Electronic Journal of Probability

Electron. J. Probab. **17** (2012), no. 86, 1–17.
ISSN: 1083-6489 DOI: 10.1214/EJP.v17-1900

Random number sequences and the first digit phenomenon

Bruno Massé* Dominique Schneider*

Abstract

The sequences of mantissa of positive integers and of prime numbers are known not to be distributed as Benford's law in the sense of the natural density. We show that we can correct this defect by selecting the integers or the primes by means of an adequate random process and we investigate the rate of convergence. Our main tools are uniform bounds for deterministic and random trigonometric polynomials. We then adapt the random process to prove the same result for logarithms and iterated logarithms of integers. Finally we show that, in many cases, the mantissa law of the nth randomly selected term converges weakly to the Benford's law.

Keywords: Benford's law ; weak convergence ; mantissa ; density.
AMS MSC 2010: Primary 60B10, Secondary 11B05 ; 11K99.
Submitted to EJP on March 23, 2012, final version accepted on September 29, 2012.

1 Introduction and definitions

We fix a numeration base $b > 1$ and denote the logarithm in base b by $\log_b$ and the set of positive integers by $\mathbb{N}^*$. Many sequences $(u_n)_n$ of positive numbers, like $u_n = 2^n$ (if b is not 2, 4 and so on), $u_n = n!$, $u_n = n^n$, $u_n = F_n$ where F_n is the nth Fibonacci number, are known to verify the so-called *first digit phenomenon*. This means that, if $FD(u_n)$ denotes the first digit of u_n, the natural density of $\Delta_k^u = \{n \in \mathbb{N}^* : FD(u_n) = k\}$ is $\log_b\left(\frac{k+1}{k}\right)$, that is to say

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\Delta_k^u}(n) = \log_b\left(\frac{k+1}{k}\right) \qquad (k = 1, \ldots, b)$$

(here and in the sequel, $\mathbb{1}_B$ 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 when $b = 10$. As Newcomb [15] and Benford [1] noticed, this occurs more or less in many real-life sets of data and this is why this phenomenon is used in fraud detection [16] and computer design [12, 9] (roundoff error estimation and data storage). Some considerations about scale-invariance have led Newcomb and Benford to use the so-called *Benford's law* μ_b (defined below) to depict this phenomenon.

*Université du Littotal Côte d'Opale, France. E-mail: bmasse,schneider@lmpa.univ-littoral.fr

The *mantissa* in base b of a positive real number x is the unique number $\mathcal{M}_b(x)$ in $[1, b[$ such that there exists an integer k verifying $x = \mathcal{M}(x)b^k$ (there exists another definition of the mantissa, but for technical reasons we shall use this one). Of course studying the mantissa of a number is more precise than studying its first digit. To state the properties concerning the distribution of the mantissa of the above sequences, we need more definitions: the Benford's law in base b is the probability measure μ_b on the interval $[1, b[$ defined by

$$\mu_b([1, t[\,) = \log_b t \qquad (1 \le t < b)\,.$$

A sequence $(v_n)_n$ of real numbers in $[1, b[$ is called *natural-Benford* in base b if it is naturally distributed as μ_b, that is to say if

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{[1,\,t[}(v_n) = \log_b t \qquad (1 \le t < b)\,.$$

The above formula means that, for each $t \in [1, b[\,$, the set $\{n \in \mathbb{N}^* : 1 \le v_n < t\}$ admits a natural density and its natural density is $\log_b t$ and this can be interpreted as the weak convergence of the sequence of probability measures $(1/N)\sum_1^N \delta_{v_n}$ to μ_b as $N \to +\infty$ (δ_x denotes the Dirac measure at point x).

A sequence $(u_n)_n$ of positive numbers is also called natural-Benford in base b when the sequence of mantissae $(\mathcal{M}_b(u_n))_n$ is natural-Benford. We can now say that the sequences $(2^n)_n$ (if b is not 2 and so on), $(n!)_n$, $(n^n)_n$ and $(F_n)_n$ are all natural-Benford.

When $u_n = n$ or $u_n = p_n$ (p_n is the nth prime number) and $b = 10$,

$$\liminf_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\Delta_1^u}(n) = \frac{1}{9} \quad \text{and} \quad \limsup_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\Delta_1^u}(n) = \frac{5}{9} \tag{1.1}$$

(see [8] and [20]). So these two sequences do not verify the first digit phenomenon in base 10 in the sense of the natural density (in fact they do not verify this phenomenon in any base and so they are not natural-Benford in any base). From [6], we know that this phenomenon is verified by $u_n = n$ in the sense of the logarithmic density, that is is to say

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

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. The defect in equation (1.1) is corrected by assigning lighter weights to large numbers. The calculations in [8] can be adapted to prove the same property for $u_n = p_n$ and a more general statement (also proved in [7]): $(n)_n$ and $(p_n)_n$ are *logarithmic-Benford* in any base b which means that, when $u_n = n$ or $u_n = p_n$, we have

$$\lim_{N\to+\infty} \frac{1}{\log N}\sum_{n=1}^{N} \frac{1}{n}\mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \log_b t \qquad (1 \le t < b)\,.$$

Some convergence rates are given in [7].

1.1 Main notations and results

We consider a sequence $(q_n)_n$ of numbers in $[0, 1]$ summing to infinity and a sequence $(X_n)_n$ of independent Bernoulli random variables such that $P(X_n = 1) = q_n$. By the Borel-Cantelli lemma, $\sum_{n=1}^{+\infty} X_n = +\infty$ almost surely (a.s. in abbreviated form), so we can suppose, without loss of generality, that $\sum_{n=1}^{+\infty} X_n = +\infty$ everywhere. The

nth number in the random set $\{k : X_k = 1\}$ is denoted Y_n (in other words, Y_n is the number of trials needed to get n successes). We shall consistently use the following notations through the paper: whenever $(u_n)_n$ is a given sequence of positive numbers, then $(U_n)_n = (u_{Y_n})_n$ will denote the random subsequence of $(u_n)_n$ whose terms are selected by means of the X_n. Moreover we set $\pi(N) = \sum_{n=1}^{N} X_n$ (of course $\pi(Y_N) = N$) and

$$A_N^t = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(U_n)) = \frac{1}{\pi(Y_N)}\sum_{n=1}^{Y_N} X_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)).$$

So the random variable A_N^t is the frequency of the U_n, among $U_1, \ldots, U_N$, whose mantissa is less than t. We want to find conditions on $(u_n)_n$ and $(q_n)_n$ ensuring that a.s. $\lim_N A_N^t = \log_b t$ for every $t \in [1, b[$, that is to say: ensuring that the random sequence $(U_n)_n$ is a.s. natural-Benford.

We also look for conditions ensuring that the law of the random variable $\mathcal{M}_b(U_n)$ converges weakly to μ_b as n tends to infinity.

