ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 41, Number 5, 2011

A SURVEY ON WEIGHTED DENSITIES AND THEIR CONNECTION WITH THE FIRST DIGIT PHENOMENON

BRUNO MASSÉ AND DOMINIQUE SCHNEIDER

ABSTRACT. This paper is a general treatment of the various notions of densities used in papers on mantissa distribution of sequences of numbers. Equivalence classes of weighted densities are identified, and their hierarchy is stated. This permits us to give clear answers to several questions about the first digit phenomenon. Moreover, however light the weights are, we exhibit an example of a sequence of positive numbers whose mantissae do not admit any distribution in the sense of the corresponding density.

1. Introduction and definitions. Following the early works of Benford and Newcomb [**1, 17**] on real life numbers, many authors have studied the distribution of the first digit in base 10 of sequences $(u_n)_n$ of positive numbers like $u_n = 2^n$, $u_n = n!$, $u_n = n^n$ and $u_n = F_n$ where F_n is the nth Fibonacci number, $u_n = n$ or $u_n = p_n$ where p_n is the nth prime number and so on (see [**19**] for a survey). In the first four cases they proved that, if $D(u_n)$ denotes the first digit of u_n and $\log_{10}$ the decimal logarithm, the natural density of $A_k^u = \{n \in \mathbf{N}^* : D(u_n) = k\}$ is $\log_{10}((k+1)/k)$, that is to say,

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{A_k^u}(n) = \log_{10}\left(\frac{k+1}{k}\right) \quad (k = 1, \dots, 9);$$

(here and in the sequel, $\mathbf{1}_B$ is the indicator function of the subset B). In particular, about 30.1 percent of the u_n have first digit 1 in the sense of the above formula. This property is known as the *first digit phenomenon.* Classical applications of this phenomenon are fraud detection [**18**] and computer design [**11, 15**].

In fact, we know a more precise property which needs three other definitions to be stated: *Benford's law* (in base 10) is the probability

2010 AMS *Mathematics subject classification.* Primary 60B10, 11B05, 11K99.
Keywords and phrases. Benford's law, first digit phenomenon, mantissa, weighted density, hierarchy.
Received by the editors on June 22, 2011.

DOI:10.1216/RMJ-2011-41-5-1395

measure μ_B on the interval $[1;10[$ defined by

$$\mu_B([1;a[\,) = \log_{10} a \quad (1 \le a < 10).$$

The *mantissa* of a positive real number x is the unique number $\mathcal{M}(x)$ in $[1;10[$ such that there exists an integer k verifying $x = \mathcal{M}(x)10^k$ (there exists another definition of the mantissa, but for technical reasons we shall use this one). A sequence $(U_n)_n$ of real numbers in $[1;10[$ is called *natural-Benford* if it is naturally distributed as μ_B, that is to say, if

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{[1;\,a[}(U_n) = \log_{10} a \quad (1 \le a < 10).$$

The above formula means that, for each $a \in [1;10[$, the set $\{n \in \mathbf{N}^* : 1 \le U_n < a\}$ admits a natural density, its natural density is $\log_{10} a$ and this can be interpreted as the weak convergence of the uniform probability measure on the set $\{U_1, \dots, U_N\}$ to μ_B as $N \to +\infty$.

A sequence $(u_n)_n$ of positive numbers is also called natural-Benford if the sequence of mantissae $(\mathcal{M}(u_n))_n$ is natural-Benford. We can now state: the sequences $(2^n)_n$, $(n!)_n$, $(n^n)_n$ and $(F_n)_n$ are all natural-Benford. The study of the mantissa is of course more general than the study of the first digit and allows us to easily derive the distribution of every digit and every string of digits of the u_n.

When $u_n = n$ or $u_n = p_n$ (see [**8, 24**]),

$$\liminf_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{A_1^u}(n) = \frac{1}{9} \tag{1}$$

and

$$\limsup_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{A_1^u}(n) = \frac{5}{9}.$$

So these two sequences do not verify the first digit phenomenon in the sense of the natural density. From [**4**], we know that they verify this phenomenon in the sense of the logarithmic density, that is to say,

$$\lim_{N\to+\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} \mathbf{1}_{A_k^u}(n) = \log_{10}\left(\frac{k+1}{k}\right) \quad (k = 1, \dots, 9)$$

where log is the natural logarithm. In a way (but not the same way as above), about 30.1 percent of the u_n have first digit 1. One corrects the defect in (1) by properly assigning lighter weights to larger numbers.

A more precise property is available in [**6**] and needs another definition to be stated. A sequence $(U_n)_n$ of real numbers in $[1;10[$ is called *logarithmic-Benford* if it is logarithmically distributed as μ_B, that is to say, if

$$\lim_{N\to+\infty}\left(\sum_{n=1}^{N}\frac{1}{n}\right)^{-1}\sum_{n=1}^{N}\frac{1}{n}\mathbf{1}_{[1;\,a[}(U_n)=\log_{10}a \quad (1\le a<10).$$

The above formula means that, for each $a\in[1;10[$, the set $\{n\in\mathbf{N}^*: 1\le U_n<a\}$ admits a logarithmic density, its logarithmic density is $\log_{10}a$ and again this can be interpreted as the weak convergence of some sequence (P_N) of discrete probability measures to μ_B as $N\to+\infty$, the atoms of P_N again being $U_1,\dots,U_N$. A sequence $(u_n)_n$ of positive numbers is also called logarithmic-Benford if the sequence of mantissae $(\mathcal{M}(u_n))_n$ is logarithmic-Benford.

We can now state: the sequences $(n)_n$ and $(p_n)_n$ are logarithmic-Benford. Note that [**8, 23**] consider the relative logarithmic density in the set of prime numbers which is strictly weaker than the logarithmic density, as we shall see below.

