^2(N)}\right)^{1/(2k+1)}\right).$$

Hence we can take any f that satisfies

$$f(N) = o\left(\left(\frac{N}{B_k^2(N)}\right)^{1/(2k+1)}\right),$$

where we recall that B_k satisfies (2.11). The theorem is then proved for $x \in [N, \exp(f(N))]$. If $x < N$, it holds

$$|\varrho_N(x)| \leq |\varrho_{\lfloor x \rfloor}(x)|,$$

hence the result we have just proved holds also whenever $x \leq \exp\big(f(\lfloor x \rfloor)\big)$. But there must exist $U > 0$ such that this holds for any $x > U$, since $\log x = o(f(x))$. □

We are now ready to prove a more general version of Theorem 2.1 of [3] for the sequence $\big(b_n^{(k)}\big)_{n\in\mathbb{N}}$.

Theorem 2.12. *Let f and a_n be defined as in Proposition 2.11. Then for all intervals $I \subseteq \mathbb{R}$ of length $|I| > \exp(-f(N))$ one has*

$$\mathbb{P}[X_N \in I] = \int_I g(x)\,\mathrm{d}x + o(|I|),$$

as $N \to \infty$, where

$$g(x) := 2\int_0^\infty \cos(2\pi u x)\prod_{n=1}^\infty \cos\left(\frac{2\pi u}{b_n^{(k)}}\right)\mathrm{d}u = 2\int_0^\infty \cos(2\pi u x)\varrho(2u)\,\mathrm{d}u.$$

Proof. The proof follows the one of Theorem 2.1 of [3]. Let $\varepsilon > 0$ be fixed. We define

$$\xi := \xi_{N,-\varepsilon} := \exp(-(1-\varepsilon)f(N)),$$
$$\xi_0 := \xi_{N,0} = \exp(-f(N)),$$

so that $\xi^{-1} < \xi_0^{-1}$ and Proposition 2.11 holds for $x \in [N, \xi_0^{-1}]$. For an interval $I = [a,b]$ with $b - a > 2\xi_0$, let us define $I^+ := [a-\xi, b+\xi]$ and $I^- := [a+\xi, b-\xi]$.

Then one can construct two smooth functions $\Phi^{\pm}_{N,\varepsilon,I}(x):\mathbb{R}\to[0,1]$ (from now on, we will drop the subscripts when they are clear by the context) such that

$$\begin{cases}\operatorname{supp}\Phi^+\subseteq I^+ & \\ \Phi^+(x)=1 & \text{for } x\in I,\\ \operatorname{supp}\Phi^-\subseteq I & \\ \Phi^-(x)=1 & \text{for } x\in I^-,\end{cases}$$

and besides

$$(\Phi^{\pm})^{(j)}(x)\ll_j \xi^{-j}\quad \text{for all } j\geq 0. \tag{2.12}$$

By (2.12), we know that the Fourier transforms of $\Phi^{\pm}$ satisfy

$$\widehat{\Phi^{\pm}}(x)\ll_B (1+|x|\xi)^{-B}\quad \text{for any } B>0 \text{ and } x\in\mathbb{R}. \tag{2.13}$$

Since

$$\mathbb{E}[\Phi^-(X_N)]\leq \mathbb{P}[X_N\in I]\leq \mathbb{E}[\Phi^+(X_N)],$$

we just need to show that

$$\mathbb{E}[\Phi^{\pm}(X_N)]=\int_{\mathbb{R}}\Phi^{\pm}(x)g(x)\,\mathrm{d}x+o_\varepsilon(|I|).$$

From now on, Φ will indicate either Φ^+ or Φ^-. By Lemma 2.3 we have

$$\mathbb{E}[\Phi(X_N)]=\frac{1}{2}\int_{\mathbb{R}}\widehat{\Phi}(x/2)\varrho_N(x)\,\mathrm{d}x=I_1+I_2+I_3,$$

where I_1, I_2 and I_3 are the integrals supported in $|x|<N^\varepsilon$, $|x|\in[N^\varepsilon,\xi^{-(1+\varepsilon)}]$ and $|x|>\xi^{-(1+\varepsilon)}$ respectively. Note that $\xi^{-(1+\varepsilon)}=\mathrm{e}^{(1-\varepsilon^2)f(N)}>\mathrm{e}^{\varepsilon\log N}=N^\varepsilon$, that $\xi^{-(1+\varepsilon)}=\xi_0^{-(1-\varepsilon^2)}<\xi_0^{-1}$, and that $\xi^{-(1+\varepsilon)}\cdot\xi=\xi^{-\varepsilon}=\xi_0^{-\varepsilon(1-\varepsilon)}\to+\infty$ for $N\to+\infty$. By Lemma 2.4 and Corollary 2.8, we have

$$\begin{aligned}I_1&=\frac{1}{2}\int_{-N^\varepsilon}^{N^\varepsilon}\widehat{\Phi}(x/2)\varrho_N(x)\,\mathrm{d}x=\frac{1}{2}\int_{-N^\varepsilon}^{N^\varepsilon}\widehat{\Phi}(x/2)\varrho(x)\,\mathrm{d}x+O\left(\|\widehat{\Phi}\|_\infty N^{-1+3\varepsilon}\right)\\&=\frac{1}{2}\int_{\mathbb{R}}\widehat{\Phi}(x/2)\varrho(x)\,\mathrm{d}x+O_A\left(\|\widehat{\Phi}\|_\infty N^{-(A-1)\varepsilon}\right)+O\left(\|\widehat{\Phi}\|_\infty N^{-1+3\varepsilon}\right)\\&=\int_{\mathbb{R}}\widehat{\Phi}(x)\varrho(2x)\,\mathrm{d}x+O_\varepsilon\left(\|\Phi\|_1 N^{-1+3\varepsilon}\right),\end{aligned}$$

where to conclude we chose $A=A(\varepsilon)$ sufficiently large. For the second integral, we use Proposition 2.11 and obtain

$$\begin{aligned}|I_2|&\leq\|\widehat{\Phi}\|_\infty\int_{N^\varepsilon}^{\xi^{-(1+\varepsilon)}}|\varrho_N(x)|\,\mathrm{d}x\leq\|\Phi\|_1\int_{N^\varepsilon}^{\xi^{-(1+\varepsilon)}}x^{-A}\,\mathrm{d}x\leq\|\Phi\|_1\int_{N^\varepsilon}^{+\infty}x^{-A}\,\mathrm{d}x\\&\ll_\varepsilon\|\Phi\|_1 N^{-A\varepsilon+\varepsilon}\ll_\varepsilon\|\Phi\|_1 N^{-1},\end{aligned}$$

where, as before, to conclude we took $A = A(\varepsilon)$ sufficiently large. For the last integral, we recall that trivially $|\varrho_N(x)| \leq 1$; using the bound (2.13), we obtain

$$|I_3| \leq \int_{|x|>\xi^{-(1+\varepsilon)}} |\widehat{\Phi}(x/2)| \,\mathrm{d}x \ll_B \int_{\xi^{-(1+\varepsilon)}}^{+\infty} (1+x\xi)^{-B} \,\mathrm{d}x = (B-1)(\xi^{-1}+\xi^{-(1+\varepsilon)})^{1-B}$$
$$\ll_B \xi_0^{B-1} = o_\varepsilon(\xi_0) = o_\varepsilon(|I|),$$

where to conclude we chose $B = B(\varepsilon)$ sufficiently large. We can now put these results together: using Parseval's theorem and the fact that $\|\Phi\|_1 = O_\varepsilon(|I|)$, we get

$$\mathbb{E}[\Phi(X_N)] = \int_{\mathbb{R}} \widehat{\Phi}(x)\varrho(2x)\,\mathrm{d}x + O_\varepsilon\left(\|\Phi\|_1 N^{-1+3\varepsilon}\right) + o_\varepsilon(|I|) = \int_{\mathbb{R}} \Phi(x)g(x)\,\mathrm{d}x + o_\varepsilon(|I|)$$

and the theorem is then proved. □

Remark 2.2. By Corollary 2.8, for any $n \in \mathbb{N}$ it holds

$$\int_{-\infty}^{+\infty} |t^n \varrho(t)| \,\mathrm{d}t < \infty,$$

which implies by standard arguments (see e.g. §5 of [18]) that the density g is a smooth strictly positive function. Besides, by the same corollary, $g(x) \ll_D x^{-D}$ for any $D > 0$.

Corollary 2.13. *For all $\tau \in \mathbb{R}$, we have*

$$\left|\left\{ (s_1,\ldots,s_N) \in \{\pm 1\}^N : \left|\tau - \sum_{n=1}^N \frac{s_n}{b_n^{(k)}}\right| < \delta \right\}\right| \sim 2^{N+1} g(\tau)\delta(1+o_\tau(1))$$

as $N \to \infty$ and $\delta \to 0$, uniformly in $\delta \geq \exp(-f(N))$, where f is defined as in Proposition 2.11. In particular, for N large enough, one has $\mathfrak{m}_N(\tau) < \exp\left(-f(N)\right)$.

Chapter 3

On the distribution of the digits of quotients of integers

3.1 Introduction and statement of the problem

It is a somewhat surprising discovery of Cesàro that if one takes two "random" positive integers n and m and considers the distribution of the first "decimal" digit of their ratio n/m, it is slightly more likely that this turns out to be 0 rather than, say, 1.

In order to be more precise, we start by giving some definitions. We consider a number basis $b \geq 2$, the set of digits $S_b = \{0, \ldots, b-1\}$ and we are concerned with the i-th place to the right of the point, where i is a positive integer. We recall that $\lfloor x \rfloor \in \mathbb{Z}$ and $\{x\} \in [0, 1)$ denote the integer and the fractional part of the real number x, respectively, so that $x = \lfloor x \rfloor + \{x\}$. For a given $x \in \mathbb{R}$, a basis b and a position i, we remark that the i-th digit of x, which we denote by $\phi(x; b; i)$, can be computed by means of

$$\phi(x; b; i) = \lfloor b\{b^{i-1}x\} \rfloor = \lfloor b^i x - b\lfloor b^{i-1}x \rfloor \rfloor.$$

In fact, this formula is correct for any $i \in \mathbb{Z}$, with the obvious interpretation if $i < 0$. With b and i as above and a digit $r \in S_b$, we recall that $\Phi(T; b, r; i) := |\mathcal{A}(T; b, r; i)|$, where

$$\mathcal{A}(T; b, r; i) := \{(n, m) \in \mathbb{N}^2 \cap [1, T]^2 \colon \phi(n/m; b; i) = r\}.$$

As we said, Gambini, Mingari Scarpello and Ritelli [11] implicitly obtained the asymptotic formula

$$\Phi(T;b,r;i) = c(b,r;i)T^2 + O\left(T^{3/2}\right),$$

as $T \to +\infty$. The constant $c(b,r;i)$ is defined as an infinite series and can be expressed by means of the digamma function ψ, as follows:

$$c(b,r;i) = \frac{1}{2b} + \frac{1}{2}b^{i-1}\Big(\psi\Big(\frac{b^i+r+1}{b}\Big) - \psi\Big(\frac{b^i+r}{b}\Big)\Big). \tag{3.1}$$

In this chapter we introduce some number-theoretic devices which allow us to improve upon this result. We remark here that we studied just the case $i = 1$. In the first section of this chapter, we prove the following

Theorem 3.1. *We have*

$$\Phi(T;b,r;1) = c(b,r;1)T^2 + O\left(T^{4/3}\right),$$

as $T \to +\infty$, where $c(b,r;1)$ is the constant defined in (3.1)*. The implicit constant may depend on b.*

In the second section of this chapter, we deal with the case of primes. Then we actually consider the function $\Phi'(T;b,r;i) := |\mathcal{A}'(T;b,r;i)|$, where

$$\mathcal{A}'(T;b,r;i) := \{(p,q) \in \mathbb{P}^2 \cap [1,T]^2 \colon \phi(p/q;b;i) = r\}$$

and $\mathbb{P}$ is the set of primes. We prove the following

Theorem 3.2. *In the case of primes, we have*

$$\Phi'(T;b,r;1) = c(b,r;1)\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right),$$

as $T \to +\infty$, where $c(b,r;1)$ is the constant defined in (3.1)*. The implicit constant may depend on b.*

3.2 The proof of the theorem for integers

We follow [11] quite closely. We first decompose the set $\mathcal{A}(T;b,r;i)$ as an appropriate union of sets. We write

$$\mathcal{A}_k(T) = \mathcal{A}_k(T;b,r;i) = \Big\{(n,m) \in \mathbb{N}^2 \cap [1,T]^2 \colon \frac{n}{m} \in \Big[\frac{kb+r}{b^i}, \frac{kb+r+1}{b^i}\Big)\Big\},$$

so that $\mathcal{A}(T;b,r;i) = \bigcup_{k\geq 0} \mathcal{A}_k(T;b,r;i)$. This is easily checked using the definition of ϕ. The sets $\mathcal{A}_k(T)$ are pairwise disjoint and correspond to the lattice points contained in triangles with a vertex at the origin and the other vertices either on the segment $[1,T]\times\{T\}$, when $k < b^{i-1}$ (i.e. $n < m$), or on $\{T\}\times[1,T]$, when $k \geq b^{i-1}$ (i.e. $n \geq m$). Therefore, we split the infinite union above accordingly as

$$\mathcal{U}(T;b,r;i) := \bigcup_{k=0}^{b^{i-1}-1} \mathcal{A}_k(T;b,r;i) \qquad \text{and} \qquad \mathcal{L}(T;b,r;i) := \bigcup_{k\geq b^{i-1}} \mathcal{A}_k(T;b,r;i).$$

At this point, it is worth looking back at the definition of $\mathcal{A}_k(T)$. We have that $n \geq T$ and $m \geq 1$, so that $n/m \leq T$. Hence, when k is such that

$$\frac{kb+r}{b^i} > T,$$

the set $\mathcal{A}_k(T)$ is empty. This means that we can limit the range for k, taking

$$k \leq b^{i-1}T - \frac{r}{b} \leq b^{i-1}T.$$

So it holds that

$$\mathcal{L}(T;b,r;i) = \bigcup_{b^{i-1}\leq k\leq b^{i-1}T} \mathcal{A}_k(T;b,r;i).$$

For k above a certain threshold depending on T, it is difficult to evaluate the cardinality of $\mathcal{A}_k(T)$ exactly: this will be the source of our first error term. To a first approximation, $|\mathcal{A}_k(T)|$ is the area of the corresponding triangle, with an error proportional to the perimeter, that is $O(T)$. However, as $T \to +\infty$, the number of such triangles tends to infinity as well, and we have to be much more careful than this. Our improvement over the previous results depends on the fact that we manage to choose a specific threshold for k: above it, we will estimate $|\bigcup_k \mathcal{A}_k(T)|$ in a trivial way, while for k less than it we will be able to keep a good precision in evaluating $|\mathcal{A}_k(T)|$.