In [11], it is proved that if $u_n = n$ and $q_n = 1/n$, then $(U_n)_n$ is a.s. natural-Benford. In Section 2, we extend this property to a larger class of probabilities q_n and to the case $u_n = p_n$. We investigate the rates of convergence in the cases $u_n = n$, $u_n = n \log n$ and $u_n = p_n$ for $q_n = (\log n)^\delta/n$ with $\delta > 0$. Our techniques are totally different from those of [11].

By theorem 2.6 in [13, p. 15] and direct calculations it is easily seen that $(\log n)_n$ and $(\log\log n)_n$ are not natural-Benford because the sequences of their logarithms in base b are not uniformly distributed modulo 1 (see Section 2 for the link between natural-Benford sequences and sequences which are uniformly distributed modulo 1). In Section 3, we prove that $(\log n)_n$ and $(\log\log n)_n$ are not logarithmic-Benford either and that $(U_n)_n$ is a.s. natural-Benford when $u_n = \log n$ and $q_n = 1/(n \log n)$ and when $u_n = \log\log n$ and $q_n = 1/((n \log n)(\log\log n))$.

In Section 4, we prove that, in many situations, if the law of $\mathcal{M}_b(U_n)$ converges weakly, then the limit must be the Benford's law. We also prove that if the sequence $(nq_n)_n$ is nonincreasing, then this law converges actually to the Benford's law in the case $u_n = n$ and, under additional conditions on q_n, in the cases $u_n = n \log n$, $u_n = n \log\log n$ and $u_n = p_n$.

1.2 Weighted densities

The above definitions use tacitly the notion of weighted density of a subset of $\mathbb{N}^*$. We recall below the definition and some useful facts.

Let $(w_n)_n$ 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 $\Delta \subset \mathbb{N}^*$ has a w_n*-density* when the sequence $\left(\sum_{n=1}^{N} \frac{w_n}{W_N} \mathbb{1}_\Delta(n)\right)_N$ converges, and in that case its limit is called the w_n-density of Δ. This is the limit of the weighted frequency of the elements of Δ among those of $\mathbb{N}^*$. In view of this definition, a sequence $(u_n)_n$ is natural-Benford (respectively logarithmic-Benford) when, for all $t \in [1, b[$, the set $\{n : \mathcal{M}_b(u_n) < t\}$ admits a 1-density (respectively a $(1/n)$-density) equal to $\log_b(t)$.

Another sequence $(v_n)_n$ of positive real numbers summing to infinity being given, we say that the w_n-density is *stronger* than the v_n-density when the existence of a w_n-density for $\Delta \subset \mathbb{N}^*$ implies the existence of a v_n-density and that the two densities are equal (as pointed out by an anonymous referee, some authors adopt the reverse point of view, and say that a density is stronger than another if it allows to define the density of more subsets of $\mathbb{N}^*$). If each density is stronger than the other one, then the two densities are said to be *equivalent*. It is known [13, 14] that the 1-density is strictly stronger than the $(1/n)$-density, which is strictly stronger than the $(1/n\log n)$-density, and so on, and that all the n^δ-densities with $\delta > -1$ are equivalent, all the the $((\log n)^\delta/n)$-densities with $\delta > -1$ are equivalent, all the the $((\log\log n)^\delta/n\log n)$-densities with $\delta > -1$ are equivalent, and so on. Moreover the $(1/n\log n)$-density and the $(1/p_n)$-density are equivalent.

In particular, if a sequence is natural-Benford, then it is logarithmic-Benford. The converse is false. Moreover, if a sequence $(u_n)_n$ is logarithmic-Benford and is not natural-Benford, then the sets $\{n : \mathcal{M}_b(u_n) < t\}$ do not admit any 1-density.

2 Random integers and primes

For every $x \in \mathbb{R}$ and $h \in \mathbb{Z}^*$, we set $e_h(x) = \exp(2i\pi hx)$ where $i^2 = -1$.

2.1 Uniform distribution modulo 1 and Weyl criterion

Here and in the sequel, the fractional part of x will be denoted $\{x\}$. A sequence $(a_n)_n$ of real numbers is said to be uniformly distributed modulo 1 when, for every $s \in [0, 1[$,

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{[0,\,s[}(\{a_n\}) = s.$$

By the Weyl criterion (see [13, p. 7] or [5, p. 15]), this happens if and only if, for every $h \in \mathbb{Z}^*$,

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} e_h(a_n) = 0\,.$$

And, since $\log_b y$ and $\log_b(\mathcal{M}_b(y))$ are equal modulo 1, the following lemma holds.

Lemma 2.1. *A sequence $(v_n)_n$ of positive numbers is natural-Benford if and only if, for every $h \in \mathbb{Z}^*$,*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} e_h(\log_b v_n) = 0\,.$$

2.2 Random sequences which are a.s. Benford

Aiming at simplicity and clarity, we state the following theorems only for sequences $(q_n)_n$ such that $(nq_n)_n$ is monotonic. The sufficient conditions obtained in the proofs are more general. Our statements imply that, if $u_n = n$ or $u_n = n\log n$ or $u_n = p_n$, the random sequence $(U_n)_n$ is a.s. natural-benford in particular when $q_n = (\log n)^\delta/n$, $q_n = (\log\log n)^\delta/(n\log n)$, $q_n = (\log\log\log n)^\delta/(n(\log n)(\log\log n))$ and so on, with $\delta \geq -1$.

The next lemma can be proved by combining two famous estimates, namely Lemma 4.10 in [19, p. 76] and Lemma 2.43 in [5, p. 253] (see lemma 8 in [7] for details). It gives a uniform bound for some trigonometric polynomial.

Lemma 2.2. *There exists a constant C_0 (depending only on b) such that, for every integer $N \geq 1$ and every $h \in \mathbb{Z}^*$,*

$$\left|\sum_{n=1}^{N} \frac{e_h(\log_b n)}{n}\right| \leq C_0 + \log|h|.$$

The following lemma characterizes the random sequences $(U_n)_n$ (where $U_n = u_{Y_n}$) which are a.s. natural-Benford by means of conditions on q_n and u_n.

Lemma 2.3. *The sequence $(U_n)_n$ is a.s. natural-Benford if and only if, for every $h \in \mathbb{Z}^*$,*

$$\lim_{N\to+\infty} \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n e_h(\log_b u_n) = 0. \tag{2.1}$$

Proof. For $h \in \mathbb{Z}^*$ and $N \in \mathbb{N}^*$, set

$$F_N^h = \frac{1}{N}\sum_{n=1}^N e_h(\log_b U_n) = \frac{1}{\pi(Y_N)}\sum_{n=1}^{Y_N} X_n e_h(\log_b u_n)$$

and

$$G_N^h = \frac{1}{\pi(N)}\sum_{n=1}^N X_n e_h(\log_b u_n)\,.$$

By lemma 2.1, $(U_n)_n$ is a.s. natural-Benford if and only if

$$\text{a.s. } \forall\, h \in \mathbb{Z}^*\,,\ \lim_{N\to+\infty} F_N^h = 0\,. \tag{2.2}$$