In the spirit of Diaconis's work about binomial coefficients [**2**], it is also proved in [**6**] that the rows of the infinite matrix $(\mathcal{M}(u_n^m))_{m,n}$, which are logarithmic-Benford and not natural-Benford, tend to be natural-Benford as m tends to infinity; that is to say, there exists an increasing function N from $\mathbf{N}^*$ to $\mathbf{N}^*$ such that

$$\lim_{m\to+\infty}\sup_{1<a<10}\left|\frac{1}{N(m)}\sum_{n=1}^{N(m)}\mathbf{1}_{[1;\,a[}(\mathcal{M}(u_n^m))-\log_{10}a\right|=0.$$

Of course, many sequences of mantissae of positive numbers are neither natural nor logarithmic-Benford, but some of them, like $(\mathcal{M}(10^n))_n$ or $(\mathcal{M}((-1)^n+2))_n$, do admit a distribution (distinct from the Benford's one) anyway.

1.1. Some questions. The quite strange facts described above generate many questions. Here are a few that shall be answered in Section 5 below:

Question 1. Are there densities which are strictly weaker (see below) than the logarithmic one?

Question 2. Are there classical sequences of positive numbers whose mantissae do not admit any distribution in the sense of natural or logarithmic densities? If yes, what about weaker densities?

Question 3. Is there an interest to consider weights lighter than $1/n$ or heavier than 1?

Question 4. Is the first digit phenomenon verified by $u_n = 2^n$, $u_n = n!$, $u_n = n^n$ or $u_n = F_n$ if we use weights heavier than 1, like n^α with $\alpha > 0$ or α^n with $\alpha > 1$ for instance? If yes, is there a maximal value for α?

Question 5. Is the first digit phenomenon verified by $u_n = n$ or $u_n = p_n$ if we use weights heavier than $1/n$, like $1/\sqrt{n}$ for instance? If not, does the choice of one of these weights have an influence anyway?

Question 6. If $u_n = n$ or $u_n = p_n$, we know from [**6**] that the rows of $(\mathcal{M}(u_n^m))_{m,n}$ are logarithmic-Benford, do not admit any distribution in the sense of the natural density and tend to be natural-Benford. What will happen if we choose intermediate weights (between 1 and $1/n$)?

1.2. Weighted densities. This leads us, in the wake of [**8, 9**], for example, to consider the general notion of weighted densities of $A \subset \mathbf{N}^*$ where $\mathbf{N}^*$ is the set of positive integers.

Let $(w_n)_{n\geq 1}$ be a sequence of positive real numbers summing to infinity and, for each $N \geq 1$, let $W_N = \sum_{n=1}^N w_n$. One says that $A \subset \mathbf{N}^*$ has a w_n-*density* when the sequence

$$\left(\sum_{n=1}^{N} \frac{w_n}{W_N} \, \mathbf{1}_A(n) \right)_N$$

converges and its limit is then called the w_n-density of A. This is the limit of the weighted frequency of the elements of A among those of

$\mathbf{N}^*$. The condition on the weights w_n is necessary to assign the density $1/2$ to the set of even numbers, for example, and the density 0 to every finite subset of $\mathbf{N}^*$.

Another sequence $(v_n)_{n\geq 1}$ of positive real numbers summing to infinity being given, we set $V_N = \sum_{n=1}^{N} v_n$ and we say that the w_n-density is *stronger* than the v_n-density when, for every $A \subset \mathbf{N}^*$,

$$\left(\left(\sum_{n=1}^{N} \frac{w_n}{W_N}\,\mathbf{1}_A(n)\right)_N \text{ converges}\right) \Longrightarrow \left(\left(\sum_{n=1}^{N} \frac{v_n}{V_N}\,\mathbf{1}_A(n)\right)_N \text{ converges}\right)$$

and when, in this case, the two limits are equal. If either density is stronger than the other one, then the two densities are said to be *equivalent*.

The most commonly used weighted densities are the two we have considered above: the 1-density called *natural* or *arithmetic density* and the $1/n$-density called *logarithmic* or *harmonic density*. In [**6**], the $1/n\log n$-density is also considered and called the *loglog-density*. In some papers on Benford's law ([**8, 23**], for example), we also find the so-called *logarithmic density conditioned* (*or relative*) *to the prime numbers* which can be seen as the $1/p_n$-density. The other weights that come immediately to mind are $w_n = \alpha^n$ where $\alpha > 1$, $w_n = n^\alpha$ or $w_n = p_n^\alpha$ with $\alpha > 0$ or $-1 < \alpha < 0$, $w_n = n^\alpha(\log n)^\beta$ or $w_n = p_n^\alpha(\log p_n)^\beta$ with $\alpha > -1$ and $\beta \in \mathbf{R}$, w_n polynomial, and $w_n = 1/(g_q(n))$ with $g_0(n) = n$, $g_1(n) = n\log n$, $g_2(n) = n(\log n)(\log\log n)$, and so on ($n$ large enough).

1.3. Contents. It is well known ([**22**, page 272] for example) that the 1-density is strictly stronger than the $1/n$-density and Kuipers and Niederreiter [**16**, page 64] mentioned a quite surprising property: all the n^α-densities with $-1 < \alpha < 0$ and $0 < \alpha$ are equivalent to the 1-density. In Section 2, we state three theorems found in Hardy's book [**12**] which give a clear view on the hierarchy between weighted densities. To the best of our knowledge, these theorems are not mentioned in papers on Benford's law. Section 3 is devoted to new results. The first one shows that equivalent weights lead to equivalent densities. The second one proves that the hierarchy between the $1/(g_q(n))$-densities ($q = 0, 1, 2, \dots$) is strict. The third one states that, however light the considered weights are, we can find a classical sequence of positive numbers whose mantissae do not admit any distribution in the sense of the corresponding density. Combining Sections 2 and 3 enables us to identify in Section 4 an infinite number of equivalence classes of weighted densities and in Section 5 to give simple and clear answers to the questions we have listed above. We give in Section 6 a short overview on other densities used in the study of the first digit phenomenon and their connections with weighted densities. Some open problems are described in Section 7. For the sake of clarity and self-contained exposition, we give in an Appendix the proofs of the theorems stated in Section 2, rewritten in the context of weighted densities.

