ACTA ARITHMETICA
163.1 (2014)

The mantissa distribution of the primorial numbers

by

Bruno Massé and Dominique Schneider (Calais and Lille)

1. Introduction. Fix a numeration base b and define the mantissa of the positive real number x as the unique number $\mathcal{M}_b(x)$ in $[1, b[$ such that $x = \mathcal{M}_b(x)b^k$ for some integer k. We will use the following notations: log is the natural logarithm, $\log_b$ is the logarithm in base b, p_n is the nth prime number and P_n is the product of the first n prime numbers (P_n is sometimes denoted $p_n\#$). We are interested in the distribution of the sequence $(\mathcal{M}_b(P_n))_n$ and, secondarily, of some sequences defined in a like manner.

1.1. Benford sequences. *Benford's law in base* b is the probability measure μ_b on $[1, b[$ defined by

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

When u_n is the nth Fibonacci number or $u_n = n^n$, $u_n = n!$, $u_n = a^n$ and $u_n = a^{p_n}$ (with $\log_b a$ irrational), the sequence $(u_n)_n$ is known to have a mantissa distributed following Benford's law in the sense of the natural density [1, 5, 11, 12]. That is,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[1,t[}(\mathcal{M}_b(u_n)) = \log_b t \quad (1 \le t < b). \tag{1.1}$$

The sequences $(u_n)_n$ satisfying (1.1) will be called *natural-Benford*. In particular, about 30.1 percent of the terms of a natural-Benford sequence have first digit 1, in the sense of (1.1), when $b = 10$. This kind of property is known as *the first digit phenomenon* and holds, more or less for many real-life data sets.

The sequences $(n)_n$ and $(p_n)_n$ are considered in particular in [2], [4], [5], [6], [16] and [18]. They are not natural-Benford and their mantissa does not admit any distribution in the sense of the natural density. However, they

2010 *Mathematics Subject Classification*: Primary 11K31; Secondary 11K06, 11A41.
Key words and phrases: uniform distribution, Benford's law, mantissa, primorial number.

DOI: 10.4064/aa163-1-4

exhibit Benford behavior in a weaker sense: if $u_n = n$ or $u_n = p_n$, then

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

Such sequences $(u_n)_n$ will be called *log-Benford*.

The sequence $(\log n)_n$ is neither natural-Benford nor log-Benford [8]. But it can be called *loglog-Benford* because

$$\lim_{N\to\infty} \frac{1}{\log\log N} \sum_{n=2}^{N} \frac{1}{n\log n} \mathbb{1}_{[1,t[}(\mathcal{M}_b(\log n)) = \log_b t \quad (1 \le t < b).$$

All the natural-Benford sequences are log-Benford and all the log-Benford sequences are loglog-Benford. The converses are false [8].

More generally, let $(w_n)_n$ be a sequence of positive real numbers summing to infinity and, for each $N \ge 1$, let $W_N = w_1 + \cdots + w_N$. The *w_n-density* of a set Δ of positive integers is the number

$$\lim_{N\to\infty} \frac{1}{W_N} \sum_{n=1}^{N} w_n \, \mathbb{1}_\Delta(n)$$

provided that it exists. This is the limit of the weighted frequency of the elements of Δ among the positive integers. We shall say that the sequence $(u_n)_n$ is *Benford in the sense of the w_n-density* when, for all $t \in [1, b[$, the set $\Delta = \{n : \mathcal{M}_b(u_n) < t\}$ has w_n-density $\log_b t$. See [8] for a general treatment of this kind of densities and their connection with Benford sequences.

As pointed out by an anonymous referee, many references on Benford's law and special sequences are available in [3] (Schatte's contributions should be mentioned), and [6], [16], [15] and [10] contain quantitative theorems, with respect to weighted means, closely related to those featuring in the present paper.

1.2. Content. In Section 2, we present some useful properties of the theory of uniform distribution. In particular, we pay a lot of attention to van der Corput's Difference Theorem and to some of its generalizations.

In Section 3 we show that the sequence $(P_n)_n$ of primorial numbers is natural-Benford (while $(p_n)_n$ is not). We provide an estimate of the convergence rate in (1.1) depending only on the numeration base b. This is done by proving the uniform distribution of $(a\vartheta(p_n))_n$, where ϑ denotes the first Chebyshev function and a any nonzero real, with convergence rate depending only on a.

In order to illustrate by a second example the utility of van der Corput's methods, we provide in Section 4 a brief discussion of the sequence $(\log 2 \times \cdots \times \log n)_n$. We show that it is log-Benford (while $(\log n)_n$ is not).

2. Preliminaries. We collect here the main tools we use. They are all connected with the theory of uniform distribution modulo 1. Most of them are known results.

The fractional part of a real number x will be denoted by $\{x\}$. For every real λ, we set $e_\lambda(x) = \exp(2i\pi\lambda x)$ with $i^2 = -1$. All along this section, $(w_n)_n$ is a sequence of positive numbers summing to infinity. Recall that $W_N = w_1 + \cdots + w_N$.