So, to prove our lemma, it suffices to show that

$$\forall\, \omega \in \Omega\,,\ \forall\, h \in \mathbb{Z}^*\,,\ \left(\lim_{N\to+\infty} F_N^h(\omega) = 0\right) \Longleftrightarrow \left(\lim_{N\to+\infty} G_N^h(\omega) = 0\right) \tag{2.3}$$

and that (2.1) is equivalent to

$$\text{a.s. } \forall\, h \in \mathbb{Z}^*\,,\ \lim_{N\to+\infty} G_N^h = 0\,. \tag{2.4}$$

Fix $\omega \in \Omega$ and $h \in \mathbb{Z}^*$. The converse part of (2.3) is evident because $(F_N^h(\omega))_N$ is a subsequence of $(G_N^h(\omega))_N$. For every $N \in \mathbb{N}^*$, there exists a unique $M(\omega)$ such that $Y_{M(\omega)}(\omega) \le N < Y_{M(\omega)+1}(\omega)$ and we get $G_N^h(\omega) = F_{M(\omega)}^h(\omega)$ because either $N = Y_{M(\omega)}$ or $X_n(\omega) = 0$ for $Y_{M(\omega)} < n \le N$. This proves the direct part of (2.3) since $M(\omega) \to +\infty$ as $N \to +\infty$.

Now we fix $h \in \mathbb{Z}^*$ and we consider the two martingales $(S_N)_N$ and $(S_N^*)_N$ defined by

$$S_N = \sum_{n=1}^N \frac{X_n - q_n}{s_n} \quad \text{and} \quad S_N^* = \sum_{n=1}^N \frac{(X_n - q_n)e_h(\log_b u_n)}{s_n}$$

where $s_n = \sum_{k=1}^n q_k$. Note that, for $n \ge 2$,

$$\frac{q_n(1-q_n)}{s_n^2} \le \frac{q_n}{s_n s_{n-1}} = \frac{1}{s_{n-1}} - \frac{1}{s_n}\,.$$

Hence $\mathbb{E}(S_N^2) \le 2/q_1$ and $\mathbb{E}(|S_N^*|^2) \le 2/q_1$ since the X_n are independent. So $(S_N)_N$ and $(S_N^*)_N$ are a.s. convergent by the second Doob's classical martingale convergence theorem. By Kronecker lemma, the convergence of $(S_N)_N$ implies

$$\lim_{N\to+\infty} \frac{1}{s_N}\sum_{n=1}^N (X_n - q_n) = 0 \ \text{ a.s.}$$

which implies

$$\lim_{N\to+\infty} \frac{\pi(N)}{s_N} = 1 \ \text{ a.s.} \tag{2.5}$$

and the convergence of $(S_N^*)_N$ implies

$$\lim_{N\to+\infty} \frac{1}{s_N}\sum_{n=1}^N (X_n - q_n)e_h(\log_b u_n) = 0 \ \text{ a.s..} \tag{2.6}$$

This completes the proof because (2.5) and (2.6) prove that (2.1) and (2.4) are equivalent. □

We are now able to treat the case $u_n = n$.

Theorem 2.4. *Let $u_n = n$. If $(nq_n)_n$ is non-increasing or if $(nq_n)_n$ is non-decreasing and*

$$\lim_{N\to+\infty} \frac{Nq_N}{\sum_{n=1}^N q_n} = 0,$$

then $(U_n)_n$ is a.s. natural-Benford.

Proof. Fix $h \in \mathbb{Z}^*$ and $N \geq 1$. Then, by Abel's transformation,

$$\sum_{n=1}^N q_n e_h(\log_b n) = Nq_N \sum_{k=1}^N \frac{e_h(\log_b k)}{k} + \sum_{n=1}^{N-1} (nq_n - (n+1)q_{n+1}) \sum_{k=1}^n \frac{e_h(\log_b k)}{k}.$$

So, by lemma 2.2,

$$\lim_{N\to+\infty} \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n e_h(\log_b n) = 0$$

when

$$\lim_{N\to+\infty} \frac{Nq_N}{\sum_{n=1}^N q_n} = 0 \quad\text{and}\quad \lim_{N\to+\infty} \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^{N-1} |nq_n - (n+1)q_{n+1}| = 0.$$

We can now conclude with lemma 2.3. □

To treat the cases $u_n = n\log n$ and $u_n = p_n$, we need to estimate another trigonometric polynomial.

Lemma 2.5. *There exists a constant C (depending only on b) such that, for every integer $N \geq 3$ and every $h \in \mathbb{Z}^*$,*

$$\left| \sum_{n=2}^N \frac{e_h(\log_b(n\log n))}{n} \right| \leq C|h|\log|h|\log\log N.$$

Proof. Fix h and N. Then, by Abel's transformation,

$$\sum_{n=2}^N \frac{e_h(\log_b(n\log n))}{n} = a_N \sum_{k=2}^N \frac{e_h(\log_b k)}{k} + \sum_{n=2}^{N-1} (a_n - a_{n+1}) \sum_{k=2}^n \frac{e_h(\log_b k)}{k}$$

where $a_n = e_h(\log_b \log n)$. By the mean value theorem, for every $n \geq 2$,

$$|a_n - a_{n+1}| \leq \frac{2\pi|h|}{\log b} \frac{1}{n\log n}.$$

We can conclude by using lemma 2.2. □

We are now able to treat the cases $u_n = n\log n$ and $u_n = p_n$.

Theorem 2.6. *Let $u_n = n\log n$ or $u_n = p_n$. If $(nq_n)_n$ is monotonic and*

$$\lim_{N\to+\infty} \frac{N(\log\log N)q_N}{\sum_{n=1}^N q_n} = 0,$$

then $(U_n)_n$ is a.s. natural-Benford.

Proof. Replacing $\log_b n$ by $\log_b(n\log n)$ and lemma 2.2 by lemma 2.5 in the proof of theorem 2.4 shows that, if $u_n = n\log n$, $(U_n)_n$ is a.s. natural-Benford when

$$\lim_{N\to+\infty} \frac{N(\log\log N)q_N}{\sum_{n=1}^N q_n} = 0 \text{ and } \lim_{N\to+\infty} \frac{\log\log N}{\sum_{n=1}^N q_n} \sum_{n=1}^{N-1} |nq_n - (n+1)q_{n+1}| = 0$$

and this concludes the proof in the case $u_n = n\log n$.

The mean value theorem shows that, for every $h \in \mathbb{Z}^*$ and $N \geq 1$,

$$\left| \sum_{n=1}^N q_n e_h(\log_b p_n) - \sum_{n=1}^N q_n e_h(\log_b(n\log n)) \right| \leq 2\pi|h| \sum_{n=1}^N q_n \left| \log_b \frac{p_n}{n\log n} \right|.$$

Since $p_n \sim n\log n$, this and the theorem of Cesàro prove that, for a given sequence $(q_n)_n$, the hypotheses of lemma 2.3 are verified by $u_n = p_n$ if and only if they are verified by $u_n = n\log n$. This completes the proof. $\square$

2.3 Rate of convergence

It is natural to look for an estimation of the rate of convergence in the limit involved in the definition of a Benford sequence of numbers. So we are seeking a.s. bounds for

$$d_N(U) = \sup_{1<c<d<b} \left| \left(\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{[c,\,d[}(\mathcal{M}_b(U_n)) \right) - \log_b \frac{d}{c} \right|$$

which is a distance, similar to the Kolmogorov one, between the Benford law μ_b and $\frac{1}{N}\sum_{n=1}^N \delta_{\mathcal{M}_b(U_n)}$. We shall treat the cases $q_n = \frac{(\log n)^\delta}{n}$ ($\delta > 0$) and $u_n = n$, $u_n = n\log n$ and $u_n = p_n$. Our methods seem inefficient when $q_n \leq \frac{1}{n}$.