3.2.1 Estimate of $\mathcal{U}(T)$

We define $d_r := (r,b)$ for $r \in S_b \cup \{b\}$. For $r = 0$ we let $\delta(T;b,0) = 0$, and for $b > 0$ we denote by $\delta(T;b,r)$ the number

$$\begin{aligned}\delta(T;b,r) &:= \left|\left\{n \leq T\colon \frac{rn}{b} \in \mathbb{N}\right\}\right| = \left|\left\{n \leq T\colon \frac{b}{d_r} \mid n\right\}\right| \\ &= \left\lfloor \frac{Td_r}{b} \right\rfloor = \frac{d_r}{b}T + O(1). \end{aligned} \tag{3.2}$$

We remark that we will treat just the case $i = 1$. First of all we notice that the edges of the sets $\mathcal{A}_k(T)$ can be ignored: we are referring to the sets

$$\partial\mathcal{A}_k(T) := \left\{(n,m) \in \mathbb{N}^2 \cap [1,T]^2 \colon \frac{n}{m} \in \left\{\frac{kb+r}{b}, \frac{kb+r+1}{b}\right\}\right\}$$

and we want to study their cardinality. It is trivial to see that $|\partial\mathcal{A}_0(T)| \ll T$. Let us suppose, then, that $k \geq 1$, and concentrate on the first condition, that is

$$\frac{n}{m} = \frac{kb+r}{b}, \tag{3.3}$$

since for the second one the reasoning is analogous. We want to count the couples of all natural numbers $n, m \leq T$ satisfying it. First of all we reduce the fraction

$$\frac{kb+r}{b} = \frac{\frac{kb+r}{(kb+r,b)}}{\frac{b}{(kb+r,b)}} = \frac{(kb+r)/(kb+r,b)}{b/(kb+r,b)} = \frac{\frac{kb+r}{(r,b)}}{\frac{b}{(r,b)}}.$$

Then all the couples (n,m) will be of the form

$$\left(a \cdot \frac{kb+r}{(r,b)}, a \cdot \frac{b}{(r,b)}\right),$$

with $a \in \mathbb{N}$ such that both elements are at most T. Looking at the first one, we get

$$a \leq \frac{T \cdot (r,b)}{kb+r} \ll \frac{T}{k}.$$

For the second condition the argument is almost the same, and so $|\partial\mathcal{A}_k(T)| \ll T/k$. This implies that

$$\left|\bigcup_{0 \leq k \leq T} \partial\mathcal{A}_k(T)\right| \ll T + \sum_{1 \leq k \leq T} \frac{T}{k} \ll T \log T.$$

This error term is far better then the one we will obtain in the theorem: edges can thus be ignored.

Now let us consider $\mathcal{U}(T)$. We have

$$\begin{aligned}
&|\mathcal{U}(T;b,r;1)| \\
&\quad = |\mathcal{A}_0(T)| = \sum_{n \in [1,T]} \left(\left[\frac{(r+1)n}{b}\right] - \left[\frac{rn}{b}\right]\right) - \delta(T;b,r+1) + \delta(T;b,r) \\
&\quad = \sum_{n \in [1,T]} \frac{n}{b} - \sum_{n \in [1,T]} \left(\left\{\frac{(r+1)n}{b}\right\} - \left\{\frac{rn}{b}\right\}\right) - \delta(T;b,r+1) + \delta(T;b,r). \quad (3.4)
\end{aligned}$$

We now exploit a simple and useful property of the fractional parts.

Lemma 3.3. *Let $a \in \mathbb{N}$ and $q \in \mathbb{N}^*$ be integers with $(a,q)=1$, and let $T \geq 1$ be a real number. Then*

$$\sum_{n\leq T}\left\{\frac{an}{q}\right\} = \frac{1}{2}\Big(1-\frac{1}{q}\Big)T + O\left(q\right).$$

Proof. We split the interval $[1,T]$ into $[T/q]$ intervals of type $[kq+1,(k+1)q]$ and an interval of length $< q$. We now exploit the periodicity of the fractional part, and the fact that the map $n \mapsto an \bmod q$ is a permutation of the integers $\{0,\dots,q-1\}$. Hence

$$\sum_{n=kq+1}^{(k+1)q}\left\{\frac{an}{q}\right\} = \sum_{n=1}^{q}\left\{\frac{n}{q}\right\} = \sum_{n=1}^{q-1}\frac{n}{q} = \frac{1}{2}(q-1)$$

and

$$\sum_{n\leq T}\left\{\frac{an}{q}\right\} = \frac{1}{2}(q-1)[T/q] + O\left(q\right) = \frac{1}{2}\Big(1-\frac{1}{q}\Big)T - \frac{1}{2}(q-1)\{T/q\} + O\left(q\right). \quad \square$$

Combining (3.2) and (3.4), if $T \in \mathbb{N}$ we have

$$\begin{aligned}
|\mathcal{U}(T;b,r;1)| &= \frac{|T|\left(|T|+1\right)}{2b} + (d_r - d_{r+1})\left(\frac{(b-1)T}{2b} + O(b)\right) \\
&\quad + \left(d_r - d_{r+1}\right)T/b + O(1) \\
&= \frac{T^2}{2b} + O\left(\frac{T}{b}\right) + \frac{d_r - d_{r+1}}{2b}(b+1)T + O\left(b^2\right) \\
&= \frac{1}{2b}T^2 + O_b\left(T\right).
\end{aligned}$$

3.2.2 Estimate of $\mathcal{L}(T)$

We now turn to the problem of estimating $\mathcal{L}(T;b,r;1)$. We choose $\beta \in (0,1)$ and estimate trivially the contribution from $k > T^{1-\beta}$, that is

$$\begin{aligned}
\left|\bigcup_{k\geq T^{1-\beta}}\mathcal{A}_k(T)\right| &\leq \left|\Big\{(n,m)\in\mathbb{N}^2\cap[1,T]^2\colon \frac{n}{m} \geq \frac{T^{1-\beta}b+r}{b}\Big)\Big\}\right| \\
&\leq \left|\Big\{(n,m)\in\mathbb{N}^2\cap[1,T]^2\colon \frac{n}{m} \geq T^{1-\beta}\Big)\Big\}\right| \\
&\leq \sum_{T^{1-\beta}\leq n\leq T}\ \sum_{m\leq n/T^{1-\beta}} 1 \\
&\leq \sum_{T^{1-\beta}\leq n\leq T}\Big(\frac{n}{T^{1-\beta}} + O(1)\Big) \\
&= O\left(\frac{1}{T^{1-\beta}}\cdot T^2\right) \\
&= O\big(T^{1+\beta}\big).
\end{aligned}$$

Hence, from now on, we assume that $1 \le k \le T^{1-\beta}$. For $r \in S_b$ we define

$$y_k := y_k(b,r) := \frac{b}{bk+r} = \frac{1}{k+\frac{r}{b}},$$
$$x_k := y_k(b,r+1) = \frac{b}{bk+r+1} = \frac{1}{k+\frac{r+1}{b}}.$$

We notice here, even though we will need this later, that for $k \ge 1$ and $r \in S_b \cup \{b\}$ we have

$$\frac{1}{3k} \le \frac{1}{k}\cdot\frac{1}{2+1/b} = \frac{1}{k}\cdot\frac{1}{1+\frac{b+1}{b}} < \frac{1}{k}\cdot\frac{1}{1+\frac{r+1}{bk}} = \frac{1}{k+\frac{r+1}{b}} = x_k < y_k = \frac{1}{k+\frac{r}{b}} \le \frac{1}{k},$$

so that

$$\begin{cases} x_k \asymp 1/k \\ y_k \asymp 1/k. \end{cases} \tag{3.5}$$

Assuming that T is an integer, the number of lattice points in $\mathcal{A}_k(T;b,r;1)$ is

$$\begin{aligned} &\sum_{n\le x_kT}\Big(\Big[\frac{n}{x_k}\Big]-\Big[\frac{n}{y_k}\Big]\Big) + \sum_{x_kT\le n\le y_kT}\Big(T-\Big[\frac{n}{y_k}\Big]\Big) \\ &\quad = \sum_{n\le x_kT}\Big(\Big[\frac{n}{x_k}\Big]-\Big[\frac{n}{y_k}\Big]\Big) + \sum_{x_kT/y_k\le m\le T}\big([my_k]-[x_kT]\big). \end{aligned} \tag{3.6}$$

The first term in (3.6) is

$$\begin{aligned} &\sum_{n\le x_kT}\Big(\frac{n}{x_k}-\frac{n}{y_k}\Big) - \sum_{n\le x_kT}\Big(\Big\{\frac{n}{x_k}\Big\}-\Big\{\frac{n}{y_k}\Big\}\Big) \\ &\quad = \Big(\frac{1}{x_k}-\frac{1}{y_k}\Big)\sum_{n\le x_kT} n - \frac{1}{2}\Big(1-\frac{d_{r+1}}{b}\Big)x_kT + \frac{1}{2}\Big(1-\frac{d_r}{b}\Big)x_kT + O\left(b\right) \\ &\quad = \frac{y_k-x_k}{x_ky_k}\sum_{n\le x_kT} n + \frac{1}{2}\frac{d_{r+1}-d_r}{b}x_kT + O\left(b\right), \end{aligned} \tag{3.7}$$

by Lemma 3.3. We further rewrite the last summand in (3.6) as

$$\begin{aligned} \sum_{x_kT/y_k\le m\le T}\big([my_k]-[x_kT]\big) &= \sum_{x_kT/y_k\le m\le T}\big(my_k-\{my_k\}\big) - [x_kT]\Big(T-\Big[\frac{x_kT}{y_k}\Big]\Big) \\ &= I_1 - I_2, \end{aligned}$$

say. We have

$$\begin{aligned} I_1 &= \sum_{x_kT/y_k\le m\le T} my_k - \sum_{m\le T}\{my_k\} + \sum_{m< x_kT/y_k}\{my_k\} \\ &= y_k\sum_{x_kT/y_k\le m\le T} m - \frac{1}{2}\Big(1-\frac{d_r}{bk+r}\Big)T + \frac{1}{2}\Big(1-\frac{d_r}{bk+r}\Big)\frac{x_k}{y_k}T + O\left(k\right) \\ &= y_k\sum_{x_kT/y_k\le m\le T} m - \frac{1}{2}\Big(1-\frac{d_r}{bk+r}\Big)\frac{1}{bk+r}T + O\left(k\right). \end{aligned} \tag{3.8}$$

We also have

$$
\begin{aligned}
I_2 &= [x_kT]\Big(T - \Big[\frac{x_kT}{y_k}\Big]\Big) = \big(x_kT + O(1)\big)\Big(T - \frac{x_kT}{y_k} + \Big\{\frac{x_kT}{y_k}\Big\}\Big) \\
&= \frac{(y_k - x_k)x_k}{y_k}T^2 + O\left(\frac{y_k - x_k}{y_k}T\right) + O(x_kT) + O(b) \\
&= \frac{(y_k - x_k)x_k}{y_k}T^2 + O\left(\frac{T}{k}\right). \qquad (3.9)
\end{aligned}
$$

