From Uniform Distributions to Benford's Law

Élise Janvresse, Thierry de la Rue

Abstract

We provide a new probabilistic explanation for the appearance of Benford's law in everyday-life numbers, by showing that it arises naturally when we consider mixtures of uniform distributions. Then we connect our result to the theorem of B. J. Flehinger ("On the probability that a random integer has initial digit A", *Amer. Math. Monthly*, 73:1056–1061, 1966), for which we provide a shorter proof and a speed of convergence.

Key-words: Benford's law, first-digit law, mantissa, uniform distribution, coupling method.
Classification AMS 2000: 60J10, 11K99

1   Introduction

1.1   Benford's law

We define the *mantissa* (base 10) of a positive real number x as the unique real number $\mathscr{M}(x) \in [1, 10[$, such that

$$x = \mathscr{M}(x)10^k,$$

for some integer $k \in \mathbb{Z}$. Benford's law describes the probability distribution of the mantissa: more precisely, it says that the proportion of numbers $x > 0$ which satisfy $\mathscr{M}(x) \in [a, b[$ is, for any $1 \le a < b \le 10$,

$$P_{Benford}([a, b[) := \log_{10} b - \log_{10} a. \tag{1}$$

In its most popular form, Benford's law is stated in the particular case where $[a, b[= [i, i + 1[$ for some $i \in \{1, \dots, 9\}$, and it gives the proportion of numbers whose first significant digit D_1 is i:

$$\log_{10}\left(1 + \frac{1}{i}\right).$$

Here we must point out that the mathematical meaning of the word "proportion" is not well defined: in fact, Benford's law is just the description of the distribution of the significant digits observed in large sets of empirical data. The astronomer Simon Newcomb was the first to observe that the probability of occurrence of digits is not uniform. The physicist Franck Benford, unaware of Newcomb's article ([8]), empirically rediscovered the law some 57 years later ([1]) and popularized it.

Because it is not clear which law arbitrary data should follow, Benford's law is quite difficult to prove! However, many mathematicians have managed to give various explanations to the natural appearance of Benford's law in everyday-life numbers. In particular, we have to mention scale-invariance (R. S. Pinkham, [9]) and base-invariance (T. Hill [5, 6]). R. A. Raimi's survey ([10]) also provides several probabilistic interpretations, such as iterations of a mixture process, which amounts to consider products of independent random variables. Berger, Bunimovich and Hill ([2]) have also shown that Benford's law arises when we consider orbits of a large class of dynamical systems. Surprisingly, in spite of the abundance of the litterature about Benford's law, it seems that our characterization, linking Benford's law to mixture of uniform distributions, has never been proposed yet.

1.2   Heuristics

As mentioned by Raimi, Benford had collected thousands of numbers from twenty different tables. Some of them obeyed Benford's law rather badly, while others, such as the street addresses of the first 342 persons in *American Men of Science*, did better. As noticed by Benford, the union of his tables is what came closest to the predicted law: the more the datas come from various sources, the better they fit the law. We propose to model these different sources by uniformly distributed random variables bounded by an unknown maximum S depending on the source. The first reason is that it is natural to introduce uniform distributions to describe datas on which we do not know anything. Besides, we can find many cases where the distribution of empirical datas are naturally modelized by some mixture of uniform distributions. A simple one is constituted by all the numbers describing the month and the day of the month for the birthdays of a list of people: The month is well approximated by a uniform random variable taking its values in $\{1, 2, \ldots, 12\}$, and given the month the day itself is uniformly distributed in $\{1, 2, \ldots, (\text{number of days in this month})\}$. The example of street addresses also presents this kind of phenomenon: Imagine you only know that the numbers of the addresses in a given street vary between 1 and S. By the principle of indifference, picking a random address in this street naturally gives rise to a random variable uniformly distributed between 1 and S. In other words, conditioned on the highest number S in the street, the street numbers follow a uniform distribution on $\{1, \ldots, S\}$. So the set of street numbers used by Benford can be seen as a mixture of uniform distributions, weighted by the law of the highest number of a street. A third example is the first-page numbers of articles in a bibliography, which, conditioned on the size S of the volume, can be considered uniformly distributed. Of course, not every set of empirical datas is well described by uniform distributions. However it is easy to check that averaging over many different such sets with varying laws amounts to considering uniformly distributed variables.

