Benford or not Benford: a systematic but not always well-founded use of an elegant law in experimental fields

Stéphane Blondeau da Silva

Benford or not Benford: a systematic but not always well-founded use of an elegant law in experimental fields

Blondeau Da Silva Stéphane

June 11, 2018

Abstract

In this paper, we will see that the proportion of d as leading digit, $d \in [\![1, 9]\!]$, in data (obtained thanks to the hereunder developed model) is more likely to follow a law whose probability distribution is determined by a specific upper bound, rather than Benford's Law. These probability distributions fluctuate around Benford's value as can often be observed in the literature in many naturally occurring collections of data (where the physical, biological or economical quantities considered are upper bounded). Knowing beforehand the value of the upper bound can be a way to find a better adjusted law than Benford's one.

Introduction

Benford's Law, also called Newcomb-Benford's Law, is noteworthy to say the least: according to it, the first digit d, $d \in [\![1, 9]\!]$, of numbers in many naturally occurring collections of data does not follow a discrete uniform distribution, as might be thought, but a logarithmic distribution. Discovered by the astronomer Newcomb in 1881 ([13]), this law was definitively brought to light by the physicist Benford in 1938 ([2]). He proposed the following probability distribution: the probability for d to be the first digit of a number seems to be equal to $\log(d+1) - \log(d)$, *i.e.* $\log(1 + \frac{1}{d})$. Benford tested it over data set from 20 different domains (surface areas of rivers, sizes of american populations, physical constants, molecular weights, entries from a mathematical handbook, numbers contained in an issue of Reader's Digest, the street addresses of the first persons listed in American Men of Science, death rates, *etc.*). Most of the empirical data, as physical data (Knuth in [12] or Burke and Kincanon in [5]), economic and demographic data (Nigrini and Wood in [14]) or genome data (Friar *et al.* in [7]), follow approximately Benford's Law. To such an extent that this law is used to detect possible frauds in lists of socio-economic data ([19]) or in scientific publications ([6]).

First restricted to the experimental field, it is now established that this law holds for various mathematical sequences (see for example [3]). In the situation, where the distribution of first digits is scale, unit or base invariant, this distribution is always given by Benford's Law ([15] and [9]). Selecting different samples in different populations, under certain constraints, leads also

to construct a sequence that follows the Benford's Law ([10]). Furthermore independant variables multiplication conducts to this law ([4]). One might add that some sequences satisfy Benford's Law exactly (for example see [17],[20] or [11]).

We can note that there also exist distributions known to disobey Benford's Law ([16] and [1]). And even concerning empirical data sets, this law appears to be a good approximation of the reality, but no more than an approximation ([8]).

In the model we build in the article, the naturally occurring data will be considered as the realizations of independant random variables following the hereinafter constraints: (a) the data is strictly positive and is upper-bounded by an integer n, constraint which is often valid in data sets, the physical, biological and economical quantities being limited ; (b) each random variable is considered to follow a discrete uniform distribution whereby the first i strictly positive integers are equally likely to occur (i being uniformly randomly selected in $[\![1, n]\!]$). This model relies on the fact that the random variables are not always the same. The article is divided into two parts. In the first one, we will accurately study the case where the leading digit is 1. In the second one, we will generalize our results to the eight last cases.

Through this article we will demonstrate that the predominance of 1 as first digit (followed by those of 2 and so on) is all but surprising, and that the observed fluctuations around the values of probability determined by Benford's Law are also predictible. The point is that, since 1938, Benford's Law probabilities became standard values that should exactly be followed by most of naturally occurring collections of data. However the reality is that the proportion of each d as leading digit, $d \in [\![1, 9]\!]$, structurally fluctuates. There is not a single Benford's Law but numerous distinct laws that we will hereafter examine.

1 The chosen probability space

1.1 Notations

In order to determine the proportion of numbers whose leading digit is $d \in [\![1, 9]\!]$, we will first build our probability space and further explain the model we choose.

Let i be a strictly positive integer. Let $\mathcal{U}_{\{i\}}$ denote the discrete uniform distribution whereby the first i strictly positive integers are equally likely to be observed.

Let n be a strictly positive integer. Let us consider the random experiment $\mathcal{E}_n$ of tossing two independent dice. The first one is a fair n-sided die showing n different numbers from 1 to n. The number i rolled on it defines the number of faces on the second die. It thus shows i different numbers from 1 to i.

Let us now define the probability space Ω_n as follows: $\Omega_n = \{(i, j) : i \in [\![1, n]\!] \text{ and } j \in [\![1, i]\!]\}$. Our probability measure is denoted by P.

Let us denote by L_n the random variable from Ω_n to $[\![1, 9]\!]$ that maps each element ω of Ω_n to the leading digit of the second component of ω.

1.2 Why such a model?

Let us imagine a perfect consumer shopping in a perfectly structured supermarket: (a) in the i^{th} (i being a strictly positive integer) section of this supermarket, the products prices range between 1 and i cents of the considered currency; (b) the prices in a section are uniformly distributed; (c) each section contains the same quantity of products and (d) the consumer randomly chooses his products in the whole store.

Under these constraining hypotheses, these perfect entities enable us to use our model. Note that conditions (c) and (d) gathered avoid us to conduct a double drawing every time: first the section then the product. In that respect, the sales receipt will verify the following results in terms of proportion of d as leading digit, $d \in [\![1, 9]\!]$.

Among the different domains studied by Benford ([2]), some could be well adapted to our model: sizes of populations (sections here gathering all the populations having the same usable areas, the geographic constraints preventing the surface area to be broader; populations being not neccessary settled on the entire area, their sizes fluctuate) or street adresses for example (sections here gathering the adresses of a selected street; the lenght of the considered streets might be uniformly distributed to fit model criteria).

Hence the defined model is relevant when the studied data can be considered as realizations of a homogeneous and expanded range of random variables approximately following discrete uniform distributions.

2 Proportion of d

Through the below proposition, we will express the probability $\mathrm{P}(L_n = d)$, for each $n \in \mathbb{N}^*$, *i.e.* the probability that the leading digit of our second throw in our random experiment is d.

Proposition 2.1. *Let k denote the positive integer such that $k = \min\{i \in \mathbb{N} : d \times 10^i > n\}$. If $n < (d+1) \times 10^{k-1}$, the value of $\mathrm{P}(L_n = d)$ is:*

$$\frac{1}{n}\Big(\sum_{l=0}^{k-2}\Big(\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1} \frac{b - \frac{(9d-1)\times 10^l - 8}{9}}{b} + \sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1} \frac{\frac{10^{l+1}-1}{9}}{a}\Big) + \sum_{b=d\times 10^{k-1}}^{n} \frac{b - \frac{(9d-1)\times 10^{k-1} - 8}{9}}{b}\Big).$$

Otherwise the value of $\mathrm{P}(L_n = d)$ is:

$$\frac{1}{n}\Big(\sum_{l=0}^{k-2}\Big(\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1} \frac{b - \frac{(9d-1)\times 10^l - 8}{9}}{b} + \sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1} \frac{\frac{10^{l+1}-1}{9}}{a}\Big) + \sum_{b=d\times 10^{k-1}}^{(d+1)\times 10^{k-1}-1} \frac{b - \frac{(9d-1)\times 10^{k-1} - 1}{9}}{b} + \sum_{a=(d+1)\times 10^{k-1}}^{n} \frac{\frac{10^k-1}{9}}{a}\Big).$$