We have focused on numeration in base 10, but all statements remain true in every base $b \geq 2$ except for the sequence $(2^n)_n$ which is not a Benford sequence in bases 2, 4, 8, and so on.

2. Summation method properties. These properties are derived from Cesàro's and Toeplitz's works and are stated and proved in [**12**, pages 42–63] in the general context of hierarchy of summation methods applied to a sequence (s_n) of finite sums of a series. But a close look at the proofs shows that the nature of the sequence (s_n) is not important. We state them below in the context of weighted densities, for instance the sequence $(\mathbf{1}_A(n))$ where $A \subset \mathbf{N}^*$ will take the place of (s_n). Theorem 1 below is also stated and proved in [**16**, pages 63–64].

Let us recall that (v_n) and (w_n) are two sequences of positive real numbers summing to infinity and that $V_N = \sum_{n=1}^N v_n$ and $W_N = \sum_{n=1}^N w_n$.

Theorem 1. *A sufficient condition for the w_n-density to be stronger than the v_n-density is*

$$\left(\frac{v_n}{w_n}\right)_n \text{ is non-increasing} \tag{2}$$

or

$$\left(\frac{v_n}{w_n}\right)_n \text{ is non-decreasing and } \left(\frac{v_N W_N}{w_N V_N}\right)_N \text{ is bounded.} \tag{3}$$

Theorem 2 below gives a clear view on the effect of the heaviness of the weights.

Theorem 2. *Let $A \subset \mathbf{N}^*$. If the sequence $(v_n/w_n)_n$ is non-increasing, then*

$$\begin{aligned}\underline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N}\,\mathbf{1}_A(n) &\le \underline{\lim}_N \sum_{n=1}^{N} \frac{v_n}{V_N}\,\mathbf{1}_A(n)\\ &\le \overline{\lim}_N \sum_{n=1}^{N} \frac{v_n}{V_N}\,\mathbf{1}_A(n)\\ &\le \overline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N}\,\mathbf{1}_A(n).\end{aligned}$$

Theorem 3 below will permit us to prove that the exponential weights are not relevant in the context of weighted densities of a subset of $\mathbf{N}^*$.

Theorem 3. *Let $A \subset \mathbf{N}^*$. If the sequence $(W_{N-1}/w_N)_N$ is bounded, then A cannot admit any w_n-density unless A is finite or cofinite.*

3. New results. Let us recall that (v_n) and (w_n) are two sequences of positive real numbers summing to infinity and that $V_N = \sum_{n=1}^N v_n$ and $W_N = \sum_{n=1}^N w_n$.

Proposition 1. *If v_n and w_n are equivalent as $n \to +\infty$, then the v_n-density and the w_n-density are equivalent too.*

Proof. It is well known that V_N and W_N are equivalent as $N \to +\infty$ since the v_n and the w_n sum to infinity. So we can write $w_n = v_n + v_n\theta_1(n)$ with $\lim_n \theta_1(n) = 0$ and $W_N = V_N + V_N\theta_2(N)$ with $\lim_N \theta_2(N) = 0$ and

$$\theta_2(N) = \sum_{n=1}^{N} \frac{v_n}{V_N}\,\theta_1(n).$$

Now consider $A \subset \mathbf{N}^*$. For every $N \geq 1$,

$$\begin{aligned}\left|\sum_{n=1}^{N} \frac{w_n}{W_N}\, \mathbf{1}_A(n) - \sum_{n=1}^{N} \frac{v_n}{V_N}\, \mathbf{1}_A(n)\right| &\leq \sum_{n=1}^{N} \frac{v_n \Big| V_N + V_N \theta_1(n) - W_N \Big|}{W_N V_N} \\ &\leq \frac{|V_N - W_N|}{W_N} + \left|\sum_{n=1}^{N} \frac{v_n}{W_N}\, \theta_1(n)\right| \\ &\leq 2 \frac{V_N}{W_N}\, |\theta_2(N)| \,.\end{aligned}$$

So A admits a w_n-density if and only if it admits a v_n-density and, in this case, the two densities are equal. □

Remark. In fact, the above calculations imply a deeper property: Define the lower and the upper w_n-densities of A by $\underline{\lim}_N \sum_{n=1}^{N} (w_n/W_N)\, \mathbf{1}_A(n)$ and $\overline{\lim}_N \sum_{n=1}^{N} (w_n/W_N)\, \mathbf{1}_A(n)$. If v_n and w_n are equivalent, then the lower (respectively upper) w_n-density and the lower (respectively upper) v_n-density of any $A \subset \mathbf{N}^*$ are equal. And it is easy to verify that, if two weighted densities are equivalent, the corresponding lower (or upper) densities of any $A \subset \mathbf{N}^*$ are not necessarily equal. For example, the upper 1-density and the upper n-density of the set of positive integers whose first digit is 1 are respectively 5/9 and 3/4, although the 1-density and the n-density are equivalent as we shall see below.

The classical example of a subset of $\mathbf{N}^*$ which admits a logarithmic density but does not admit any natural density is $B = \cup_{m=0}^{+\infty} \{n : e^m \leq n < 2e^m\}$. We now generalize this example to densities lighter than the natural and the logarithmic ones.