2.1. Basic properties. A sequence $(x_n)_n$ of real numbers is said to be *uniformly distributed modulo* 1 (u.d. mod 1) *in the sense of the* w_n*-density* when, for all $t \in [0,1[$,

$$\lim_{N\to\infty} \frac{1}{W_N} \sum_{n=1}^{N} w_n \mathbb{1}_{[0,t[}(\{x_n\}) = t.$$

Due to Dini's Theorem, this is equivalent to

$$\lim_{N\to\infty} \frac{1}{W_N} \sup_{t\in[0,1[} \Big| \sum_{n=1}^{N} w_n \mathbb{1}_{[0,t[}(\{x_n\}) - t \Big| = 0.$$

The simple fact that $\log_b x$ and $\log_b(\mathcal{M}_b(x))$ are equal modulo 1 yields the following.

Lemma 2.1. *A sequence $(u_n)_n$ of positive numbers is Benford in the sense of the w_n-density if and only if $(\log_b u_n)_n$ is u.d. mod 1 in the sense of the same w_n-density.*

The next statement is known as the Weyl Criterion. A simple proof in the case $w_n = 1$ is available in [7, p. 7]. The proof in the general case proceeds along the same lines.

Lemma 2.2. *A sequence $(x_n)_n$ of real numbers is u.d. mod 1 in the sense of the w_n-density if and only if for all integers $k \neq 0$,*

$$\lim_{N\to\infty} \frac{1}{W_N} \sum_{n=1}^{N} w_n e_k(x_n) = 0.$$

As a direct consequence of this lemma, if a sequence $(u_n)_n$ of positive numbers is Benford in the sense of the w_n-density, then so is $(\lambda u_n^m)_n$ for all integer $m \neq 0$ and positive λ.

Lemma 2.3 is elementary but crucial and we did not find any reference for it.

Lemma 2.3. *Let $(a_n)_n$ and $(b_n)_n$ be two bounded equivalent sequences of complex numbers, and let $(w'_n)_n$ be a sequence of positive numbers summing* *to infinity and equivalent to* $(w_n)_n$. *Set* $W'_N = w'_1 + \cdots + w'_N$ $(N \geq 1)$. *Then*

$$\lim_{N\to\infty} \left| \frac{1}{W_N} \sum_{n=1}^{N} w_n a_n - \frac{1}{W'_N} \sum_{n=1}^{N} w'_n b_n \right| = 0.$$

Proof. Set $b_n = a_n + a_n \varepsilon_n$ and $w'_n = w_n + w_n \varepsilon'_n$ with $\lim_n \varepsilon_n = \lim_n \varepsilon'_n = 0$. Then, for all $N \geq 1$,

$$\left| \frac{1}{W_N} \sum_{n=1}^{N} w_n a_n - \frac{1}{W'_N} \sum_{n=1}^{N} w'_n b_n \right| \leq \left| \frac{W'_N - W_N}{W'_N} \frac{1}{W_N} \sum_{n=1}^{N} w_n a_n \right| + \left| \frac{1}{W'_N} \sum_{n=1}^{N} w_n a_n (\varepsilon_n + \varepsilon'_n + \varepsilon_n \varepsilon'_n) \right|.$$

By the Stolz–Cesàro Theorem, the sequences $(W_N)_N$ and $(W'_N)_N$ are equivalent. So the first term in the above sum tends to 0 as N tends to infinity because the sequence $((1/W_N)\sum_{n=1}^{N} w_n a_n)_N$ is bounded. Moreover, by the classical generalization of the Cesàro Theorem, the second term tends to 0 too, because $(W_N)_N$ and $(W'_N)_N$ are equivalent and $\lim_n a_n(\varepsilon_n + \varepsilon'_n + \varepsilon_n \varepsilon'_n) = 0$. ■

If two sequences of positive numbers $(u_n)_n$ and $(v_n)_n$ are equivalent, then for every integer k the sequences of complex numbers $(e_k(\log_b(u_n)))_n$ and $(e_k(\log_b(v_n)))_n$ are equivalent too. This and Lemmas 2.1, 2.2 and 2.3 prove Lemma 2.4 below.

LEMMA 2.4. *Two equivalent sequences of positive numbers are simultaneously natural-Benford* (*respectively log-Benford, loglog-Benford*).

2.2. Van der Corput's Difference Theorem. Van der Corput's Difference Theorem [7, p. 26] says that, in the context of the natural density, a sequence $(x_n)_n$ is u.d. mod 1 when, for every positive integer h, the sequence $(x_{n+h} - x_n)_n$ is u.d. mod 1. It derives from Lemma 2.5 below. We present here three generalizations of this theorem. The first one is due to Tsuji [17] and extends van der Corput's Difference Theorem to the general context of weighted densities. The second and the third say that, again in the context of weighted densities, $(x_n)_n$ is u.d. mod 1 if $(x_{n+h} - x_n)_n$ tends to be u.d. mod 1 as h tends to infinity in a sense made precise in the statements of Lemmas 2.9 and 2.10. To the best of our knowledge, Lemmas 2.8–2.10 are new results.