Summing up from (3.6), (3.7), (3.8) and (3.9), we have

$$
\begin{aligned}
&|\mathcal{A}_k(T; b, r; 1)| \\
&\quad = \frac{y_k - x_k}{x_ky_k}\sum_{n \le x_kT} n + y_k \sum_{x_kT/y_k \le m \le T} m - \frac{(y_k - x_k)x_k}{y_k}T^2 \\
&\qquad + \frac{1}{2}\frac{d_{r+1} - d_r}{b}x_kT - \frac{1}{2}\Big(1 - \frac{d_r}{bk + r}\Big)\frac{1}{bk + r}T + O(k) + O\left(\frac{T}{k}\right) \\
&\quad = \frac{y_k - x_k}{x_ky_k}\frac{1}{2}\big(x_k^2T^2 + O(x_kT)\big) + y_k\frac{1}{2}\Big(T^2 - \frac{x_k^2}{y_k^2}T^2 + O(T)\Big) - \frac{(y_k - x_k)x_k}{y_k}T^2 \\
&\qquad + \frac{1}{2}\Big(\frac{d_{r+1} - d_r}{bk + r + 1} - \Big(1 - \frac{d_r}{bk + r}\Big)\frac{1}{bk + r}\Big)T + O(k) + O\left(\frac{T}{k}\right) \\
&\quad = \frac{1}{2}(y_k - x_k)T^2 + O(y_kT) + O(k) + O\left(\frac{T}{k}\right). \qquad (3.10)
\end{aligned}
$$

By definition, $y_k \le k^{-1}$ so that the first error term can be absorbed into the last. We finally sum (3.10) over $k \le T^{1-\beta}$, obtaining

$$
\sum_{k \le T^{1-\beta}} |\mathcal{A}_k(T; b, r; 1)| = \frac{1}{2}T^2 \sum_{k \le T^{1-\beta}} (y_k - x_k) + O\left(\sum_{k \le T^{1-\beta}} (k + \frac{T}{k})\right).
$$

Since $y_k - x_k \le k^{-2}$, the error term arising from extending the first sum over k to all positive integers is $\ll T^{\beta-1}$. The other error terms contribute $\ll T^{2(1-\beta)} + T\log T$. Hence

$$
\sum_{k \le T^{1-\beta}} |\mathcal{A}_k(T; b, r; 1)| = \frac{1}{2}T^2 \sum_{k \ge 1}(y_k - x_k) + O\big(T^{2(1-\beta)} + T^{1+\beta}\big),
$$

since $\beta > 0$. We choose $\beta = \frac{1}{3}$ and the proof is complete, since

$$
\sum_{k \ge 1}(y_k - x_k) = \sum_{k \ge 1}\frac{b}{(bk + r)(bk + r + 1)} = \psi\Big(\frac{b + r + 1}{b}\Big) - \psi\Big(\frac{b + r}{b}\Big), \qquad (3.11)
$$

by (4.1).

3.3 The case of primes

We now follow the same ideas to study the distribution of the digits of quotients of primes. Recalling the definition of $\mathcal{A}_k(T)$, we adjust it to the case of primes in the following way:

$$\mathcal{A}'_k(T) := \mathcal{A}'_k(T;b,r;i) := \Big\{(p,q) \in \mathbb{P}^2 \cap [1,T]^2 \colon \frac{p}{q} \in \Big[\frac{kb+r}{b^i}, \frac{kb+r+1}{b^i}\Big)\Big\},$$

where $\mathbb{P}$ is the set of primes. It clearly holds that $\mathcal{A}'(T;b,r;i) = \bigcup_{k\geq 0} \mathcal{A}'_k(T;b,r;i)$. As before, we split $\mathcal{A}'(T;b,r;i)$ in two subsets, namely

$$\mathcal{U}'(T;b,r;i) := \bigcup_{k=0}^{b^{i-1}-1} \mathcal{A}'_k(T;b,r;i) \quad \text{and} \quad \mathcal{L}'(T;b,r;i) := \bigcup_{k\geq b^{i-1}} \mathcal{A}'_k(T;b,r;i)$$
$$= \bigcup_{b^{i-1}\leq k\leq b^{i-1}T} \mathcal{A}_k(T;b,r;i),$$

so that $\mathcal{A}'(T) = \mathcal{L}'(T) \cup \mathcal{U}'(T)$. If not differently specified, the definitions given for the case of integers hold for all the other quantities.

3.3.1 Estimate of $\mathcal{L}'(T)$

As before, we just consider the case $i = 1$ and estimate trivially the contribution for $k > T^{1-\beta}$ for $\beta \in (0,1)$, that gives an error term which is $O\big(T^{1+\beta}\big)$. In this problem with primes, the value of β will not be crucial, since the final error term will be much bigger. Its role is now to assure that k cannot have the same order of magnitude as T. So, from now on, we will assume that $1 \leq k \leq T^{1-\beta}$. First of all, we show that the contribution of the points on the oblique edges of $\mathcal{A}'_k(T)$ is small. We notice that the points (x,y) belonging to them must satisfy one of the following two equations: $y = x/x_k$ or $y = x/y_k$, which are

$$x(bk+r+1) = by,$$
$$x(bk+r) = by.$$

In either case, since x is prime, it must hold that $x \mid by$. But

$$\sum_{\substack{x,y\in\mathbb{P},\ x,y\leq T \\ x\mid b}} 1 \ll b \sum_{\substack{y\in\mathbb{P} \\ y\leq T}} 1 \ll \frac{T}{\log T}$$

and

$$\sum_{\substack{x,y\in\mathbb{P},\ x,y\le T\\ x\mid y}} 1 = \sum_{\substack{x,y\in\mathbb{P},\ x,y\le T\\ x=y}} 1 \ll \frac{T}{\log T},$$

hence the contribution of the points on the oblique sides is $O(T/\log T)$.

Now we can compute $|\mathcal{A}'_k(T)|$ without paying too much attention to its edges. With (3.6) in mind, we easily realize that for the case of primes we have

$$|\mathcal{A}'_k(T;b,r;1)| = \sum_{n\le x_kT}\left(\pi\Big(\frac{p}{x_k}\Big)-\pi\Big(\frac{p}{y_k}\Big)\right) + \sum_{x_kT/y_k\le p\le T}\big(\pi(py_k)-\pi(x_kT)\big). \quad (3.12)$$

Before proceeding, we recall some basic definitions and results. We define the logarithmic integral as

$$\mathrm{li}(x) := \lim_{\varepsilon\to 0+}\left(\int_0^{1-\varepsilon}\frac{\mathrm{d}t}{\log t}+\int_{1+\varepsilon}^{x}\frac{\mathrm{d}t}{\log t}\right) \quad \text{for } x>0 \text{ and } x\ne 1.$$

Then we know that

$$\pi(x) = \mathrm{li}(x) + O\left(x\mathrm{e}^{-c\sqrt{\log x}}\right) \quad \text{as } x\to+\infty,$$

which is the prime number theorem with de la Vallée Poussin's error term. Even though more precise versions have been proved, using the best-known error term would give no improvement in our result. Instead, sometimes we just need and use the less precise formula

$$\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right) \quad \text{as } x\to+\infty,$$

which is an immediate consequence of previous one.

Let us now start with the first sum in (3.12). We have

$$\begin{aligned}\sum_{p\le x_kT}&\left(\pi\left(\frac{p}{x_k}\right)-\pi\left(\frac{p}{y_k}\right)\right)\\ &= \sum_{p\le x_kT}\left[\mathrm{li}\left(\frac{p}{x_k}\right)-\mathrm{li}\left(\frac{p}{y_k}\right)+O\left(\frac{p}{x_k}\mathrm{e}^{-c\sqrt{\log(p/x_k)}}\right)\right] \qquad (3.13)\end{aligned}$$

for some positive constant c, by the prime number theorem. Since $y\mathrm{e}^{-c\sqrt{\log y}}$ is increasing from some point on, the error term is

$$\begin{aligned}\sum_{p\le x_kT} O\left(\frac{p}{x_k}\mathrm{e}^{-c\sqrt{\log(p/x_k)}}\right) &\ll \left(\sum_{p\le x_kT}1\right)\left(T\mathrm{e}^{-c\sqrt{\log T}}\right)\\ &\ll T^2x_k\mathrm{e}^{-c'\sqrt{\log T}} \ll \frac{T^2}{k}\mathrm{e}^{-c'\sqrt{\log T}},\end{aligned}$$

where we exploited the fact that, for $\beta > 0$ fixed and $k \leq T^{1-\beta}$, we have

$$\log T \ll \log(T^{\beta}) \leq \log\left(\frac{T}{k}\right) \ll \log(x_k T) < \log T \tag{3.14}$$

by (3.5).

For the main term, we integrate by parts and obtain

$$\begin{aligned}
&\left(\sum_{p \leq x_k T} 1\right)\left(\mathrm{li}(T) - \mathrm{li}\left(\frac{Tx_k}{y_k}\right)\right) \\
&\qquad - \int_2^{Tx_k} \left(\sum_{p \leq z} 1\right)\left(\frac{1}{x_k \log(z/x_k)} - \frac{1}{y_k \log(z/y_k)}\right) \mathrm{d}z \\
&=: A - B,
\end{aligned}$$

say.

We have

$$A = \pi(x_k T) \int_{Tx_k/y_k}^{T} \frac{1}{\log z} \mathrm{d}z,$$

and by the mean value theorem there exists $z' \in (x_k/y_k, 1)$ such that

$$A = \pi(x_k T) \cdot T\left(1 - \frac{x_k}{y_k}\right)\frac{1}{\log(z'T)}. \tag{3.15}$$

But $0 < |\log z'| < |\log(x_k/y_k)|$ and

$$\frac{x_k}{y_k} = \frac{bk+r}{bk+r+1} = \frac{1}{1 + \frac{1}{bk+r}} = 1 - \frac{1}{bk+r} + O\left(\frac{1}{k^2}\right),$$

so

$$\log\left(\frac{x_k}{y_k}\right) \asymp \frac{1}{k}; \tag{3.16}$$

hence

$$\log(z'T) = \log T + O(1/k).$$

Then (3.15) becomes

$$\begin{aligned}
A &= T\pi(x_k T)\left(1 - \frac{x_k}{y_k}\right)\frac{1}{\log T}\left(1 + O\left(\frac{1}{k \log T}\right)\right)^{-1} \\
&= \frac{x_k T^2}{\log^2 T}\left(1 - \frac{x_k}{y_k}\right)\left(1 + O\left(\frac{1}{k \log T}\right)\right) \\
&= \frac{1}{bk+r+1}\frac{x_k T^2}{\log^2 T}\left(1 + O\left(\frac{1}{k \log T}\right)\right) \\
&= \frac{x_k^2 T^2}{b \log^2 T} + O\left(\frac{T^2}{k^3 \log^3 T}\right),
\end{aligned}$$

where we used the prime number theorem and (3.14). To study B, we recall (3.5) and (3.16), and observe that

$$\begin{aligned}
\frac{1}{x_k \log(z/x_k)} - \frac{1}{y_k \log(z/y_k)} &= \frac{1}{x_k \log(z/x_k)} - \frac{1}{y_k\big(\log(z/x_k) + \log(x_k/y_k)\big)} \\
&= \frac{1}{\log(z/x_k)}\left(\frac{1}{x_k} - \frac{1}{y_k\left(1 + \frac{\log(x_k/y_k)}{\log(z/x_k)}\right)}\right) \\
&= \frac{1}{\log(z/x_k)}\left(\frac{1}{x_k} - \frac{1}{y_k\left(1 + O\left(\frac{1}{k\log z}\right)\right)}\right) \\
&= \frac{1}{\log(z/x_k)}\left(\frac{1}{x_k} - \frac{1}{y_k}\left(1 + O\left(\frac{1}{k\log z}\right)\right)\right) \\
&= \frac{1}{\log(z/x_k)}\left(\frac{1}{x_k} - \frac{1}{y_k}\right) + O\left(\frac{1}{\log^2 z}\right) \\
&= \frac{1}{b(\log z + \log(1/x_k))} + O\left(\frac{1}{\log^2 z}\right) \\
&= \frac{1}{b\log z} + O\left(\frac{\log k}{\log^2 z}\right),
\end{aligned}$$

since

$$\frac{1}{x_k} - \frac{1}{y_k} = \frac{1}{b} \tag{3.17}$$

and for any $a, b > 0$ we have

$$\frac{1}{a+b} = \frac{1}{a} + O\left(\frac{b}{a^2}\right).$$

We know that, for any constant $C > \mathrm{e}$ and for $U \to +\infty$, we have

$$\begin{aligned}
\int_C^U \frac{y}{(\log y)^m}\,\mathrm{d}y &= \left[\frac{y^2}{2(\log y)^m}\right]_C^U + m\int_C^U \frac{y}{(\log y)^{m+1}}\,\mathrm{d}y \\
&= \frac{U^2}{2(\log U)^m} + O_C(1) + O_m\left(\frac{U^2}{(\log U)^{m+1}}\right) \\
&= \frac{U^2}{2(\log U)^m} + O\left(\frac{U^2}{(\log U)^{m+1}}\right).
\end{aligned}$$