Set $\log^{(1)} = \log$, $\exp^{(1)} = \exp$ and, for $q \geq 1$, $\log^{(q+1)} = \log^{(q)} \circ \log$, $\exp^{(q+1)} = \exp^{(q)} \circ \exp$ and

$$B_q = \bigcup_{m=0}^{+\infty} \{n : \exp^{(q)}(e^m) \leq n < \exp^{(q)}(2e^m)\}.$$

With these notations, the numbers $g_q(n)$ (see the introduction for definitions) can be defined by $g_0(n) = n$ and $g_q(n) = g_{q-1}(n) \log^{(q)} n$ ($q \geq 1$ and n large enough).

Proposition 2. *For $q \geq 1$, B_q admits a $1/(g_q(n))$-density, but does not admit any $1/(g_{q-1}(n))$-density.*

Proof. We shall use techniques like those of Fuchs and Letta in [**4**]. Let $q \geq 1$, $w_n = 1/(g_{q-1}(n))$, $v_n = 1/(g_q(n))$, $W_N = \sum_1^N w_n$, $V_N = \sum_1^N v_n$, $a_m = \exp^{(q)}(e^m)$, $b_m = \exp^{(q)}(2e^m)$,

$$C_m = \{n : a_m \leq n < b_m\},$$
$$D_m = \{n : a_{m-1} \leq n < a_m\}$$

and

$$E_m = \{n : b_{m-1} \leq n < b_m\}.$$

Then

$$\underline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_{B_q}(n) = \lim_M \frac{\sum_{m=1}^{M} \sum_{n \in C_{m-1}} w_n}{\sum_{m=0}^{M} \sum_{n \in D_m} w_n} = \lim_m \frac{\sum_{n \in C_{m-1}} w_n}{\sum_{n \in D_m} w_n}$$

because the numbers in the numerators sum to infinity and so do the ones in the denominators. For the same reasons,

$$\overline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_{B_q}(n) = \lim_m \frac{\sum_{n \in C_m} w_n}{\sum_{n \in E_m} w_n}.$$

Using classical integral calculations, we get $\sum_{n \in C_m} w_n \sim e^m$, $\sum_{n \in D_m} w_n \sim e^m - e^{m-1}$ and $\sum_{n \in E_m} w_n \sim 2(e^m - e^{m-1})$ as $m \to +\infty$. Hence,

$$\underline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_{B_q}(n) = \frac{1}{e-1} \quad \text{and} \quad \overline{\lim}_N \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_{B_q}(n) = \frac{e}{2(e-1)}.$$

Since $\lim_m \sum_{n \in C_m} v_n = \log 2$ and $\lim_m \sum_{n \in D_m} v_n = \lim_m \sum_{n \in E_m} v_n = 1$, the same arguments give

$$\underline{\lim}_N \sum_{n=1}^{N} \frac{v_n}{V_N} \mathbf{1}_{B_q}(n) = \overline{\lim}_N \sum_{n=1}^{N} \frac{v_n}{V_N} \mathbf{1}_{B_q}(n) = \log 2. \qquad \square$$

The following theorem gives a general view of the $1/g_q(n)$-densities ($q = 1, 2, \dots$) in connection with the first digit phenomenon. For any real x, $\{x\}$ denotes the fractional part of x, that is to say, $\{x\} = x - [x]$ where $[x]$ is the greatest integer smaller than x and δ_x is the Dirac measure at x. For $x > 0$, we set $f(x) = \{\log_{10} x\}$.

Theorem 4. *Let $q \geq 0$. The sequence $(\mathcal{M}(\log^{(q+2)} n))_n$ does not admit any distribution in the sense of the $1/(g_q(n))$-density.*

Proof. In the sequel, n_0 is any integer greater than $\exp^{(q+2)}(1)$. For $n \geq n_0$, let $w_n = 1/(g_q(n))$, $x_n = \mathcal{M}(\log^{(q+2)} n)$ and $y_n = \log_{10}(\log^{(q+2)} n)$. For $N \geq n_0$, let

$$P_N = \sum_{n=n_0}^{N} \frac{w_n}{W_N} \delta_{x_n} \quad \text{and} \quad Q_N = \sum_{n=n_0}^{N} \frac{w_n}{W_N} \delta_{\{y_n\}},$$

where $W_N = \sum_{n=n_0}^{N} w_n$, and let

$$G_N(t) = \sum_{n=n_0}^{N} \frac{w_n}{W_N} e^{it\{y_n\}} \quad (t \text{ real}).$$

Note that W_N is equivalent to $\log^{(q+1)} N$ as $N \to +\infty$ and that G_N is the Fourier transform of Q_N. Since $f(x) = f(\mathcal{M}(x))$, we get $Q_N = P_N f^{-1}$, that is to say, $Q_N(I) = P_N(f^{-1}(I))$ for every interval $I \subset [0, 1[$. It is easy to verify that the weak convergence of the sequence $(P_N)_N$ to some probability measure μ is equivalent to the weak convergence of $(Q_N)_N$ to μf^{-1} and, by Lévy's theorem on weak convergence, this is equivalent to pointwise convergence of $(G_N)_N$ to the Fourier transform of μf^{-1}. So, to prove our theorem, it suffices to verify that the sequence $(G_N(2\pi))_N$ diverges.