The so-called *discrepancy* of a sequence $(v_n)_n$ of real numbers is defined by

$$D_N(v) = \sup_{0<x<y<1} \left| \left(\frac{1}{N} \sum_{n=1}^N \mathbb{1}_{[x,\,y[}(\{v_n\}) \right) - (y-x) \right|.$$

This is the distance between the Lebesgue measure on $[0,1[$ and $\frac{1}{N}\sum_{n=1}^N \delta_{\{v_n\}}$.

If $v_n = \log_b u_n$, then $d_N(u) = D_N(v)$ because $\log_b(\mathcal{M}_b(x)) = \{\log_b x\}$. Considering the disprepancy will permit us to use the following lemma, known as the *Erdős-Turán inequality* and available, among many other sources, in [18].

Lemma 2.7. *Let $(v_n)_n$ be a sequence of real numbers and let N be a natural number. Then, for every natural number H, we have*

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

The second lemma is in the wake of lemmas 2.2 and 2.5. We denote $\mathcal{O}$ the standard big O of Landau.

Lemma 2.8. *There exists a constant C (depending only on b) such that, for every integer $N \geq 3$ and every $h \in \mathbb{Z}^*$,*

$$\left| \sum_{n=1}^N \frac{e_h(\log_b(p_n))}{n} \right| \leq C|h|\log|h|(\log\log N)^2.$$

Proof. The celebrated relation

$$p_n = n\log n + \mathcal{O}(n\log\log n)$$

and the mean value theorem imply

$$\begin{aligned}\left|\sum_{n=3}^{N}\frac{e_h(\log_b p_n)}{n}-\frac{e_h(\log_b(n\log n))}{n}\right| &= \mathcal{O}\left(\frac{|h|}{\log b}\sum_{n=3}^{N}\frac{\log\log n}{n\log n}\right)\\ &= \mathcal{O}\left(\frac{|h|}{\log b}(\log\log N)^2\right).\end{aligned}$$

Lemma 2.5 completes the proof. □

The last lemma gives an estimation of some random trigonometric polynomial. Recall from Section 1.1 that Y_N is the nth randomly selected integer.

Lemma 2.9. *If $v_n = \log_b n$ or $v_n = \log_b(n\log n)$ or $v_n = \log_b p_n$ there exists a positive a.s. finite random variable C such that, for all positive integers N and h,*

$$\left|\sum_{n=1}^{Y_N}(X_n-q_n)e_h(v_n)\right| \le C\sqrt{\log(1+2\pi h)\log(Y_N)\sum_{n=1}^{Y_N}q_n}.$$

Proof. Applying theorem 1.1 (2) (ii) in [3] to the sequence of independent centered random variables $(X_n-q_n)_n$ yields the existence of a positive a.s. finite random variable C_1 such that

$$\left|\sum_{n=1}^{Y_N}(X_n-q_n)e_h(v_n)\right|^2 \le C_1\log(1+2\pi h)\log(\max(Y_N,v_{Y_N}))\sum_{n=1}^{Y_N}((X_n-q_n)^2+\mathbb{E}(X_n-q_n)^2)$$

for all positive integers N and h. This proves our statement because, by (2.5),

$$(X_n-q_n)^2+\mathbb{E}(X_n-q_n)^2 = |X_n-q_n| \le C_2\, q_n$$

and because $Y_N \ge v_{Y_N}$ when $v_n = \log_b n$ or $v_n = \log_b(n\log n)$ or $v_n = \log_b p_n$. □

We can now give an estimation of the rate of convergence in theorems 2.4 and 2.6 when $u_n = n$, $u_n = n\log n$ or $u_n = p_n$ and for a large family of sequences of probabilities $(q_n)_n$.

Theorem 2.10. *Let $\delta > 0$ and $q_n = \dfrac{(\log n)^\delta}{n}$. Then*

$$d_N(U) = \mathcal{O}\left(\frac{(\log N)^{\frac{3}{2}}}{N^\beta}\right) \quad \text{a.s.}$$

where $\beta = \dfrac{\min(\delta,2)}{2(\delta+1)}$ if $u_n = n$ and $\beta = \dfrac{\min(\delta,1)}{2(\delta+1)}$ if $u_n = n\log n$ or $u_n = p_n$.

Proof. In this proof, C denotes a positive a.s. finite random variable which may vary from line to line. Here $u_n = n$ or $u_n = n\log n$ or $u_n = p_n$. By lemma 2.9,

$$\begin{aligned}\left|\sum_{n=1}^{N}e_h(\log_b U_n)\right| &= \sum_{n=1}^{Y_N}X_n e_h(\log_b u_n)\\ &\le C\sqrt{\log(1+2\pi h)\log(Y_N)\sum_{n=1}^{Y_N}q_n}+\left|\sum_{n=1}^{Y_N}q_n e_h(\log_b u_n)\right| \quad (N\ge 1).\end{aligned}$$

Since $q_n = \dfrac{(\log n)^\delta}{n}$, (2.5) yields

$$\log Y_N \le CN^{\frac{1}{\delta+1}} \quad \text{and} \quad Y_N q_{Y_N} \le CN^{\frac{\delta}{\delta+1}}. \tag{2.7}$$

Hence, for every $H \ge 1$ and $N \ge 1$,

$$\sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{N} e_h(\log_b U_n) \right| \le C(\log H)^{\frac{3}{2}} N^{\frac{\delta+2}{2(\delta+1)}} + \sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{Y_N} q_n e_h(\log_b u_n) \right| \tag{2.8}$$

because, by (2.5) again, $\sum_{n=1}^{Y_N} q_n \sim N$ a.s..

Using the Abel transformation, lemma 2.2, lemma 2.5, lemma 2.8 and (2.7), we get for every H and N

$$\sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{Y_N} q_n e_h(\log_b n) \right| \le CN^{\frac{\delta}{\delta+1}} (\log H)^2, \tag{2.9}$$

$$\sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{Y_N} q_n e_h(\log_b(n \log n)) \right| \le CN^{\frac{\delta}{\delta+1}} (\log\log N) H \log H \tag{2.10}$$

and

$$\sum_{h=1}^{H} \frac{1}{h} \left| \sum_{n=1}^{Y_N} q_n e_h(\log_b p_n) \right| \le CN^{\frac{\delta}{\delta+1}} (\log\log N)^2 H \log H. \tag{2.11}$$

When $u_n = n$, by choosing $H = [N^\alpha]$ with $\alpha = \frac{\delta}{2(\delta+1)}$ if $\delta \le 2$ and $\alpha = \frac{1}{(\delta+1)}$ if $\delta > 2$ in (2.8) and (2.9), we get

$$\frac{1}{H+1} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e_h(\log_b U_n) \right| \le C \frac{(\log N)^{\frac{3}{2}}}{N^\beta}$$

with $\beta = \frac{\min(\delta,2)}{2(\delta+1)}$.