Then, for B, we have

$$\begin{aligned} B &= \int_2^{Tx_k} \left(\frac{z}{\log z} + O\left(\frac{z}{\log^2 z}\right)\right) \left(\frac{1}{b \log z} + O\left(\frac{\log k}{\log^2 z}\right)\right) \mathrm{d}z \\ &= \frac{1}{b} \int_2^{Tx_k} \frac{z}{\log^2 z} \,\mathrm{d}z + O\left(\int_2^{Tx_k} \frac{z \log k}{\log^3 z} \,\mathrm{d}z \right) \\ &= \frac{1}{2b} \frac{T^2 x_k^2}{\log^2(Tx_k)} + O\left(\frac{T^2 x_k^2}{\log^3 (Tx_k)}\right) + O\left(\frac{T^2 x_k^2 \log k}{\log^3 (Tx_k)}\right) \\ &= \frac{1}{2b} \frac{T^2 x_k^2}{\log^2(Tx_k)} \left(1 + O\left(\frac{\log k}{\log(Tx_k)}\right)\right). \end{aligned}$$

But

$$\frac{1}{\log(Tx_k)} = \frac{1}{\log T + O(\log k)} = \frac{1}{\log T} + O\left(\frac{\log k}{\log^2 T}\right),$$

so that

$$\begin{aligned} B &= \frac{1}{2b} T^2 x_k^2 \left(\frac{1}{\log^2 T} + O\left(\frac{\log k}{\log^3 T} + \frac{\log^2 k}{\log^4 T}\right)\right) \left(1 + O\left(\frac{1}{\log T} + \frac{\log k}{\log^2 T}\right)\right) \\ &= \frac{1}{2b} \frac{T^2 x_k^2}{\log^2 T} + O\left(\frac{T^2}{k^2} \left(\frac{\log k}{\log^3 T} + \frac{\log^2 k}{\log^4 T} + \frac{\log^3 k}{\log^6 T}\right)\right) \\ &= \frac{1}{2b} \frac{T^2 x_k^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T} \frac{\log^3 k}{k^2}\right). \end{aligned}$$

Summing up, we have that (3.13) equals

$$\begin{aligned} &\frac{x_k^2 T^2}{b \log^2 T} - \frac{1}{2b} \frac{T^2 x_k^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T} \frac{\log^3 k}{k^2}\right) + O\left(\frac{T^2}{k} \mathrm{e}^{-c' \sqrt{\log T}}\right) \\ &\qquad = \frac{x_k^2 T^2}{2b \log^2 T} + O\left(\frac{T^2}{\log^3 T} \frac{\log^3 k}{k^2}\right) + O\left(\frac{T^2}{k} \mathrm{e}^{-c' \sqrt{\log T}}\right). \end{aligned} \tag{3.18}$$

Now, let us study the second sum in (3.12). We have

$$\begin{aligned} &\sum_{x_k T / y_k \le p \le T} \big(\pi(p y_k) - \pi(x_k T) \big) \\ &\quad = \sum_{x_k T / y_k \le p \le T} \left(\mathrm{li}(p y_k) - \mathrm{li}(x_k T) + O\left(\frac{T}{k} \mathrm{e}^{-c \sqrt{\log T}}\right)\right) \\ &\quad = \sum_{x_k T / y_k \le p \le T} \int_{x_k T}^{p y_k} \frac{1}{\log z} \,\mathrm{d}z + O\left(\frac{T^2}{k} \mathrm{e}^{-c' \sqrt{\log T}}\right) \\ &\quad = \int_{x_k T}^{y_k T} \frac{1}{\log z} \left(\sum_{z / y_k \le p \le T} 1 \right) \mathrm{d}z + O\left(\frac{T^2}{k} \mathrm{e}^{-c' \sqrt{\log T}}\right) \\ &\quad = \int_{x_k T}^{y_k T} \frac{\mathrm{li}(T) - \mathrm{li}(z / y_k) + O(T \mathrm{e}^{-c \sqrt{\log T}})}{\log z} \,\mathrm{d}z + O\left(\frac{T^2}{k} \mathrm{e}^{-c' \sqrt{\log T}}\right) \\ &\quad = \int_{x_k T}^{y_k T} \frac{\mathrm{li}(T) - \mathrm{li}(z / y_k)}{\log z} \,\mathrm{d}z + O\left(\frac{T^2}{k} \mathrm{e}^{-c'' \sqrt{\log T}}\right). \end{aligned} \tag{3.19}$$

But

$$\begin{aligned}\mathrm{li}(T) - \mathrm{li}(z/y_k) &= \int_{z/y_k}^{T} \frac{1}{\log w}\,\mathrm{d}w \\ &= \int_{z/y_k}^{T} \frac{1}{\log T + O(1/k)}\,\mathrm{d}w \\ &= \left(T - \frac{z}{y_k}\right)\left(\frac{1}{\log T} + O\left(\frac{1}{k\log^2 T}\right)\right),\end{aligned}$$

since

$$T \geq w \geq \frac{z}{y_k} \geq T\frac{x_k}{y_k} = T\left(1 + O\left(\frac{1}{k}\right)\right).$$

Hence the integral in (3.19) is equal to

$$\left(\frac{1}{\log T} + O\left(\frac{1}{k\log^2 T}\right)\right)\int_{x_kT}^{y_kT}\left(T - \frac{z}{y_k}\right)\frac{\mathrm{d}z}{\log z}$$

and

$$\begin{aligned}\int_{x_kT}^{y_kT}\left(T - \frac{z}{y_k}\right)\frac{\mathrm{d}z}{\log z} &= y_k\int_{\frac{x_k}{y_k}T}^{T}\frac{T-z}{\log z + O(\log k)}\,\mathrm{d}z \\ &= y_k\int_{\frac{x_k}{y_k}T}^{T}\frac{T-z}{\log T + O(\log k)}\,\mathrm{d}z \\ &= \frac{y_k}{\log T + O(\log k)}\left[Tz - \frac{z^2}{2}\right]_{Tx_k/y_k}^{T} \\ &= \frac{y_k}{\log T + O(\log k)}\frac{T^2}{2}\left(1 - \frac{x_k}{y_k}\right)^2,\end{aligned}$$

so that (3.19) equals

$$\begin{aligned}&\left(\frac{1}{\log T} + O\left(\frac{1}{k\log^2 T}\right)\right)\frac{y_k}{\log T + O(\log k)}\frac{T^2}{2}\left(1 - \frac{x_k}{y_k}\right)^2 + O\left(\frac{T^2}{k}\mathrm{e}^{-c''\sqrt{\log T}}\right) \\ &\quad= \frac{y_kT^2}{2\log^2 T}\left(1 - \frac{x_k}{y_k}\right)^2\left(1 + O\left(\frac{\log k}{\log T}\right)\right) + O\left(\frac{T^2}{k}\mathrm{e}^{-c''\sqrt{\log T}}\right) \\ &\quad= \frac{y_kT^2}{2\log^2 T}\left(1 - \frac{x_k}{y_k}\right)^2 + O\left(\frac{T^2\log k}{k^3\log^3 T}\right) + O\left(\frac{T^2}{k}\mathrm{e}^{-c''\sqrt{\log T}}\right).\end{aligned} \tag{3.20}$$

Collecting our results in (3.18) and (3.20), and recalling (3.17), we have showed that

$$\begin{aligned}|\mathcal{A}'_k(T;b,r;1)| &= \frac{T^2}{2\log^2 T}\left(x_k^2\left(\frac{1}{x_k} - \frac{1}{y_k}\right) + y_k\left(1 - \frac{x_k}{y_k}\right)^2\right) \\ &\qquad + O\left(\frac{T^2\log^3 k}{k^2\log^3 T}\right) + O\left(\frac{T^2}{k}\mathrm{e}^{-c\sqrt{\log T}}\right) \\ &= \frac{(y_k - x_k)T^2}{2\log^2 T} + O\left(\frac{T^2\log^3 k}{k^2\log^3 T}\right) + O\left(\frac{T^2}{k}\mathrm{e}^{-c\sqrt{\log T}}\right).\end{aligned}$$

Now we must sum over $k \le T^{1-\beta}$ to obtain

$$\begin{aligned}
|\mathcal{L}'(T)| &= \sum_{k \le T^{1-\beta}} |\mathcal{A}'_k(T;b,r;1)| \\
&= \frac{T^2}{2\log^2 T} \sum_{k \le T^{1-\beta}} (y_k - x_k) + O\left(\frac{T^2}{\log^3 T}\right) \\
&= \frac{T^2}{2\log^2 T} \sum_{k \ge 1} (y_k - x_k) + O\left(\frac{T^2}{\log^2 T} \sum_{k > T^{1-\beta}} \frac{1}{k^2}\right) + O\left(\frac{T^2}{\log^3 T}\right) \\
&= \frac{T^2}{2\log^2 T} \sum_{k \ge 1} (y_k - x_k) + O\left(\frac{T^{1+\beta}}{\log^2 T}\right) + O\left(\frac{T^2}{\log^3 T}\right) \\
&= \frac{T^2}{2\log^2 T} \sum_{k \ge 1} (y_k - x_k) + O\left(\frac{T^2}{\log^3 T}\right).
\end{aligned}$$

At this point, we notice that the sum is the same as the one in (3.11). Hence

$$|\mathcal{L}'(T)| = \left(\psi\Big(\frac{b+r+1}{b}\Big) - \psi\Big(\frac{b+r}{b}\Big)\right) \frac{T^2}{2\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right).$$

3.3.2 Estimate of $\mathcal{U}'(T)$

As before, we assume that $r \in S_b \cup \{b\}$. Let first suppose that $r \ge 1$. The reasoning about edges holds true for $\mathcal{U}'(T)$ as well. So we recall that we are dealing with the case $i = 1$ and compute

$$\begin{aligned}
|\mathcal{U}'(T;b,r;1)| = |\mathcal{A}'_0(T)| &= \sum_{p \le T} \left(\pi\left(\frac{(r+1)p}{b}\right) - \pi\left(\frac{rp}{b}\right)\right) \\
&= \sum_{p \le T} \left(\mathrm{li}\left(\frac{(r+1)p}{b}\right) - \mathrm{li}\left(\frac{rp}{b}\right) + O\left(p\, \mathrm{e}^{-c\sqrt{\log((r+1)p/b)}}\right)\right).
\end{aligned}$$

The error term is

$$O\left(\sum_{p \le T} p\, \mathrm{e}^{-c\sqrt{\log((r+1)p/b)}}\right) = O\left(\pi(T) T\, \mathrm{e}^{-c'\sqrt{\log T}}\right) = O\left(T^2 \mathrm{e}^{-c''\sqrt{\log T}}\right).$$

For the main term, we have

$$\begin{aligned}
&\sum_{p\le T}\left(\mathrm{li}\left(\frac{(r+1)p}{b}\right)-\mathrm{li}\left(\frac{rp}{b}\right)\right)\\
&\qquad=\left(\sum_{p\le T}1\right)\left(\mathrm{li}\left(\frac{(r+1)T}{b}\right)-\mathrm{li}\left(\frac{rT}{b}\right)\right)\\
&\qquad\quad-\int_2^T\left(\sum_{p\le z}1\right)\left(\frac{r+1}{b\log((r+1)z/b)}-\frac{r}{b\log(rz/b)}\right)\mathrm{d}z\\
&\qquad=: A'-B',
\end{aligned}$$

say. Then

$$A'=\pi(T)\int_{rT/b}^{(r+1)T/b}\frac{1}{\log z}\,\mathrm{d}z,$$

and by the mean value theorem there exists $z'\in(r/b,(r+1)/b)$ such that

$$A'=\pi(T)\cdot\frac{T}{b}\cdot\frac{1}{\log(z'T)}.$$

But $|\log z'|\le\max(\log b,\log(1+1/b))$, so that $\log(z'T)=\log T+O(1)$ and

$$\begin{aligned}
A'&=\pi(T)\cdot\frac{T}{b}\cdot\frac{1}{\log T+O(1)}\\
&=\left(\frac{T}{\log T}+O\left(\frac{T}{\log^2 T}\right)\right)\frac{T}{b}\cdot\frac{1}{\log T}\left(1+O\left(\frac{1}{\log T}\right)\right)\\
&=\frac{T^2}{b\log^2 T}+O\left(\frac{T^2}{\log^3 T}\right).
\end{aligned}$$

For what concerns B', we have

$$\begin{aligned}
\frac{r+1}{b\log\left(\frac{r+1}{b}z\right)}-\frac{r}{b\log\left(\frac{r}{b}z\right)}&=\frac{1}{b}\left(\frac{r+1}{\log\left(\frac{r+1}{b}z\right)}-\frac{r}{\log\left(\frac{r+1}{b}z\right)+\log\left(\frac{r}{r+1}\right)}\right)\\
&=\frac{1}{b\log\left(\frac{r+1}{b}z\right)}\left(r+1-\frac{r}{1+\frac{\log(r/(r+1))}{\log((r+1)z/b)}}\right)\\
&=\frac{1}{b\log\left(\frac{r+1}{b}z\right)}\left(r+1-\frac{r}{1+O(1/\log z)}\right)\\
&=\frac{1}{b(\log z+O(1))}\left(r+1-r\left(1+O\left(\frac{1}{\log z}\right)\right)\right)\\
&=\frac{1}{b\log z}\left(1+O\left(\frac{1}{\log z}\right)\right)^2\\
&=\frac{1}{b\log z}\left(1+O\left(\frac{1}{\log z}\right)\right).
\end{aligned}$$

