Department of Applied Mathematics, University of Venice

WORKING PAPER SERIES

Marco Corazza, Andrea Ellero and
Alberto Zorzi

What Sequences obey Benford's Law ?

Working Paper n. 185/2008
November 2008

ISSN: 1828-6887

What Sequences obey Benford's Law?

MARCO CORAZZA
<corazza@unive.it>

ANDREA ELLERO
<ellero@unive.it>

ALBERTO ZORZI
<albzorzi@unive.it>

Dipartimento di Matematica Applicata
Università Ca' Foscari di Venezia

Abstract. We propose a new necessary and sufficient condition to test whether a sequence is Benford (base-b) or not and apply this characterization to some kinds of sequences (re)obtaining some well known results, as the fact that the sequence of powers of 2 is Benford (base-10).

Keywords: Benford's law, equidistributed sequences, ergodic endomorphisms

JEL Classification Numbers: C02, C83

MathSci Classification Numbers: 28D99, 60F99

Correspondence to:

Marco Corazza	Dept. of Applied Mathematics, University of Venice
	Dorsoduro 3825/e
	30123 Venezia, Italy
Phone:	[++39] (041)-234-6921
Fax:	[++39] (041)-522-1756
E-mail:	corazza@unive.it

1 The importance of being one

If we consider the most significant digit of the powers of two, $2^1, 2^2, 2^3 ... 2^n$ it turns out that the frequencies are not the same for all the figures: for example, among the first 1000 powers of two the ones which start with digit 1 appear more often (30.1 %), the powers which have 2 as first digit follow (17.6 %), then the ones with 3 (12.5 %), and so on the frequencies decrease until 4.5 % for digit 9. In fact, it is possible to prove (see Section 3) that the probability for a generic term of the sequence to display d as the most significant digit is

$$\log_{10}\left(1+\frac{1}{d}\right),$$

it means that the sequence of powers of 2 obeys Benford's Law [5].

In general a sequence of real numbers represented in base b is said to be *Benford (base-b)* if the probability to observe digit d as the first digit of a term of the sequence is

$$\log_{b}\left(1+\frac{1}{d}\right),$$

for each integer d such that $1 \leq d < b$ [8]. For an overview of Benford's Law and a discussion of its possible applications see e.g. [5]. [1]

The main result we propose in this paper is a new necessary and sufficient condition to test whether a sequence is Benford (base-b) or not: we then apply this characterization to some kinds of sequences (re)obtaining, for example, that the sequence of powers of 2 is Benford (base-10) but not Benford (base-4).

We also show how the proposed characterization is related in a natural way to Birkhoff's ergodic theorem.

2 Benford (base-b) sequences: a necessary and sufficient condition

We first show how to define the most significant digit of a number by means of elementary functions. Given a real number x, we use the floor and ceiling functions defined, respectively, by $\lfloor x \rfloor = \max\{n \in \mathrm{Z} : \ n \leq x\}$ and $\lceil x \rceil = \min\{n \in \mathrm{Z} : \ n \geq x\}$. this way the fractional part of x is $x \bmod 1 = x - \lfloor x \rfloor$.

Lemma 1 *If x is a positive real number, then its first digit in base b is $\lfloor b^{(\log_b x) \mathrm{mod} 1} \rfloor$.*

Proof. Since $\lfloor \log_b x \rfloor \leq \log_b x < \lfloor \log_b x \rfloor + 1$ we have $b^{\lfloor \log_b x \rfloor} \leq x < b^{\lfloor \log_b x \rfloor + 1}$ and $1 \leq \frac{x}{b^{\lfloor \log_b x \rfloor}} < b$. Therefore $\lfloor \frac{x}{b^{\lfloor \log_b x \rfloor}} \rfloor$ is the most significant digit of x. To complete the proof it is sufficient to observe that $x = b^{\log_b x} = b^{\lfloor \log_b x \rfloor} b^{(\log_b x) \mathrm{mod} 1}$, i.e., each real number is the sum of its integer and fractional parts; this way $\frac{x}{b^{\lfloor \log_b x \rfloor}} = b^{(\log_b x) \mathrm{mod} 1}$. ◇

[1] As observed in [8] (Remark 9.2.5), the definition of Benford (base-b) sequence we use here is less restrictive than the definition of *b-Benford* sequence given for example in [1, 7].

The next result provides a possible way to count the number of terms of a sequence which display a first significant digit which is lower than d: it will be used later on to prove a general criterion to test whether a sequence is Benford (base-b) or not.