What can we expect for the distribution of mantissae in such a model? First, it is easy to observe that if X is uniformly distributed in $[0, S]$, the law of $\mathscr{M}(X)$ only depends on $\mathscr{M}(S)$ (see (2) below). Therefore if we want to study the law of $\mathscr{M}(X)$ when X is distributed according to a mixture of

uniform distributions, it only remains to answer the following question: Which law should the mantissa of the source-depending maximum S follow? Now, let's assume that each time we collect a huge number of datas coming from numerous origins, their mantissae are distributed according to a fixed distribution. Then, since the maxima themselves come from various origins, we expect both the mantissae of the whole datas and those of the maxima to conform to this fixed distribution.

1.3 Brief description of the content

In Section 2, we derive from the preceding heuristics an equation which should be satisfied by the law of the mantissa of X in our model. Then we prove that it characterizes Benford's law (Theorem 2.1). In Section 3, we construct a Markov chain $(M_n)_n$ taking its value in $[1, 10[$, such that M_{n+1} conditioned on M_n follows the law of the mantissa of a uniformly distributed random variable in $[0, M_n]$. By coupling techniques, we show that the law of M_n converges exponentially fast to Benford's law, which also provides an alternative proof of Theorem 2.1. In the last section we use our results to give a simpler, probabilistic proof of Flehinger's Theorem about the initial digit of a random integer, together with a speed of convergence.

It is to be noticed that our argument is given for the base 10, but carries over automatically to other bases.

2 Characterization of Benford's law via uniform distributions

In the sequel, we place ourselves on a probability space $(\Omega, \mathscr{A}, P)$. For any $S > 0$, we denote by U_S the uniform distribution in $[0, S]$. Let us suppose X is a random variable with law U_S. We can compute the probability distribution of $\mathscr{M}(X)$ as a function of S. Let k be the greatest integer such that $10^k \leq S$, so that $S = \mathscr{M}(S)10^k$. We have

$$
\begin{aligned}
P(\mathscr{M}(X) \leq t) &= P(\mathscr{M}(X) \leq t \text{ and } X \leq 10^k) + P(\mathscr{M}(X) \leq t \text{ and } X > 10^k) \\
&= P(X \leq 10^k)P(\mathscr{M}(X) \leq t | X \leq 10^k) + P(10^k < X \leq t10^k).
\end{aligned}
$$

Conditionally to $X \leq 10^k$, $\mathscr{M}(X)$ is uniformly distributed in $[1, 10[$. Therefore, the probability distribution of $\mathscr{M}(X)$ only depends on $\mathscr{M}(S)$ and is given, for $t \in [1, 10[$, by

$$
P(\mathscr{M}(X) \leq t) = \begin{cases} \dfrac{1}{\mathscr{M}(S)} \dfrac{t-1}{9} + \dfrac{t-1}{\mathscr{M}(S)} & \text{if } t \leq \mathscr{M}(S), \\ \dfrac{1}{\mathscr{M}(S)} \dfrac{t-1}{9} + \dfrac{\mathscr{M}(S)-1}{\mathscr{M}(S)} & \text{if } t \geq \mathscr{M}(S). \end{cases} \tag{2}
$$

When S varies, this quantity oscillates between $\frac{t-1}{9}$ and $\frac{10}{9}\frac{t-1}{t}$. As we proposed in the previous section, our modelization consists in regarding S as a random variable. That is to say, we now have two random variables X and S: X

represents the generic data in our collection and S is the source-depending maximum, so that X conditioned on S follows the uniform distribution U_S.