Hence

$$\begin{aligned} B' &= \int_2^T \left(\frac{z}{\log z} + O\left(\frac{z}{\log^2 z}\right)\right)\left(\frac{1}{b\log z} + O\left(\frac{1}{\log^2 z}\right)\right)\mathrm{d}z \\ &= \frac{1}{b}\int_2^T \frac{z}{\log^2 z}\,\mathrm{d}z + O\left(\int_2^T \frac{z}{\log^3 z}\,\mathrm{d}z\right) \\ &= \frac{1}{2b}\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right). \end{aligned}$$

So we obtain

$$|\mathcal{U}'(T)| = A' - B' + O\left(T^2\mathrm{e}^{-c''\sqrt{\log T}}\right) = \frac{1}{2b}\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right).$$

To conclude, we must consider the case $r = 0$. We have

$$\begin{aligned} |\mathcal{U}'(T;b,0;1)| &= |\mathcal{A}_0'(T)| = \sum_{p\le T}\pi\left(\frac{p}{b}\right) = \sum_{2b\le p\le T}\pi\left(\frac{p}{b}\right) = \sum_{3b\le p\le T}\pi\left(\frac{p}{b}\right) + O(1) \\ &= \sum_{3b\le p\le T}\mathrm{li}\left(\frac{p}{b}\right) + O\left(T^2\mathrm{e}^{-c\sqrt{\log T}}\right) \\ &= \pi(T)\,\mathrm{li}\left(\frac{T}{b}\right) - \pi(3b)\,\mathrm{li}(3) - \int_{3b}^T \frac{\pi(z)}{b\log(z/b)}\,\mathrm{d}z \\ &= \left(\frac{T}{\log T} + O\left(\frac{T}{\log^2 T}\right)\right)\left(\frac{T}{b\log(T/b)} + O\left(\frac{T}{\log^2 T}\right)\right) \\ &\quad + O(1) - \int_{3b}^T \frac{\pi(z)}{b\log(z/b)}\,\mathrm{d}z. \end{aligned}$$

Since for $z \in [3b, T]$ it holds

$$\begin{aligned} \frac{1}{\log(z/b)} &= \frac{1}{\log z(1 + O(1/\log z))} = \frac{1}{\log z}\left(1 + O\left(\frac{1}{\log z}\right)\right) \\ &= \frac{1}{\log z} + O\left(\frac{1}{\log^2 z}\right), \end{aligned}$$

we have

$$\begin{aligned} |\mathcal{U}'(T;b,0;1)| &= \left(\frac{T}{\log T} + O\left(\frac{T}{\log^2 T}\right)\right)\left(\frac{T}{b\log T} + O\left(\frac{T}{\log^2 T}\right)\right) \\ &\quad - \frac{1}{b}\int_{3b}^T \left(\frac{z}{\log z} + O\left(\frac{z}{\log^2 z}\right)\right)\left(\frac{1}{\log z} + O\left(\frac{1}{\log^2 z}\right)\right)\mathrm{d}z \\ &= \frac{T^2}{b\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right) - \frac{1}{b}\int_{3b}^T \left(\frac{z}{\log^2 z} + O\left(\frac{z}{\log^3 z}\right)\right)\mathrm{d}z \\ &= \frac{1}{2b}\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right), \end{aligned}$$

exactly as before. Summing up, we obtain

$$|\mathcal{A}'(T)| = |\mathcal{L}'(T)| + |\mathcal{U}'(T)| = c(b, r; 1)\frac{T^2}{\log^2 T} + O\left(\frac{T^2}{\log^3 T}\right),$$

where

$$c(b, r; 1) = \frac{1}{2b} + \frac{1}{2}\Big(\psi\Big(\frac{b+r+1}{b}\Big) - \psi\Big(\frac{b+r}{b}\Big)\Big),$$

as in the case of integers. This completes the proof of Theorem 3.2.

Chapter 4

Appendix

4.1 The digamma function

We collect here some properties of the digamma function that we need: for the proofs, the reader is referred to Chapter 5 of Olver et al. [16]. We recall that

$$\psi(z) = \frac{\Gamma'}{\Gamma}(z) \qquad \text{for } z \notin \{0, -1, -2, \dots\}.$$

We need the representation

$$\psi(z) = -\gamma - \frac{1}{z} + \sum_{k \geq 1} \frac{z}{k(z+k)} = -\gamma + \sum_{k \geq 0} \Big(\frac{1}{k+1} - \frac{1}{k+z} \Big), \tag{4.1}$$

which is (5.7.6) of [16], and immediately implies (3.11), since

$$\psi(z) - \psi(w) = \sum_{k \geq 0} \Big(\frac{1}{k+w} - \frac{1}{k+z} \Big)$$

and we need just to take $z = (b+r+1)/b = 1/x_k$ and $w = (b+r)/b = 1/y_k$.

We recall also the following connection between the Hurwitz zeta function and the digamma function: for any positive integer m, it holds that

$$\psi^{(m)}(z) = (-1)^{m+1} m! \, \zeta(m+1, z);$$

in particular, we have that

$$\psi'(z) = \zeta(2, z).$$

4.2 Ramanujan's theorem

For the sake of completeness, we state here a weak form of Ramanujan's theorem [17], as it can be found in [3].

Theorem 4.1 (Ramanujan)**.** *Be $\sigma_s(m)$ the sum of the s-th powers of the divisors of $m \in \mathbb{N}$. Then for each fixed $\varepsilon > 0$ there exists $C_1 > 0$ such that*

$$\sigma_{-s}(m) < \exp\left(C_1 \frac{(\log m)^{1-s}}{\log\log m}\right), \tag{4.2}$$

for all integers $m \geq 3$ and for all $s \in [\varepsilon, 1-\varepsilon]$.

4.3 Numerical data

We collect here some numerical data regarding the problem of signed harmonic sums, obtained with the collaboration of Mattia Cafferata. Trying to improve on the previous available data required both reformulating the problem for computational purposes (for example, breaking computations into two parts, as suggested in [3]) and optimizing the code in a very subtle way, because very large numbers and matrices are involved as soon as one explores values for $N > 27$, say. Both PARI/GP and MATLAB were used for these purposes.

N	$\mathfrak{m}_N(0) \cdot \prod_{j=1}^N p_j$	
1	1	1
2	1	1
3	1	1
4	23	23
5	43	43
6	251	251
7	263	263
8	21013	21013
9	1407079	41*34319
10	4919311	1423*3457
11	818778281	27059*30259
12	2402234557	139*2647*6529
13	379757743297	306899*1237403
14	3325743954311	89*37367909599
15	54237719914087	83*653466504989
16	903944329576111	38699*102763*227303
17	46919460458733911	101*277*1087*1542845089
18	367421942920402841	13313*27598733788057
19	17148430651130576323	3121*10543807*521114509
20	1236225057834436760243	743*1663829149171516501
21	4190310920096832376289	3517*6703*177747974707739
22	535482916756698482410061	107*176923*28286391485091301
23	29119155169912957197310753	9500963*3064863548033284331
24	443284248908491516288671253	3001*3301259141*44744193836633
25	28438781483496930396689638231	151*571*110051*2997118645488555761

Table 4.1: The values, multiplied by $\prod_{j=1}^N p_j$, of the smallest signed harmonic sums with the first N primes, for N up to 25. On the right, their prime number factorisations, which were computed to search for any possible arithmetical regularity; actually, no pattern seems to emerge.

N	$\mathfrak{m}_N(0) \cdot \prod_{j=1}^{N} p_j$
26	10196503226925713726754541885481
27	137512198125317766267968137765087
28	5572821202475305606211985553786081
29	77833992457426020006787481021085581
30	24244850423688161715955346535954790877
31	2030349334778419995324119439659994086131
32	76860130392109667765387079377871685276909
33	5191970624445760882844533168270184721318637
34	329643209271348431895096550792159132283920307
35	19171590315567357340242017182966253037383120953
36	58192378490977430486851365332352874578233287403
37	837477642920747839191618216897250374978659503996169
38	130665466261033919414441892800025408642432364448372023
39	7541550169407232608689149525984967898398947805296216009
40	23868339955752715692132986729285170427530832996153507207
41	3343165792500492306892396976512891068137770193474133826457
42	47233268931962642510303169511493601517566800154537867238057
43	93915329439868205746156163805290441755151986127947916375626793
44	50313439148416324581127610155641150127987318260569172331033593181
45	2035703788246113211455753014584246782664737720644793016891955087197

Table 4.2: The values, multiplied by $\prod_{j=1}^{N} p_j$, of the smallest signed harmonic sums with the first N primes, with N between 26 and 45.

N	$\mathfrak{m}_N(0) \cdot \prod_{j=1}^{N} p_j$
46	1937688615891780440916248774686275817721164643503688338812098644 12247
47	4664128549520402650533030541013467806288648880741654578068005845271177
48	25229409968071098806367386200315218884168013574116192401844690408603954 1
49	164152705533632496799540344537262942048356425519773153500697538193607 3433
50	2543642450545133244192831947465647133687416765504736677470218788227489 4064063
51	1780024077761328763318128562703299120404666081323149178582620236480827415289259
52	11553364375146609761969934518303398078666123048462189253113162991092436404 0946261
53	346445205731766592295370811989346241267385291503362454494739411253204971046538 17109
54	7369668963051661582966392617319633009625522375611294051784365401090471220946387592789
55	199963258224846876335793874247507216756651341869412816388166951273778698828707537 4795317
56	151351981933638637742621357138936533979590998748883750430193460129876391573603481014628429
57	15302724902698188450027684974980553939987991074013402437579866232981371846926226684458406969
58	6269085432675155135477773589250562149563926327373176617473379555222137615792922214195964225281
59	429918790837116674905123858093668694474961832761345115366942177591943696826657060080682245858603
60	115809464188499233574522294110279752895686365776568444548440426304978721966632473743873345620708313

Table 4.3: The values, multiplied by $\prod_{j=1}^{N} p_j$, of the smallest signed harmonic sums with the first N primes, with N between 46 and 60.

N	$\min\{\lvert S_{N,1} - S_{N,2}\rvert : S_{N,i} \in \mathfrak{S}_N\} \cdot \prod_{j=1}^{N} p_j$
1	1
2	1
3	1
4	2
5	22
6	35
7	263
8	4675
9	24871
10	104006
11	2356081
12	6221080
13	141769355
14	6096082265
15	6928889495
16	367231143235
17	1283811918935
18	78312527055035
19	5246939312687345
20	372532691200801495
21	8815359347599933286
22	223849990729887044174
23	6148176498383067879445

Table 4.4: The values, multiplied by $\prod_{j=1}^{N} p_j$, of the shortest distances between two different signed harmonic sums with the first N primes, with N up to 23.

N	$\mathfrak{m}'_N(0) \cdot L_N$	N	$\mathfrak{m}'_N(0) \cdot L_N$	N	$\mathfrak{m}'_N(0) \cdot L_N$
1	1	24	24319	47	3103390393
2	1	25	71559	48	3103390393
3	1	26	4261	49	421936433719
4	1	27	13703	50	175378178867
5	7	28	13703	51	8643193037
6	3	29	872843	52	8643193037
7	11	30	872843	53	461784703049
8	13	31	17424097	54	461784703049
9	11	32	13828799	55	461784703049
10	11	33	902339	56	461784703049
11	23	34	7850449	57	514553001783
12	23	35	7850449	58	116096731427
13	607	36	7850449	59	2810673355099
14	251	37	10683197	60	2810673355099
15	251	38	68185267	61	4723651835663
16	125	39	37728713	62	136420009515743
17	97	40	37728713	63	136420009515743
18	97	41	740674333	64	23093515509397
19	3767	42	740674333	65	23093515509397
20	3767	43	1774907231	66	23093515509397
21	3767	44	1774907231	67	3786341162179960
22	2285	45	1774907231	68	3786341162179960
23	24319	46	1699239271	69	313407851480621

Table 4.5: The values, multiplied by $L_N = \mathrm{lcm}(1, \dots, \mathrm{N})$, of the numerators of the smallest signed harmonic sums, with N up to 69. The first 64 values were obtained by Bettin, Molteni and Sanna [3]; we computed the values from 65 to 69.

4.4 Plots

4.4.1 Classical harmonic series

We collect here some histograms, created using Wolfram Mathematica, obtained with the numerical computations that we performed for the problem of digits. Some variants were considered trying to understand if a different counting function could help to regularize the problem and lower the error term. In every histogram, on the x-axis we have the digit r, and the height of the bar represents $\Phi(T;b,r;1)$ (or one of its variants, which will be defined later) normalized by dividing by $\sum_{r=0}^{b-1}\Phi(T;b,r;1)$. We recall that $\Phi(T;b,r;1) := |\mathcal{A}(T;b,r;1)|$. The continuous line, instead, is the graph of $c(b,r;i)$ as a function of r; we indicated with a dot the values corresponding to integer values of r, which are the ones appearing in the theorems. Let us start with $b=10$ in Figures 4.1, 4.2 and 4.3.