When $u_n = n \log n$ or $u_n = p_n$, by choosing $H = [N^\alpha]$ with $\alpha = \frac{1}{2(\delta+1)}$ in (2.8) and (2.10) or (2.11), we get

$$\frac{1}{H+1} + \sum_{h=1}^{H} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e_h(\log_b U_n) \right| \le C \frac{(\log N)^{\frac{3}{2}}}{N^\beta}$$

with $\beta = \frac{\min(\delta,1)}{2(\delta+1)}$.

Lemma 2.7 completes the proof.

□

3 Random logarithms

The techniques used in Section 2.2 can be adapted to the case $u_n = \log n$, but they do not give useful results. Indeed, we can adapt lemma 2.5 and prove

$$\left| \sum_{n=2}^{N} \frac{e_h(\log_b(\log n))}{n} \right| \le C|h| \log|h| \log N$$

and then adapt theorem 2.6 and prove that, if $u_n = \log n$, $(U_n)_n$ is a.s. natural-Benford when

$$\lim_{N \to +\infty} \frac{N q_N \log N}{\sum_{n=1}^{N} q_n} = 0 \text{ and } \lim_{N \to +\infty} \frac{\log N}{\sum_{n=1}^{N} q_n} = 0.$$

But these two conditions seem incompatible at least for the kind of sequences $(q_n)_n$ we are interested in: $q_n = 1/n^\alpha$ with $\alpha \in]0,1[$, $q_n = (\log n)^\delta/n$, $q_n = (\log\log n)^\delta/(n\log n)$, $q_n = (\log\log\log n)^\delta/(n(\log n)(\log\log n))$ and so on, with $\delta \geq -1$.

Hence we must use different techniques when u_n is equal to $\log n$ or $\log\log n$. The following lemma is quite similar to lemma 2.3. It states that the random sequence $(u_{Y_n})_n$ is a.s. natural-Benford if and only if the set $\{n : \mathcal{M}_b(u_n) < t\}$ has q_n-density $\log_b(t)$ for any $t \in [1, b[$ (see Section 1.2).

Lemma 3.1. *The random sequence $(U_n)_n$ is a.s. natural-Benford if and only if*

$$\forall\, t \in [1, b[\,,\ \lim_{N\to+\infty} \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \log_b t.$$

Proof. For $t \in [1, b[$ and $N \in \mathbb{N}^*$, set

$$B_N^t = \frac{1}{\pi(N)} \sum_{n=1}^N X_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)).$$

The sequences $(A_N^t(\omega))_N$ (see Section 1.1 for definition) and $(B_N^t(\omega))_N$ have the same role as $(F_N^h(\omega))_N$ and $(G_N^h(\omega))_N$ in the proof of lemma 2.3 and the arguments we used there prove the following equivalence

$$\forall\,\omega \in \Omega\,,\ \forall\, t \in [1, b[\,,\ \left(\lim_{N\to+\infty} A_N^t(\omega) = 0\right) \Longleftrightarrow \left(\lim_{N\to+\infty} B_N^t(\omega) = 0\right).$$

Since we already know that

$$\lim_{N\to+\infty} \frac{\pi(N)}{s_N} = 1 \ \text{ a.s.}$$

where $s_n = \sum_{k=1}^n q_k$ and since $\log_b$ is continuous, we only need to prove that, for every $t \in [1, b[$,

$$\lim_{N\to+\infty} \frac{1}{s_N} \sum_{n=1}^N (X_n - q_n) \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = 0 \ \text{ a.s..}$$

But this follows from the Doob theorem and the Kronecker lemma applied to the martingale $(S_N^{**})_N$ defined by

$$S_N^{**} = \sum_{n=1}^N \frac{(X_n - q_n)}{s_n} \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)).$$

□

The following general lemma happens to be very useful when $u_n = f(n)$ and $q_n = g(n)$ with f and g such that the inverse of f and a primitive integral of g are explicitly known. This is only a restatement in the context of mantissa distribution of techniques that Fuchs and Letta have used in [8] to study the digit distribution. Recall that the q_n are positive numbers summing to infinity.

Lemma 3.2. *Fix a real $t \in [1, b[$. Suppose that $(u_n)_n$ is increasing and unbounded. For $m \in \mathbb{N}$, set*

$$C_m = \{n : b^m \leq u_n < tb^m\}\,,\ D_m = \{n : b^{m-1} \leq u_n < b^m\}$$

and

$$E_m = \{n : tb^{m-1} \leq u_n < tb^m\}\,.$$

Then

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \lim_{m\to+\infty} \frac{\sum_{n\in C_{m-1}} q_n}{\sum_{n\in\, D_m} q_n}$$

and

$$\overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \lim_{m\to+\infty} \frac{\sum_{n\in\, C_m} q_n}{\sum_{n\in\, E_m} q_n}$$

provided that the two limits exist.

Proof. First note that the sequence

$$\left(\frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) \right)_N$$

is increasing on the sets C_m and decreasing on the sets $\{n : tb^{m-1} \le u_n < b^m\}$. Then

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \lim_{M\to+\infty} \frac{\sum_{m=1}^M \sum_{n\in C_{m-1}} q_n}{\sum_{m=0}^M \sum_{n\in D_m} q_n}$$

and

$$\overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \lim_{M\to+\infty} \frac{\sum_{m=1}^M \sum_{n\in C_m} q_n}{\sum_{m=0}^M \sum_{n\in E_m} q_n}.$$

But, since the numbers q_n sum to infinity, so do the numbers $\alpha_m = \sum_{n\in D_m} q_n$ and the numbers $\beta_m = \sum_{n\in E_m} q_n$. So, when the limits

$$\lim_{m\to+\infty} \frac{\sum_{n\in C_{m-1}} q_n}{\sum_{n\in\, D_m} q_n} \quad \text{and} \quad \lim_{m\to+\infty} \frac{\sum_{n\in\, C_m} q_n}{\sum_{n\in\, E_m} q_n}$$

exist, they are respectively equal to

$$\lim_{M\to+\infty} \frac{\sum_{m=1}^M \sum_{n\in C_{m-1}} q_n}{\sum_{m=0}^M \sum_{n\in D_m} q_n} \quad \text{and} \quad \lim_{M\to+\infty} \frac{\sum_{m=1}^M \sum_{n\in C_m} q_n}{\sum_{m=0}^M \sum_{n\in E_m} q_n}$$

by the Stolz-Cesàro theorem. □

We have already seen in Section 1.1 that the deterministic sequences $(\log n)_n$ and $(\log\log n)_n$ are not natural-Benford. We prove now that they are not logarithmic-Benford either.