Lemma 2.5 is a direct consequence of van der Corput's Fundamental Inequality [7, p. 25] and is crucial in the proof of Theorem 3.4 below.

LEMMA 2.5. *Let* N *be a positive integer greater than* 1, *and* $a_1, \ldots, a_N$ *be* N *complex numbers of modulus* 1. *Then there exists an absolute constant*

C such that for all positive integers $H < N$,

$$\left|\frac{1}{N}\sum_{n=1}^{N} a_n\right|^2 \leq \frac{C}{H} + \frac{C}{H}\sum_{h=1}^{H}\left|\frac{1}{N-h}\sum_{n=1}^{N-h} a_n\overline{a}_{n+h}\right|.$$

Lemma 2.6 is Tsuji's extension of the above lemma to the case of general weighted densities. The real part of a complex number z is denoted by $\mathcal{R}(z)$.

Lemma 2.6. *Let N be a positive integer greater than 1, and $a_1, \dots, a_N$ be N complex numbers. Then for all positive integers $H < N$,*

$$\frac{|\sum_{n=1}^{N} w_n a_n|^2}{W_{N+H-1}} \leq \frac{1}{H^2}\sum_{n=1}^{N} w_n^2 |a_n|^2 \sum_{j=0}^{H-1}\frac{1}{w_{n+j}} + 2\mathcal{R}\left(\frac{1}{H^2}\sum_{h=1}^{H-1}\sum_{n=1}^{N-h} w_n w_{n+h} a_n\overline{a}_{n+h}\sum_{j=h}^{H-1}\frac{1}{w_{n+j}}\right).$$

Tsuji [17, Theorem 10] used this lemma to prove the following generalization of van der Corput's Difference Theorem.

Lemma 2.7. *Let $(x_n)_n$ be a sequence of real numbers. Suppose that, for all positive integers h, $(w_n/w_{n+h})_n$ is decreasing and $(x_{n+h} - x_n)_n$ is u.d. mod 1 in the sense of the w_n-density. Then $(x_n)_n$ is u.d. mod 1 in the sense of the w_n-density.*

We now present the second generalization. The next lemma uses the bound in Lemma 2.6 more precisely than Lemma 2.7.

Lemma 2.8. *Let $(a_n)_n$ be a sequence of complex numbers bounded in modulus by 1. Assume that $(w_n)_n$ and $(w_{n+1})_n$ are equivalent. For $h \geq 1$, set*

$$l_h = \limsup_{N\to\infty}\left|\frac{1}{W_N}\sum_{n=1}^{N} w_n a_n\overline{a}_{n+h}\right|.$$

Then

$$\lim_{H\to\infty}\frac{1}{H}\sum_{h=1}^{H-1} l_h = 0 \;\Rightarrow\; \lim_{N\to\infty}\left|\frac{1}{W_N}\sum_{n=1}^{N} w_n a_n\right| = 0.$$

Proof. Note that the hypotheses on $(w_n)_n$ and the Stolz–Cesàro Theorem imply that for all $j \geq 1$ the sequences $(W_N)_N$ and $(W_{N+j})_N$ are equivalent.

Since $W_N \geq W_{N-h}$ and $H \geq H - h$, Lemma 2.6 leads to

$$\frac{|\sum_{n=1}^{N} w_n a_n|^2}{W_N^2}\frac{W_N}{W_{N+H-1}} \leq \frac{1}{HW_N}\sum_{n=1}^{N}\frac{w_n}{H}\sum_{j=0}^{H-1}\frac{w_n}{w_{n+j}} + \frac{2}{H}\sum_{h=1}^{H-1}\frac{1}{W_{N-h}}\left|\sum_{n=1}^{N-h}\frac{w_n}{H-h}a_n\overline{a}_{n+h}\sum_{j=h}^{H-1}\frac{w_{n+h}}{w_{n+j}}\right|.$$

Fix h and H. Then W_N and W_{N+H-1} are equivalent as $N \to \infty$. Moreover $\sum_{j=0}^{H-1} \frac{w_n}{w_{n+j}}$ converges to H and $\sum_{j=h}^{H-1} \frac{w_{n+h}}{w_{n+j}}$ converges to $H - h$ as $n \to \infty$, since $(w_n)_n$ and $(w_{n+1})_n$ are supposed to be equivalent.

Hence, by Lemma 2.3, for all H,

$$\limsup_{N\to\infty} \left| \frac{1}{W_N} \sum_{n=1}^{N} w_n a_n \right|^2 \le \frac{1}{H} + \frac{2}{H} \sum_{h=1}^{H-1} l_h,$$

which implies

$$\limsup_{N\to\infty} \left| \frac{1}{W_N} \sum_{n=1}^{N} w_n a_n \right|^2 \le 2 \limsup_{H\to\infty} \left[\frac{1}{H} \sum_{h=1}^{H-1} l_h \right]. \ \blacksquare$$

Lemmas 2.8 and 2.2 prove the following generalization of Tsuji's Theorem. The condition "$(x_{n+h} - x_n)_n$ is u.d. mod 1 for all h" is replaced by a weaker one which may be interpreted as "$(x_{n+h} - x_n)_n$ tends to be u.d. mod 1 as $h \to \infty$".