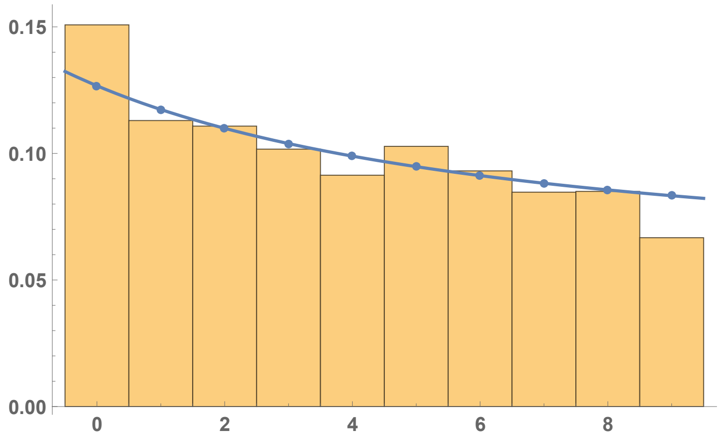

Figure 4.1: The histogram for $|\mathcal{A}(100;10,r;1)|$.

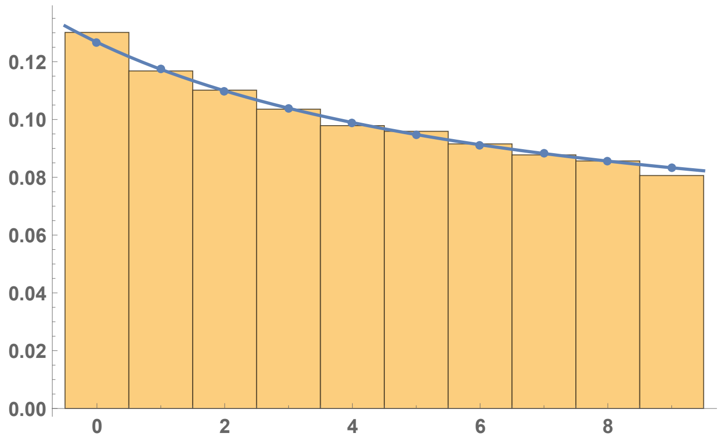

Figure 4.2: The histogram for $|\mathcal{A}(1000;10,r;1)|$.

What we can notice is that for $T=10000$ the histogram is really close to the expected value. But for smaller T we can see some phenomenon taking place: for $T=100$, the irregularities seem all but random. For $r=0$ and $r=5$ the numerical values are noticeably exceeding the expected ones. One could conjecture some

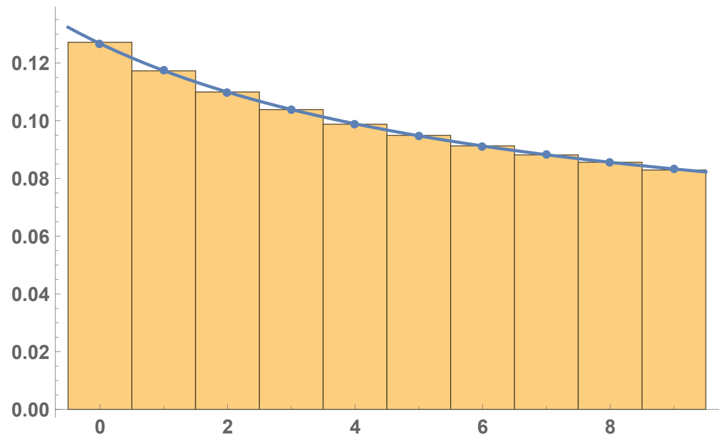

Figure 4.3: The histogram for $|\mathcal{A}(10000; 10, r; 1)|$.

connection with the greater common divisor of b and r or some other phenomenon related to arithmetic properties, like divisibility. Actually, the main problem for small T is a simple fact of multiplicities: fractions with small numerator and denominator (like 1, 1/2, 1/3, 2/3, etc.) will be counted many times. This is clear by the histogram for $b = 17$, which is prime, and the one for $b = 30$, which has many divisors: in both of them we can recognize high counting values for the fractions that we have just mentioned. See Figures 4.4, 4.5, 4.6 and 4.7.

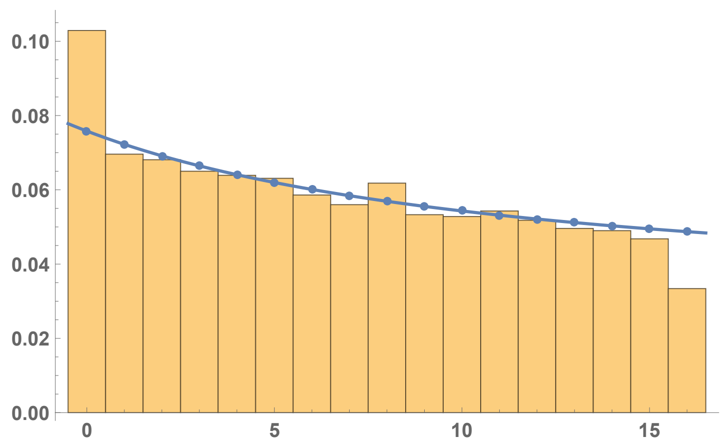

Figure 4.4: The histogram for $|\mathcal{A}(100; 17, r; 1)|$.

To avoid this phenomenon, we can just count the fractions with multiplicity one, which means taking just the reduced ones: in Figures 4.8, 4.9 and 4.10, we represent $\Phi_{\mathcal{B}}(T; b, r; 1)$, the cardinality of

$$\mathcal{B}(T; b, r; 1) := \{(n, m) \in \mathbb{N}^2 \cap [1, T]^2 \colon (n, m) = 1,\ \phi(n/m; b; i) = r\}.$$

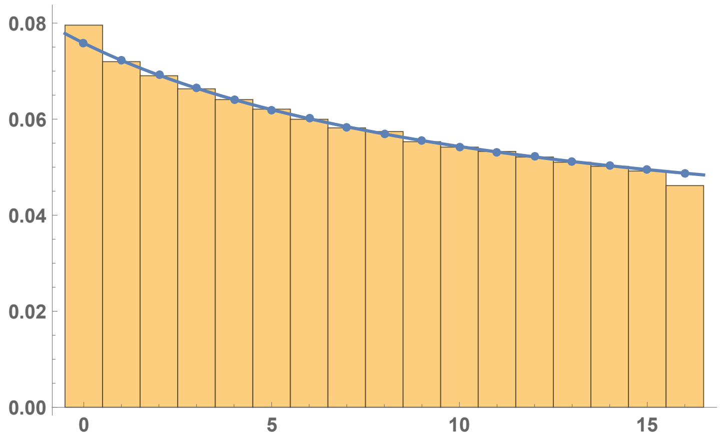

Figure 4.5: The histogram for $|\mathcal{A}(1000; 17, r; 1)|$.

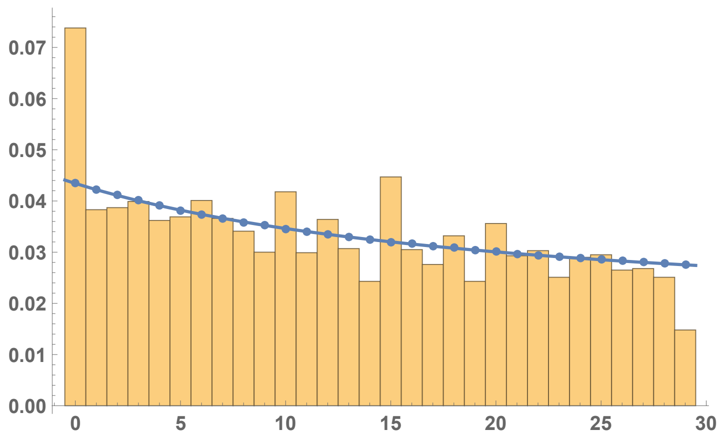

Figure 4.6: The histogram for $|\mathcal{A}(100; 30, r; 1)|$.

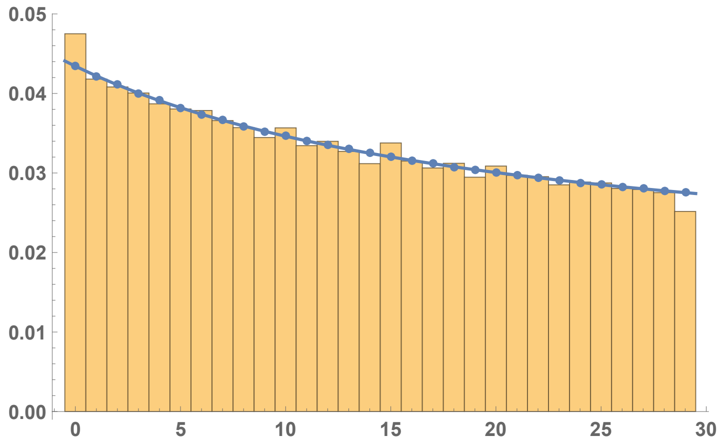

Figure 4.7: The histogram for $|\mathcal{A}(1000; 30, r; 1)|$.

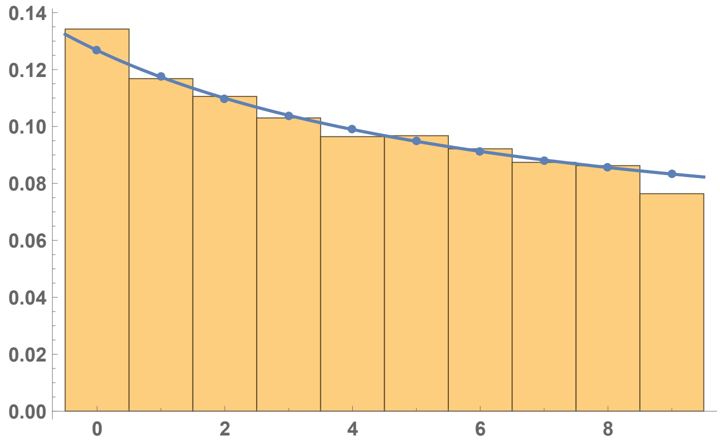

Figure 4.8: The histogram for $|\mathcal{B}(100; 10, r; 1)|$ (with $(n, m) = 1$).

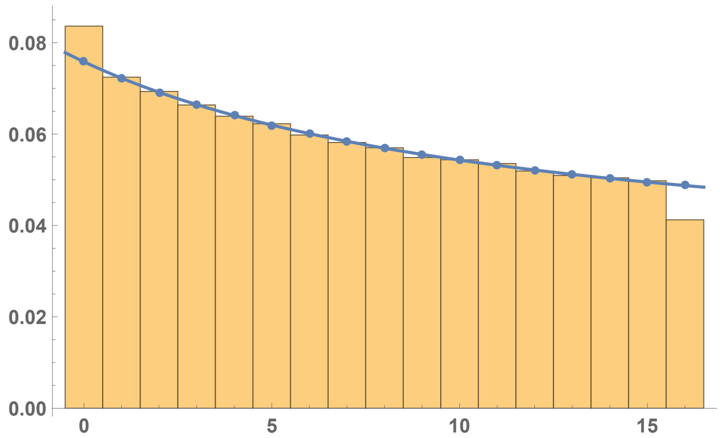

Figure 4.9: The histogram for $|\mathcal{B}(100; 17, r; 1)|$ (with $(n, m) = 1$).

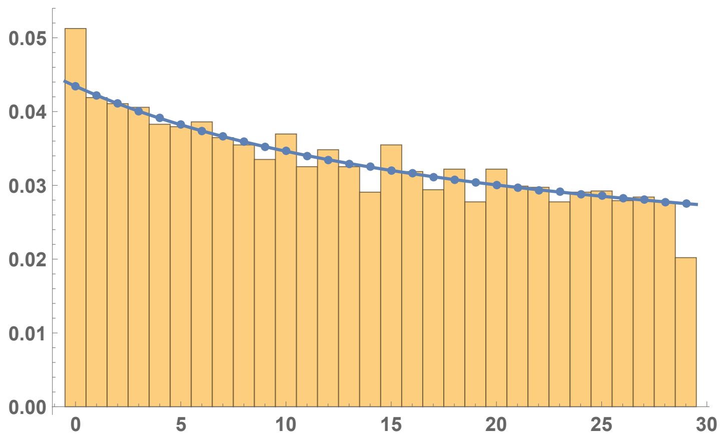

Figure 4.10: The histogram for $|\mathcal{B}(100; 30, r; 1)|$ (with $(n, m) = 1$).

Now it seems that for the prime base 17 the regularization is really good, except for the digits 0 and 16. If we look at the case $b = 30$, it has been regularized as well, but some other phenomenon emerged: it seems that the divisors and in general the numbers with prime factors in common with 30 tend to be have higher values. The explanation for this lies in the discontinuity of the system of digits: if we perturbed just a bit a number that has no digits to the right of the first one, we could lower its first digit by one. To be more formal, we know that the representation of a number in base b is unique just by convention: indeed, if a rational number admits a finite representation (when, after reducing the fraction, the denominator divides the base), it admits also a periodic infinite one. Just to make an example, if we think about the base 10, we have $1 = 0.\bar{9}$. So, in our case, any number that admits any ambiguity in its first digit in base b should morally be divided into the two digits with an equal weight: we call $\widetilde{\Phi_{\mathcal{B}}}(T; b, r; 1)$ the counting function that we obtained as $\Phi_{\mathcal{B}}$, but the convention of assigning half weight to two digits if $b\{n/m\} \in \mathbb{Z}$. Now the rate of convergence is very good already for small values of T. See Figures 4.11, 4.12 and 4.13.