Proof. Let us denote by D_n the random variable from Ω_n to $[\![1, n]\!]$ that maps each element ω of Ω_n to the first component of ω. It returns the number obtained on the first throw of the unbiased n-sided die. For each $i \in [\![1, n]\!]$, we have:

$$\mathrm{P}(D_n = i) = \frac{1}{n}\,. \tag{1}$$

According to the Law of total probability, we state:

$$\mathrm{P}(L_n = d) = \sum_{i=1}^{n} \mathrm{P}(L_n = d | D_n = i)\, \mathrm{P}(D_n = i)\,. \tag{2}$$

Thereupon two cases appear in determining the value of $\mathrm{P}(L_n = d|D_n = i)$, for $i \in [\![1, n]\!]$. Let us study the first case where the leading digit of i is d. Let k_i be the positive integer such that $k_i = \min\{k \in \mathbb{N} : d \times 10^k > i\}$ in both cases. Among the first $d \times 10^{k_i-1} - 1$ non-zero integers (all lower than i), the number of integers whose leading digit is d is (if $k_i \geq 2$):

$$\sum_{t=0}^{k_i-2} 10^t = 1 \times \frac{1 - 10^{k_i-1}}{1 - 10} = \frac{10^{k_i-1} - 1}{9}\,.$$

This equality still holds true for $k_i = 1$. From $d \times 10^{k_i-1}$ to i, there exist $i - d \times 10^{k_i-1} + 1$ additional integers whose leading digit is d. It may be inferred that:

$$\mathrm{P}(L_n = d|D_n = i) = \frac{1}{i}(\frac{10^{k_i-1} - 1}{9} + i - d \times 10^{k_i-1} + 1) = \frac{i - \frac{(9d-1)10^{k_i-1}-8}{9}}{i}, \tag{3}$$

the leading digit of i being here d.

In the second case, we consider the integers i whose leading digits are different from d. Among the first $(d+1) \times 10^{k_i-1} - 1$ non-zero integers (i is greater than or equal to $(d+1) \times 10^{k_i-1}$), the number of integers whose leading digit is d is:

$$\sum_{t=0}^{k_i-1} 10^t = \frac{10^{k_i} - 1}{9}\,.$$

From $2 \times 10^{k_i-1}$ to i, there exists no additional integers whose leading digit is d. It can be concluded that:

$$\mathrm{P}(L_n = d|D_n = i) = \frac{\frac{10^{k_i}-1}{9}}{i}, \tag{4}$$

the leading digit of i being here different from d.
Using equalities (1), (2), (3) and (4), we get our result. □

For example, we get:

Examples 2.2. If $n = 20$, we have $k = 2$. The value of $\mathrm{P}(L_{20} = 1)$ is then (second case of Proposition 2.1):

$$\begin{aligned}
\mathrm{P}(L_{20} = 1) &= \frac{1}{20}\Big(\frac{1 - 8\frac{10^0-1}{9}}{1} + \sum_{a=2}^{9} \frac{\frac{10^{0+1}-1}{9}}{a} + \sum_{b=10}^{19} \frac{b - 8\frac{10^1-1}{9}}{b} + \frac{\frac{10^2-1}{9}}{20}\Big) \\
&= \frac{1}{20}\Big(1 + \frac{1}{2} + ... + \frac{1}{9} + \frac{2}{10} + ... + \frac{11}{19} + \frac{11}{20}\Big) \\
&\approx 0.381\,.
\end{aligned}$$

The value of $\mathrm{P}(L_{86}=8)$ is (first case of Proposition 2.1):

$$\begin{aligned}\mathrm{P}(L_{86}=8) &= \frac{1}{806}\Big(\sum_{l=0}^{1}\big(\sum_{b=8\times 10^l}^{9\times 10^l-1}\frac{b-\frac{(9\times 8-1)10^l-8}{9}}{b}+\sum_{a=9\times 10^l}^{8\times 10^{l+1}-1}\frac{\frac{10^{l+1}-1}{9}}{a}\big)+\sum_{b=800}^{806}\frac{b-\frac{71\times 10^2-8}{9}}{b}\Big)\\ &= \frac{1}{86}\Big(\frac{1}{8}+\frac{1}{9}+\ldots+\frac{1}{79}+\frac{2}{80}+\ldots+\frac{11}{89}+\frac{11}{90}+\ldots+\frac{11}{799}+\frac{12}{800}+\ldots+\frac{18}{806}\Big)\\ &\approx 0,034\,.\end{aligned}$$

3 Study of two subsequences

It is natural that we take a specific look at the values of n positioned just before a long sequence of numbers whose leading digit is d ; or conversely, at those positioned just before a long sequence of numbers whose leading digit is all but d.

To this end we will consider the sequence $(\mathrm{P}(L_n=d))_{n\in\mathbb{N}^*}$. In the interests of simplifying notation, we will denote by $(P_{(d,n)})_{n\in\mathbb{N}^*}$ this sequence. Let us study two of its subsequences.

3.1 The first subsequence

The first one is the subsequence $(P_{(d,\phi_d(n))})_{n\in\mathbb{N}^*}$ where ϕ_d is the function from $\mathbb{N}^*$ to $\mathbb{N}$ that maps n to $d\times 10^n-1$. We get the below result:

Proposition 3.1. *The subsequence $(P_{(d,\phi_d(n))})_{n\in\mathbb{N}^*}$ converges to:*

$$\frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d}=\frac{9+10\ln 10+9(d+1)\ln(1-\frac{1}{d+1})}{81d}\,.$$

Proof. Let n be a positive integer such that $n\geq 2$. According to Proposition 2.1, we have:

$$\begin{aligned}P_{(d,\phi_d(n))} &= P_{(d,d\times 10^n-1)}\\ &= \frac{1}{d\times 10^n-1}\sum_{l=0}^{n-1}\big(\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1}\frac{b-\frac{(9d-1)\times 10^l-8}{9}}{b}+\sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1}\frac{\frac{10^{l+1}-1}{9}}{a}\big)\,.\end{aligned}$$

Let us first find an appropriate lower bound of $P_{(d,\phi_d(n))}$:

$$\begin{aligned}P_{(d,\phi_d(n))} &\geq \frac{1}{d\times 10^n}\sum_{l=1}^{n-1}\big(\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1}1-\frac{(9d-1)10^l}{9}\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1}\frac{1}{b}+\frac{10^{l+1}-1}{9}\sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1}\frac{1}{a}\big)\\ &\geq \frac{1}{d\times 10^n}\sum_{l=1}^{n-1}\Big(10^l-\frac{9d-1}{9}10^l\ln(\frac{(d+1)\times 10^l-1}{d\times 10^l-1})+\frac{10^{l+1}-1}{9}\ln(\frac{d\times 10^{l+1}}{(d+1)\times 10^l})\Big),\end{aligned}$$

knowing that for all integers (p,q), such that $1<p<q$:

$$\ln(\frac{q+1}{p})\leq\sum_{k=p}^{q}\frac{1}{k}\leq\ln(\frac{q}{p-1})\,. \tag{5}$$