LEMMA 2.9. *Let $(x_n)_n$ be a sequence of real numbers. For integers $h \ge 1$ and $k \ne 0$, set*

$$l_{h,k} = \limsup_{N\to\infty} \left| \frac{1}{W_N} \sum_{n=1}^{N} w_n e_k(x_{n+h} - x_n) \right|$$

and suppose that, for all k, $(l_{h,k})_h$ converges to 0 as $h \to \infty$ or, more generally, that

$$\lim_{H\to\infty} \frac{1}{H} \sum_{h=1}^{H-1} l_{h,k} = 0.$$

Then $(x_n)_n$ is u.d. mod 1 in the sense of the w_n-density.

Lemma 2.10 is a kind of Paul Lévy's Theorem for arrays of distribution functions and is essential in the proof of Theorem 4.1.

LEMMA 2.10. *Let $(y_{h,n})_{h\ge1,n\ge1}$ be an array of numbers in $[0,1[$. For positive integers N and h, for integers k different from zero and $t \in [0,1]$, set*

$$\alpha_h(k) = \limsup_{N\to\infty} \left| \frac{1}{W_N} \sum_{n=1}^{N} w_n e_k(y_{h,n}) \right|,$$

$$F_{N,h}(t) = \frac{1}{W_N} \sum_{n=1}^{N} w_n \mathbb{1}_{[0,t[}(y_{h,n}),$$

$$\beta_h(t) = \limsup_{N\to\infty} |F_{N,h}(t) - t|.$$

Then

$$\Big(\forall t \lim_{h\to\infty} \beta_h(t) = 0\Big) \;\Rightarrow\; \Big(\forall k \lim_{h\to\infty} \alpha_h(k) = 0\Big).$$

Proof. We first apply classical methods used in the proof of Dini's Theorem. For a positive integer h and $t \in [0,1]$, set $\gamma_h = \limsup_N \sup_t |F_{N,h}(t) - t|$, and let M be a positive integer. Fix $t \in [0,1]$ and $m \in \{0, \ldots, M-1\}$ such that $m/M \le t < (m+1)/M$, and positive integers N and h. Then

$$F_{N,h}(t) - t \le F_{N,h}\left(\frac{m+1}{M}\right) - \frac{m+1}{M} + \frac{1}{M},$$
$$F_{N,h}(t) - t \ge F_{N,h}\left(\frac{m}{M}\right) - \frac{m}{M} - \frac{1}{M},$$

because all the functions considered here are nondecreasing. Hence

$$\sup_{t\in[0,1]} |F_{N,h}(t) - t| \le \max_{m=0,\ldots,M} \left| F_{N,h}\left(\frac{m}{M}\right) - \frac{m}{M} \right| + \frac{1}{M}.$$

This implies, for all h and M,

$$\begin{aligned}\gamma_h &\le \limsup_{N\to\infty} \max_{m=0,\ldots,M} \left| F_{N,h}\left(\frac{m}{M}\right) - \frac{m}{M} \right| + \frac{1}{M} \\ &\le \max_{m=0,\ldots,M} \beta_h\left(\frac{m}{M}\right) + \frac{1}{M},\end{aligned}$$

since the set $\{0, \ldots, M\}$ is finite. It follows that

$$\Big(\forall t \lim_{h\to\infty} \beta_h(t) = 0\Big) \Rightarrow \Big(\lim_{h\to\infty} \gamma_h = 0\Big).$$

Fix now $k \in \mathbb{Z}^*$. For each $h \in \mathbb{N}^*$, let N_h be an integer such that

$$\left|\left|\frac{1}{W_{N_h}} \sum_{n=1}^{N_h} w_n e_k(y_{h,n})\right| - \alpha_h(k)\right| \le \frac{1}{h} \tag{2.1}$$

and large enough to ensure that

$$\sup_{t\in[0,1[} |F_{N_h,h}(t) - t| \le \gamma_h + 1/h.$$

The integer N_h exists because at least one subsequence of

$$\left(\left|\frac{1}{W_N} \sum_{n=1}^{N} w_n e_k(y_{h,n})\right|\right)_N$$

converges to $\alpha_h(k)$ and every subsequence of

$$\Big(\sup_{t\in[0,1[} |F_{N,h}(t) - t|\Big)_N$$

has upper limit lower than or equal to γ_h.