Lemma 2 *If $x_1, x_2, ...x_n$ are positive real numbers and if $1 \leq d < b$, then*

$$\sharp\{x_k : 1 \leq k \leq n, \text{ first digit of } x_k \leq d\} = \sum_{k=1}^{n} \left\lceil \frac{1}{b^{(\log_b x_k)\mathrm{mod}1}} - \frac{1}{d+1} \right\rceil .$$

Proof. According to Lemma 1, the first digit of x_k is $\leq d$ if and only if $b^{(\log_b x_k)\mathrm{mod}1} < d+1$, that is, if and only if $(\log_b x_k)\mathrm{mod}1 < \log_b(d+1)$. To complete the proof we just need to prove that the function

$$f_{d+1}(y) = \left\lceil \frac{1}{b^y} - \frac{1}{d+1} \right\rceil$$

can be rewritten as

$$f_{d+1}(y) = \begin{cases} 1 & \text{if } y \in [0, \log_b(d+1)) \\ 0 & \text{if } y \in [log_b(d+1), 1] \end{cases} .$$

In fact, function $g(y) = \frac{1}{b^y} - \frac{1}{d+1}$ is strictly decreasing on $[0,1]$ since $g'(y) = -\frac{1}{b^y}\log b < 0$. Moreover $0 < g(0) = \frac{d}{d+1} < 1$, $-1 < g(1) = \frac{1}{b} - \frac{1}{d+1} \leq 0$ and $g(\log_b(d+1)) = 0$. ◇

Now it is possible to prove the following theorem, which characterizes a Benford (base-b) sequence.

Theorem 1 *The sequence x_k of positive real numbers is Benford (base-b) if and only if*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \left\lceil \frac{1}{b^{(\log_b x_k)\mathrm{mod}1}} - \frac{1}{d+1} \right\rceil = \log_b(d+1)$$

for each integer d such that $1 \leq d < b$.

Proof. Consider a generic term x_k of the sequence. Using Lemma 2, we have that the probability that the first digit of x_k is not greater than d is given by

$$Prob(\text{ first digit of } x_k \text{ is } \leq d) = \lim_{n\to+\infty} \frac{1}{n} \sum_{k=1}^{n} \left\lceil \frac{1}{b^{(\log_b x_k)\mathrm{mod}1}} - \frac{1}{d+1} \right\rceil .$$

Observe now that the definition of Benford (base-b) sequence given in Section 1, since

$$Prob(\text{ first digit of } x_k \leq d) = \sum_{i=1}^{d} Prob(\text{ first digit of } x_k = i)$$

and

$$Prob(\text{ first digit of } x_k = i) = Prob(\text{ first digit of } x_k \leq i) - Prob(\text{ first digit of } x_k \leq i-1) ,$$

allows to claim that x_k is b-Benford if and only if

$$Prob(\text{ first digit of } x_k \le d) = \log_b(d+1)$$

for each integer d such that $1 \le d < b$. This completes the proof. $\diamond$

The theorem above, suggests a way to verify whether a sequence is Benford or not, computing a rather complicated limit, it seems to be a difficult task. The following Lemma suggests, as a way to deal with the limit, to compute a suitably defined integral: we will show how to use this idea in the next Section.

Lemma 3 *Let be d an integer such that $1 \le d < b$ and $f_{d+1}(y) = \left\lceil \frac{1}{b^y} - \frac{1}{d+1} \right\rceil$. If the sequence y_k is contained in the interval $[0,1]$ and if*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} f_{d+1}(y_k) = \int_0^1 f_{d+1}(y)dy$$

then

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \left\lceil \frac{1}{b^{y_k}} - \frac{1}{d+1} \right\rceil \;=\; \log_b(d+1)\ .$$

Proof. It is sufficient to observe that

$$f_{d+1}(y) = \left\lceil \frac{1}{b^y} - \frac{1}{d+1} \right\rceil = \begin{cases} 1 & \text{if } y \in [0, \log_b(d+1)) \\ 0 & \text{if } y \in [\log_b(d+1), 1] \end{cases} ,$$

thus $\int_0^1 f_{d+1}(y)dy = \log_b(d+1)$. $\diamond$

3 Some Benford (base-b) sequences

As a possible application of Theorem 1, we give a new kind of proof of a well known theorem, due to Diaconis [3]: a sequence of positive real numbers x_k is Benford (base-b) if the sequence of their logarithms $\log_b x_k$ is uniformly distributed modulo 1. We recall that a sequence y_k is uniformly distributed modulo 1, or equidistributed in $[0,1]$, if

$$\lim_{n\to+\infty} \frac{\sharp\{y_k :\ 1 \le k \le n,\ a \le y_k \le b\}}{n} \;=\; b-a$$

for each interval $[a,b] \subseteq [0,1]$. Another, and equivalent way to define an equidistributed sequence is the following (see for example [9, Problem 162]): the sequence y_k is equidistributed if and only if

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} f(y_k) = \int_0^1 f(y)dy \tag{1}$$

for every function f which is Riemann integrable on $[0,1]$. It is now an easy matter to prove the following well known result, stating that the exponentials of equidistributed sequences are Benford.