Fix $N > n_0$. Since $e^{2i\pi\{y_n\}} = e^{2i\pi y_n}$, the classical Abel transform of a sum gives

$$G_N(2\pi) = e^{2i\pi y_N} + \frac{1}{W_N} \sum_{n=n_0}^{N-1} W_n (e^{2i\pi y_n} - e^{2i\pi y_{n+1}}).$$

Now, the mean value inequality gives

$$\left|e^{2i\pi y_n} - e^{2i\pi y_{n+1}}\right| \leq \frac{2\pi(\log 10)^{-1}}{g_{q+2}(n)} \quad (n = n_0, \dots, N-1).$$

Moreover,

$$\frac{W_n}{g_{q+2}(n)} \sim \frac{w_n}{\log^{(q+2)} n} \quad (n \to +\infty).$$

Since the numbers above sum to infinity,

$$\sum_{n=n_0}^{N-1} \frac{W_n}{g_{q+2}(n)} \sim \sum_{n=n_0}^{N-1} \frac{w_n}{\log^{(q+2)} n} \quad (N \to +\infty).$$

The classical generalizations of Cesàro's theorem (see [**12**, page 43] or the lemma in Section 8 of the present paper) show that

$$\lim_{N\to+\infty} \sum_{n=n_0}^{N-1} \frac{w_n}{W_N \log^{(q+2)} n} = 0$$

because $\lim_n \frac{1}{\log^{(q+2)} n} = 0$. So $(G_N(2\pi))_N$ diverges since $(e^{2i\pi y_N})_N$ diverges and

$$\lim_{N\to+\infty} \frac{1}{W_N} \sum_{n=n_0}^{N-1} W_n(e^{2i\pi y_n} - e^{2i\pi y_{n+1}}) = 0. \qquad \square$$

4. Hierarchy of weighted densities. Combining Sections 2 and 3 gives a clear vision of the hierarchy between weighted densities.

General principle: Theorem 2 shows clearly that the heavier weights w_n are, the rarer are the subsets A of $\mathbf{N}^*$ which admit a w_n-density. Moreover, when A does not admit any density, w_n's heaviness affects the lower and the upper w_n-densities of A (see the above remark).

On this subject, we must mention [**9**] which contains a study of the continuity of the function $\alpha \mapsto (\underline{d}_\alpha(A), \overline{d}_\alpha(A))$ where $A \subset \mathbf{N}^*$ is fixed,

α is varying in $[-1; +\infty]$ and $\underline{d}_\alpha(A)$ and $\overline{d}_\alpha(A)$ are, respectively, the lower and the upper n^α-densities. For example, the set A of positive integers whose first digit is 1 does not admit any n^α-density for $\alpha > -1$, but, since it verifies the conditions of Theorem 2 in [**9**], the function $\alpha \mapsto (\underline{d}_\alpha(A), \overline{d}_\alpha(A))$ is continuous at point -1 and so $\underline{d}_\alpha(A)$ and $\overline{d}_\alpha(A)$ monotonically tend to $\log_{10} 2$ as $\alpha \to -1^+$.

Exponential weights: The weights $w_n = \alpha^n$ ($\alpha \neq 1$) are not relevant because the condition $\sum_n w_n = +\infty$ implies $\alpha > 1$ and then, by Theorem 3, the only subsets of $\mathbf{N}^*$ which admit a w_n-density are the finite and the cofinite ones. The densities of these kinds of subsets of $\mathbf{N}^*$ are respectively equal to 0 and 1, whatever is the value of $\alpha > 1$.

Strict hierarchy: It is evident (see directly above) that the α^n-densities with $\alpha > 1$ are strictly stronger than the 1-density, and it is well known that the 1-density is strictly stronger than the $1/n$-density ([**22**, page 272] for example). We can now state: the $1/n$-density is strictly stronger than the $1/(n \log n)$-density which is strictly stronger than the $1/[n(\log n)(\log\log n)]$-density and more generally, for $q \geq 0$, the $1/(g_q(n))$-density is strictly stronger than the $1/(g_{q+1}(n))$-density. Indeed, the $1/(g_q(n))$-density is stronger than the $1/(g_{q+1}(n))$-density by condition (2) in Theorem 1, and the $1/(g_{q+1}(n))$-density is not stronger than the $1/(g_q(n))$-density as Proposition 2 shows.

Natural density equivalence class: Most of the weights which come naturally to mind lead to densities which are equivalent to the 1-density. Firstly, as Kuipers and Niederreiter noticed [**16**, page 64], all the n^α-densities with $-1 < \alpha$ are equivalent. Indeed, if $-1 < \alpha_1 < \alpha_2$, condition (2) of Theorem 1 proves that the n^{α_2}-density is stronger than the n^{α_1}-density and condition (3) shows the converse. Secondly, for $-1 < \alpha$ and $\beta \in \mathbf{R}$, the $n^\alpha(\log n)^\beta$-density is equivalent to the n^α-density and then to the 1-density because condition (2) proves that it is stronger than the $\sqrt{n^{\alpha-1}}$-density and weaker than the $n^{\alpha+1}$-density. Moreover, Proposition 1 shows that all $P(n)Q(\log n)$-densities (P and Q polynomials) belong to this equivalence class too and so do the p_n^α-densities with $\alpha > -1$ because $p_n \sim n \log n$ as $n \to +\infty$.

Logarithmic density equivalence class: Somehow, this equivalence class is smaller than the previous one (although it possesses infinitely many elements). Condition (3) of Theorem 1 shows that, if $w_n = (\log n)^\alpha / n$ with $\alpha > -1$, then the w_n-density is equivalent to the $1/n$-density. This and Proposition 1 prove that the $P(\log n)/n$ (P polynomial) and the $(\log p_n)/p_n$-density (utilized in [**14**]) belong to this equivalence class too.

Weaker densities equivalence classes: When $w_n = (\log^{(q+1)} n)^\alpha / g_q(n)$ ($q \geq 1$) with $\alpha > -1$, the same arguments show that the w_n-density is equivalent to the $1/g_q(n)$-density and again the only simple way to construct other equivalent densities seems to use Proposition 1. For example, the $1/p_n$-density (utilized in [**8, 23**]) and the $1/n \log n$-density are equivalent (case $q = 1$).