If we suppose $\lim_h \gamma_h = 0$ and denote the Dirac measure at $y_{h,n}$ by $\delta_{y_{h,n}}$, then the sequence of probability measures

$$\left(\frac{1}{W_{N_h}}\sum_{n=1}^{N_h} w_n \delta_{y_{h,n}}\right)_h$$

converges weakly to the uniform distribution on $[0,1[$. So Paul Lévy's Theorem, characterizing weak convergence by means of Fourier transform, implies

$$\lim_{h\to\infty}\left|\frac{1}{W_{N_h}}\sum_{n=1}^{N_h} w_n e_k(y_{h,n})\right| = 0.$$

The proof is completed by using (2.1). ∎

Remark 2.11. The converse of Lemma 2.10 is true and is a consequence of the Erdős–Turán inequality (see Lemma 3.1 for the version of this inequality involving the natural density and [3] for the versions involving weighted densities). Applying Lemma 2.9 and Lemma 2.10 with $y_{h,n} = x_{n+h} - x_n$ leads to a new criterion, more telling than Lemma 2.9: if for all $t \in [0,1[$,

$$\lim_{h\to\infty}\limsup_{N\to\infty}\left|\frac{1}{W_N}\sum_{n=1}^{N} w_n \mathbb{1}_{[0,t[}(x_{n+h}-x_n) - t\right| = 0,$$

then $(x_n)_n$ is u.d. mod 1 in the sense of the w_n-density.

3. The primorial numbers. Throughout this section we set, for $n \geq 2$, $P_n = p_1 \times \cdots \times p_n$, where p_n denotes the nth prime number, and $Q_n = 2\log 2 \times \cdots \times n \log n$.

Recall that the sequence $(p_n)_n$ is log-Benford and is not natural-Benford. We shall now prove that $(P_n)_n$ is natural-Benford and provide a convergence rate estimate. We shall make use of Lemma 2.5 above and of Lemmas 3.1 and 3.3 below.

Lemma 3.1 is available, among many other references, in [14] and known as the Erdős–Turán inequality. Let $(x_n)_n$ be a real sequence in $[0,1[$ and N be a positive integer. The number $D_N(x_n)$, defined by

$$D_N(x_n) = \sup_{0<c<d<1}\left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}_{[c,d[}(x_n) - (d-c)\right|,$$

is called the *discrepancy* of $(x_n)_{n=1}^N$ (see [3] for more information on this subject).

Lemma 3.1. *For every positive integer K,*

$$D_N(x_n) \leq \frac{1}{K+1} + \sum_{k=1}^{K}\frac{1}{k}\,\frac{1}{N}\Big|\sum_{n=1}^{N} e_k(x_n)\Big|.$$

We point out that, in the above inequality, the choice of the integer $K \geq 1$ is free. Lemma 3.2 is known as van der Corput's Theorem [7, p. 17].

Lemma 3.2. *Let a and b be two integers with $a < b$, let $\rho > 0$, and let f be twice differentiable in $[a,b]$ and such that $f''(x) \geq \rho > 0$ or $-f''(x) \geq \rho > 0$ for all $x \in [a,b]$ and some positive real ρ. Then*

$$\Big|\sum_{n=a}^{b} \exp\left(2i\pi f(n)\right)\Big| \leq (|f'(a) - f'(b)| + 2)\Big(\frac{4}{\sqrt{\rho}} + 3\Big).$$

Lemma 3.3 is a direct consequence of Lemma 3.2.

Lemma 3.3. *There exists an absolute constant C such that for all $\theta > 0$ and all positive integers L, M and h with $L < M$ and $h \leq M$,*

$$\Big|\sum_{n=L}^{M} e_\theta(\log(Q_{n+h}/Q_n))\Big| \leq C\Big(\frac{M\sqrt{h\theta}}{L} + \frac{M}{\sqrt{h\theta}} + \frac{h\theta}{L} + 1\Big).$$

Proof. Let L, M, θ and h satisfy the hypothesis. If

$$f(x) = \theta \sum_{j=1}^{h} \log\big((x+j)\log(x+j)\big),$$

then for all $x \in [L, M]$,

$$0 < f'(x) \leq f'(L) \leq \frac{2h\theta}{L} \quad \text{and} \quad -f''(x) \geq -f''(M) \geq \frac{h\theta}{4M^2}.$$

Thus, by Lemma 3.2,

$$\Big|\sum_{n=L}^{M} e_\theta(\log(Q_{n+h}/Q_n))\Big| \leq \Big(\frac{2h\theta}{L} + 2\Big)\Big(\frac{8M}{\sqrt{h\theta}} + 3\Big). \quad \blacksquare$$

The next theorem is stated in terms of mantissa distribution of $(P_n)_n$. It could have been equivalently stated in terms of distribution modulo 1 of $(a\vartheta(p_n))_n$ where ϑ denotes the first Chebyshev function and a any nonzero real.

Theorem 3.4. *There exists a positive constant C_b, depending only on the numeration base b, such that for every positive integer N,*

$$\sup_{t \in [1,b[} \Big|\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{[1,t[}(\mathcal{M}_b(P_n)) - \log_b t\Big| \leq C_b \frac{(\log\log N)^{1/2}}{(\log N)^{1/9}}.$$

In particular the sequence $(P_n)_n$ is natural-Benford.

Proof. In what follows, C denotes an absolute positive constant which may vary from line to line. It is written C_b when it depends on b. Recall that for all $n \geq 1$, $n \log n \leq p_n \leq n \log n + Cn \log\log n$ [13].

