On the mantissa distribution of powers of natural and prime numbers

Shalom Eliahou, Bruno Massé, Dominique Schneider
L.M.P.A. Joseph Liouville,
Université du Littoral Côte d'Opale

November 15, 2011

Abstract

Given a fixed integer exponent $r \geq 1$, the mantissa sequences of $(n^r)_n$ and of $(p_n^r)_n$, where p_n denotes the nth prime number, are known not to admit any distribution with respect to the natural density. In this paper however, we show that, when r goes to infinity, these mantissa sequences *tend* to be distributed following Benford's law in an appropriate sense, and we provide convergence speed estimates. In contrast, with respect to the log-density and the loglog-density, it is known that the mantissa sequences of $(n^r)_n$ and of $(p_n^r)_n$ *are* distributed following Benford's law. Here again, we provide previously unavailable convergence speed estimates for these phenomena. Our main tool is the Erdős-Turán inequality.

1 Introduction

Given a sequence $(u_n)_n$ of positive real numbers, a classical question consists in determining whether the sequence of the first significant digit of u_n, in base 10 say, follows some specific distribution, such as Benford's law for instance. More generally, one may consider the mantissa sequence of the u_n, and ask again whether it satisfies any distribution. Recall that the *mantissa* of a positive real number x is the unique number $\mathcal{M}(x) \in [1, 10[$ such that $x = \mathcal{M}(x)10^k$ for some integer k.

To be more specific, let ν be a probability measure on the interval $[1, 10[$. We say that the mantissa of $(u_n)_n$ are *distributed following ν, in the sense of the natural density*, if

$$\lim_{N\to+\infty}(1/N)\sum_{n=1}^{N}\mathbb{1}_{[1,t[}(\mathcal{M}(u_n)) \;=\; \nu([1,t[) \tag{1}$$

for all $t \in [1, 10[$, where $\mathbb{1}_A$ denotes the indicator function of the subset A.

A probability measure on $[1, 10[$ of particular interest is *Benford's law* μ_B, characterized by

$$\mu_B\,([1,t[) = \log_{10} t \qquad (1 \leq t < 10)$$

where $\log_{10}$ denotes the logarithm in base 10. Several sequences are known whose mantissa are distributed following Benford's law, such as α^n whenever $\log(\alpha)$ is irrational, or $n!$, or n^n, or the nth Fibonacci number F_n (see [9]).

On the other hand, it is known that the mantissa of n^r for a fixed exponent $r \geq 1$, or of p_n^r where p_n is the nth prime number, do not follow any distribution in the sense of the natural density.

However, when r goes to infinity, experimental observation suggests that the mantissa of n^r and of p_n^r *tend* to be distributed following Benford's law (see Figure 1). Our first purpose in this paper is to provide a setting where this phenomenon can be formally defined (see next section) and established.

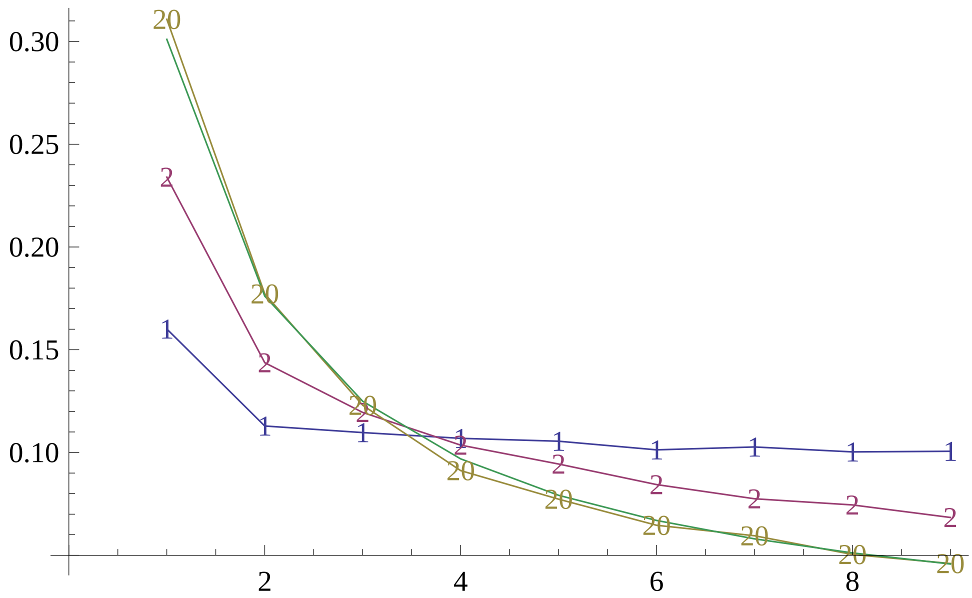

Figure 1: Frequency of $1, 2, \ldots, 9$ as first digit of p_i^r for $i \leq 10000$ and $r = 1$, 2 and 20. The unmarked curve is Benford's law.

Now, the fact that the mantissa of n^r and of p_n^r do not follow any distribution can be fixed by considering weaker distribution conditions.