Maybe the weakest density: Let $w_n^{(q)} = 1/g_q(n)$. Then, by Theorem 2, the limits below exist, and we can set

$$\underline{W}_\infty(A) = \lim_{q \to +\infty} \left(\underline{\lim}_N \sum_{n=1}^{N} \frac{w_n^{(q)}}{W_N^{(q)}} \mathbf{1}_A(n) \right)$$

and

$$\overline{W}_\infty(A) = \lim_{q \to +\infty} \left(\overline{\lim}_N \sum_{n=1}^{N} \frac{w_n^{(q)}}{W_N^{(q)}} \mathbf{1}_A(n) \right).$$

And so we have defined a new lower density and a new upper density which define a density when they are equal. This new density can be named the W_∞-density. It is not, apparently, a weighted density and is weaker than all the densities we have considered in the present paper.

5. Consequences for Benford sequences.

We give here some answers to the questions we have listed in Section 1.

Answer to Question 1: The $1/g_q(n)$-densities for $q \geq 1$ are all strictly weaker than the logarithmic one.

Answer to Question 2: Assuming that everybody agrees to consider $(\log \log n)_n$ as a classical sequence, Theorem 4 shows that the correct answer to Question 2 is yes. It also shows that, however light the weights are, we can find a sequence of positive numbers whose mantissae do not admit any distribution in the sense of the corresponding density.

Answer to Question 3: Proposition 2 and Theorem 4 show that there is an interest in considering densities which are strictly weaker than the logarithmic one. The question of the interest of densities strictly stronger than the natural one remains open because of the second point of Section 4.

Answer to Question 4: Yes, the first digit phenomenon is verified by $u_n = 2^n$, $u_n = n!$, $u_n = n^n$ or $u_n = F_n$ if we choose heavier weights like n^α with $\alpha > 0$ and no, there is no maximal value for α. The first digit phenomenon is not verified if we choose weights like α^n with $\alpha > 1$.

Answer to Question 5: No, the first digit phenomenon is not verified by $u_n = n$ or $u_n = p_n$ if we choose weights like $1/\sqrt{n}$ because, if that was the case, it would be verified in the sense of the natural density. However, Theorem 2 shows that, if we choose weights like n^α with $-1 < \alpha < 0$, that will bring the upper and the lower densities (see the remark in 3.1) together.

Answer to Question 6: No row of $(\mathcal{M}(u_n^m))_{m,n}$ admits a distribution in the sense of the natural density and then in the sense of the n^α-density for $-1 < \alpha < 0$. But these rows tend (as $m \to +\infty$) to be distributed as μ_B in the sense of the natural density [**6**]. Somehow, there is a quantum leap from $\alpha = -1$ to $\alpha = 0$ (and even $\alpha = 1$, and so on) as $m \to +\infty$. Of course, the last sentence of the answer to Question 5 is still true here.

6. Other densities. Some authors [**8, 14, 23**] have introduced the notion of *conditional densities relative to the set of prime numbers*. The natural conditional density [**8**] is in fact the 1-density, the logarithmic conditional density [**8, 14**] is the $1/p_n$-density and is equivalent to the $1/n \log n$-density by Proposition 1, and the $\log p_n/p_n$-density [**23**] is equivalent to the $1/n$-density.

In [**21**], Serre claimed that Bombieri proved the following result: the *analytic density* of $A = \{n : D(p_n) = 1\}$ is $\log_{10} 2$, that is to say,

$$\lim_{\sigma \to 1^+} \zeta(\sigma)^{-1} \sum_{n \in A} n^{-\sigma} = \lim_{\sigma \to 1^+} (\sigma - 1) \sum_{n \in A} n^{-\sigma} = \log_{10} 2$$

where ζ is the Riemann zeta function. This density is sometimes called *Dirichlet density* or *ζ-density* and is equivalent to the logarithmic density [**22**, page 274]. Moreover, the *analytic density relative to prime* *numbers* of $B \subset \mathbf{N}^*$ is defined by

$$\lim_{\sigma \to 1^+} (-\log(\sigma - 1))^{-1} \sum_{p_n \in B} p_n^{-\sigma}$$

(Dirichlet used it to prove his theorem on arithmetic progressions), but we have not found it in papers on the first digit phenomenon.

In [**3, 5, 7, 13, 20**], for instance, the H_∞-*density* is defined in the following manner. A subset $A \subset \mathbf{N}^*$ being given, set $H_{0,n} = \mathbf{1}_A(n)$ $(n \geq 1)$ and for $m \geq 1$,

$$H_{m,n} = \frac{1}{n} \sum_{j=1}^{j=n} H_{m-1,j} \quad (n \geq 1).$$

When the sequence $(H_{m,n})_n$ converges, its limit is called the H_m-density of A. One says that A admits a H_∞-density when

$$\lim_{m \to +\infty} \underline{\lim}_n H_{m,n} = \lim_{m \to +\infty} \overline{\lim}_n H_{m,n}$$

and then its H_∞-density is the common value of these two limits. Since the sequences $(\mathbf{1}_A(n))_n$ are bounded, all H_m-densities are equivalent to the natural density [**12**, page 62]. By Cesàro's theorem, the natural density is stronger than the H_∞-density. In [**5**], Duran shows that the H_∞-density is stronger than the logarithmic density. In [**3**], Diaconis exhibits examples which prove that the converses of these two properties are false.

In short, each conditional weighted density can be viewed as a classical weighted density and belongs to one of the equivalence classes we have listed above, and the natural density is strictly stronger than the H_∞-density which is strictly stronger than the logarithmic density which is equivalent to the analytic density.

We have focused our attention on densities employed in papers about the first digit phenomenon, but many other densities are considered by Analytic Number Theory specialists. Many are listed in [**10**].

7. Conclusion. We list in this section a few open (as far as we know) questions about weighted densities and their hierarchy.