Fix $N > 1$. By Lemma 3.1, for every positive integer K,

$$\sup_{t\in[1,b[}\left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}_{[1,t[}(\mathcal{M}_b(P_n)) - \log_b t\right| \le D_N(\{\log_b P_n\}) \le \frac{1}{K+1} + \sum_{k=1}^{K}\frac{1}{k}\frac{1}{N}\left|\sum_{n=1}^{N} e_k(\log_b P_n)\right| \tag{3.1}$$

and by Lemma 2.5, for all positive integers h, k and $H < N$,

$$\left|\frac{1}{N}\sum_{n=1}^{N} e_k(\log_b P_n)\right|^2 \le \frac{C}{H} + \frac{C}{H}\sum_{h=1}^{H}\left|\frac{1}{N-h}\sum_{n=1}^{N-h} e_k(\log_b(P_{n+h}/P_n))\right|. \tag{3.2}$$

By the Triangular Inequality and the Mean Value Theorem, for all h, k and M,

$$\left|\sum_{n=2}^{M} e_k(\log_b(P_{n+h}/P_n))\right| \le \sum_{n=2}^{M} 2k\pi\log_b\left(\frac{P_{n+h}/P_n}{Q_{n+h}/Q_n}\right) + \left|\sum_{n=2}^{M} e_k(\log_b(Q_{n+h}/Q_n))\right|. \tag{3.3}$$

We are going to calculate a bound for each term of (3.3) and then choose the best possible H in (3.2) and the best possible K in (3.1).

Fix $M \ge 3$ and $h \ge 1$. For $n \ge 3$,

$$1 \le \frac{P_{n+h}/P_n}{Q_{n+h}/Q_n} \le \prod_{j=1}^{h}\frac{(n+j)\log(n+j) + C(n+j)\log\log(n+j)}{(n+j)\log(n+j)} \le \left(1 + C\left(\frac{\log\log n}{\log n}\right)\right)^h.$$

If we set $\theta = k/\log b$, this leads to

$$\sum_{n=3}^{M} 2k\pi\log_b\left(\frac{P_{n+h}/P_n}{Q_{n+h}/Q_n}\right) \le Ch\theta\sum_{n=3}^{M}\frac{\log\log n}{\log n} \le Ch\theta(\log\log M)\frac{M}{\log M} \tag{3.4}$$

by the classical properties of the logarithmic integral function. On the other hand, by Lemma 3.3, if $h \le M$,

$$\left|\sum_{n=3}^{M} e_k(\log_b(Q_{n+h}/Q_n))\right| \le \sqrt{M} + \left|\sum_{n=\lfloor\sqrt{M}\rfloor}^{M} e_\theta(\log(Q_{n+h}/Q_n))\right| \le \sqrt{M} + C\left(\sqrt{M}\sqrt{h\theta} + \frac{M}{\sqrt{h\theta}} + \frac{h\theta}{\sqrt{M}} + 1\right). \tag{3.5}$$

Combining (3.3)–(3.5) gives

$$\Big|\sum_{n=1}^{M} e_k(\log_b(P_{n+h}/P_n))\Big| \le C_b\bigg(hk\frac{M\log\log M}{\log M} + \sqrt{hk}\sqrt{M} + \frac{M}{\sqrt{hk}} + \frac{hk}{\sqrt{M}}\bigg).$$

Since $N-(\log N)^{2/3}$ and N are equivalent, since $\log(N-(\log N)^{2/3})$ and $\log N$ are equivalent, and so on, we get, for all $h \le (\log N)^{2/3}$,

$$\frac{1}{N-h}\Big|\sum_{n=1}^{N-h} e_k(\log_b(P_{n+h}/P_n))\Big| \le C_b\bigg(hk\frac{\log\log N}{\log N} + \frac{\sqrt{hk}}{\sqrt{N}} + \frac{1}{\sqrt{hk}} + \frac{hk}{N^{3/2}}\bigg).$$

Hence, if $H = \lfloor(\log N)^{2/3}\rfloor$, then

$$\frac{1}{H}\sum_{h=1}^{H}\frac{1}{N-h}\Big|\sum_{n=1}^{N-h} e_k(\log_b(P_{n+h}/P_n))\Big| \le C_b k\frac{\log\log N}{(\log N)^{1/3}},$$

and so, by (3.2),

$$\begin{aligned}\bigg|\frac{1}{N}\sum_{n=1}^{N} e_k(\log_b P_n)\bigg| &\le C_b\bigg(\frac{1}{(\log N)^{2/3}} + k\frac{\log\log N}{(\log N)^{1/3}}\bigg)^{1/2}\\ &\le C_b\sqrt{k}\frac{(\log\log N)^{1/2}}{(\log N)^{1/6}}.\end{aligned}$$

Thus, for every positive integer K,

$$\sum_{k=1}^{K}\frac{1}{k}\,\frac{1}{N}\Big|\sum_{n=1}^{N} e_k(\log_b P_n)\Big| \le C_b\sqrt{K}\frac{(\log\log N)^{1/2}}{(\log N)^{1/6}}.$$

The proof is completed by taking $K = \lfloor(\log N)^{1/9}\rfloor$ in (3.1). ∎

4. Product of the first n logarithms. Recall that the sequence $(\log n)_n$ is loglog-Benford and is not log-Benford [8]. We shall now prove that the sequence $(\log 2\times\cdots\times\log n)_n$ is log-Benford. Lemma 3.2 is inefficient in dealing with $\log 2\times\cdots\times\log n$ because of the value of the second derivative of $x\mapsto\log\log x$. Fortunately, we can use direct calculations which are themselves fruitless for sequences like $(P_n)_n$.