Proposition 3.3. *The sequences* $(\log n)_n$ *and* $(\log\log n)_n$ *are not logarithmic-Benford.*

Proof. We want to verify that, for some $t \in [1, b[$,

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) \neq \overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n))$$

when $q_n = 1/n$ and $u_n = \log n$ or $u_n = \log\log n$. Fix $t \in [1, b[$ and $q_n = 1/n$. If $u_n = \log n$, then $C_m = \{n : e^{b^m} \le n < e^{tb^m}\}$ and we can write D_m and E_m in the same way (C_m, D_m and E_m are defined in the proof of lemma 3.2). Direct calculations and lemma 3.2 give

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \frac{t-1}{b-1}$$

and

$$\overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \frac{b(t-1)}{t(b-1)}\,.$$

If $u_n = \log\log n$, then $C_m = \{n : e^{e^{b^m}} \leq n < e^{e^{tb^m}}\}$ and we can write D_m and E_m in the same way. We get

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = 0$$

and

$$\overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = 1\,.$$

□

We prove now that, if the numbers q_n are chosen in an adequate manner, the random sequence $(U_n)_n$ corresponding to $(\log n)_n$ or $(\log\log n)_n$ is a.s. natural-Benford.

Proposition 3.4. *If $u_n = \log n$ and $q_n = 1/(n\log n)$, then $(U_n)_n$ is a.s. natural-Benford.*

Proof. Set $u_n = \log n$ and $q_n = 1/(n\log n)$. By lemma 3.1 it suffices to verify that, for every $t \in [1, b[$,

$$\underline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \overline{\lim}_N \frac{1}{\sum_{n=1}^N q_n} \sum_{n=1}^N q_n \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(u_n)) = \log_b t.$$

Fix $t \in [1, b[$. Again $C_m = \{n : e^{b^m} \leq n < e^{tb^m}\}$ and, writing D_m and E_m in the same way, direct calculations give

$$\lim_{m\to+\infty} \sum_{n\in\, C_{m-1}} q_n = \lim_{m\to+\infty} \sum_{n\in\, C_m} q_n = \log t$$

and

$$\lim_{m\to+\infty} \sum_{n\in\, D_m} q_n = \lim_{m\to+\infty} \sum_{n\in\, E_m} q_n = \log b\,.$$

We now conclude with lemma 3.1 and lemma 3.2. □

A slight adaptation of the above arguments prove the following proposition.

Proposition 3.5. *If $u_n = \log\log n$ and $q_n = 1/((n\log n)(\log\log n))$, then $(U_n)_n$ is a.s. natural-Benford.*

Remark 3.6. *If $u_n = \log n$ or $u_n = \log\log n$ and if $(q_n)_n$ is non-decreasing or $q_n = 1/n^\alpha$ with $0 < \alpha < 1$ or $q_n = (\log n)^\delta/n$ with $\delta > -1$, then the random sequence $(U_n)_n$ is not a.s. natural-Benford, because, if it were, the sequence $(u_n)_n$ would be natural-Benford or logarithmic-Benford by the general properties of summation methods (see [10, p. 68] or [13, p. 63]) and lemma 3.1. The same arguments and the above propositions show that if $q_n = (n(\log n)(\log\log n))^{-1}$ or $q_n = (n(\log n)(\log\log n)(\log\log\log n))^{-1}$ and so on and if $u_n = \log n$ or $u_n = \log\log n$, then the random sequence $(U_n)_n$ is a.s. natural-Benford. Moreover, these techniques can be used to show that $(U_n)_n$ is not a.s. natural-Benford when $u_n = \log\log n$ and $q_n = 1/(n\log n)$. See [14] for a general treatement of weighted densities in connection with Benford's sequences of numbers.*

4 Mantissa limit law

For a sequence of positive random variables, there exists no general link between having a mantissa whose law converges to μ_b and being almost surely natural-Benford. Indeed, let $T_n = n!$ a.s.. Then the law of $\mathcal{M}_b(T_n)$ does not converge weakly, but the sequence $(T_n)_n$ is a.s. natural-Benford. Conversely, let $T_n = T$ where the law of T is μ_b. Then the law of $\mathcal{M}_b(T_n)$ converges weakly to μ_b, but the sequence $(T_n)_n$ is a.s. not natural-Benford.

Recall that $U_n = u_{Y_n}$ is the nth randomly selected term in the sequence $(u_n)_n$. In [11] it is proved that when $u_n = n$ and $q_n = 1/n$, the law of $\mathcal{M}_{10}(U_n)$ converges weakly to μ_{10} as n tends to infinity. We give below a broad generalization of this property. Our main tool is the following elementary lemma.

Lemma 4.1. *Let $(Z_n)_n$ be a sequence of positive random variables. Then the law of $\mathcal{M}_b(Z_n)$ converges weakly to μ_b as $n \to +\infty$ if and only if, for every $h \in \mathbb{Z}^*$,*

$$\lim_{n\to+\infty} \mathbb{E}\left[e_h(\log_b(Z_n))\right] = 0.$$

Proof. The weak convergence of a sequence of probability measures on the torus is entirely characterized by the convergence of its Fourier coefficients (see [2, p. 363] for example). So a sequence of probability measures $(Q_n)_n$ on $[0, 1[$ converges weakly to the uniform probability if and only if, for every $h \in \mathbb{Z}^*$,

$$\lim_{n\to+\infty} \int_0^1 e_h(x)dQ_n(x) = 0.$$

But the law of $\mathcal{M}_b(Z_n)$ converges weakly to μ_b if and only if the law of $\{\log_b Z_n\}$ converges weakly to the uniform probability on $[0, 1[$ and $e_h(\{\log_b Z_n\}) = e_h(\log_b Z_n)$ for every $h \in \mathbb{Z}$. □

We first prove that, in many cases and in particular under the hypotheses of theorems 2.4 and 2.6 and of propositions 3.4 and 3.5, the only possible limit for the law of $\mathcal{M}_b(U_n)$ is μ_b .