Open question 1: What is the exact influence of the weights w_n over the discrepancy

$$\sup_{1<a<10} \left| \left(\sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_{[1;\,a[}(\{\mathcal{M}(u_n)\}) \right) - \log_{10} a \right| ?$$

Open question 2: Is the natural density the strongest weighted density among those which are relevant in the study of mantissae distributions?

Open question 3: Can we find two weighted densities such that none of them is stronger than the other? If yes, can we find a subset of $\mathbf{N}^*$ admitting two distinct weighted densities?

Open question 4: We can replace $w_n = 1$ by $w_n = n^\alpha$ with $-1 < \alpha < 0$ in the study of the rows of the infinite matrix $(\mathcal{M}(u_n^m))_{m,n}$ where $u_n = n$ or $u_n = p_n$. How does the choice of α influence the choice of the function N in the formula

$$\lim_{m\to+\infty} \sup_{1<a<10} \left| \sum_{n=1}^{N(m)} \frac{w_n}{W_{N(m)}} \mathbf{1}_{[1;\,a[}(\mathcal{M}(u_n^m)) - \log_{10} a \right| = 0$$

and the convergence rate?

Open question 5: Is the analytic density relative to prime numbers equivalent to the $1/p_n$-density and then equivalent to the $1/n\log n$-density?

Open question 6: Is the W_∞-density (see the last paragraph of Section 4) strictly weaker than any $1/g_q(n)$-density?

Open question 7: Does the sequence $(\mathcal{M}(p_n))$ admit a distribution in the sense of the H_∞-density?

APPENDIX

For the sake of clarity and self-contained presentation, here are the proofs of the three theorems stated in Section 2, rewritten in the context of weighted densities.

Lemma. *Let $(S_n)_{n\geq 1}$ be a convergent sequence of real numbers and $(C_{N,n})_{N,n}$ a triangular array $(N \geq 1,\ 1 \leq n \leq N)$ of real numbers* *verifying*

$$\text{the sequence } \left(\sum_{n=1}^{N} |C_{N,n}| \right)_N \text{ is bounded,} \tag{4}$$

for all $n \geq 1$,

$$\lim_{N \to +\infty} C_{N,n} = 0 \tag{5}$$

and for all $N \geq 1$,

$$\sum_{n=1}^{N} C_{N,n} = 1. \tag{6}$$

For $N \geq 1$, we set

$$T_N = \sum_{n=1}^{N} S_n C_{N,n}.$$

Then the sequence $(T_N)_{N \geq 1}$ converges and $(S_n)_{n \geq 1}$ and $(T_N)_{N \geq 1}$ have the same limit.

Proof. Let $K > 0$ be a bound evoked in (4). Let $\varepsilon > 0$ and N_0 be such that, for every $n > N_0$ and every $N > N_0$, $|S_n - S_N| \leq \varepsilon K^{-1}$, and let $N > N_0$. Then

$$\begin{aligned} |T_N - S_N| &= \left| \sum_{n=1}^{N} S_n C_{N,n} - S_N \sum_{n=1}^{N} C_{N,n} \right| \\ &= \left| \sum_{n=1}^{N_0} (S_n - S_N) C_{N,n} + \sum_{n=N_0+1}^{N} (S_n - S_N) C_{N,n} \right| \\ &\leq \left| \sum_{n=1}^{N_0} (S_n - S_N) C_{N,n} \right| + \varepsilon \end{aligned}$$

by (6) and the definition of N_0. It remains to remark that, by (5),

$$\lim_{N \to +\infty} \left| \sum_{n=1}^{N_0} (S_n - S_N) C_{N,n} \right| = 0$$

since all the sequences $(S_n - S_N)_{N \geq 1}$ $(n = 1, \ldots, N_0)$ converge. $\square$

8.1. Proof of Theorem 1. Let $A \subset \mathbf{N}^*$ and, for every $N \geq 1$,

$$S_N = \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_A(n) \quad \text{and} \quad T_N = \sum_{n=1}^{N} \frac{v_n}{V_N} \mathbf{1}_A(n).$$

Then $w_1 s_1 = W_1 S_1$ and $w_N s_N = W_N S_N - W_{N-1} S_{N-1}$ $(N = 2, \dots)$ and then, for every $N \geq 1$,

$$\begin{aligned} T_N V_N = \frac{v_1}{w_1} W_1 S_1 &+ \frac{v_2}{w_2}(W_2 S_2 - W_1 S_1) + \cdots \\ &+ \frac{v_N}{w_N}(W_N S_N - W_{N-1} S_{N-1}), \end{aligned}$$

that is to say,

$$T_N = \sum_{n=1}^{N} S_n C_{N,n}$$

with

$$C_{N,N} = \frac{v_N}{w_N} \frac{W_N}{V_N}$$

and, for $n = 1, \dots, N-1$,

$$C_{N,n} = \left(\frac{v_n}{w_n} - \frac{v_{n+1}}{w_{n+1}} \right) \frac{W_n}{V_N}.$$

So condition (5) of the lemma is verified. Condition (6) is, too, since, if $\mathbf{1}_A(n)$ are all equal to 1, so are S_N and T_N.

If (2) is verified, then $C_{N,n}$ are nonnegative, and so the sequence

$$\left(\sum_{n=1}^{N} |C_{N,n}| \right)_N$$

is constant and this gives the lemma's condition (4).

If (3) is verified, then $C_{N,N}$ is nonnegative and $C_{N,n}$ is negative for $n = 1, \dots, N-1$. Then

$$\sum_{n=1}^{N} |C_{N,n}| = C_{N,N} - \sum_{n=1}^{N-1} C_{N,n}.$$

Since, as we have seen above,

$$C_{N,N} + \sum_{n=1}^{N-1} C_{N,n} = 1,$$

this and the second part of the second condition of Theorem 1 prove that

$$\sum_{n=1}^{N} |C_{N,n}| = 2C_{N,N} - 1 = 2\frac{v_N}{w_N}\frac{W_N}{V_N} - 1$$

is bounded. Condition (4) of the lemma is verified.