Therefore we have:

$$\begin{aligned}
P_{(d,\phi_d(n))} &\geq \frac{1}{d\times 10^n}\Big(\sum_{l=1}^{n-1}10^l - \frac{9d-1}{9}\sum_{l=1}^{n-1}10^l\big(\ln(\frac{d+1}{d})+\ln(\frac{10^l-\frac{1}{d+1}}{10^l-\frac{1}{d}})\big)\\
&\quad + \frac{\ln(\frac{10d}{d+1})}{9}\sum_{l=1}^{n-1}(10^{l+1}-1)\Big)\\
&\geq \frac{1}{d\times 10^n}\Big(\frac{10(10^{n-1}-1)}{9} - \frac{(9d-1)\ln(\frac{d+1}{d})}{9}\times\frac{10(10^{n-1}-1)}{9}\\
&\quad - \frac{9d-1}{9}\sum_{l=1}^{n-1}10^l\ln(1+\frac{\frac{1}{d(d+1)}}{10^l-\frac{1}{d}}) + \frac{\ln(\frac{10d}{d+1})}{9}\times(\frac{10^2(10^{n-1}-1)}{9}-(n-1))\Big)\\
&\geq \frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d}\\
&\quad - \frac{9+10\ln(\frac{10d}{d+1})+\frac{9\ln(\frac{10d}{d+1})}{10}n}{81d\times 10^{n-1}} - \frac{9d-1}{9}\sum_{l=1}^{n-1}\frac{10^l\ln(1+\frac{\frac{1}{d+1}}{d\times 10^l-1})}{d\times 10^n}
\end{aligned}$$

we know that for all $x\in]-1;+\infty[$ we have: $\ln(1+x)\leq x$, thus:

$$-\frac{9d-1}{9}\sum_{l=1}^{n-1}\frac{10^l\ln(1+\frac{\frac{1}{d+1}}{d\times 10^l-1})}{d\times 10^n} \geq -\frac{1}{10^n}\sum_{l=1}^{n-1}\frac{\frac{1}{d+1}\times 10^l}{d\times 10^l-1} \geq -\frac{1}{10^n}\sum_{l=1}^{n-1}1 \geq -\frac{n}{10^n}\,.$$

Consequently, we obtain this lower bound:

$$P_{(d,\phi_d(n))} \geq \frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d} - \frac{90+100\ln(\frac{10d}{d+1})+9\ln(\frac{10d}{d+1})n+81dn}{81d\times 10^n}\,. \quad (6)$$

Let us now find an appropriate upper bound of $P_{(d,\phi_d(n))}$:

$$\begin{aligned}
P_{(d,\phi_d(n))} &\leq \frac{1}{d\times 10^n-1}\sum_{l=0}^{n-1}\Big(\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1}1 - \frac{(9d-1)10^l-8}{9}\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1}\frac{1}{b}\\
&\quad + \frac{10^{l+1}}{9}\sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1}\frac{1}{a}\Big)\\
&\leq \frac{1}{d\times 10^n-1}\Big(\frac{10^n-1}{9} - \frac{1}{9}\sum_{l=1}^{n-1}\big((9d-1)10^l-8\big)\ln(\frac{(d+1)\times 10^l}{d\times 10^l})\\
&\quad + \sum_{l=0}^{n-1}\frac{10^{l+1}}{9}\ln(\frac{d\times 10^{l+1}-1}{(d+1)\times 10^l-1})\Big),
\end{aligned}$$

thanks to inequalities (5). Thus we get:

$$
\begin{aligned}
P_{(d,\phi_d(n))} &\leq \frac{1}{d\times 10^n-1}\Big(\frac{10^n}{9}-\frac{\ln(\frac{d+1}{d})}{9}\sum_{l=1}^{n-1}\big((9d-1)10^l-8\big)\\
&\quad+\sum_{l=0}^{n-1}\frac{10^{l+1}}{9}(\ln(\frac{10d}{d+1})+\ln(\frac{10^l-\frac{1}{10d}}{10^l-\frac{1}{d+1}}))\Big)\\
&\leq \frac{1}{d\times 10^n-1}\Big(\frac{10^n}{9}-\frac{\ln(\frac{d+1}{d})}{9}\big(\frac{(9d-1)10(10^{n-1}-1)}{9}-8n\big)+\frac{10\ln(\frac{10d}{d+1})}{9}\times\frac{10^n-1}{9}\\
&\quad+\frac{10}{9}\sum_{l=0}^{n-1}10^l\ln(1+\frac{\frac{1}{d+1}-\frac{1}{10d}}{10^l-\frac{1}{d+1}})\Big)\\
&\leq \frac{1}{d(10^n-\frac{1}{d})}\Big(\frac{10^n-\frac{1}{d}+\frac{1}{d}}{9}-\frac{\ln(\frac{d+1}{d})}{9}\times\big(\frac{(9d-1)(10^n-\frac{1}{d}+\frac{1}{d}-10)}{9}-8n\big)\\
&\quad+\frac{10\ln(\frac{10d}{d+1})}{9}\times\frac{10^n-\frac{1}{d}+\frac{1}{d}-1}{9}+\frac{10}{9}\sum_{l=0}^{n-1}10^l\ln(1+\frac{\frac{1}{d+1}-\frac{1}{10d}}{10^l-\frac{1}{d+1}})\Big)\\
&\leq \frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d}+\frac{\frac{9}{d}+90d\ln(\frac{d+1}{d})+72n\ln(\frac{d+1}{d})+\frac{10\ln(\frac{10d}{d+1})}{d}}{81(d\times 10^n-1)}\\
&\quad+\frac{10}{9(d\times 10^n-1)}\sum_{l=0}^{n-1}\frac{10^l(1-\frac{d+1}{10d})}{(d+1)10^l-1}\\
&\leq \frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d}+\frac{\frac{9}{d}+90d\ln(\frac{d+1}{d})+72n\ln(\frac{d+1}{d})+\frac{10\ln(\frac{10d}{d+1})}{d}}{81(d\times 10^n-1)}\\
&\quad+\frac{10}{9(d\times 10^n-1)}\sum_{l=0}^{n-1}1\,.
\end{aligned}
$$

The last step is easy to demonstrate even for $l=0$. Consequently, we obtain this upper bound:

$$
\begin{aligned}
P_{(d,\phi_d(n))} \leq\ & \frac{9-(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81d}\\
&+\frac{\frac{9}{d}+90d\ln(\frac{d+1}{d})+72n\ln(\frac{d+1}{d})+\frac{10\ln(\frac{10d}{d+1})}{d}+90n}{81(d\times 10^n-1)}.
\end{aligned}
$$

The bound just above and the one brought to light in inequality (6), added to the following limits:

$$
\lim_{n\to+\infty}\Big(\frac{90+100\ln(\frac{10d}{d+1})+9\ln(\frac{10d}{d+1})n+81dn}{81d\times 10^n}\Big)=0\quad\text{and}
$$

$$
\lim_{n\to+\infty}\Big(\frac{\frac{9}{d}+90d\ln(\frac{d+1}{d})+72n\ln(\frac{d+1}{d})+\frac{10\ln(\frac{10d}{d+1})}{d}+90n}{81(d\times 10^n-1)}\Big)=0
$$

lead to the expected result. □

Let us denote by α_d the limit of $(P_{(d,\phi_d(n))})_{n\in\mathbb{N}^*}$:

$$
\alpha_d=\frac{9+10\ln 10+9(d+1)\ln(1-\frac{1}{d+1})}{81d}\,.
$$

Here is the first values of $P_{(1,\phi(n))}$ ($\alpha_1\approx 0.241$):
Here is a few values of $P_{(d,\phi_d(n))}$, for $d\in[\![2,9]\!]$:

n	$\phi(n)$	$P_{(i,\phi(n))}$	$P_{(i,\phi(n))} - \alpha_1$
1	9	0.314	7.30×10^{-2}
2	99	0.253	1.12×10^{-2}
3	999	0.243	1.55×10^{-3}
4	9999	0.242	1.99×10^{-4}
5	99999	0.241	2.43×10^{-5}

Table 1: First five values $P_{(1,\phi(n))}$ and $P_{(1,\phi(n))} - \alpha_1$. We round off these values to three significant digits.

d	$P_{(d,\phi_d(1))}$	$P_{(d,\phi_d(2))}$	$P_{(d,\phi_d(3))}$	$P_{(d,\phi_d(4))}$	α_d
2	0.134	0.131	0.130	0.130	0.130
3	0.085	0.089	0.089	0.089	0.089
4	0.062	0.067	0.068	0.068	0.068
5	0.049	0.054	0.055	0.055	0.055
6	0.040	0.045	0.046	0.046	0.046
7	0.034	0.039	0.039	0.040	0.040
8	0.030	0.034	0.035	0.035	0.035
9	0.026	0.030	0.031	0.031	0.031

Table 2: Values of $P_{(d,\phi_d(n))}$ and α_d, for $n \in [\![1, 4]\!]$. These values are rounded to the nearest thousandth.