Proposition 4.2. *If a sequence $(Z_n)_n$ of positive random variables is a.s. natural-Benford and if the law of $\mathcal{M}_b(Z_n)$ converges weakly to a probabilty measure Q as $n \to +\infty$, then $Q = \mu_b$.*

Proof. Fix $h \in \mathbb{Z}^*$. For $N \geq 1$, set

$$M_N = \frac{1}{N}\sum_{n=1}^{N} e_h(\log_b Z_n)$$

By lemma 2.1, $(Z_n)_n$ is a.s. natural-Benford if and only if $(M_N)_N$ converge a.s. to 0 for every $h \in \mathbb{Z}^*$. Hence, by the dominated convergence theorem,

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left[e_h(\log_b(Z_n))\right] = \lim_{N\to+\infty} \mathbb{E}(M_N) = 0. \tag{4.1}$$

If moreover the law of $\mathcal{M}_b(Z_n)$ converges weakly to a probability measure on $[1, b[$ as $n \to +\infty$, then the law of $\{\log_b(\mathcal{M}_b(Z_n))\} = \{\log_b(Z_n)\}$ converges weakly to a probability measure on $[0, 1[$ and its Fourier coefficient $\mathbb{E}\left[e_h(\{\log_b(Z_n)\})\right] = \mathbb{E}\left[e_h(\log_b(Z_n))\right]$ admits a limit. But, by (4.1), this limit must be 0 and so, by lemma 4.1, the law of $\mathcal{M}_b(Z_n)$ converges weakly to μ_b. □

Remark 4.3. *This remains true when $(Z_n)_n$ is only supposed to be a.s. logarithmic-Benford instead of a.s. natural-Benford.*

The theorem 4.6 below seems inefficient when $u_n = \log n$ and $u_n = \log\log n$, but it shows in particular that, when $(nq_n)_n$ is nonincreasing and $\sum_{n=1}^{+\infty} q_n = +\infty$, the law of $\mathcal{M}_b(U_n)$ converges weakly to μ_b in the cases listed below:

- when $u_n = n$ (this contains the case considered in [11]);
- when $u_n = n\log n$ or $u_n = p_n$ and $q_n = \mathcal{O}(1/(n\log^\delta n))$ where $1 > \delta > 0$;
- when $u_n = n\log\log n$ and $q_n = \mathcal{O}(1/(n\log\log^\delta n))$ where $1 > \delta > 0$

($\mathcal{O}$ denotes the Landau's big O). The proof of theorem 4.6 uses the following estimation.

Lemma 4.4. *Let $(u_n)_n$ be a sequence of positive numbers. Suppose that the sequence $(nq_n)_n$ is nonincreasing and that*

$$\sum_{n=1}^{+\infty} nq_n \left|\log\left[\frac{w_n}{w_{n+1}}\right]\right| < +\infty \tag{4.2}$$

where $w_n = u_n/n$. Then there exists a constant C' (depending only on b) such that, for every integer $N \geq 1$ and every $h \in \mathbb{Z}^*$,

$$\left|\sum_{n=1}^{N} q_n e_h(\log_b(u_n))\right| \leq C'|h|\log|h|\,.$$

Proof. Fix h and N. The Abel's transformation gives

$$\sum_{n=1}^{N} q_n e_h(\log_b(u_n)) = Nq_N e_h(\log_b(w_N))\sum_{j=1}^{N} a_j + \sum_{n=2}^{N-1}\left((c_n - c_{n+1})\sum_{j=1}^{n} a_j\right)$$

where $a_j = (1/j)e_h(\log_b j)$ and $c_n = nq_n e_h(\log_b w_n)$.

Lemma 2.2 yields that $\left(\left|\sum_{j=1}^n a_j\right|\right)_n$ is bounded by $C\log|h|$. The sequence $(nq_n)_n$ is bounded too since it is positive and nonincreasing. It remains to verify that, for all $N \geq 1$,

$$\sum_{k=2}^{N-1} |c_k - c_{k+1}| \leq C''|h|$$

where C'' is a constant. But

$$|c_k - c_{k+1}| \leq |kq_k - (k+1)q_{k+1}| + (k+1)q_{k+1}|e_h(\log_b w_k) - e_h(\log_b w_{k+1})|\,.$$

By the mean value theorem, we get

$$|c_k - c_{k+1}| \leq (kq_k - (k+1)q_{k+1}) + 2\pi|h|kq_k\left|\log_b\left[\frac{w_k}{w_{k+1}}\right]\right|\,.$$

We can conclude with (4.2) because $\sum_{k=1}^{N}(kq_k - (k+1)q_{k+1}) \leq q_1$. □

Remark 4.5. *Using lemma 2.2 and the Abel transformation, it is easily seen that, when $w_n = 1$, $|h|\log|h|$ can be replaced by $\log|h|$ in lemma 4.4.*

Theorem 4.6. *Suppose that the sequences $(u_n)_n$ and $(q_n)_n$ verify the hypotheses of lemma 4.4 and that the q_n sum to infinity. Then the law of $\mathcal{M}_b(U_n)$ converges weakly to μ_b.*

Proof. Recall that $\pi_n = X_1 + \cdots + X_n$. Fix $h \in \mathbb{Z}^*$. We aim at showing that

$$\lim_{n\to+\infty} \sum_{k=n}^{+\infty} P(Y_n = k) e_h(\log_b u_k) = 0$$

because of lemma 4.1. For every $n \geq 1$ and every $N > n$, the Abel's transformation gives

$$\sum_{k=n}^{N} P(Y_n = k) e_h(\log_b u_k) = d_N^{(n)} \sum_{j=n}^{N} a_j + \sum_{k=n}^{N-1} \left((d_k^{(n)} - d_{k+1}^{(n)}) \sum_{j=n}^{k} a_j \right)$$

where

$$a_j = q_j e_h(\log_b u_j) \quad \text{and} \quad d_k^{(n)} = (1/q_k) P(Y_n = k) = P(\pi_{k-1} = n-1)\,.$$

By lemma 4.4, since h and b are fixed, $|\sum_{j=n}^{N} a_j|$ and $|\sum_{j=n}^{k} a_j|$ are bounded. Moreover, for every fixed n, $\lim_{N\to+\infty} d_N^{(n)} = 0$. Thus it remains only to verify that

$$\lim_{n\to+\infty} \lim_{N\to+\infty} \sum_{k=n}^{N-1} |d_k^{(n)} - d_{k+1}^{(n)}| = 0\,.$$

Fix $n \geq 1$. Then, for every $k \geq n$,

$$d_k^{(n)} - d_{k+1}^{(n)} = q_k (P(\pi_{k-1} = n-1) - P(\pi_{k-1} = n-2))$$

because $P(\pi_k = n-1) = q_k P(\pi_{k-1} = n-2) + (1-q_k) P(\pi_{k-1} = n-1)$. Now, it is proved in [4] that, for every $k \geq 1$ the law of π_k is unimodal with mode $m = [q_1 + \cdots + q_k]$ or $m = [q_1 + \cdots + q_k] + 1$ or with two modes $m_1 = [q_1 + \cdots + q_k]$ and $m_2 = [q_1 + \cdots + q_k] + 1$ with $P(\pi_k = m_1) = P(\pi_k = m_2)$ (here $[x]$ denotes the integer part of x). Hence $d_k^{(n)} - d_{k+1}^{(n)} \geq 0$ when $n - 1 \leq [q_1 + \cdots + q_{k-1}]$ and $d_k^{(n)} - d_{k+1}^{(n)} \leq 0$ when $n - 1 \geq [q_1 + \cdots + q_{k-1}] + 1$. This yields, for every $N > n$,

$$\begin{aligned} \sum_{k=n}^{N-1} |d_k^{(n)} - d_{k+1}^{(n)}| &\leq 2 \sup_{k\geq n} P(\pi_{k-1} = n-1) \\ &\leq 2 \sup_{k\geq n} \max_{0\leq j\leq k-1} P(\pi_{k-1} = j). \end{aligned}$$

