DYNAMICS AND DIGITS: ON THE UBIQUITY OF BENFORD'S LAW

A. BERGER
Institute of Mechanics, Vienna University of Technology
e-mail: arno.berger@tuwien.ac.at

The distribution of digits and mantissae in dynamical systems (both in continuous and discrete time) is discussed in light of two simple yet fundamental results. By utilizing shadowing and uniform distribution techniques, it is shown that systems with regular long-term behavior are surprisingly likely to exhibit Benford's logarithmic mantissa distribution — much in contrast to systems with stationary statistical properties. The results complement and extend recent work.

Let b be an integer larger than one. Every real $x > 0$ can be written uniquely as $x = M_b(x)b^l$ where $M_b : \mathbb{R}^+ \to [1, b[$ denotes the (*base* b) *mantissa function*, and l is the appropriate integer. For convenience set $M_b(0) = 0$. Also, let λ symbolize Lebesgue measure on $\mathbb{R}$, and, for $x \in \mathbb{R}$, denote by $\lfloor x \rfloor$ the largest integer not larger than x.

Intuitively, one might expect that mantissae M_b and first significant digits $\lfloor M_b \rfloor$ should be more or less uniformly distributed on $[1, b[$ and $\{1, \ldots, b-1\}$, respectively, for sufficiently large and diverse aggregations of numerical data. Quite often, however, this intuition is mistaken: a *logarithmic* mantissa distribution turns out to be of fundamental importance.

Definition 1. A function $f : [0, +\infty[\to \mathbb{R}$ is a b-*Benford function* if

$$\lim_{T\to\infty} T^{-1}\lambda\big(\{0 \le t \le T \,:\, M_b(|f(t)|) \le s\}\big) = \log_b s\,, \quad \forall s \in [1, b[\,;$$

it is called a (*strict*) *Benford function* if it is b-Benford for all $b \in \mathbb{N}\backslash\{1\}$.

For simplicity, real-valued sequences (x_n) will be treated as functions $t \mapsto x_{\lfloor t \rfloor}$. It is well-known[1,2] that any measurable function f is b-Benford if and only if $\big(\log_b |f(t)|\big)$ is continuously uniformly distributed (c.u.d.) mod 1.

Following an article by Benford,[3] the emergence of the logarithmic mantissa distribution, termed *Benford's Law* (*BL*), has been discussed extensively.[2,4,5] It was, however, only recently that dynamical systems have been studied as potential sources of that distribution.[1,6,7,8,9,10]

BL has been characterized as the only continuous mantissa distribution which is *base-invariant*.[4] It is natural to require that a general pattern of mantissa distribution, if one exists at all, does not depend on the particular choice of the base. Various classical systems (e.g. Lorenz flow, Hénon and logistic maps) have been studied for the emergence of BL.[9,10] Even though trajectories of such "chaotic" systems may coincidentally be Benford functions, strict Benford behavior is completely unlikely. (For simplicity the following result is stated for the one-dimensional discrete-time case only; see Ref. 7 for details and generalizations.)

Theorem 2. *Let $T:\mathbb{R}\to\mathbb{R}$ preserve the probability measure μ. Then*

$$\mu\big(\{x\,:\,\textit{the } T\textit{-orbit of } x \textit{ is strict Benford}\}\big) = 0\,.$$

A key idea in the proof is the simple observation that for any probability μ on $\mathbb{R}$ and $\Phi_b : x \mapsto \log_b |x|$ the measure $\mu\circ\Phi_b^{-1}$ cannot be uniform for *all* b. (See Ref. 11 for a related result.) Obviously, this need not be true if μ is infinite (take $d\mu = x^{-1}dx$, $x>0$ as an example), and systems preserving an infinite measure may well generate strict Benford functions.[7]

In light of Theorem 2 it is natural to focus on transient dynamics which may indeed generate Benford data in abundance.[1,8] As a generalization of earlier results[1,6], this is illustrated here through an analysis of the initial value problem,

$$\dot{x} = a_1(t) + a_2(t)x + a_3(x,t)\,, \quad x(0) = x_0\,, \tag{1}$$

where a_1, a_2 are C^0, and a_3 is C^1 such that $|a_3(x,t)| \le \Gamma(|x|)$ for all $t\ge 0$, $|x|\ge\xi\ge 0$ and a non-increasing function Γ. Let $A_2(t) = \int_0^t a_2(\tau)d\tau$. The unique solution of (1) is denoted by $(\varphi_t x_0)$ provided it exists for all $t\ge 0$.