3.2 The second subsequence

The second subsequence we will consider is $(P_{(d,\psi_d(n))})_{n\in\mathbb{N}^*}$ where ψ_d is the function from $\mathbb{N}^*$ to $\mathbb{N}$ that maps n to $(d+1)\times 10^n - 1$. We get the following result:

Proposition 3.2. *The subsequence $(P_{(d,\psi_d(n))})_{n\in\mathbb{N}^*}$ converges to:*

$$\frac{10\big(9-(9d-1)\ln(\frac{d+1}{d})+\ln(\frac{10d}{d+1})\big)}{81(d+1)} = \frac{10\big(9+\ln 10+9d\ln(1-\frac{1}{d+1})\big)}{81(d+1)}\,.$$

Proof. Let n be a positive integer such that $n \geq 2$. According to Proposition 2.1, we have:

$$\begin{aligned} P_{(d,\psi_d(n))} &= P_{(d,(d+1)\times 10^n-1)} \\ &= \frac{1}{(d+1)\times 10^n - 1}\Big(\sum_{l=0}^{n}\sum_{b=d\times 10^l}^{(d+1)\times 10^l-1} \frac{b-\frac{(9d-1)10^l-8}{9}}{b} + \sum_{l=0}^{n-1}\sum_{a=(d+1)\times 10^l}^{d\times 10^{l+1}-1}\frac{\frac{10^{l+1}-1}{9}}{a}\Big)\,. \end{aligned}$$

Let us first find an appropriate lower bound of $P_{(d,\psi_d(n))}$ in a way very similar to that used in the proof of Proposition 3.1:

$$\begin{aligned} P_{(d,\psi_d(n))} &\geq \frac{1}{(d+1)\times 10^n}\Big(\frac{10(10^n-1)}{9} - \frac{(9d-1)\ln(\frac{d+1}{d})}{9}\times\frac{10(10^n-1)}{9} \\ &\quad - \frac{9d-1}{9}\sum_{l=1}^{n} 10^l \ln\big(1+\frac{\frac{1}{d(d+1)}}{10^l-\frac{1}{d}}\big) + \frac{\ln(\frac{10d}{d+1})}{9}\times\big(\frac{10^2(10^{n-1}-1)}{9}-(n-1)\big)\Big) \\ &\geq \frac{90-10(9d-1)\ln(\frac{d+1}{d})+10\ln(\frac{10d}{d+1})}{81(d+1)} - \frac{90+100\ln(\frac{10d}{d+1})+9n\ln(\frac{10d}{d+1})}{81(d+1)\times 10^n} \\ &\quad - d\sum_{l=1}^{n}\frac{10^l\frac{1}{d+1}}{d\times 10^l-1}\,. \end{aligned}$$

Then:

$$P_{(d,\psi_d(n))} \geq \frac{90 - 10(9d-1)\ln(\frac{d+1}{d}) + 10\ln(\frac{10d}{d+1})}{81(d+1)} - \frac{90 + 100\ln(\frac{10d}{d+1}) + 9n\ln(\frac{10d}{d+1}) + 81dn}{(d+1)\times 10^n}\,. \tag{7}$$

Let us now find an appropriate upper bound of $P_{(d,\psi_d(n))}$ using the proof of Proposition 3.1:

$$\begin{aligned}
P_{(d,\psi_d(n))} &\leq \frac{1}{(d+1)\times 10^n - 1}\Big(\sum_{l=0}^{n} 10^l - \frac{\ln(\frac{d+1}{d})}{9}\sum_{l=0}^{n}\big((9d-1)10^l - 8\big) \\
&\quad + \sum_{l=0}^{n-1}\frac{10^{l+1}}{9}\big(\ln(\frac{10d}{d+1}) + \ln(1 + \frac{\frac{1}{d+1} - \frac{1}{10d}}{10^l - \frac{1}{d+1}})\big)\Big) \\
&\leq \frac{1}{(d+1)(10^n - \frac{1}{d+1})}\Big(\frac{10^{n+1}-1}{9} - \frac{\ln(\frac{d+1}{d})}{9}\big(\frac{(9d-1)(10^{n+1}-1)}{9} - 8(n+1)\big) \\
&\quad + \frac{10\ln(\frac{10d}{d+1})}{9}\times\frac{10^n-1}{9} + \frac{10}{9}\sum_{l=0}^{n-1}\frac{10^l(1-\frac{d+1}{10d})}{(d+1)10^l - 1}\Big) \\
&\leq \frac{1}{(d+1)(10^n - \frac{1}{d+1})}\Big(\frac{10(10^n - \frac{1}{d+1}) - 1 + \frac{10}{d+1}}{9} + \frac{10\ln(\frac{10d}{d+1})(10^n - \frac{1}{d+1} + \frac{1}{d+1} - 1)}{81} \\
&\quad - \frac{\ln(\frac{d+1}{d})}{9}\big(\frac{(9d-1)(10(10^n - \frac{1}{d+1}) - 1 + \frac{10}{d+1})}{9} - 8(n+1)\big) + \frac{10}{9}\sum_{l=0}^{n-1} 1\Big)\,.
\end{aligned}$$

Thereby:

$$P_{(d,\psi_d(n))} \leq \frac{90 - 10(9d-1)\ln(\frac{d+1}{d}) + 10\ln(\frac{10d}{d+1})}{81(d+1)} + \frac{\frac{90}{d+1} + 72(n+1)\ln(\frac{d+1}{d}) + 90n}{81\big((d+1)\times 10^n - 1\big)}\,. \tag{8}$$

Bounds brought to light in inequalities (7) and (8) and the fact that:

$$\lim_{n\to+\infty}\big(\frac{90 + 100\ln(\frac{10d}{d+1}) + 9n\ln(\frac{10d}{d+1}) + 81dn}{(d+1)\times 10^n}\big) = 0 \quad \text{and}$$
$$\lim_{n\to+\infty}\big(\frac{\frac{90}{d+1} + 72(n+1)\ln(\frac{d+1}{d}) + 90n}{81\big((d+1)\times 10^n - 1\big)}\big) = 0$$

lead to the expected result. □

Let us denote by β_d the limit of $(P_{(d,\psi_d(n))})_{n\in\mathbb{N}^*}$:

$$\beta_d = \frac{10\big(9 + \ln 10 + 9d\ln(1 - \frac{1}{d+1})\big)}{81(d+1)}\,.$$

Here is the first values of $P_{(1,\psi(n))}$ ($\beta_1 \approx 0.313$):
Here is a few values of $P_{(d,\psi_d(n))}$, for $d \in [\![2, 9]\!]$:

4 The graph of $\big(P_{(d,n)}\big)_{n\in\mathbb{N}^*}$

Let us first plot the graph of the sequence $(P_{(1,n)})_{n\in\mathbb{N}^*}$ for values of n from 1 to 1200 (Figure 1). Then we plot a second graph of $P_{(1,n)}$ *versus* $\log(n)$, for $n \in [\![1, 32000]\!]$ (Figure 2). On this graph, the four dots represented by red circles are associated with the first values of $(P_{(1,\phi(n))})_{n\in\mathbb{N}^*}$ and the blue ones are

n	$\psi(n)$	$P_{(i,\psi(n))}$	$P_{(i,\psi(n))} - \beta_1$
1	19	0.373	6.00×10^{-2}
2	199	0.321	7.93×10^{-3}
3	1999	0.314	1.01×10^{-3}
4	19999	0.313	1.23×10^{-4}
5	199999	0.313	1.45×10^{-5}