By a classical property of the Levy's concentration function (lemma 1 in [17, page 38]), for every $k \geq n$,

$$\max_{0\leq j\leq k-1} P(\pi_{k-1} = j) \leq \max_{0\leq j\leq n-1} P(\pi_{n-1} = j)$$

because $\pi_{k-1} = \pi_{n-1} + X_n + \cdots + X_{k-1}$ and π_{n-1} and $X_n + \cdots + X_{k-1}$ are independent. Applying the Kolmogorov-Rogozin inequality (theorem 4 in [17, page 44]) to the concentration functions of the random variables X_l with $l = 1, \ldots, n-1$, we get

$$\max_{0\leq j\leq n-1} P(\pi_{n-1} = j) = \mathcal{O}\left(1/\sqrt{q_1 + \cdots + q_{n-1}}\,\right)$$

where $\mathcal{O}$ denotes the Landau's big O. This completes the proof since the q_n sum to infinity. □

It is quite surprising to find sequences of random variables whose mantissa in base b converges in law to μ_b for every base b since it is known (see [12, pages 238–247]) that there does not exist any random variable X such that the law of $\mathcal{M}_b(X)$ is μ_b for every base b.

5 Concluding remark

It is natural to wonder what happens when $q_n = q \in]0,1[$ (this is the i.i.d. case) or when $q_n = n^{-\delta}$ with $0 < \delta < 1$.

Is $(U_n)_n$ a.s. natural-Benford? Lemma 3.1 shows that, when $(q_n)_n$ is constant or nondecreasing, $(U_n)_n$ is a.s. natural-Benford if and only if $(u_n)_n$ is natural Benford. We know that this is not true in all the cases we have considered above: $u_n = n$, $u_n = n\log n$, $u_n = p_n$, $u_n = \log n$ and $u_n = \log\log n$. Lemma 3.1 yields the same conclusion in the case $q_n = n^{-\delta}$ with $0 < \delta < 1$ because the weights 1 and the weights $n^{-\delta}$ with $0 < \delta < 1$ lead to equivalent weighted densities (see [13, page 64]). But if $(q_n)_n$ is nondecreasing, we can adapt our techniques to prove that $(U_n)_n$ is a.s. *logarithmic-Benford*, that is to say: a.s.

$$\forall\, t \in [1, b[\lim_{N\to+\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{1}_{[1,\,t[}(\mathcal{M}_b(U_n)) = 0.$$

Does the law of $\mathcal{M}_b(U_n)$ converge weakly to μ_b? All we can say is that our methods seem inefficient when $(q_n)_n$ is constant. Indeed, if we want to use lemma 2.2 to treat the case $u_n = n$ and $q_n = 1/2$, the bound $2\sup_{k\geq n} P(\pi_{k-1} = n-1)$ in the proof of theorem 4.6 is replaced by $n\sup_{k\geq n} P(\pi_k = n)$ which does not tend to 0 as n tends to infinity since $P(\pi_{2n} = n)$ is equivalent to $1/\sqrt{n\pi}$. The same difficulties appear when we choose $q_n = n^{-\delta}$ with $0 < \delta < 1$. Anyway, some computer simulations we have made suggest that the law of $\mathcal{M}_b(U_n)$ does not converge weakly when $(q_n)_n$ is constant or when $q_n = 1/\sqrt{n}$.

References

[1] Benford, F.: The law of anomalous numbers. *Proceedings of the American Philosophical Society* **78**, (1938), 551–572.

[2] Billingsley, P.: Probability and measure. Wiley Series in Probability and Mathematical Statistics. *John Wiley & Sons*, New York-Chichester-Brisbane, 1979. xiv+515 pp. MR-0534323

[3] Cohen, G. and Cuny, C.: On random almost periodic trigonometric polynomials and applications to ergodic theory. *Ann. Probab.* **34-1**, (2006), 39–79. MR-2206342

[4] Darroch, J. N.: On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* **35**, (1964), 1317–1321. MR-0164359

[5] Drmota, M. and Tichy, R. F.: Sequences, discrepancies and applications. Lecture Notes in Mathematics, 1651. *Springer-Verlag*, Berlin, 1997. xiv+503 pp. MR-1470456

[6] Duncan, R. L.: Note on the initial digit. *Fibonacci Quart.* **7-5**, (1969), 474–475.

[7] Eliahou, S., Massé, B. and Schneider, D.: On the mantissa distribution of powers of natural and prime numbers. To appear in *Acta Mathematica Hungarica*, (2012), `http://www.springerlink.com/content/37m42117885580mk/`

[8] Fuchs, A. and Letta, G.: Le problème du premier chiffre décimal pour les nombres premiers. (French) [The first digit problem for primes] The Foata Festschrift. *Electron. J. Combin.* **3-2**, (1996), Research Paper 25, approx. 7 pp. (electronic). MR-1392510

[9] Hamming, R. W.: On the distribution of numbers. *Bell System Tech. J.* **49**, (1970), 1609–1625. MR-0267809

[10] Hardy, G. H.: Divergent Series. *Oxford University Press*, 1949. xvi+396 pp. MR-0030620

[11] Janvresse E. and de la Rue T.: Averaging along uniform random integers. *Uniform Distribution Theory* **7-2**, (2012), 61–76.

[12] Knuth, D. E.: The art of computer programming. Vol. 2. Seminumerical algorithms. Second edition. Addison-Wesley Series in Computer Science and Information Processing. *Addison-Wesley*, 1981. xiii+688 pp. MR-0633878

[13] Kuipers L. and Niederreiter H.: Uniform Distribution of Sequences. *Dover Publications*, New-York, 2006. xviii+390 pp.

[14] Massé, B. and Schneider, D.: A survey on weighted densities and their connection with the first digit phenomenon. *Rocky Mountain J. Math.* **41-5**, (2011), 1395–1415. MR-2838069

[15] Newcomb, S.: Note on the Frequency of Use of the Different Digits in Natural Numbers. *Amer. J. Math.* **4-1**, (1881), 39–40. MR-1505286

[16] Nigrini M. J. and Mittermaier L. J.: The use of Benford's law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory* **16-2**, (1997), 52–57.

[17] Petrov, V.: Sums of independent random variables. Translated from the Russian by A. A. Brown. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82. *Springer-Verlag*, New York-Heidelberg, 1975. x+346 pp. MR-0388499

[18] Rivat, J. and Tenenbaum, G.: Constantes d'Erdos-Turán. (French) [Erdos-Turan constants] *Ramanujan J.* **9**, (2005), 111–121. MR-2166382

[19] Titchmarsh, E. C.: The theory of the Riemann zeta-function. Second edition. Edited and with a preface by D. R. Heath-Brown. *Oxford University Press*, 1986. x+412 pp. MR-0882550

[20] Wintner A.: On the cyclical distribution of the logarithms of the prime numbers. *The Quarterly Journal of Mathematics* **6-1**, (1935), 65–68.