Consider again a probability measure ν on $[1, 10[$. We say that the mantissa of $(u_n)_n$ is *log-distributed following* ν if, for all $t \in [1, 10[$,

$$\lim_{N \to +\infty} (1/\log N) \sum_{n=1}^{N} (1/n) \mathbb{1}_{[1,t[}(\mathcal{M}(u_n)) \;=\; \nu([1, t[) \tag{2}$$

where log denotes the natural logarithm. Similarly, we say that the mantissa of $(u_n)_n$ is *loglog-distributed following* ν if, for all $t \in [1, 10[$,

$$\lim_{N \to +\infty} (1/\log \log N) \sum_{n=1}^{N} (1/n \log n) \mathbb{1}_{[1,t[}(\mathcal{M}(u_n)) \;=\; \nu([1, t[). \tag{3}$$

In (3) $n \log n$ can be replaced by p_n. Conditions (1), (2) and (3) are gradually weaker in a strict sense. See [8] for a survey on these questions. When $u_n = n$ and $u_n = p_n$, it is known that

$$(\mathcal{M}(u_n))_n \text{ is log-distributed (and so loglog-distributed) following } \mu_B \tag{4}$$

(see [4], [5] and [14]). Our second purpose in this paper is to provide previously unavailable convergence speed estimates for these phenomena.

Remark. *More precisely, it is proved in [4], [5] and [14] that, as regards the first digit,* $(\mathcal{M}(n))_n$ *is log-distributed following* μ_B *and that* $(\mathcal{M}(p_n))_n$ *is distributed following* μ_B *in the sense of the so-called* logarithmic density relative to the prime numbers. *This relative density is equivalent to the loglog-density defined in* (3). *The calculations used in these papers can be quite easily adapted to prove* (4). *See also [1] for closely related questions.*

2 Notations and definitions

The fractional part of a real number x will be denoted by $\{x\}$.

Let $S = (v_n)_n$ be a sequence of real numbers, $N \in \mathbb{N}^*$ and $(w_n)_n$ a sequence of non-negative weights summing to infinity. The discrepancy modulo 1 of order N of S (see [3, 12]), associated to $(w_n)_n$, is the number

$$D_N^{(w_n)}(S) := \sup_{0<a<b<1} \left| \left(1/\sum_{n=1}^{N} w_n \right) \sum_{n=1}^{N} w_n \mathbb{1}_{[a,\,b[}(\{v_n\}) - (b-a) \right| .$$

It represents the distance between the uniform distribution in $[0,1[$ and the distribution, with respect to $(w_n)_n$, of the first N terms of $(\{v_n\})_n$. For the weights $w_n = 1$, $w_n = 1/n$ and $w_n = 1/p_n$ (or $w_n = 1/n \log n$), these numbers will be denoted $D_N(S)$, $D_N^{log}(S)$ and $D_N^{loglog}(S)$, respectively. If we set $u_n = 10^{v_n}$, this number is also the distance between μ_B and the distribution, with respect to $(w_n)_n$, of the first N terms of $(\mathcal{M}(u_n))_n$. This is because $\{\log_{10} u_n\} = \log_{10} \mathcal{M}(u_n)$. So

$$D_N^{(w_n)}(S) = \sup_{1<c<d<10} \left| \left(1/\sum_{n=1}^{N} w_n \right) \sum_{n=1}^{N} w_n \mathbb{1}_{[c,\,d[}(\mathcal{M}(u_n)) - \log_{10}(d/c) \right| .$$

See [7, p. 100–131] and [3, p. 252–259] for two examples of the study of the discrepancy.

We say that *the mantissa of the terms in the rows of the array of positive real numbers* $(u_{r,n})_{r,n}$ *tends to be distributed following* μ_B *when* $r \to +\infty$ if there exists a non-decreasing function ϕ from $\mathbb{N}^*$ to $\mathbb{N}^*$, of infinite limit, such that

$$\lim_{r \to +\infty} D_{\phi(r)}(S_r) = 0$$

where $S_r = (u_{r,n})_n$. This definition is most useful when the mantissa of the terms in the rows of the array do not admit any distribution (like in the sequel). The introduction of a truncation function ϕ as above is then necessary.

The last notation we need is $e_h(x) := \exp(2i\pi hx)$ with $i^2 = -1$.

3 Results

We present here the main results of this paper. See Section 2 for the link between the discrepancy of $(\log_{10} u_n)_n$ and the convergence speed in the study of the distribution of $(\mathcal{M}(u_n))_n$. We denote by $\mathcal{O}$ the standard big O of Landau.

To the best of our knowledge, the only preceding results on Benford's law concerning sequences of sequences, or arrays of numbers, are due to Diaconis [2] and concern the binomial coefficients.

Theorem 1. *The mantissa of the terms in the rows of the arrays $(n^r)_{r,n}$ and $(p_n^r)_{r,n}$ tends to be distributed following μ_B when $r \to +\infty$. Moreover, setting $\phi(r) = [e^r]$ and $S_r = (\log_{10} n^r)_n$, $S_r = (\log_{10}(n \log n)^r)_n$ or $S_r = (\log_{10} p_n^r)_n$ $(r \in \mathbb{N}^*)$, we have*

$$D_{\phi(r)}(S_r) = \mathcal{O}(r^{-1})\,.$$

The speed of convergence of D_N^{log} cannot be better than $1/\log N$ and the speed of convergence of D_N^{loglog} cannot be better than $1/\log\log N$ (see [3, p. 252]). As a particular case of theorem 2.40 in [3] we get that, if $S_1' = (\log_{10} n)_n$, then $D_N^{log}(S_1') = \mathcal{O}\left((\log N)^{-1}\right)$. The following theorem gives a bound for $D_N^{loglog}(S_1')$ close to the best possible speed.