Table 3: First five values $P_{(1,\psi(n))}$ and $P_{(1,\psi(n))} - \beta_1$. We round off these values to three significant digits.

d	$P_{(d,\psi_d(1))}$	$P_{(d,\psi_d(2))}$	$P_{(d,\psi_d(3))}$	$P_{(d,\psi_d(4))}$	β_d
2	0.176	0.166	0.165	0.165	0.165
3	0.110	0.109	0.109	0.109	0.109
4	0.078	0.080	0.081	0.0.081	0.081
5	0.060	0.063	0.064	0.064	0.064
6	0.049	0.052	0.052	0.053	0.053
7	0.041	0.044	0.045	0.045	0.045
8	0.035	0.038	0.039	0.039	0.039
9	0.031	0.034	0.034	0.034	0.034

Table 4: Values of $P_{(d,\psi_d(n))}$ and β_d, for $n \in [\![1, 4]\!]$. These values are rounded to the nearest thousandth.

associated with the subsequence $(P_{(1,\psi(n))})_{n\in\mathbb{N}^*}$. Their distances to the same-coloured horizontal dotted asymptote tend towards 0. The linear equations of these lines are $y = \alpha_1$ and $y = \beta_1$ respectively (see Propositions 3.1 and 3.2).

Through Figure 2, it is clear that the proportion of 1 as leading digit structurally fluctuate and does not follow Benford's Law.

Let us additionally plot graphs of sequences $(P_{(d,n)})_{n\in\mathbb{N}^*}$ for values of n from 1 to 400 (Figure 3). Then we plot graphs of $P_{(d,n)}$ *versus* $\log(n)$, for $n \in [\![1, 32000]\!]$ (Figure 4).

Through Figure 4, it is once more clear that the proportion of each d as leading digit, $d \in [\![1, 9]\!]$, structurally fluctuate and does not follow Benford's Law.

For each $d \in [\![1, 9]\!]$ the values seem to fluctuate between two values, under the following constraint:

Proposition 4.1. *For all $n \in \mathbb{N}^*$ such that $n \geq 10$ and for all $(p, q) \in [\![1, 9]\!]^2$ such that $p < q$, we have:*

$$P_{(p,n)} > P_{(q,n)}\,.$$

The relative position of graphs of $P_{(d,n)}$, for $d \in [\![1, 9]\!]$, can be observed on Figures 3 and 4.

Proof. For all $n \in \mathbb{N}^*$ such that $n \geq 10$ and for all $(p, q) \in [\![1, 9]\!]^2$ such that $p < q$, we denote by k_p and k_q the positive integers such that $k_p = \min\{i \in \mathbb{N} : p\times 10^i > n\}$ and $k_q = \min\{i \in \mathbb{N} : q \times 10^i > n\}$. We note that $k_p \geq k_q$. For $(r, l) \in \mathbb{N}^2$, let A_l and $B_{r,l}$ be real numbers such that $A_l = \frac{10^{l+1}-1}{9}$ and $B_{r,l} = \frac{(9r-1)\times 10^l - 8}{9}$. We consider that $A_{-1} = 0$. Third cases can be distinguished:

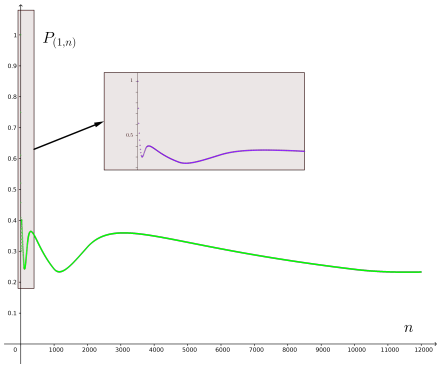

Figure 1: Graph of $(P_{(1,n)})_{n\in\mathbb{N}^*}$.

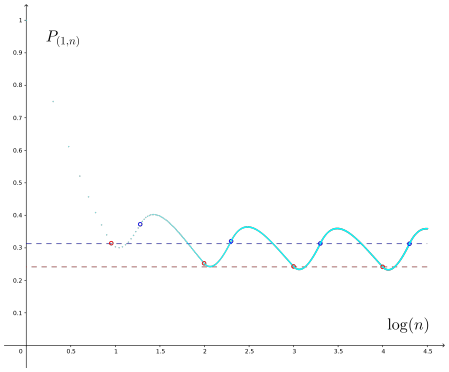

Figure 2: Graph of $P_{(1,n)}$ *versus* $\log(n)$, values of $P_{(1,\phi(n))}$ being in red and those of $P_{(1,\psi(n))}$ being in deep blue. Limits α_1 and β_1 of these two subsequences are represented by horizontal asymptotes.

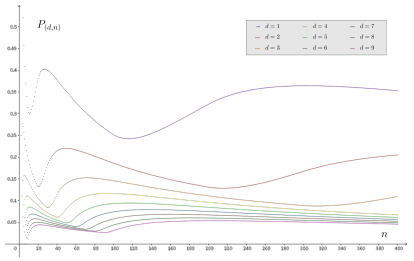

Figure 3: Graphs of $(P_{(d,n)})_{n\in\mathbb{N}^*}$, for $d \in [\![1, 9]\!]$.

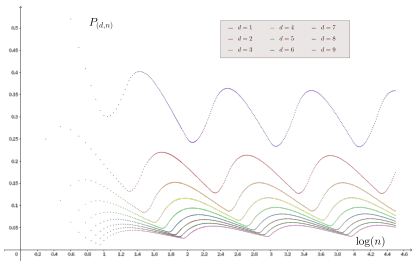

Figure 4: For $d \in [\![1, 9]\!]$, graphs of $P_{(d,n)}$ *versus* $\log(n)$. Note that points have not been all represented.

In the first case $(q+1)\times 10^{k_p-2} \leq p \times 10^{k_p-1} \leq n < (p+1)\times 10^{k_p-1} \leq$

$q \times 10^{k_p-1}$. Thus we have $k_p = k_q + 1$. Thanks to Proposition 2.1 we obtain:

$$
\begin{aligned}
P_{(p,n)} - P_{(q,n)} = \frac{1}{n}\Big(\sum_{l=0}^{k_p-2} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l}}{b} + \sum_{a=(p+1)\times 10^l}^{p\times 10^{l+1}-1} \frac{A_l}{a} \Big) + \sum_{b=p\times 10^{k_p-1}}^{n} \frac{b - B_{p,k_p-1}}{b} \Big) \\
& - \frac{1}{n}\Big(\sum_{l=0}^{k_p-3} \Big(\sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{b - B_{q,l}}{b} + \sum_{a=(q+1)\times 10^l}^{q\times 10^{l+1}-1} \frac{A_l}{a} \Big) \\
& + \sum_{b=q\times 10^{k_p-2}}^{(q+1)\times 10^{k_p-2}-1} \frac{b - B_{q,k_p-2}}{b} + \sum_{a=(q+1)\times 10^{k_p-2}}^{n} \frac{A_{k_p-2}}{a} \Big) \\
= \frac{1}{n}\Big(\sum_{l=0}^{k_p-3} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l}}{b} + \sum_{a=(p+1)\times 10^l}^{p\times 10^{l+1}-1} \frac{A_l}{a} - \sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{b - B_{q,l}}{b} \\
& - \sum_{a=(q+1)\times 10^l}^{q\times 10^{l+1}-1} \frac{A_l}{a} \Big) + \sum_{b=p\times 10^{k_p-2}}^{(p+1)\times 10^{k_p-2}-1} \frac{b - B_{p,k_p-2}}{b} + \sum_{a=(p+1)\times 10^{k_p-2}}^{p\times 10^{k_p-1}-1} \frac{A_{k_p-2}}{a} \\
& + \sum_{b=p\times 10^{k_p-1}}^{n} \frac{b - B_{p,k_p}}{b} - \sum_{b=q\times 10^{k_p-2}}^{(q+1)\times 10^{k_p-2}-1} \frac{b - B_{q,k_p}}{b} \\
& - \sum_{a=(q+1)\times 10^{k_p-2}}^{n} \frac{A_{k_p-2}}{a} \Big) \\
= \frac{1}{n}\Big(\sum_{l=0}^{k_p-2} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l} - A_{l-1}}{b} + \sum_{a=(p+1)\times 10^l}^{q\times 10^l-1} \frac{A_l - A_{l-1}}{a} \\
& + \sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{A_l - (b - B_{q,l})}{b} + \sum_{a=(q+1)\times 10^l}^{p\times 10^{l+1}-1} \frac{A_l - A_l}{a} \Big) \\
& + \sum_{b=p\times 10^{k_p-1}}^{n} \frac{b - B_{p,k_p-1} - A_{k_p-2}}{b} \Big) .
\end{aligned}
$$

Furthermore, for all $(r,l) \in \mathbb{N}^2$, we have $A_l > A_{l-1}$ and:

$$
\begin{aligned}
\forall s \in [\![r \times 10^l, (r+1) \times 10^{l+1} - 1]\!], \\
s - B_{r,l} - A_{l-1} &= s - \frac{(9r-1)\times 10^l - 8}{9} - \frac{10^l - 1}{9} \\
&= s - (r \times 10^l - 1) > 0 \\
A_l - (s - B_{r,l}) &= \frac{10^{l+1} - 1}{9} - s + \frac{(9r-1)\times 10^l - 8}{9} \\
&= (r+1) \times 10^l - 1 - s \geq 0 .
\end{aligned}
$$

Consequently, in this case, $P_{(p,n)} - P_{(q,n)} > 0$.

In the second case $(p+1) \times 10^{k_p-1} \leq n < q \times 10^{k_p-1}$. Thus we have

$k_p = k_q + 1$. Thanks to Proposition 2.1 we obtain:

$$\begin{aligned}
P_{(p,n)} - P_{(q,n)} = \frac{1}{n}\Big(\sum_{l=0}^{k_p-2} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l}}{b} + \sum_{a=(p+1)\times 10^l}^{p\times 10^{l+1}-1} \frac{A_l}{a} \Big) \\
& + \sum_{b=p\times 10^{k_p-1}}^{(p+1)\times 10^{k_p-1}-1} \frac{b - B_{p,k_p-1}}{b} + \sum_{a=(p+1)\times 10^{k_p-1}}^{n} \frac{A_{k_p-1}}{a} \Big) - \frac{1}{n}\Big(\sum_{l=0}^{k_p-3} \\
& \Big(\sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{b - B_{q,l}}{b} + \sum_{a=(q+1)\times 10^l}^{q\times 10^{l+1}-1} \frac{A_l}{a} \Big) + \sum_{b=q\times 10^{k_p-2}}^{(q+1)\times 10^{k_p-2}-1} \frac{b - B_{q,k_p-2}}{b} \\
& + \sum_{a=(q+1)\times 10^{k_p-2}}^{n} \frac{A_{k_p-2}}{a} \Big) \\
= \frac{1}{n}\Big(\sum_{l=0}^{k_p-2} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l} - A_{l-1}}{b} + \sum_{a=(p+1)\times 10^l}^{q\times 10^l-1} \frac{A_l - A_{l-1}}{a} \\
& + \sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{A_l - (b - B_{q,l})}{b} \Big) + \sum_{b=p\times 10^{k_p-1}}^{(p+1)\times 10^{k_p-1}-1} \frac{b - B_{p,k_p-1} - A_{k_p-2}}{b} \\
& + \sum_{a=(p+1)\times 10^{k_p-1}}^{n} \frac{A_{k_p-1} - A_{k_p-2}}{a} \Big) .
\end{aligned}$$

Consequently, in this case, $P_{(p,n)} - P_{(q,n)} > 0$.

In the third case $q \times 10^{k_p-1} \leq n < p \times 10^{k_p}$. Thus we have $k_p = k_q$. Thanks to Proposition 2.1 we obtain:

$$\begin{aligned}
P_{(p,n)} - P_{(q,n)} = \frac{1}{n}\Big(\sum_{l=0}^{k_p-2} \Big(& \sum_{b=p\times 10^l}^{(p+1)\times 10^l-1} \frac{b - B_{p,l} - A_{l-1}}{b} + \sum_{a=(p+1)\times 10^l}^{q\times 10^l-1} \frac{A_l - A_{l-1}}{a} \\
& + \sum_{b=q\times 10^l}^{(q+1)\times 10^l-1} \frac{A_l - (b - B_{q,l})}{b} \Big) + \sum_{b=p\times 10^{k_p-1}}^{(p+1)\times 10^{k_p-1}-1} \frac{b - B_{p,k_p-1} - A_{k_p-2}}{b} \\
& + \sum_{a=(p+1)\times 10^{k_p-1}}^{q\times 10^{k_p-1}} \frac{A_{k_p-1} - A_{k_p-2}}{a} + \sum_{b=q\times 10^{k_p-1}}^{\min((q+1)\times 10^{k_p-1}-1,n)} \frac{A_l - (b - B_{q,l})}{b} \Big) .
\end{aligned}$$

Consequently, in this latter case, $P_{(p,n)} - P_{(q,n)} > 0$. □

Remark 4.2. For $n \in \mathbb{N}^*$, we have, if $d > n$, $P_{(d,n)} = 0$. Hence for all $n \in \mathbb{N}^*$ and for all $(p, q) \in [\![1, 9]\!]^2$ such that $p < q$, we have:

$$P_{(p,n)} \geq P_{(q,n)} .$$

Let us from now on denote by k_d the positive integer such that $k_d = \min\{i \in \mathbb{N} : d \times 10^i > n\}$. Through Figures 2 and 4, "delayed effects" appear to exist, in particular after a long sequence of numbers whose leading digit is d. Let us examine these features in more detail. We study for this purpose the increasing or decreasing nature of sequences $(P_{(d,n)})_{n\in\mathbb{N}^*}$, for $d \in [\![1, 9]\!]$:

Proposition 4.3. $i \in \{0, 1\}$.

$$P_{(d,n+1)} - P_{(d,n)} = \frac{1}{n+1}\Big(\frac{c_i}{n+1} - P_{(d,n)}\Big),$$

where:

$$c_i = \begin{cases} n + 1 - \frac{(9d-1)10^{k_d}-8}{9} & \text{if} \quad n+1 \in [\![d \times 10^{k_d}, (d+1) \times 10^{k_d} - 1]\!] \\ \frac{10^{k_d+1}-1}{9} & \text{if} \quad n+1 \in [\![(d+1) \times 10^{k_d}, d \times 10^{k_d+1} - 1]\!] \end{cases}$$

Proof. It is based on the formulas of Proposition 2.1. Indeed:

$$P_{(d,n+1)} = \frac{1}{n+1}(nP_{(d,n)} + \frac{c_i}{n+1}) = P_{(d,n)} + \frac{1}{n+1}(\frac{c_i}{n+1} - P_{(d,n)}).$$