Let's suppose that the mantissa of S follows some probability distribution μ. Then, as mentioned at the end of section 1.2, we expect the mantissa of X to behave in the same way. Therefore, for any $t \in [1, 10[$, the probability $\mu([1, t])$ that $\mathscr{M}(X) \leq t$ should satisfy

$$\mu([1, t]) \;=\; \int_1^{10} P\Big(\mathscr{M}(X) \leq t \mid \mathscr{M}(S) = a\Big)\, d\mu(a).$$

Using (2), we get

$$\mu([1, t]) = \int_1^t \left(1 - \frac{t}{y}\right) d\mu(y) + \frac{10}{9}(t - 1) \int_1^{10} \frac{d\mu(y)}{y}. \tag{3}$$

Theorem 2.1 *Benford's law is the unique probability distribution satisfying equation (3). In other words, Benford's law is the unique probability distribution μ on $[1, 10[$ such that, if $\mathscr{M}(S)$ follows μ and X conditioned on S is uniformly distributed on $[0, S]$, then $\mathscr{M}(X)$ still follows μ.*

Proof $-$ Let μ be a probability measure on $[1, 10[$ satisfying (3). Considering the measure ν defined by $\frac{d\nu}{d\mu}(y) = 1/y$, (3) can be rewritten as

$$0 \;=\; -t\nu([1, t]) + \frac{10}{9}(t - 1)\nu([1, 10[).$$

Therefore,

$$\nu([1, t]) \;=\; \frac{10}{9}\frac{t - 1}{t}\nu([1, 10[),$$

hence ν has density proportional to y^{-2} with respect to the Lebesgue measure. Since

$$\frac{d\mu}{d\nu}(y) \;=\; y,$$

we get that μ has density proportional to y^{-1} with respect to the Lebesgue measure. Therefore μ is equal to Benford's law.

Conversely, an easy computation yields

$$\int_1^t \left(1 - \frac{t}{y}\right) dP_{Benford}(y) + \frac{10}{9}(t - 1) \int_1^{10} \frac{dP_{Benford}(y)}{y}$$
$$= \log_{10}(t) = P_{Benford}([1, t]). \tag{4}$$

This means Benford's law is the unique solution to (3). $\qquad\qquad \Box$

3   Construction of a Markov chain

Following the method developed in [3], we will now construct a Markov chain $(M_n)_n$ taking its value in $[1, 10[$, such that M_{n+1} conditioned on M_n follows the law of the mantissa of a uniformly distributed random variable in $[0, M_n]$.

Let $(U_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ be two independent sequences of independent random variables uniformly distributed in $[0,1]$. We consider $F : [1,10[\times[0,1]^2 \to [1,10[$, given by

$$F(m,u,v) = \begin{cases} um & \text{if } um \in [1,10[, \\ \mathscr{M}(v) & \text{otherwise.} \end{cases}$$

We now define a Markov chain $(M_n)_{n\in\mathbb{N}}$ on $[1,10[$ starting from M_0 by

$$M_n = F(M_{n-1}, U_n, V_n), \quad \text{for all positive integers } n.$$

Observe that the transition probability of our Markov chain $(M_n)_{n\in\mathbb{N}}$ is given, for $t \geq 1$, by

$$\begin{aligned} P(M_n \leq t \mid M_{n-1}) &= P(F(M_{n-1}, U_n, V_n) \leq t \mid M_{n-1}) \\ &= P(\mathscr{M}(V_n) \leq t, \ M_{n-1}U_n \in [0,1[\mid M_{n-1}) \\ &\quad + P(M_{n-1}U_n \leq t, M_{n-1}U_n \in [1,10[\mid M_{n-1}). \end{aligned}$$

Since U_n and V_n are independent and uniformly distributed in $[0,1]$, this expression can be rewritten as

$$P(\mathscr{M}(V_n) \leq t) \, P(M_{n-1}U_n \in [0,1[\mid M_{n-1})$$
$$+ \frac{t-1}{M_{n-1}}\mathbb{1}_{M_{n-1}\geq t} + \frac{M_{n-1}-1}{M_{n-1}}\mathbb{1}_{M_{n-1}<t},$$

which yields

$$P(M_n \leq t \mid M_{n-1}) = \begin{cases} \dfrac{t-1}{9M_{n-1}} + \dfrac{t-1}{M_{n-1}} & \text{if } t \leq M_{n-1}, \\ \dfrac{t-1}{9M_{n-1}} + \dfrac{M_{n-1}-1}{M_{n-1}} & \text{otherwise.} \end{cases} \tag{5}$$

Comparing with (2), we see that the law of M_n given M_{n-1} is exactly the law of $\mathscr{M}(X)$, where X is a uniform random variable on $[0, M_{n-1}]$. Therefore, a probability measure is invariant for this Markov chain if and only if it satisfies (3). This is the case of Benford's law.