8.2. Proof of Theorem 2. With the same notations as in Theorem 1, the coefficients $C_{N,n}$ in the formula

$$T_N = \sum_{n=1}^{N} S_n C_{N,n}$$

are nonnegative (see the calculations above) if the condition of Theorem 2 is verified. Let $\varepsilon > 0$, $I = \underline{\lim}_n S_n$ and N_0 an integer such that for all $n > N_0$,

$$S_n > I - \varepsilon.$$

Then, for each $N > N_0$,

$$T_N > \sum_{n=1}^{N_0} S_n C_{N,n} + (I - \varepsilon) \sum_{n=N_0+1}^{N} C_{N,n}.$$

This implies $\underline{\lim}_N T_N \geq I - \varepsilon$ because

$$\lim_{N \to +\infty} C_{N,n} = 0 \qquad (n = 1, \dots, N_0)$$

and so

$$\lim_{N \to +\infty} \sum_{n=N_0+1}^{N} C_{N,n} = 1$$

(recall that the $C_{N,n}$ verify the lemma's condition (6)).

The superior limits can be investigated in the same way.

8.3. Proof of Theorem 3. Again let $A \subset \mathbf{N}^*$ and

$$S_N = \sum_{n=1}^{N} \frac{w_n}{W_N} \mathbf{1}_A(n).$$

Then, for each N,

$$\begin{aligned}\mathbf{1}_A(N) - S_N &= \frac{W_N S_N - W_{N-1} S_{N-1}}{w_N} - \frac{W_N - W_{N-1}}{w_N} S_N \\ &= \frac{W_{N-1}}{w_N}(S_N - S_{N-1}).\end{aligned}$$

We can conclude using Cauchy criterion for convergence because the sequence $(W_{N-1}/w_N)_N$ is bounded and the sequence $(\mathbf{1}_A(N))_N$ cannot converge unless A is finite or cofinite.

REFERENCES

1. F. Benford, *The law of anomalous numbers*, Proc. Amer. Philos. Soc. **78** (1938), 551–572.

2. P. Diaconis, *The distribution of leading digits and uniform distribution* mod 1, Annals Probab. **5** (1977), 72–81.

3. ———, *Examples for the theory of infinite iteration of summability methods*, Canad. J. Math. **29** (1977), 489–497.

4. R.L. Duncan, *Note on the initial digit problem*, Fibonacci Quart. **7** (1969), 474–475.

5. J.P. Duran, *Almost convergence, summability and ergodicity*, Canad. J. Math. **26** (1974), 372–387.

6. S. Eliahou, B. Massé and D. Schneider, *Benford's law and the mantissa distribution of natural and prime number powers*, Tech. Rept. L.M.P.A. **447**, 2011.

7. B.J. Flehinger, *On the probability that a random integer has initial digit A*, Amer. Math. Month. **73** (1966), 1056–1061.

8. A. Fuchs and G. Letta, *Le problème du premier chiffre décimal pour les nombres premiers*, Electr. J. Combin. **3** (1996), Research paper 25 (electronic).

9. R. Giuliano-Antonini, G. Grekos and L. Mišík, *On weighted densities*, Czech. Math. J. **57** (2007), 947–962.

10. G. Grekos, *On various definitions of density* (*survey*), Tatra Mountains Math. Publ. **31** (2005), 17–27.

11. R. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1976), 1609–1625.

12. G.H. Hardy, *Divergent Series*, Oxford University Press, Oxford, 1949.

13. E. Janvresse and T. de la Rue, *From uniform distributions to Benford's law*, J. Appl. Probab. **41** (2004), 1203–1210.

14. S. Kanemitsu, K. Nagasaka, G. Rauzy and J-S. Shiue, *On Benford's law: The first digit problem*, Lect. Notes Math. **1299**, 158–169, Springer, Berlin, 1988.

15. D.E. Knuth, *The art of computer programming*, volume 2, Addison-Wesley, Reading, Massachusetts, 1968.

16. L. Kuipers and H. Niederreiter , *Uniform distribution of sequences*, Dover Publications, New York, 2006.

17. S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39–40.

18. M.J. Nigrini and L.J. Mittermaier, *The use of Benford's law as an aid in analytical procedures*, Auditing: A Journal of Practice and Theory **16** (1997), 52–57.

19. P.N. Posch, *A survey of sequences and distribution functions satisfying the first-digit-law*, J. Statist. Manage. Syst. **11** (2008), 1–19.

20. P. Schatte, *Some estimates of the H_∞-uniform distribution*, Monats. Math. **103** (1987), 233–240.

21. J-P. Serre, *A course in arithmetic*, Springer, New York, 1996.

22. G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambr. Stud. Adv. Math. **46**, Cambridge University Press, Cambridge, 1995.

23. R.E. Whitney, *Initial digits for the sequence of primes*, Amer. Math. Month. **79** (1972), 150–152.

24. A. Wintner, *On the cyclical distribution of the logarithms of the prime numbers*, Quart. J. Math. **6** (1935), 65–68.

Univ. Lille Nord de France, F-59000 Lille, France; ULCO, LMPA J. Liouville, B.P. 699, F-62228 Calais, France and CNRS, FR 2956, France
Email address: bmasse@lmpa.univ-littoral.fr

Univ. Lille Nord de France, F-59000 Lille, France; ULCO, LMPA J. Liouville, B.P. 699, F-62228 Calais, France and CNRS, FR 2956, France
Email address: dominique.schneider@lmpa.univ-littoral.fr