Theorem 2. *If we set $S_1' = (\log_{10} n)_n$, we have*

$$D_N^{loglog}(S_1') = \mathcal{O}\left((\log\log N)^{-1}(\log\log\log N)^2\right)\,.$$

We now treat the rates of convergence for $u_n = p_n$ in place of $u_n = n$.

Theorem 3. *If we set $S_2' = (\log_{10} p_n)_n$, we have*

$$D_N^{log}(S_2') = \mathcal{O}\left((\log N)^{-\frac{1}{2}}(\log\log N)^2\right)$$

and

$$D_N^{loglog}(S_2') = \mathcal{O}\left((\log\log N)^{-\frac{1}{2}}(\log\log\log N)^2\right)\,.$$

4 Preliminaries

We present here the properties, already known or purely technical, used in the proofs of the results in the preceding section.

The following result appears in various guises in the literature, in particular due to the still ongoing search for the best possible constants. We shall use those found in [3] and [12]. We want to point out the fact that, in the next three inequalities, the choice of the integer $H \geq 1$ is free.

The Erdős-Turán inequality. *Let $S = (v_n)_n$ be a sequence of elements in $[0, 1[$ and let N be a natural number. Then, for every natural number H, we have*

$$D_N(S) \leq \frac{1}{H+1} + \sum_{h=1}^{H} \frac{1}{h}\frac{1}{N}\left|\sum_{n=1}^{N} e_h(v_n)\right|\,,$$

$$D_N^{\log}(S) \le \frac{3}{2}\left(\frac{2}{H+1} + \sum_{h=1}^{H}\frac{1}{h}\frac{1}{(\sum_{n=1}^{N}\frac{1}{n})}\left|\sum_{n=1}^{N}\frac{e_h(v_n)}{n}\right|\right)$$

and

$$D_N^{\log\log}(S) \le \frac{3}{2}\left(\frac{2}{H+1} + \sum_{h=1}^{H}\frac{1}{h}\frac{1}{(\sum_{n=1}^{N}\frac{1}{p_n})}\left|\sum_{n=1}^{N}\frac{e_h(v_n)}{p_n}\right|\right).$$

We now give four elementary lemmas. The first one may be found, for instance, in [11].

Lemma 1. *For all $n \in \mathbb{N}^*$, $p_n \ge n\log n$ and there exists a real number $C_0 > 0$ such that, for all integer $n \ge 3$,*

$$|p_n - n\log n| \le C_0 n \log\log n\,.$$

Lemma 2. *For all integer $n \ge 1$ and all $\theta \ne 0$, we have*

$$\left|\sum_{j=1}^{n}\exp(2i\pi\theta\log j)\right| \le \frac{n}{2\pi|\theta|} + 1 + \pi|\theta|\log n\,.$$

Proof. Fix $n \ge 1$ and $\theta \ne 0$. Then

$$\frac{1}{n}\sum_{j=1}^{n} j^{2i\pi\theta} = n^{2i\pi\theta}R_n(f)$$

where $R_n(f)$ is the Riemann sum of $f : t \mapsto t^{2i\pi\theta}$ on $[0,1]$ with n regular steps of length n^{-1}. Since $\int_0^1 f(t)\,dt = \frac{1}{2i\pi\theta+1}$, thanks to the mean value inequality, we have

$$\begin{aligned}
\left|R_n(f) - \frac{1}{2i\pi\theta+1}\right| &= \left|\sum_{j=0}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}\left(\left(\frac{j+1}{n}\right)^{2i\pi\theta} - t^{2i\pi\theta}\right)dt\right| \\
&\le \left|\int_0^{\frac{1}{n}}\left(\left(\frac{1}{n}\right)^{2i\pi\theta} - t^{2i\pi\theta}\right)dt\right| + \sum_{j=1}^{n-1}\int_{\frac{j}{n}}^{\frac{j+1}{n}}\left(\frac{j+1}{n} - t\right)\frac{2\pi|\theta|}{\frac{j}{n}}\,dt \\
&\le \frac{1}{n}\left|\frac{2i\pi\theta}{2i\pi\theta+1}\right| + 2\pi|\theta|\sum_{j=1}^{n-1}\frac{1}{2jn}\,.
\end{aligned}$$

□

Lemma 3. *There exists $C_2 > 0$ such that, for all integer $n \ge 3$ and all $\theta \in \mathbb{R}$, we have*

$$\big|\exp(2i\pi\theta\log p_n) - \exp(2i\pi\theta\log(n\log n))\big| \le C_2|\theta|\frac{\log\log n}{\log n}\,.$$

Proof. For all integer $n \ge 3$ and all $\theta \in \mathbb{R}$, we have

$$\begin{aligned}
\big|\exp(2i\pi\theta\log p_n) - \exp(2i\pi\theta\log(n\log n))\big| &= \left|\left(\frac{p_n}{n\log n}\right)^{2i\pi\theta} - 1\right| \\
&= \left|\left(1 + \frac{p_n - n\log n}{n\log n}\right)^{2i\pi\theta} - 1\right| \\
&\le 2\pi|\theta|\frac{|p_n - n\log n|}{n\log n}
\end{aligned}$$

as follows from the mean value inequality and taking Lemma 1 into account. Therefore, we may choose $C_2 = 2\pi C_0$, where C_0 is the constant involved in Lemma 1. □