Proposition 3.1 *Let μ be an invariant measure for the Markov chain with transition probability (5). Then, for any $a \in [1,10[$ and any $B \subset [1,10[$,*

$$|P(M_n^a \in B) - \mu(B)| \leq (9/10)^n,$$

where $(M_n^a)_n$ is the Markov chain with transition probability (5) and starting from a.
Consequently, Benford's law is the unique invariant probability measure for the Markov chain with transition probability (5).

Proof – To prove unicity of the invariant measure and obtain an estimate of the speed of convergence, we will use a coupling method inspired by [3],

[11]. We consider two chains M_n^a and M_n^b with same transition probability and starting from a and b respectively: $M_0^a = a$, $M_0^b = b$, and for any $n \geq 1$,

$$M_n^a = F(M_{n-1}^a, U_n, V_n), \quad M_n^b = F(M_{n-1}^b, U_n, V_n).$$

We point out the fact that the *same* sequences $(U_n)_{n \geq 1}$ and $(V_n)_{n \geq 1}$ are used in the definition of both M_n^a and M_n^b. Let $\tau^{a,b}$ be the coupling time, that is the first time the two chains M_n^a and M_n^b meet:

$$
\begin{aligned}
\tau^{a,b} &= \min\{n \geq 1 | M_n^a = M_n^b\} \\
&\leq \min\{n \geq 1 | U_n M_{n-1}^a < 1; U_n M_{n-1}^b < 1\}.
\end{aligned}
$$

Notice that for any $n \geq \tau^{a,b}$, $M_n^a = M_n^b$. Moreover, we can check $P(\tau^{a,b} > n)$ decreases exponentially fast in n:

$$
\begin{aligned}
P(\tau^{a,b} > n) &\leq P(U_k \max(M_{k-1}^a; M_{k-1}^b) \geq 1, \ \forall 1 \leq k \leq n) \\
&\leq P(10\, U_k \geq 1, \ \forall 1 \leq k \leq n) \\
&= (9/10)^n.
\end{aligned}
$$

Let μ be an invariant probability distribution for the Markov chain. Then, for any $B \subset [1, 10[$,

$$
\begin{aligned}
|P(M_n^a \in B) - \mu(B)| &= \left| \int_1^{10} \left(P(M_n^a \in B) - P(M_n^b \in B) \right) d\mu(b) \right| \\
&\leq \int_1^{10} d\mu(b) E\left[\left| \mathbb{1}_{M_n^a \in B} - \mathbb{1}_{M_n^b \in B} \right| \right] \\
&\leq \int_1^{10} d\mu(b) E\left[\mathbb{1}_{\tau^{a,b} > n} \right] \\
&\leq \sup_{1 \leq b < 10} P(\tau^{a,b} > n) \ \leq \ (9/10)^n.
\end{aligned}
$$

Hence,

$$\sup_{a,B} |P(M_n^a \in B) - \mu(B)| \leq (9/10)^n. \tag{6}$$

We have already seen in (4) that Benford's law is invariant for the Markov chain. Hence, (6) proves that for each $a \in [1, 10[$, the law of M_n^a converges exponentially fast to Benford's law, which is therefore the unique invariant probability distribution. ❏

4   A probabilistic proof of Flehinger's theorem

In [4], Flehinger is interested in the distribution of the first significant digit of a random number in $\mathbb{N} \setminus \{0\}$. She tries to make sense of the heuristic question "What proportion of the positive integers have their initial digit less than or equal to i, for $i \in \{1, \ldots, 9\}$?" According to Benford's law, this should happen with probability

$$P_{Benford}([1, i+1[) = \log_{10}(1+i).$$

The set L_i of the positive integers with initial digit less than i has no natural density among positive integers:

$$P_n^1(i) = \frac{1}{n}|L_i \cap \{1, 2, \ldots, n\}| = \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{L_i}(m) \tag{7}$$

oscillates between $i/9$ and $10\,i/9(i+1)$ as n varies. Flehinger thus proposes to iterate this averaging process (Cesaro average). For $k \geq 1$, she considers

$$P_n^k(i) = \frac{1}{n}\sum_{m=1}^{n} P_m^{k-1}(i).$$

She then proves that the amplitude of the oscillations of the functions $P_n^k(i)$ decreases and the averaging process converges to Benford's law in the following way.