THEOREM 4.1. *The sequence* $(\log 2\times\cdots\times\log n)_n$ *is log-Benford.*

Proof. Set $u_n = \log 2\times\cdots\times\log n$ and, for $N\ge 2$ and $s\in[1,b]$,

$$F_{N,h}(s) = \frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}\mathbb{1}_{[1,s[}\big(\mathcal{M}_b((\log n)^h)\big).$$

For all h, u_{n+h}/u_n is equivalent to $(\log n)^h$ as $n\to\infty$, and this implies that, for all integers k, $e_k(\log_b(u_{n+h}/u_n))$ is equivalent to $e_k(\log_b((\log n)^h))$. So, using Lemma 2.3 with $w_n = w'_n = 1/n$, for all k and all h we obtain

$$\lim_{N\to\infty}\left(\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}e_k\left(\log_b\left(\frac{u_{n+h}}{u_n}\right)\right)\right.$$
$$\left.-\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}e_k(\log_b((\log n)^h))\right)=0. \tag{4.1}$$

If we prove that for all $s\in[1,b]$,

$$\lim_{h\to\infty}\limsup_{N\to\infty}F_{N,h}(s)=\lim_{h\to\infty}\liminf_{N\to\infty}F_{N,h}(s)=\log_b s, \tag{4.2}$$

this will imply

$$\lim_{h\to\infty}\limsup_{N\to\infty}\left|\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}\mathbb{1}_{[0,\log_b s[}(\{\log_b((\log n)^h)\})-\log_b s\right|=0.$$

Then Lemma 2.10 will prove that, for all k,

$$\lim_{h\to\infty}\limsup_{N\to\infty}\left|\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}e_k(\log_b((\log n)^h))\right|=0,$$

and (4.1), Lemma 2.9 with $w_n=1/n$, and Lemmas 2.2 and 2.1 will conclude the proof.

But direct calculations, using Fuchs' and Letta's methods (see [5] and [9, p. 11]), show that

$$\liminf_{N\to\infty}F_{N,h}(s)=\frac{s^{1/h}-1}{b^{1/h}-1}\quad\text{and}\quad\limsup_{N\to\infty}F_{N,h}(s)=\frac{b^{1/h}(s^{1/h}-1)}{s^{1/h}(b^{1/h}-1)}.$$

These two limits tend to $\log_b s$ as $h\to\infty$. This proves (4.2). ∎

5. Concluding remarks. Van der Corput's methods, described in Section 2.2, can also be used to prove that the product sequence $(U_n)_n=(u_1\times\cdots\times u_n)_n$ and the corresponding iterated product sequence $(U_1\times\cdots\times U_n)_n$ are log-Benford or natural-Benford when $u_n=\log 2\times\cdots\times\log n$, $u_n=n!$, $u_n=n^n$ and $u_n=P_n$. When $u_n=n!$ (respectively $u_n=n^n$), the numbers U_n are called *superfactorials* (respectively *hyperfactorials*). A detailed discussion of this subject is in preparation.

It is known that

$$P_n=e^{(1+\varepsilon_n)n\log n}$$

with $\lim_n\varepsilon_n=0$. But this does not provide an equivalent of P_n. In contrast, the hyperfactorial sequence is equivalent to

$$An^{n^2/2+n/2+1/12}\times e^{-n^2/4}$$

where A is a constant, and this can be used together with Lemma 2.4 to prove that the hyperfactorial sequence is natural-Benford.

We do not know if the sequence $(n\#)_n$, defined by $n\# = \prod_{p_r \le n} p_r$, is Benford or not because of the irregularity of the prime gaps. So we do not know if the values $\vartheta(n)$ of the first Chebyshev function at integers are u.d. mod 1 or not. Of course, since they are respectively equivalent to n and $n \log n$, the sequences $(\vartheta(n))_n$ and $(\vartheta(p_n))_n$ are log-Benford and are not natural-Benford. But this is off topic.

The fact that the sequence $(P_n)_n$ is natural-Benford can be proved quickly by combining Lemmas 3.3, 2.3 and 2.9. Indeed, Lemma 3.3 shows that

$$\lim_{h\to\infty} \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^{N} e_k(Q_{n+h}/Q_n) \right| = 0$$

where $Q_n = 2\log 2 \times \cdots \times n \log n$. Since Q_{n+h}/Q_n and P_{n+h}/P_n are equivalent as $n \to \infty$, Lemma 2.3 shows that this is also true when we replace Q_{n+h}/Q_n by P_{n+h}/P_n. Lemma 2.9 with $w_n = 1$ concludes the proof.