Theorem 2 ([3, 8]) A sequence x_k of positive real numbers is Benford (base-b) if the sequence $y_k = \log_b x_k$ is uniformly distributed modulo 1.

Proof. For a fixed integer d consider function $f_{d+1}(y) = \left\lceil \frac{1}{b^y} - \frac{1}{d+1} \right\rceil$, which is Riemann integrable on $[0, 1]$, and the sequence y_k. Applying (1) we obtain

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} f_{d+1}(y_k) = \int_0^1 f_{d+1}(y)dy \quad .$$

By Lemma 3 we have therefore

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} \left\lceil \frac{1}{b^{y_k}} - \frac{1}{d+1} \right\rceil \ = \ \log_b(d+1) \ .$$

Thus the sequence x_k is Benford (base-b) due to Theorem 1. ◇

The previous theorem allows to easily use known results on equidistributed sequences to state that related geometric sequences are Benford (base-b).

For example, the sequence $y_k = (k\alpha) \bmod 1$ is equidistributed if α is an irrational number, as independently proved by Bohl, Sierpinski and Weyl (see e.g. [8], Theorem 12.3.2); hence, the sequence $x_k = r^k$ is Benford (base-b) if $\log_b r$ is irrational ([8], Theorem 9.2.6). Incidentally, this property allows to prove the claim made at the beginning of Section 1: the sequence $x_k = 2^k$ is Benford (base-10), since $\log_{10} 2$ is irrational. The same sequence is clearly not Benford (base-4), instead, as one can easily observe [2].

A more general context where Lemma 3 applies concerns ergodic theory. Consider an ergodic endomorphism T defined on a probability space with domain $[0, 1]$, and function f_{d+1} defined as in Lemma 3. Birkhoff's ergodic theorem (see e.g. [2], Appendix 3) claims that for almost every $y_1 \in [0, 1]$,

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n} f_{d+1}(T^k(y_1)) = \int_0^1 f_{d+1}(y)dy$$

while Lemma 3 tells us that

$$\int_0^1 f_{d+1}(y)dy = \log_b(d+1) \ .$$

Thus, by Theorem 1, we obtain that a sequence x_k of positive real is Benford (base-b) if $y_k = \log_b x_k$ is generated via an ergodic endomorphism. More precisely, x_k is Benford (base-b) as soon as there exists an ergodic endomorphism T such that $T^k y_1 = y_k \ \forall k$ and the equality of Birkhoff's ergodic theorem holds for the initial term y_1 and for each integer d such that $1 \leq d < b$.

[2] The sequence $x_k = 2^k$ written using base 4 reads

$$1, 2, 10, 20, 100, 200, ...$$

References

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

[2] I.P. Cornfeld, S.V. Fomin and Ya.G. Sinai (1982), *Ergodic Theory*, Springer Verlag, New York.

[3] P. Diaconis (1977), *The Annals of Probabilty*, **5** (1), 72–81.

[4] G.H. Hardy, E.M. Wright (1979), *An Introduction to the Theory of Numbers*, fifth ed., Oxford Univ. Press, New York.

[5] T.P. Hill (1995), "The Significant-Digit Phenomenon", *The Amer. Mathematical Monthly*, **102** (4), 322–327.

[6] T.P. Hill (1995), "A Statistical Derivation of the Significant-Digit Law", *Statistical Sc.*, **10** (4), 354–363.

[7] A.V. Kontorovich, S.J. Miller (2005), "Benford's law, values of L-finctions and the 3x+1 problem", *Acta Arithmetica*, **120**, 269–297.

[8] S.J. Miller, R. Takloo-Bighash (2006), *An Invitation to Modern Number Theory*, Princeton Univ. Press, Princeton and Oxford.

[9] G. Polya, G. Szegö,(1972), *Problems and Theorems in Analysis vol.I*, Springer-Verlag, Heidelberg.