Lemma 4. *There exists $C_3 > 0$ such that, for all integer $N \geq 3$ and all sequence $(\theta_n)_n$ of real numbers, we have*

$$\sum_{n=3}^{N} \left| \frac{\exp(i\theta_n)}{n \log n} - \frac{\exp(i\theta_n)}{p_n} \right| \leq C_3 \,.$$

Proof. Let $N \geq 3$ and a sequence $(\theta_n)_n$ be fixed. Then we have

$$\sum_{n=3}^{N} \left| \frac{\exp(i\theta_n)}{n \log n} - \frac{\exp(i\theta_n)}{p_n} \right| \leq C_0 \sum_{n=3}^{N} \frac{\log \log n}{p_n \log n} \,,$$

where C_0 is the constant involved in Lemma 1. And $\dfrac{\log \log n}{p_n \log n} \sim \dfrac{\log \log n}{n \log^2 n}$ is the general term of a convergent series. □

The proofs of theorems 2 and 3 rely mainly on Lemma 8 below. To prove it we need two famous estimates that we now recall. The first one is Lemma 4.10 in [13, p. 76].

Lemma 5. *Let $f(x)$ and $g(x)$ be two functions with continuous derivatives in the interval $[a, b]$ such that $f'(x)$ and $|g'(x)|$ are non-increasing and g is positive and decreasing. Then*

$$\sum_{a \leq j \leq b} g(j) \exp(2i\pi f(j)) = \sum_{\alpha - \eta < \nu < \beta + \eta} \int_a^b g(x) \exp(2i\pi(f(x) - \nu x))\, dx \\ + \mathcal{O}(g(a) \log(\beta - \alpha + 2)) + \mathcal{O}(|g'(a)|)$$

where $\beta = f'(a)$, $\alpha = f'(b)$ and η is an arbitrary constant such that $0 < \eta \leq 1$.

Here is the second estimate. To prove it, it suffices to rewrite the proof of Lemma 2.43 in [3, p. 253] for a non-integer parameter θ.

Lemma 6. *Fix $\theta \neq 0$. For all integer ν and all real number $B > 1$, we have*

$$\left| \int_1^B \frac{\exp(2i\pi(\theta \log x - \nu x))}{x}\, dx \right| \leq C \left(|\theta|^{-1} + |\theta|^{-\frac{1}{2}} \right) \,,$$

where C is an absolute constant.

We can now prove a property from which we will derive Lemma 8.

Lemma 7. *There exists $C_1 > 0$ and an integer $n_0 \geq 1$ such that, for all integer $n \geq n_0$ and all $\theta \neq 0$, we have*

$$\left| \sum_{j=n_0}^{n} \frac{\exp(2i\pi\theta \log j)}{j} \right| \leq C_1 \left(|\theta|^{-1} + |\theta|^{-\frac{1}{2}} + \frac{|\theta|^{\frac{1}{2}}}{n_0} \right) \,.$$

Proof. Without loss of generality, we may assume $\theta > 0$. We apply Lemma 5 with $a = n_0$, $b = n$, $g(x) = 1/x$ and $f(x) = \theta \log x$. This gives

$$\sum_{j=n_0}^{n} \frac{1}{j} \exp(2i\pi\theta \log j) = \sum_{\frac{\theta}{n}-\frac{1}{2}<\nu<\frac{\theta}{n_0}+\frac{1}{2}} \int_{n_0}^{n} \frac{\exp(2i\pi(\theta \log x - \nu x))}{x}\, dx$$
$$+ \mathcal{O}\left((1/n_0) \log\left(\theta(1/n_0 - 1/n) + 2\right)\right) + \mathcal{O}(1/n_0^2)\,.$$

Observing that $\log\left(\theta(1/n_0 - 1/n) + 2\right) \le \theta^{\frac{1}{2}}/n_0^{\frac{1}{2}} + 2^{\frac{1}{2}}/n_0^{\frac{1}{2}}$, that $1 \le \theta^{-\frac{1}{2}} + \theta^{\frac{1}{2}}$ and that $n_0 \ge 1$ we obtain

$$\mathcal{O}\left((1/n_0) \log\left(\theta(1 - 1/n) + 2\right)\right) + \mathcal{O}(1/n_0^2) \le C\left(\theta^{-\frac{1}{2}} + \frac{\theta^{\frac{1}{2}}}{n_0}\right),$$

where C is a constant. Applying Lemma 6, and observing that the number of terms in the above sum does not exceed $1 + \theta(1/n_0 - 1/n)$, it follows that there exists an absolute constant, still denoted C, such that

$$\left| \sum_{\frac{\theta}{n}-\frac{1}{2}<\nu<\frac{\theta}{n_0}+\frac{1}{2}} \int_{n_0}^{n} \frac{\exp(2i\pi(\theta \log x - \nu x))}{x}\, dx \right| \le C\left(\theta^{-1} + \theta^{-\frac{1}{2}}\right)\left(1 + \theta\left(\frac{1}{n_0} - \frac{1}{n}\right)\right)$$
$$\le 2C\left(\theta^{-1} + \theta^{-\frac{1}{2}} + \frac{\theta^{\frac{1}{2}}}{n_0}\right).$$