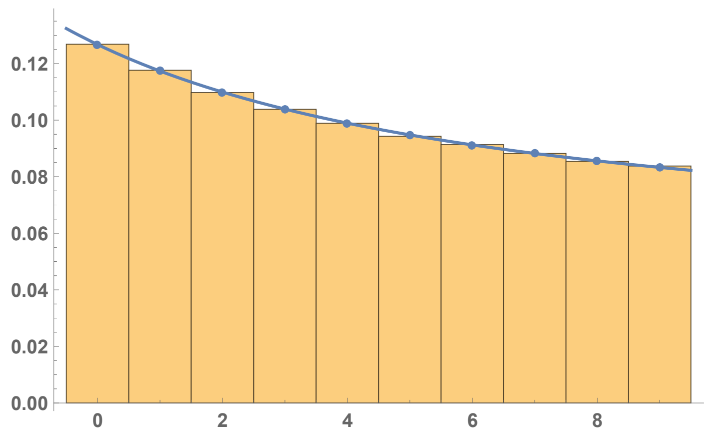

Figure 4.11: The histogram for $\widetilde{\Phi_{\mathcal{B}}}(100; 10, r; 1)$ (with $(n, m) = 1$ and assigning half weight to two digits if $b\{n/m\} \in \mathbb{Z}$).

Another way to obtain a similar result is to take m and n coprime and also coprime with the base: in Figures 4.14 and 4.15, we represented $\Phi_{\mathcal{C}}(T; b, r; 1)$, the cardinality of

$$\mathcal{C}(T; b, r; 1) := \{(n, m) \in \mathbb{N}^2 \cap [1, T]^2 \colon (n, m) = 1,\ (mn, b) = 1,\ \phi(n/m; b; i) = r\}.$$

But this does not solve the ambiguity for the integers and we can still notice the high value for the digit 0.

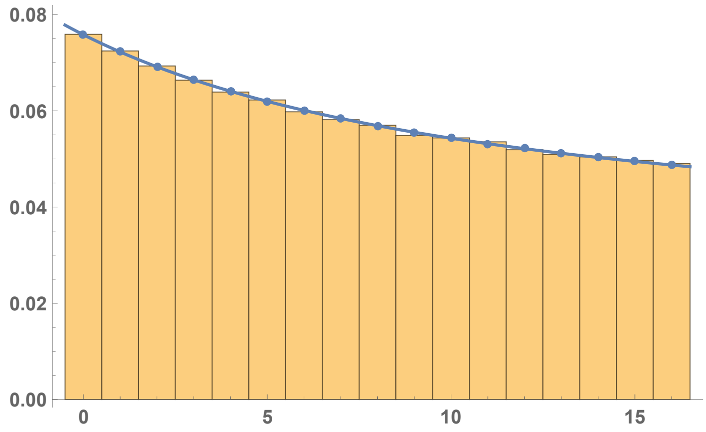

Figure 4.12: The histogram for $\widetilde{\Phi_{\mathcal{B}}}(100; 17, r; 1)$ (with $(n, m) = 1$ and assigning half weight to two digits if $b\{n/m\} \in \mathbb{Z}$).

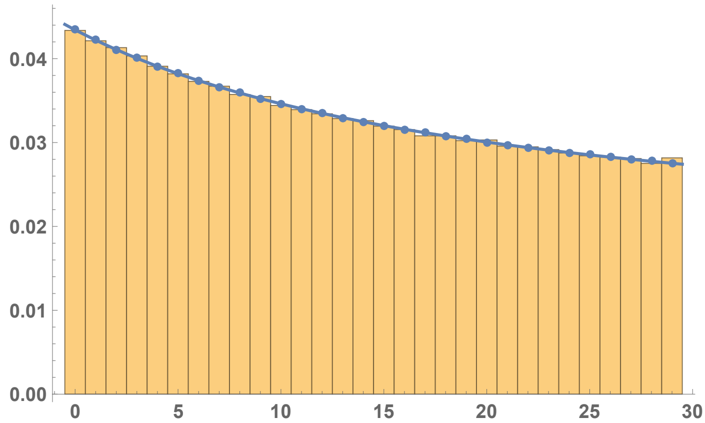

Figure 4.13: The histogram for $\widetilde{\Phi_{\mathcal{B}}}(100; 30, r; 1)$ (with $(n, m) = 1$ and assigning half weight to two digits if $b\{n/m\} \in \mathbb{Z}$).

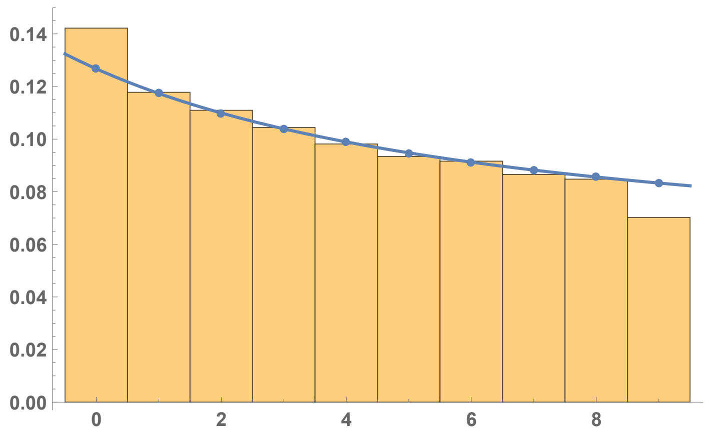

Figure 4.14: The histogram for $|\mathcal{C}(100; 10, r; 1)|$, where we take n and m coprime and coprime with the base.

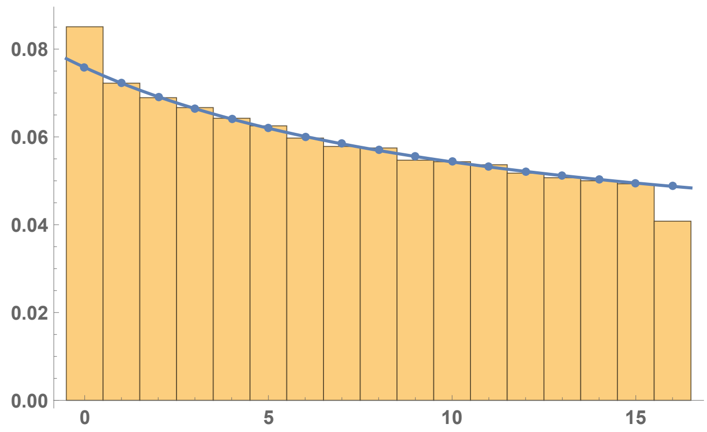

Figure 4.15: The histogram for $|\mathcal{C}(100; 17, r; 1)|$, where we take n and m coprime and coprime with the base.

4.4.2 Harmonic series with primes

We made similar computations also in the case of prime numbers. Two positive primes are not coprime if and only if they are equal: to avoid such a possibility, we just ignore the diagonal $p = q$ in every histogram. From Figure 4.16 to Figure 4.21 we represent $\Phi'_{\mathcal{B}}(T; b, r; 1)$, the cardinality of

$$\mathcal{B}'(T; b, r; 1) := \{(p, q) \in \mathbb{P}^2 \cap [1, T]^2 \colon \phi(p/q; b; i) = r,\ p \neq q\},$$

while from Figure 4.22 to Figure 4.28 we represented $\widetilde{\Phi'_{\mathcal{B}}}(T; b, r; 1)$, the counting function that we obtained as $\Phi'_{\mathcal{B}}$, but with the convention of assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

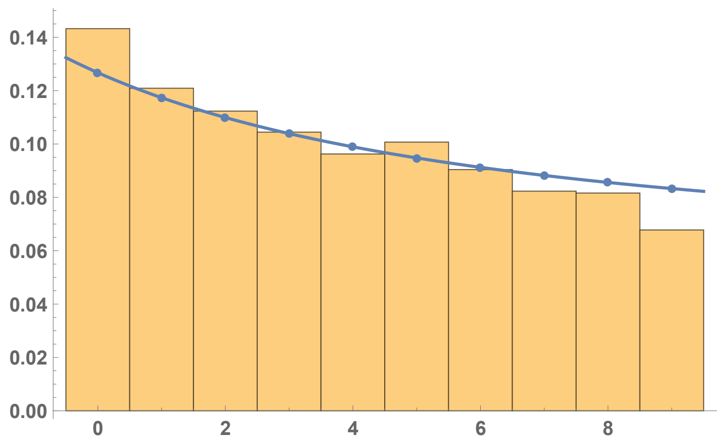

Figure 4.16: The histogram for $\Phi'_{\mathcal{B}}(100; 10, r; 1)$ for primes $p \neq q$.

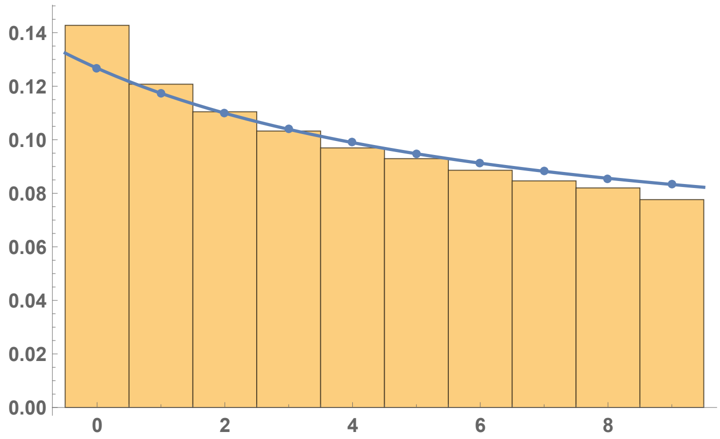

Figure 4.17: The histogram for $\Phi'_{\mathcal{B}}(1000; 10, r; 1)$ for primes $p \neq q$.

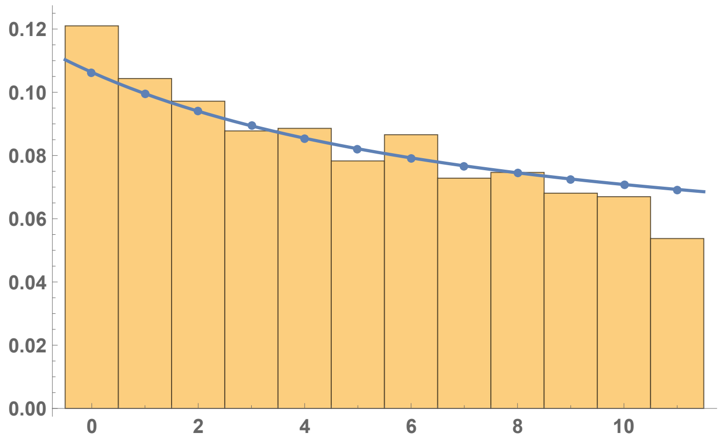

Figure 4.18: The histogram for $\Phi'_{\mathcal{B}}(100; 12, r; 1)$ for primes $p \neq q$.

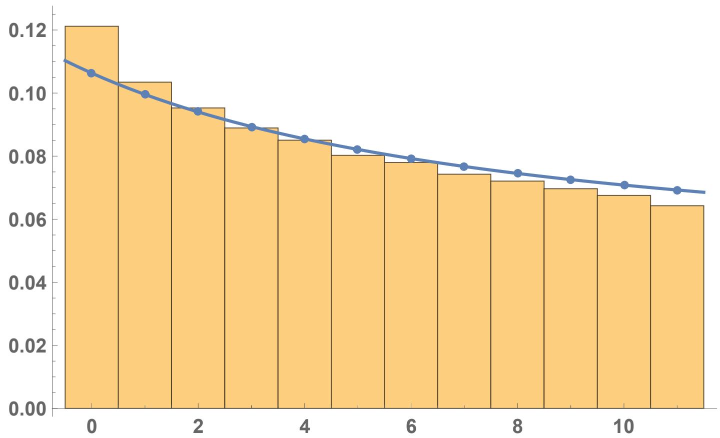

Figure 4.19: The histogram for $\Phi'_{\mathcal{B}}(1000; 12, r; 1)$ for primes $p \neq q$.