Theorem 4.1 (Flehinger)

$$\lim_k \liminf_n P_n^k(i) = \lim_k \limsup_n P_n^k(i) = \log_{10}(1 + i).$$

Now, we will show how $P_n^k(i)$ corresponds to k steps of our Markov chain with transition (5). Heuristically, this appears quite natural when considering that the iteration of the Cesaro averaging in Flehinger's process amounts to a repeated, inductive drawing of uniform discrete random variables.

Note that (7) works out to

$$P_n^1(i) = \begin{cases} \dfrac{i}{9n}(10^{j+1} - 1) & \text{if } n \in [(i+1)10^j, 10^{j+1}[\text{ for } j \in \mathbb{N}, \\ 1 - \dfrac{(9-i)(10^j - 1)}{9n} & \text{if } n \in [10^j, (i+1)10^j[\text{ for } j \in \mathbb{N}. \end{cases}$$

Let us consider integers n with a "fixed" mantissa: this means that we fix a real $a \in [1, 10[$, and we consider n of the form $\lfloor a10^j \rfloor$ (integer part of $a10^j$).

$$P_{\lfloor a10^j \rfloor}^1(i) = \begin{cases} \dfrac{i}{9\mathscr{M}(\lfloor a10^j \rfloor)} \dfrac{10^{j+1} - 1}{10^j} & \text{if } \mathscr{M}(\lfloor a10^j \rfloor) \geq i + 1, \\ 1 - \dfrac{(9 - i)}{9\mathscr{M}(\lfloor a10^j \rfloor)} \dfrac{10^j - 1}{10^j} & \text{if } \mathscr{M}(\lfloor a10^j \rfloor) < i + 1. \end{cases}$$

As j goes to infinity, the above term converges to

$$Q^1(a, i) = \begin{cases} \dfrac{10i}{9a} & \text{if } a \geq i + 1, \\ 1 - \dfrac{(9 - i)}{9a} & \text{if } a < i + 1. \end{cases}$$

We recognize the expression (2) with $t = i + 1$ and $S = a10^j$. Hence,

$$Q^1(a, i) = \lim_{j \to \infty} P_{\lfloor a10^j \rfloor}^1(i) = P(M_1^a < i + 1),$$

where M_1^a has been defined in the statement of Proposition 3.1.

In the same way, Flehinger defines $Q^k(a, i)$ by

$$Q^k(a, i) = \lim_{j \to \infty} P^k_{\lfloor a10^j \rfloor}(i),$$

and proves (see [4], p. 1059) that for any $k \geq 2$,

$$Q^k(a, i) = \frac{1}{a} \left[\frac{1}{9} \int_1^{10} Q^{k-1}(u, i) du + \int_1^a Q^{k-1}(u, i) du \right]. \tag{8}$$

Proposition 4.2 *For all $k \geq 1$ and all $a \in [1, 10[$,*

$$Q^k(a, i) = P(M_k^a < i + 1),$$

where M_k^a has been defined in the statement of Proposition 3.1.

Proof – We already have the result for $k = 1$. It is then enough to establish that $P(M_k^a < i + 1)$ satisfies a recursive equation analogous to (8). For $k \geq 2$, conditioning on M_1^a and denoting by μ_1 the law of M_1^a, we have

$$P(M_k^a < i + 1) = \int_1^{10} P(M_k^a < i + 1 | M_1^a = u) \, d\mu_1(u).$$

¿From (5), we can write

$$d\mu_1(u) = \begin{cases} \frac{1}{9a} du & \text{if } u \geq a, \\ \left(\frac{1}{9a} + \frac{1}{a} \right) du & \text{if } u < a. \end{cases}$$

Finally, using the fact that

$$P(M_k^a < i + 1 | M_1^a = u) = P(M_{k-1}^u < i + 1),$$

we get

$$P(M_k^a < i + 1) = \frac{1}{a} \left[\frac{1}{9} \int_1^{10} P(M_{k-1}^u < i + 1) \, du + \int_1^a P(M_{k-1}^u < i + 1) \, du \right]$$

❏

By Proposition 2.1 we obtain an estimation on the speed of convergence of Flehinger's averaging process.