This concludes the proof. □

Lemma 8. *There exists $C > 0$ (depending only on $b > 1$) such that, for all integer $n \ge 1$ and all $h \in \mathbb{Z}^*$, we have*

$$\left| \sum_{j=1}^{n} \frac{\exp(2i\pi h \log_b j)}{j} \right| \le C + \log|h|.$$

Proof. Without loss of generality, we assume that $n \ge |h| + 1$. Then

$$\left| \sum_{j=1}^{n} \frac{\exp(2i\pi h \log_b j)}{j} \right| \le 1 + \log|h| + \left| \sum_{j=|h|+1}^{n} \frac{\exp(2i\pi h \log_b j)}{j} \right|.$$

Using Lemma 7 with $\theta = h/\log b$, we obtain

$$\left| \sum_{j=|h|+1}^{n} \frac{\exp(2i\pi h \log_b j)}{j} \right| \le C_1\left(\log b + \log b^{\frac{1}{2}} + \log b^{-\frac{1}{2}}\right)$$

because $|h| \ge 1$. This concludes the proof. □

5 Proofs

We shall now prove the results stated in Section 3. We set $\theta_h = \frac{h}{\log 10}$,

$$\gamma_h = C_1 \log|h| \left(\log 10 + \sqrt{\log 10} + \frac{1}{\sqrt{\log 10}} \right)$$

and $B_n = \sum_{j=1}^{n} j^{2i\pi\theta_h - 1}$ for $h \in \mathbb{Z}^*$ and $n \geq 1$. The following estimates will be useful in the proofs of the theorems 2 et 3:

$$\sum_{h=1}^{H} \frac{\theta_h}{h} = \mathcal{O}(H) \ , \ \sum_{h=1}^{H} \frac{\gamma_h}{h} = \mathcal{O}(\log H)^2 \quad \text{and} \quad \sum_{h=1}^{H} \frac{\theta_h \gamma_h}{h} = \mathcal{O}\left(H \log H\right) . \tag{5}$$

Recall that Lemma 8 gives an upper bound of $|B_n|$ by γ_h which is independent of n.

5.1 Proof of Theorem 1

We have divided the proof in three stages.

1) Let us first study the rows of the array $(n^r)_{r,n}$ with respect to the natural density (and hence, here, $S_r = (\log_{10} n^r)_n$ for $r \in \mathbb{N}^*$). Let us fix the integers $N \geq 1$, $h \geq 1$ and $r \geq 1$. From Lemma 2, we have

$$\left| \sum_{n=1}^{N} e_h(\log_{10} n^r) \right| \leq \frac{N \log 10}{2\pi h r} + 1 + \frac{\pi h r \log N}{\log 10} .$$

Thus, for all $N \geq 1$ and all $r \geq 1$,

$$\sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{N} e_h(\log_{10} n^r) \right| = \mathcal{O}(1) \frac{N}{r} + \mathcal{O}(\log H) + \mathcal{O}(H) r \log N .$$

The Erdős-Turán inequality then gives

$$D_N(S_r) \leq \mathcal{O}(H^{-1}) + \mathcal{O}(1) \frac{1}{r} + \frac{\mathcal{O}(\log H)}{N} + \mathcal{O}(H) r \frac{\log N}{N}. \tag{6}$$

We have infinitely many choices for the values of $N = \phi(r)$ and values of H depending on r that lead to

$$\lim_{r \to +\infty} D_{\phi(r)}(S_r) = 0.$$

In particular, taking $H = r$ and $N = \phi(r) = \lfloor e^r \rfloor$, we obtain $D_{\phi(r)}(S_r) = \mathcal{O}(r^{-1})$, which is the best that can be expected from relation (6).

2) Let us now study the rows of the array $(n^r \log^r n)_{r,n}$ with respect to the natural density (and hence, here, $S_r = (\log_{10}(n^r \log^r n))_n$). Let us fix the integers $N \geq 3$, $h \geq 1$ and $r \geq 1$, and set $\theta = rh(\log 10)^{-1}$ and, for $n \geq 1$, $A_n = \sum_{j=1}^{n} j^{2i\pi\theta}$ and $v_n = \log_{10}(n^r \log^r n)$. We first remark that, for $n \geq 2$, we have

$$\left|(\log n)^{2i\pi\theta} - (\log(n+1))^{2i\pi\theta}\right| = 2\pi|\theta| \left| \int_n^{n+1} \frac{(\log x)^{2i\pi\theta}}{x \log x} dx \right| \leq \frac{2\pi|\theta|}{n \log n} .$$

Therefore, because of Abel's transformation and Lemma 2, we have

$$\begin{aligned}\left|\sum_{n=2}^{N} e_h(v_n)\right| &= \left|\sum_{n=2}^{N} n^{2i\pi\theta}(\log n)^{2i\pi\theta}\right| \\ &= \left|A_N(\log N)^{2i\pi\theta} - A_1(\log 2)^{2i\pi\theta} + \sum_{n=2}^{N-1} A_n\left((\log n)^{2i\pi\theta} - (\log(n+1))^{2i\pi\theta}\right)\right| \\ &\leq \frac{N}{2\pi|\theta|} + 2 + \pi|\theta|\log N + \sum_{n=2}^{N-1}\left(\frac{n}{2\pi|\theta|} + 1 + \pi|\theta|\log n\right)\frac{2\pi|\theta|}{n\log n}\,.\end{aligned}$$