Theorem 3. *Suppose that*

$$\int_0^\infty \big(|a_1(\tau)| + \Gamma(e^{A_2(\tau)})\big)e^{-A_2(\tau)}d\tau < \infty \quad \textit{and} \quad \inf_{\tau\ge 0} A_2(\tau) > -\infty\,. \tag{2}$$

Then there exists $\xi'\ge 0$ such that (1) *has a unique solution $(\varphi_t x_0)$ whenever $|x_0|\ge\xi'$, and $(\varphi_t x_0)$ is a b-Benford function if and only if $\big(A_2(t)/\log b\big)$ is c.u.d.* mod 1.

Basically, this follows from the fact that $h : x \mapsto x + \int_0^\infty a_3(\varphi_\tau x, \tau)e^{-A_2(\tau)}d\tau$ defines a continuous function on $\mathbb{R}\backslash[-\xi',\xi']$ with $\sup_{|x|\ge\xi'}|h(x)-x| < \infty$, and $\log e^{A_2(t)}h(x) - \log\varphi_t x$ tends to a finite limit as $t\to\infty$. The question as to whether $(\varphi_t x_0)$ is b-Benford is thus reduced to a problem of (continuous) uniform distribution theory for which extensive knowledge is on available.[12]

Remark 4. (i) Clearly, $(\varphi_t x_0)$ is b-Benford for either all or no sufficiently large $|x_0|$. This all-or-nothing situation is contrasted by fully non-linear, discrete-time systems for which (small) exceptional sets exist.[1,6]

(ii) To see that Theorem 3 is strictly stronger than all results for linearly dominated systems in Ref. 1 (which correspond to $a_1 = 0$, $\Gamma = 1$), consider the initial value problem

$$\dot{x} = -a_2(t)x\big(1 - g(x)\big)\,, \quad x(0) = x_0\,, \tag{3}$$

where $a_2(t) = (1+t)^{-1}$, and g is C^1 with $g(0) = 0$. The Benford behavior of $(\varphi_t x_0)$ for small $|x_0|$, undecidable by Ref. 1, is clarified by (a "reciprocal" version of) Theorem 3: no such solution is b-Benford for any b.

(iii) If the coefficients a_1, a_2 in (1) do not depend on t, then conditions (2) reduce to $a_2 > 0$, in which case $(\varphi_t x_0)$ is a strict Benford function for all large $|x_0|$. BL is thus a common phenomenon for those differential equations which are linearly-dominated in a sense and whose time-dependence is weak. A similar statement holds true in higher dimensions.[6]

(iv) Although Theorem 3 appears to be close to optimal for non-decreasing A_2, it is hardly surprising that interesting systems with oscillatory A_2 lie beyond the scope of that result. Take $A_2(t) = t(\sin t)^2$ in (3) as an example: For $g = 0$, $(\varphi_t x_0)$ is strict Benford if $x_0 \neq 0$, but $\int_0^\infty e^{-A_2(\tau)}d\tau$ diverges, and the general case $g \neq 0$ remains open.

References

1. Berger, A., Bunimovich, L. A. and Hill, T.P. (2003) One-dimensional Dynamical Systems and Benford's Law. To appear in *Trans. Amer. Math. Soc.*
2. Diaconis, P. (1979) The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* **5**, 72–81.
3. Benford, F. (1938) The law of anomalous numbers. *Proceedings of the American Philosophical Society* **78**, 551–572.
4. Hill, T. (1995) Base-invariance implies Benford's Law. *Proc Amer. Math. Soc.* **123**, 887–895.
5. Raimi, R. (1976) The first digit problem. *Amer. Math. Monthly* **102**, 322–327.
6. Berger, A. (2002) Multi-dimensional Dynamical Systems and Benford's Law. To appear in *Trans. Amer. Math. Soc.*
7. Berger, A. (2003) Base invariance and Benford's Law in dynamics. Preprint.
8. Gottwald, G. and Nicol, M. (2001) On the nature of Benford's Law. *Phys. A* **303**, 387–396.
9. Snyder, M., Curry, J. and Dougherty, A. (2001) Stochastic aspects of one-dimensional discrete dynamical systems: Benford's law. *Physical Review E* **64**, 1–5.
10. Tolle, C., Budzien, J. and LaViolette, R. (2000) Do dynamical systems follow Benford's law? *Chaos* **10**, 331–337.
11. Schatte, P. (1981) On Random Variables with Logarithmic Mantissa Distribution Relative to Several Bases. *Elektron. Informationsverarb. Kybernet.* **17**, 293–295.
12. Kuipers, L. and Niederreiter, H. (1974) *Uniform Distribution of Sequences.* Wiley, New York.