□

Through this proposition the obvious condition regarding the increasing or decreasing nature of the sequence is underscored: whether $P_{(d,n)}$ value, for $d \in [\![1,9]\!]$, is less or greater than the appropriate $\frac{c_i}{n+1}$ value, for $n \in \mathbb{N}^*$ and $i \in \{0,1\}$. Finding the approximate values of n for which the increasing or decreasing nature of the sequence $(P_{(d,n)})_{n\in\mathbb{N}^*}$ appears is henceforth the aim of that section. We first provide a proposition similar to the previous one:

Proposition 4.4. *If $n \in [\![d \times 10^{k_d}, (d+1) \times 10^{k_d} - 1]\!]$, then we have:*

$$P_{(d,n)} = \frac{1}{n}\Big(P_{(d,d\times 10^{k_d}-1)} \times (d \times 10^{k_d} - 1) + \sum_{b=d\times 10^{k_d}}^{n} \frac{b - \frac{(9d-1)10^{k_d}-8}{9}}{b}\Big).$$

If $n \in [\![(d+1) \times 10^{k_d}, d \times 10^{k_d+1} - 1]\!]$, then we have:

$$P_{(d,n)} = \frac{1}{n}\Big(P_{(d,(d+1)\times 10^{k_d}-1)} \times ((d+1) \times 10^{k_d} - 1) + \sum_{a=(d+1)\times 10^{k_d}}^{n} \frac{\frac{10^{k_d+1}-1}{9}}{a}\Big).$$

Proof. Results are directly derived from Proposition 2.1. □

We now consider the sequence $(\widehat{P}_{(d,n)})_{n\in\mathbb{N}^*}$ defined as follows.
If $n \in [\![d \times 10^{k_d}, (d+1) \times 10^{k_d} - 1]\!]$:

$$\widehat{P}_{(d,n)} = d(\alpha_d - 1)\frac{10^{k_d}}{n} + 1 + \frac{9d-1}{9}\frac{10^{k_d}}{n}\big(\ln(\frac{10^{k_d}}{n}) + \ln d\big).$$

If $n \in [\![(d+1) \times 10^{k_d}, d \times 10^{k_d+1} - 1]\!]$:

$$\widehat{P}_{(d,n)} = (d+1)\beta_d\frac{10^{k_d}}{n} - \frac{10}{9}\frac{10^{k_d}}{n}\big(\ln(\frac{10^{k_d}}{n}) + \ln(d+1)\big).$$

We denote by γ_d a real number such that $\gamma_d \in]1; 1+\frac{1}{d}[$ and $I_{(d,\gamma_d)}$ the set such that $I_{(d,\gamma_d)} = \bigcup_{i=1}^{+\infty}\big([\![d\gamma_d \times 10^i, (d+1)\times 10^i - 1]\!] \cup [\![(d+1)\gamma_d \times 10^i, d\times 10^{i+1} - 1]\!]\big)$. The below proposition can thereupon be stated:

Proposition 4.5.

$$P_{(d,n)} \underset{\substack{n\to+\infty \\ n\in I_{(d,\gamma_d)}}}{\sim} \widehat{P}_{(d,n)}.$$

Proof. Let us study both cases. In the first one, $n \in [\![\gamma_d d \times 10^{k_d}, (d+1)\times 10^{k_d} - 1]\!]$. Let I_1 be the interval such that, $I_1 = \bigcup_{i=1}^{+\infty}[\![d\gamma_d \times 10^i, (d+1)\times 10^i - 1]\!]$. We have:

$$\frac{1}{d+1} \leq \frac{10^{k_d}}{n} \leq \frac{1}{\gamma_d d}. \tag{9}$$

Before we go any further, let us prove the following lemma:

Lemma 4.6. *For $n > d \times 10^{k_d}$:*

$$\ln(\frac{n}{10^{k_d}}) - \ln d + \ln(1+\frac{1}{n}) \leq \sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b} \leq \ln(\frac{n}{10^{k_d}}) - \ln d + \ln(1+\frac{1}{d\times 10^{k_d}-1}).$$

Proof. This result is directly related to inequalities 5. Indeed for $n > d\times 10^{k_d} > 1$, we have:

$$\ln(\frac{n+1}{d\times 10^{k_d}}) \leq \sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b} \leq \ln(\frac{n}{d\times 10^{k_d}-1}).$$

□

Thanks to Proposition 4.4 we get:

$$\begin{aligned} P_{(d,n)} &= P_{(d,d\times 10^{k_d}-1)} \frac{d\times 10^{k_d}-1}{n} + \frac{n-d\times 10^{k_d}+1}{n} - \frac{(9d-1)10^{k_d}-8}{9n} \sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b} \\ &= d(P_{(d,d\times 10^{k_d}-1)} - 1)\frac{10^{k_d}}{n} + 1 - \frac{(9d-1)10^{k_d}}{9n} \sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b} \\ &\quad - \frac{P_{(1,d\times 10^{k_d}-1)}}{n} + \frac{1}{n} + \frac{8}{9n} \sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b}. \end{aligned}$$

Thanks to inequalities (9) and knowing that $\sum_{b=d\times 10^{k_d}}^{n} \frac{1}{b} = \ln(\frac{n}{10^{k_d}}) - \ln d + \underset{\substack{n\to+\infty \\ n\in I_1}}{o}(1)$ (thanks to Lemma 4.6, $\frac{n}{d\times 10^{k_d}}$ being greater than or equal to $\gamma_d > 1$) and that $P_{(d,d\times 10^{k_d}-1)} \underset{+\infty}{\sim} \alpha_d$ (see Proposition 3.1), we finally have:

$$P_{(d,n)} \underset{\substack{n\to+\infty \\ n\in I_1}}{\sim} d(\alpha_d - 1)\frac{10^{k_d}}{n} + 1 + \frac{9d-1}{9}\frac{10^{k_d}}{n}\big(\ln(\frac{10^{k_d}}{n}) + \ln d\big).$$

In the second case $n \in [\![(d+1)\gamma_d \times 10^{k_d}, d\times 10^{k_d+1}-1]\!]$. Let I_2 be the interval such that, $I_2 = \bigcup_{i=1}^{+\infty} [\![(d+1)\gamma_d \times 10^i, d\times 10^{i+1}-1]\!]$. We have:

$$\frac{1}{10d} \leq \frac{10^{k_d}}{n} \leq \frac{1}{(d+1)\gamma_d}. \tag{10}$$

Before we go any further, let us prove this additional lemma:

Lemma 4.7. *For $n > (d+1)\times 10^{k_d}$:*

$$\ln(\frac{n}{10^{k_d}}) - \ln(d+1) + \ln(1+\frac{1}{n}) \leq \sum_{a=(d+1)\times 10^{k_d}}^{n} \frac{1}{a} \qquad \textit{and,}$$

$$\sum_{a=(d+1)\times 10^{k_d}}^{n} \frac{1}{a} \leq \ln(\frac{n}{10^{k_d}}) - \ln(d+1) + \ln(1+\frac{1}{(d+1)\times 10^{k_d}-1}).$$

Proof. This result is directly related to inequalities 5.
Indeed for $n > (d+1)\times 10^{k_d} > 1$, we have:

$$\ln(\frac{n+1}{(d+1)\times 10^{k_d}}) \leq \sum_{a=(d+1)\times 10^{k_d}}^{n} \frac{1}{a} \leq \ln(\frac{n}{(d+1)\times 10^{k_d}-1}).$$