Thus, from the properties of the integral logarithm function, there exists a constant $K > 0$ such that

$$\left|\sum_{n=1}^{N} e_h(v_n)\right| \leq \frac{N}{2\pi|\theta|} + 3 + \frac{KN}{\log N} + \pi|\theta|(\log N + 2\log\log N) + 2\pi^2\theta^2\log N\,.$$

Therefore, for all integer $r \geq 1$ and all $N \geq 3$,

$$\sum_{h=1}^{H}\frac{1}{h}\left|\sum_{n=1}^{N} e_h(v_n)\right| = \mathcal{O}(1)\frac{N}{r} + \mathcal{O}(\log H)\frac{N}{\log N} + \mathcal{O}(H)r\log N + \mathcal{O}(H^2)r^2\log N\,. \quad (7)$$

This, together with the Erdős-Turán inequality, gives

$$D_N(S_r) \leq \mathcal{O}(H^{-1}) + \mathcal{O}(1)\frac{1}{r} + \frac{\mathcal{O}(\log H)}{\log N} + \mathcal{O}(H)r\frac{\log N}{N} + \mathcal{O}(H^2)r^2\frac{\log N}{N}\,.$$

And, for $H = r$ and $N = \phi(r) = \lfloor e^r \rfloor$, we again obtain

$$D_{\phi(r)}(S_r) = \mathcal{O}(r^{-1}).$$

3) Finally, let us study the rows of the array $(p_n^r)_{r,n}$ with respect to the natural density (and hence, $S_r = (\log_{10}(p_n^r))_n$ here). Let us fix the integers $N \geq 3$, $h \geq 1$ and $r \geq 1$, and set again $v_n = \log_{10}(n^r \log^r n)$. Applying Lemma 3 to $\theta = rh(\log 10)^{-1}$ and using the properties of the integral logarithm, we obtain

$$\sum_{h=1}^{H}\frac{1}{h}\left|\sum_{n=1}^{N}(e_h(\log_{10} p_n) - e_h(v_n))\right| = \mathcal{O}(H)r\frac{N\log\log N}{\log N}\,.$$

Relation (7) and the Erdős-Turán inequality then imply

$$D_N(S_r) \leq \mathcal{O}(H^{-1}) + \mathcal{O}(1)\frac{1}{r} + \frac{\mathcal{O}(\log H)}{\log N} + \mathcal{O}(H)r\left(\frac{\log N}{N} + \frac{N\log\log N}{\log N}\right) + \mathcal{O}(H^2)r^2\frac{\log N}{N}\,.$$

And, for $H = r$ and $N = \phi(r) = \lfloor e^r \rfloor$, we again obtain

$$D_{\phi(r)}(S_r) = \mathcal{O}(r^{-1}).$$

This concludes the proof of Theorem 1.

5.2 Proof of Theorem 2

We study the sequence $(\mathcal{M}(n))_n$ with respect to the loglog-density, that is with $w_n = \frac{1}{p_n}$. Let us fix $h \in \mathbb{Z}^*$ and the integer $N \geq 3$. Observe first that, for all $n \geq 2$, we have

$$\left|\frac{1}{\log n} - \frac{1}{\log(n+1)}\right| = \left|\int_n^{n+1} \frac{-1}{x(\log x)^2}\,dx\right| \leq \frac{1}{n(\log n)^2}\,. \tag{8}$$

Then, thanks to Abel's transformation and Lemma 8, we have

$$\begin{aligned}
\left|\sum_{n=2}^{N} \frac{e_h(\log_{10} n)}{n\log n}\right| &= \left|\sum_{n=2}^{N} n^{2i\pi\theta_h - 1}(\log n)^{-1}\right| \\
&= \left|B_N(\log N)^{2i\pi\theta_h} - B_1(\log 2)^{2i\pi\theta_h} + \sum_{n=2}^{N-1} B_n\left(\frac{1}{\log n} - \frac{1}{\log(n+1)}\right)\right| \\
&\leq |B_N| + 1 + \sum_{n=2}^{N-1} |B_n|\frac{1}{n(\log n)^2} \\
&\leq \gamma_h + 1 + \frac{\gamma_h}{\log 2}\,.
\end{aligned}$$

This, together with Lemma 4, implies

$$\left|\sum_{n=2}^{N} \frac{e_h(\log_{10} n)}{p_n}\right| \leq C_3 + \gamma_h + 1 + \frac{\gamma_h}{\log 2}\,. \tag{9}$$

Relations (9) and (5) show that, for all $N \in \mathbb{N}^*$,

$$\sum_{h=1}^{H} \frac{1}{h}\left|\sum_{n=1}^{N} \frac{e_h(\log_{10} n)}{p_n}\right| = \mathcal{O}(\log H)^2$$

and the Erdős-Turán inequality then gives

$$D_N^{\log\log}(S_1') \leq \frac{3}{2}\left(\mathcal{O}(H^{-1}) + \frac{\mathcal{O}(\log H)^2}{\log\log N}\right)$$

where $S_1' = (\log_{10} n)_n$. Choosing $H = \lfloor \log\log N \rfloor$ leads to the announced bound. This concludes the proof of Theorem 2.