Corollary 4.3

$$\sup_{i, a} \left| Q^k(a, i) - \log_{10}(1 + i) \right| \leq (9/10)^k.$$

Since $\liminf_n P_n^k(i) = \min_{1 \leq a < 10} Q^k(a, i)$ and $\limsup_n P_n^k(i) = \max_{1 \leq a < 10} Q^k(a, i)$, this implies in particular the result stated in Theorem 4.1.

Remark

In [7], D.E. Knuth proves a slight generalization of Flehinger's theorem, stating that the proportion of the positive integers whose mantissa is less than $r \in [1, 10[$ is $\log_{10} r$. More precisely, Knuth considers

$$P_n^1(\mathscr{M} \leq r) \; = \; \frac{1}{n}|\{m \in \{1, \ldots, n\} : \; \mathscr{M}(m) \leq r\}|,$$

and iterates a similar averaging process, setting inductively

$$P_n^k(\mathscr{M} \leq r) = \frac{1}{n} \sum_{m=1}^{n} P_m^{k-1}(\mathscr{M} \leq r).$$

Knuth's result states that

$$\lim_k \liminf_n P_n^k(\mathscr{M} \leq r) = \lim_k \limsup_n P_n^k(\mathscr{M} \leq r) = \log_{10}(r).$$

A straightforward adaptation of our argument above also gives a speed of convergence in that case : For all $k \geq 1$

$$\sup_{r \in [1,10[} \left| \liminf_n P_n^k(\mathscr{M} \leq r) - \log_{10}(r) \right| \; \leq \; (9/10)^k,$$

and

$$\sup_{r \in [1,10[} \left| \limsup_n P_n^k(\mathscr{M} \leq r) - \log_{10}(r) \right| \; \leq \; (9/10)^k.$$

Acknowledgements

References

[1] F. BENFORD. The law of anomalous numbers. *Proc. Amer. Phil. Soc.*, 78:551–572, 1938.

[2] A. BERGER, L. BUNIMOVICH and T. HILL. One-dimensional dynamical systems and Benford's Law. To appear in *Trans. Amer. Math. Soc.*

[3] P. A. FERRARI, A. GALVES. Coupling and regeneration for stochastic processes Notes for a minicourse presented in XIII Escuela Venezolana de Matematicas, 2000. http://www.ime.usp.br/~pablo/publications.html

[4] B.J. FLEHINGER. On the probability that a random integer has initial digit A. *Amer. Math. Monthly*, 73:1056–1061, 1966.

[5] T. HILL. Base-invariance implies Benford's law. *Proc. Amer. Math. Soc.*, 123:887–895, 1995.

[6] T. HILL. A statistical derivation of the significant-digit law. *Stat. Science*, 10:354–363, 1996.

[7] D. KNUTH. *The Art of Computer Programming - Vol. 2*. Addison-Wesley Publishing Company, 1981.

[8] S. NEWCOMB. Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math.*, 4:39–40, 1881.

[9] R.S. PINKHAM. On the distribution of first significant digits. *Ann. Math. Statist.*, 32:1223–1230, 1961.

[10] R.A. RAIMI. The first digit problem. *Amer. Math. Monthly*, 83:521–538, 1976.

[11] H. THORISSON. *Coupling, Stationarity, and Regeneration.* Probability and its Applications, Springer-Verlag, New York, 2000.

Élise Janvresse, Thierry de la Rue
Université de Rouen, LMRS, UMR 6085 - CNRS
76 821 Mont Saint Aignan, France
Elise.Janvresse@univ-rouen.fr
thierry.de-la-rue@univ-rouen.fr