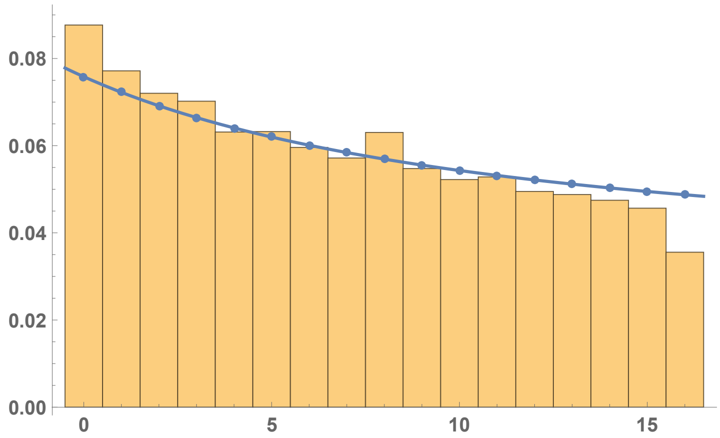

Figure 4.20: The histogram for $\Phi'_{\mathcal{B}}(100; 17, r; 1)$ for primes $p \neq q$.

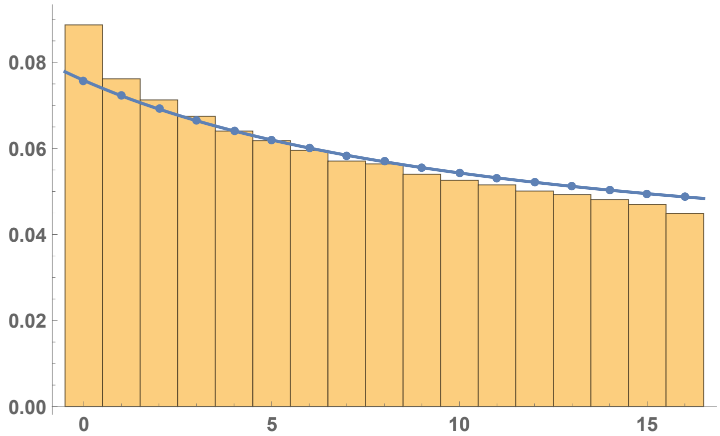

Figure 4.21: The histogram for $\Phi'_{\mathcal{B}}(1000; 17, r; 1)$ for primes $p \neq q$.

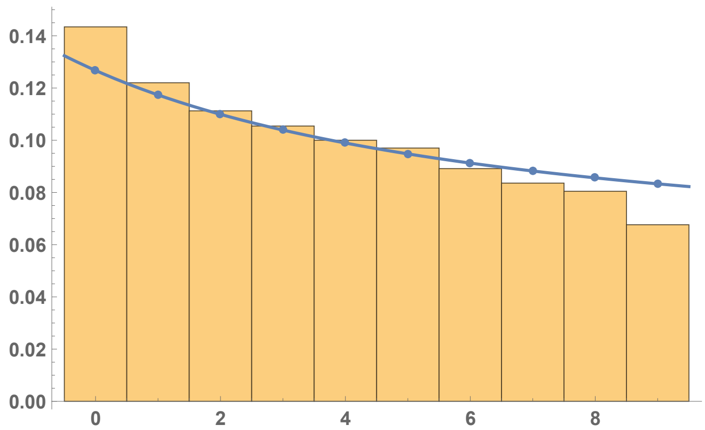

Figure 4.22: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(100; 10, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

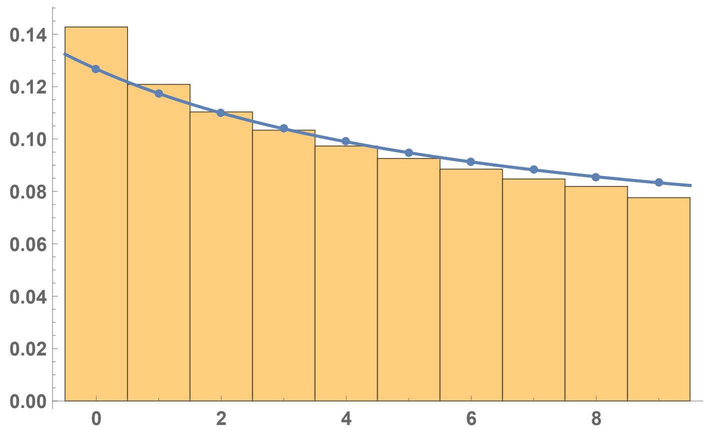

Figure 4.23: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(1000; 10, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

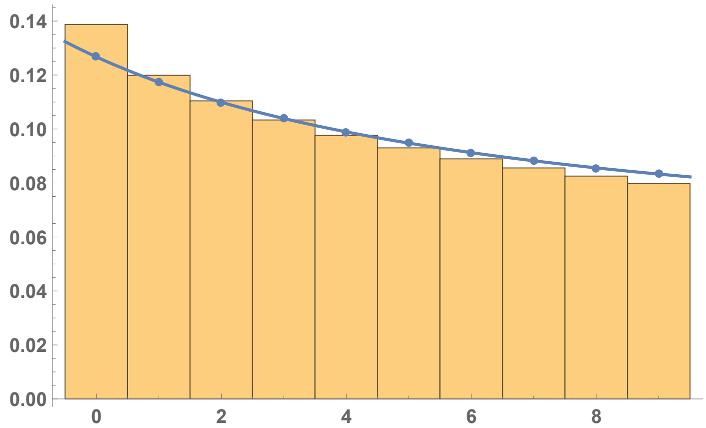

Figure 4.24: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(10000; 10, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

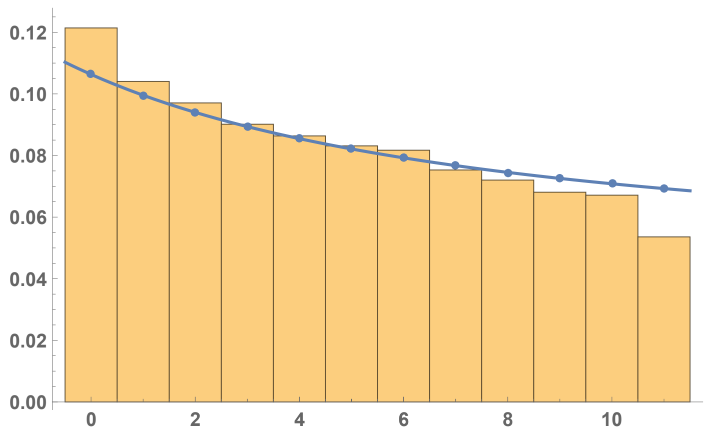

Figure 4.25: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(100; 12, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

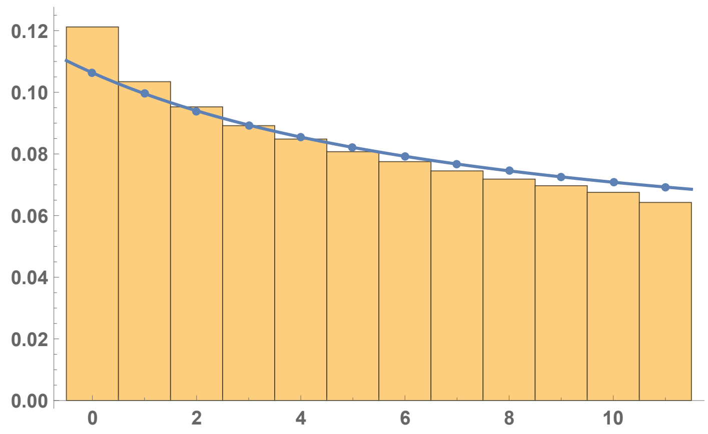

Figure 4.26: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(1000; 12, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

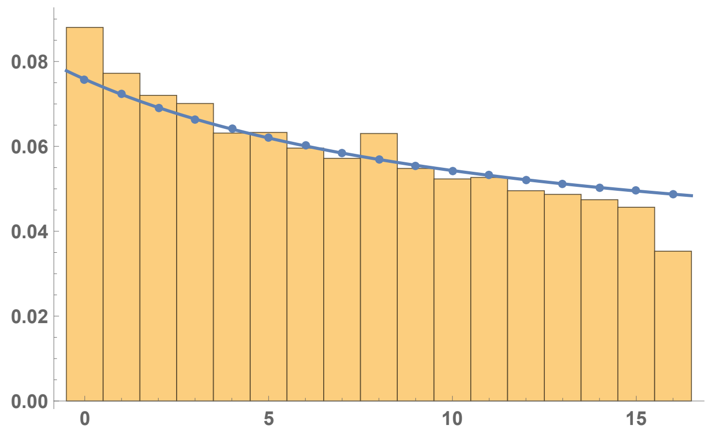

Figure 4.27: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(100; 17, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

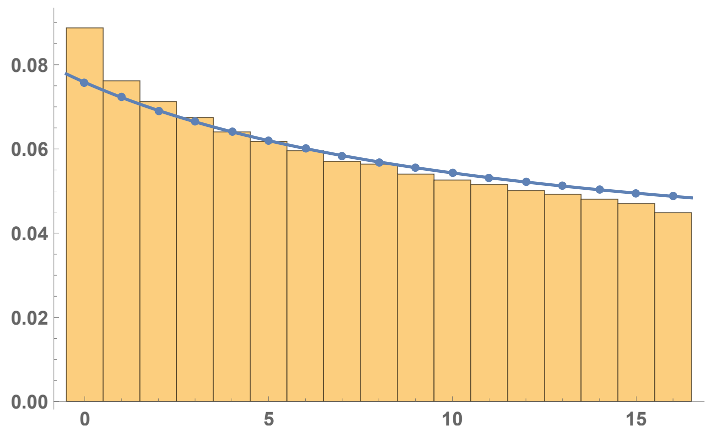

Figure 4.28: The histogram for $\widetilde{\Phi'_{\mathcal{B}}}(1000; 17, r; 1)$ for primes $p \neq q$, assigning half weight to two digits if $b\{p/q\} \in \mathbb{Z}$.

We can notice that this variant gives some regularity to the problem, but still the convergence is very slow: after all, the error term is just a logarithmic factor less than the main term, and its order of magnitude seems consistent with what we have found. At the same time the distance from the expected value reveals some regularity. This suggests that it may be possible to isolate a secondary main term of size about $T^2/\log^3 T$ and leading constant which is decreasing in r and is about 0 for r around $b/2$. This would explain why, in the above figures, for T sufficiently large, the first blocks are always above the trend line, while the latest are below.

Bibliography

[1] F. Benford. “The law of anomalous numbers”. In: *Proceedings of the American philosophical society* (1938), pp. 551–572.

[2] S. Bettin, G. Molteni, and C. Sanna. “Greedy approximations by signed harmonic sums and the Thue–Morse sequence”. In: *arXiv e-prints* (Apr. 2018). arXiv: `1805.00075`.

[3] S. Bettin, G. Molteni, and C. Sanna. “Small values of signed harmonic sums”. In: *C. R. Math. Acad. Sci. Paris* 356.11-12 (2018), pp. 1062–1074.

[4] M. N. Bleicher and P. Erdős. “Denominators of Egyptian fractions. II”. In: *Illinois J. Math.* 20.4 (1976), pp. 598–613.

[5] M. N. Bleicher and P. Erdős. “The number of distinct subsums of $\sum_1^N 1/i$”. In: *Math. Comp.* 29 (1975), pp. 29–42.

[6] M. Cafferata, A. Gambini, R. Tonon, and A. Zaccagnini. “On the distribution of the digits of quotients of integers”. In preparation. 2019.

[7] E. Cesàro. “Eventualités de la division arithmétique”. In: *Ann. Mat. Pura Appl.* 13.1 (1885), pp. 269–290.

[8] E. Cesàro. *Excursions arithmétiques à l’infini.* A. Hermann, 1885.

[9] R. E. Crandall. “Theory of ROOF walks”. Unpublished manuscript. 2007. URL: `https://www.reed.edu/physics/faculty/crandall/papers/ROOF12.pdf`.

[10] L. Euler. “Variae observationes circa series infinitas”. In: *Commentarii academiae scientiarum imperialis Petropolitanae* (1737), pp. 160–188.

[11] A. Gambini, G. Mingari Scarpello, and D. Ritelli. “Probability of digits by dividing random numbers: a ψ and ζ functions approach”. In: *Expo. Math.* 30.3 (2012), pp. 223–238.

[12] A. Gambini, R. Tonon, and A. Zaccagnini. “Signed harmonic sums of integers with k distinct prime factors”. Accepted for publication in *Rendiconti del Seminario Matematico di Torino*. 2019.

[13] E. Landau. “Sur quelques problèmes relatifs à la distribution des nombres premiers”. In: *Bull. Soc. Math. France* 28 (1900), pp. 25–38.

[14] K. E. Morrison. “Cosine products, Fourier transforms, and random sums”. In: *Amer. Math. Monthly* 102.8 (1995), pp. 716–724.

[15] K. E. Morrison. “Random walks with decreasing steps”. Unpublished manuscript, California Polytechnic State University. 1998.

[16] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark. *NIST Handbook of mathematical functions hardback and CD-ROM*. Cambridge University Press, 2010.

[17] S. Ramanujan. “Highly composite numbers [Proc. London Math. Soc. (2) **14** (1915), 347–409]”. In: *Collected papers of Srinivasa Ramanujan*. AMS Chelsea Publ., Providence, RI, 2000, pp. 78–128.

[18] B. Schmuland. “Random harmonic series”. In: *Amer. Math. Monthly* 110.5 (2003), pp. 407–416.