5.3 Proof of Theorem 3

Set $S_2' = (\log_{10} np_n)_n$. We have divided the proof in four stages.

1) We first study the sequence $(\mathcal{M}(n\log n))_n$ with respect to the log-density. Let us fix $h \in \mathbb{Z}^*$ and the integer $N \geq 3$. Then, thanks to Abel's transformation, relation (8) and Lemma 8, we have

$$
\begin{aligned}
\left|\sum_{n=2}^{N}\frac{e_h(\log(n\log n))}{n}\right| &= \left|\sum_{n=2}^{N} n^{2i\pi\theta_h-1}(\log n)^{2i\pi\theta_h}\right| \\
&= \left|B_N(\log N)^{2i\pi\theta_h} - B_1(\log 2)^{2i\pi\theta_h}\right| \\
&+ \left|\sum_{n=2}^{N-1} B_n\left((\log n)^{2i\pi\theta_h} - (\log(n+1))^{2i\pi\theta_h}\right)\right| \\
&\leq |B_N| + 1 + \sum_{n=2}^{N-1}|B_n|\frac{2\pi|\theta_h|}{n\log n} \\
&\leq \gamma_h + 1 + 2\pi|\theta_h|\gamma_h(\log\log N - \log\log 2)\,. \qquad (10)
\end{aligned}
$$

2) We now study the sequence $(\mathcal{M}(n\log n))_n$ with respect to the loglog-density. Let us fix $h \in \mathbb{Z}^*$ and the integer $N \geq 3$ and observe that, for all $n \geq 2$, we have

$$
\begin{aligned}
\left|(\log n)^{2i\pi\theta_h-1} - (\log(n+1))^{2i\pi\theta_h-1}\right| &= \left|\int_n^{n+1}(2i\pi\theta_h - 1)\frac{(\log x)^{2i\pi\theta_h-2}}{x}\,dx\right| \\
&\leq \frac{|2\pi\theta_h - 1|}{n(\log n)^2}\,.
\end{aligned}
$$

Therefore, because of Abel's transformation and Lemma 8, we have

$$
\begin{aligned}
\left|\sum_{n=2}^{N}\frac{e_h(\log_{10}(n\log n))}{n\log n}\right| &= \left|\sum_{n=2}^{N} n^{2i\pi\theta_h-1}(\log n)^{2i\pi\theta_h-1}\right| \\
&= \left|B_N(\log N)^{2i\pi\theta_h-1} - B_1(\log 2)^{2i\pi\theta_h-1}\right| \\
&+ \left|\sum_{n=2}^{N-1} B_n\left((\log n)^{2i\pi\theta_h-1} - (\log(n+1))^{2i\pi\theta_h-1}\right)\right| \\
&\leq \frac{|B_N|}{\log N} + \frac{1}{\log 2} + \sum_{n=2}^{N-1}|B_n|\frac{|2\pi\theta_h - 1|}{n(\log n)^2} \\
&\leq \frac{\gamma_h}{\log N} + \frac{1}{\log 2} + |2\pi\theta_h - 1|\frac{\gamma_h}{\log 2}\,.
\end{aligned}
$$

This, together with Lemma 4, implies

$$
\left|\sum_{n=2}^{N}\frac{e_h(\log_{10}(n\log n))}{p_n}\right| \leq C_3 + \frac{\gamma_h}{\log N} + \frac{1}{\log 2} + |2\pi\theta_h - 1|\frac{\gamma_h}{\log 2}\,. \qquad (11)
$$

3) We know study the sequence $(\mathcal{M}(p_n))_n$ with respect to the log-density. Let us fix $h \in \mathbb{Z}^*$ and the integer $N \geq 3$. Then the Abel's transformation, relation (10) and Lemma 3 give

$$\left|\sum_{n=3}^{N} \frac{e_h(\log_{10} p_n)}{n}\right| \leq \gamma_h + 1 + \gamma_h 2\pi|\theta_h|(1+\log\log N) + C_2|\theta_h| \sum_{n=3}^{N} \frac{\log\log n}{n\log n}. \qquad (12)$$

The Erdős-Turán inequality and relations (12) and (5) yield the bound

$$\frac{3}{2}\left(\mathcal{O}(H^{-1}) + \frac{\mathcal{O}(\log H)^2 + \mathcal{O}(\log H) + \mathcal{O}\left(H\log H\right)\left(1+\log\log N\right) + \mathcal{O}(H)(\log\log N)^2}{\log N}\right)$$

for $D_N^{\log}(S_2')$. Choosing $H = \left\lfloor (\log N)^{\frac{1}{2}} \right\rfloor$ leads to the announced bound.