The main term in the calculations in the proof of Theorem 3.4 comes from the term $Cn \log\log n$ in the formula $p_n \le n \log n + Cn \log\log n$. If we replace p_n by $n \log n$ in Theorem 3.4, we obtain a far better convergence rate:

$$\sup_{t\in[1,b[} \left| \frac{1}{N} \sum_{n=2}^{N} \mathbb{1}_{[1,t[}(\mathcal{M}_b(Q_n)) - \log_b t \right| \le C_b N^{-1/10}$$

where C_b is a positive constant depending only on b. The proof proceeds along the same lines as the proof of Theorem 3.4: it suffices to replace P_n by Q_n in (3.1) and (3.2), to use (3.5) and to choose $H = \lfloor \sqrt{N} \rfloor$ and $K = \lfloor N^{1/10} \rfloor$.

The numbers $P_n - 1$ and $P_n + 1$ have few divisors and many of them are prime numbers (in this case they are called *primorial primes*). Yet $(P_n - 1)_n$ and $(P_n + 1)_n$ are natural-Benford since they are equivalent to $(P_n)_n$.

We cannot prove or disprove that $(\log 2 \times \cdots \times \log n)_n$ is actually natural-Benford with our methods. Indeed, if we replace $1/n$ by 1 in the proof of Theorem 4.1, we get $\liminf_N F_{N,h}(s) = 0$ and $\limsup_N F_{N,h}(s) = 1$ for all s and all h. So $(x_n - x_{n+h})_n$ does not tend to be u.d. mod 1 as $h \to \infty$ when $x_n = \log 2 \times \cdots \times \log n$ (see the remark at the end of Section 2).

Fuchs' and Letta's methods (see [5] and [9, p. 11]) and direct calculations prove that the sequence $(\log\log n)_n$ is not loglog-Benford. The arguments used in the proof of Theorem 4.1 show that the sequence $(u_1 \times \cdots \times u_n)_n$ is loglog-Benford when $u_n = \log\log n$. Of course we can also consider $(\log\log\log n)_n$ with the weights $1/(n \log n \log\log n)$ and so on.

References

[1] A. Berger, L. A. Bunimovich and T. P. Hill, *One-dimensional dynamical systems and Benford's law*, Trans. Amer. Math. Soc. 357 (2005), 197–219.

[2] D. I. Cohen and T. M. Katz, *Prime numbers and the first digit phenomenon*, J. Number Theory 18 (1984), 261–268.
[3] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer, Berlin, 1997.
[4] S. Eliahou, B. Massé and D. Schneider, *On the mantissa distribution of powers of natural and prime numbers*, Acta Math. Hungar. 139 (2013), 49–63.
[5] A. Fuchs et G. Letta, *Le problème du premier chiffre décimal pour les nombres premiers*, Electron. J. Combin. 3 (1996), no. 2, res. paper 25, 7 pp.
[6] E. Hlawka, *Gleichverteilung und das Konvergenzverhalten von Potenzreihen am Rande des Konvergenzkreises*, Manuscripta Math. 44 (1983), 231–263.
[7] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Dover Publ., 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 (2011), 1395–1415.
[9] B. Massé and D. Schneider, *Random number sequences and the first digit phenomenon*, Electron. J. Probab. 17 (2012), 1–17.
[10] H. Niederreiter und R. F. Tichy, *Beiträge zur Diskrepanz bezüglich gewichteter Mittel*, Manuscripta Math. 42 (1983), 85–99.
[11] M. Pollicott, *A note on uniform distribution for primes and closed orbits*, Israel J. Math. 55 (1986), 199–212.
[12] P. N. Posch, *A survey of sequences and distribution functions satisfying the first-digit-law*, J. Statist. Management Systems 11 (2008), 1–19.
[13] P. Ribenboim, *The Little Book of Bigger Primes*, 2nd ed., Springer, Berlin, 2004, p. 185.
[14] J. Rivat et G. Tenenbaum, *Constantes d'Erdős–Turán*, Ramanujan J. 9 (2005), 111–121.
[15] R. F. Tichy, *Diskrepanz bezüglich gewichteter Mittel und das Konvergenzverhalten von Potenzreihen*, Manuscripta Math. 44 (1983), 265–277.
[16] R. F. Tichy and G. Turnwald, *Logarithmic uniform distribution of* $(\alpha n + \beta \log n)$, Tsukuba J. Math. 10 (1986), 351–366.
[17] M. Tsuji, *On the uniform distribution of numbers mod. 1*, J. Math. Soc. Japan 4 (1952), 313–322.
[18] R. E. Whitney, *Initial digits for the sequence of primes*, Amer. Math. Monthly 79 (1972), 150–152.

Bruno Massé, Dominique Schneider
Université du Littoral Côte d'Opale
L.M.P.A. J. Liouville
B.P. 699, F-62228 Calais, France
and
Université Lille Nord de France
F-59000 Lille, France
CNRS, FR 2956, France
E-mail: bmasse@lmpa.univ-littoral.fr
schneider@lmpa.univ-littoral.fr

Received on 14.5.2013
and in revised form on 6.9.2013 (7443)