4) We now study the sequence $(\mathcal{M}(p_n))_n$ with respect to the loglog-density. Let us fix $h \in \mathbb{Z}^*$ and the integer $N \geq 3$. Then, from relation (11) and Lemma 3, we have

$$\left|\sum_{n=3}^{N} \frac{e_h(p_n)}{p_n}\right| \leq C_3 + \frac{\gamma_h}{\log N} + \frac{1}{\log 2} + |2\pi\theta_h - 1|\frac{\gamma_h}{\log 2} + C_2|\theta_h| \sum_{n=3}^{N} \frac{\log\log n}{p_n \log n}. \qquad (13)$$

The Erdős-Turán inequality and relations (5) and (13) give

$$D_N^{\log\log}(S_2') \leq \frac{3}{2}\left(\mathcal{O}(H^{-1}) + \frac{\mathcal{O}(\log H)^2 + \mathcal{O}(\log H) + \mathcal{O}\left(H\log H\right) + \mathcal{O}(H)}{\log\log N}\right).$$

Then, by fixing $H = \left\lfloor (\log\log N)^{\frac{1}{2}} \right\rfloor$, we get the announced bound. The proof of Theorem 3 is now complete.

6 Conclusion

All our results and proofs remain valid with an arbitrary numeration base $b > 1$, even a non-integral one. It suffices to replace $\log_{10}$ by $\log_b$, the usual mantissa $\mathcal{M}$ by the mantissa $\mathcal{M}_b$ in base b, and μ_B by the probability $\mu_{B,b}$ on $[1, b[$ defined by $\mu_{B,b}([1, a[) = \log_b a$. This is all the more remarkable since, for example, the sequence $(\mathcal{M}_b(2^n))_n$ is distributed following $\mu_{B,b}$ only if b is not a rational power of 2 and there exists no random variable X such that the law of $\mathcal{M}_b(X)$ is $\mu_{B,b}$ for all $b > 1$ ([10] and [6, p. 19–29]).

Note that, in our proofs, we have shown that, when $r \to +\infty$, the mantissa of the row terms of the array $((n\log n)^r)_{r,n}$ tends to be distributed following μ_B.

By the Weyl criterion (see [7, p. 7] or [3, p. 14]), it is easy to verify that if $(\mathcal{M}(u_n))_n$ is distributed (log-distributed and loglog-distributed, respectively) following μ_B, then $(\mathcal{M}(u_n^r))_n$ also is for all $r \in \mathbb{N}^*$. Therefore, for all positive integer r, the sequences $(\mathcal{M}(n^r))_n$, $(\mathcal{M}(p_n^r))_n$ and $(\mathcal{M}((n\log n)^r))_n$ are log-distributed and loglog-distributed following μ_B.

It is well known to the arithmeticians that, for a real number α, the sequence of fractional parts $(\{\alpha n\})_n$ is distributed following U if and only if α is an irrational number. Because of the properties recalled in Section 4, we deduce from this that the mantissa of the terms of each column of the array $(p_n^r)_{r,n}$ (and, up to rare exceptions, of each column of $(n^r)_{r,n}$) is distributed following μ_B with respect to the natural density. Moreover, the diagonal of the array $(n^r)_{r,n}$ is the sequence $(n^n)_n$ whose mantissa is distributed, still with respect to the natural density, following μ_B [9].

References

[1] D.I. Cohen and T.M. Katz, Prime Numbers and the First Digit Phenomenon, J. Number Theory **18-3** (1984), 261–268.

[2] P. Diaconis, The Distribution of Leading Digits and Uniform Distribution mod 1, Ann. Probab. **5-1** (1977), 72–81.

[3] M. Drmota and R.F. Tichy, Sequences, discrepancies and applications, Lecture Notes in Math., 1651 (Springer-Verlag, Berlin, 1997).

[4] R.L. Duncan, Note on the Initial Digit Problem, Fibonacci Quart. **7-5** (1969), 474–475.

[5] A. Fuchs and G. Letta, Le problème du premier chiffre décimal pour les nombres premiers, (French) [The first digit problem for primes], Electron. J. Combin. **3-2**, R25 (1996).

[6] D.E. Knuth, The Art of Computer Programming volume 2 (Addison-Wesley, Reading, Massachusetts, 1969).

[7] L. Kuipers and H. Niederreiter, Uniform Distribution of Sequences (Dover Publications, New-York, 2006).

[8] B. Massé and D. Schneider, On weighted densities and their connection with the first digit phenomenon, Rocky Mountain J. Math., 51(5) (2011), pp 1395–1415.

[9] P.N. Posch, A Survey of Sequences and Distribution Functions satisfying the First-Digit-Law, J. Stat. and Management Systems **11-1** (2008), 1–19.

[10] R. Raimi, The first digit problem, Amer. Math. Monthly **83** (1976), 521–538.

[11] P. Ribenboim, The Little Book of Bigger Primes, 2nd ed. (Springer, Berlin, 2004), p. 185.

[12] J. Rivat and G. Tenenbaum, Inégalités d'Erdős-Turán, Ramanujan J. **9** (2005), 111–121.

[13] E.C. Titchmarsh, The theory of the Riemann zeta-function, 2nd ed. (The Clarendon Press, Oxford University Press, New York, 1986).

[14] R.E. Whitney, Initial digits for the sequence of primes, Amer. Math. Monthly **79** (1972), 150–152.

Authors' addresses:

Shalom Eliahou[a,b,c], Bruno Massé[a,b,c], Dominique Schneider[a,b,c],
[a]Univ Lille Nord de France, F-59000 Lille, France
[b]ULCO, LMPA J. Liouville, B.P. 699, F-62228 Calais, France
[c]CNRS, FR 2956, France
e-mail: {eliahou, bmasse, dominique.schneider}@lmpa.univ-littoral.fr