□

Thanks to Proposition 4.4 we get:

$$\begin{aligned}
P_{(d,n)} &= P_{(d,(d+1)\times 10^{k_d}-1)}\frac{(d+1)\times 10^{k_d}-1}{n} + \frac{1}{9}\times\frac{10^{k_d+1}-1}{n}\sum_{a=(d+1)\times 10^{k_d}}^{n}\frac{1}{a} \\
&= (d+1)P_{(d,(d+1)\times 10^{k_d}-1)}\frac{10^{k_d}}{n} + \frac{10}{9}\times\frac{10^{k_d}}{n}\sum_{a=(d+1)\times 10^{k_d}}^{n}\frac{1}{a} \\
&\quad - \frac{P_{(1,(d+1)\times 10^{k_d}-1)}}{n} - \frac{1}{9n}\sum_{a=(d+1)\times 10^{k_d}}^{n}\frac{1}{a}\,.
\end{aligned}$$

Thanks the inequalities (10) and knowing that $\sum_{a=(d+1)\times 10^{k_d}}^{n}\frac{1}{a} = \ln(\frac{n}{10^{k_d}}) - \ln(d+1) + \underset{\substack{n\to+\infty\\ n\in I_2}}{o}(1)$ (thanks to Lemma 4.7, $\frac{n}{(d+1)\times 10^{k_d}}$ being greater than or equal to $\gamma_d > 1$) and that $P_{(d,(d+1)\times 10^{k_d}-1)} \underset{+\infty}{\sim} \beta_d$ (see Proposition 3.2), we finally have:

$$P_{(d,n)} \underset{\substack{n\to+\infty\\ n\in I_2}}{\sim} (d+1)\beta_d\frac{10^{k_d}}{n} - \frac{10}{9}\frac{10^{k_d}}{n}\big(\ln(\frac{10^{k_d}}{n}) + \ln(d+1)\big)\,.$$

□

To find the approximate values of n for which the increasing or decreasing nature of $(P_{(d,n)})_{n\in\mathbb{N}^*}$ appears, we need to study two distinct functions:

Lemma 4.8. *The minimum m_d of the function f_d from $[\frac{1}{d+1}, \frac{1}{d}]$ that maps x onto $d(\alpha_d - 1)x + 1 + \frac{9d-1}{9}x(\ln(x) + \ln d)$ is reached when:*

$$x = \frac{10^{-\frac{10}{9(9d-1)}}\big(1-\frac{1}{d+1}\big)^{-\frac{d+1}{9d-1}}}{d}\,.$$

Its value is $m_d = 1 - \frac{(9d-1)10^{-\frac{10}{9(9d-1)}}\big(1-\frac{1}{d+1}\big)^{-\frac{d+1}{9d-1}}}{9d}$.

Proof. $\forall x \in [\frac{1}{d+1}, \frac{1}{d}]$, $f_d'(x) = d(\alpha_d - 1) + \frac{9d-1}{9}(1 + \ln(x) + \ln d)$. By solving the equation $d(\alpha_d - 1) + \frac{9d-1}{9}(1 + \ln(x) + \ln d) = 0$, we have:

$$x = \frac{10^{-\frac{10}{9(9d-1)}}\big(1-\frac{1}{d+1}\big)^{-\frac{d+1}{9d-1}}}{d}\,.$$

Finally $f_d(\frac{10^{-\frac{10}{9(9d-1)}}\big(1-\frac{1}{d+1}\big)^{-\frac{d+1}{9d-1}}}{d}) = 1 - \frac{(9d-1)10^{-\frac{10}{9(9d-1)}}\big(1-\frac{1}{d+1}\big)^{-\frac{d+1}{9d-1}}}{9d}$. □

Remark 4.9. We note that:

$$\begin{aligned}
f_d(\frac{1}{d}) &= \alpha_d - 1 + 1 = \alpha_d \qquad \text{and} \\
f_d(\frac{1}{d+1}) &= (\alpha_d - 1)\times\frac{d}{d+1} + 1 + \frac{9d-1}{9}\times\frac{1}{d+1}\ln(\frac{d}{d+1}) \\
&= \frac{9 + 10\ln 10 + 9(d+1)\ln(1-\frac{1}{d+1}) - 81d}{81(d+1)} + 1 + \frac{9d-1}{9(d+1)}\ln(\frac{d}{d+1}) \\
&= \frac{10\big(9 + \ln 10 + 9d\ln(1-\frac{1}{d+1})\big)}{81(d+1)} = \beta_d\,.
\end{aligned}$$

Using the approximation of Proposition 4.5, the definition of $(\widehat{P}_{(d,n)})_{n\in\mathbb{N}^*}$ (for $n \in [\![d \times 10^{k_d}, (d+1) \times 10^{k_d} - 1]\!]$) and the properties of f_d (Lemma 4.8), we can approximate the ranks of local minima. Let i be a strictly positive integer. Let us denote by $m_{(d,i)}$ the rank of the local minimum of $P_{(d,n)}$ in $[\![d \times 10^i, (d+1) \times 10^i - 1]\!]$ and by $\widehat{m}_{(d,i)}$ the estimate of this rank. Let us gather in the following table the first values of those ranks and values of local minima, for $d = 1$.

i	$\widehat{m}_{(1,i)}$	$m_{(1,i)}$	$P_{(1,m_{(1,i)})}$
1	12	11	0.300
2	116	116	0.242
3	1158	1158	0.234
4	11578	11579	0.232

Table 5: First four values of above defined ranks and associated values of local minima of $P_{(1,n)}$. We round off values of $P_{(1,n)}$ to three significant digits and values of $\widehat{m}_{(1,i)}$ to unity. Note that $m_1 \approx 0.232$.

Indeed, for $n \in [\![10, 19]\!]$, our approximation of the rank $m_{(1,1)}$ for which the minimum is reached verifies: $\frac{10^1}{\widehat{m}_{(1,1)}} = 2^{\frac{1}{9}} \times 0.2^{\frac{5}{36}}$, *i.e.* $\widehat{m}_{(1,1)} \approx 12$.

Both values of minima (approximately 0.232 according to Lemma 4.8) and ranks for which these minima are reached are correctly approximated by our results as illustrated in Table 5, for $d = 1$.

Let us similarly gather in the below table the first values of those ranks and values of local minima, for $d \in [\![2, 9]\!]$.

d	$\widehat{m}_{(d,4)}$	$m_{(d,4)}$	m_d	$P_{(d,m_{(d,4)})}$
2	21643	21642	0.127	0.127
3	31669	31668	0.088	0.088
4	41683	41681	0.067	0.067
5	51692	51690	0.054	0.054
6	61698	61696	0.046	0.046
7	71703	71701	0.039	0.039
8	81706	81704	0.034	0.034
9	91709	91707	0.031	0.031

Table 6: Values of above defined ranks and associated values of local minima of $P_{(d,n)}$, for $n \in [\![d \times 10^4, (d+1) \times 10^4 - 1]\!]$. We round off values of $P_{(d,n)}$ and m_d to three significant digits and values of $\widehat{m}_{(d,4)}$ to unity.

Indeed, for $n \in [\![20000, 29999]\!]$, our approximation of the rank $m_{(2,4)}$ for which the minimum is reached verifies: $\frac{10^4}{\widehat{m}_{(2,4)}} = \frac{10^{-\frac{10}{9(9\times 2-1)}}\left(1-\frac{1}{2+1}\right)^{-\frac{2+1}{9\times 2-1}}}{2} = \frac{10^{-\frac{10}{153}}(\frac{2}{3})^{-\frac{3}{17}}}{2}$, *i.e.* $\widehat{m}_{(2,4)} \approx 21643$.

Both values of minima and ranks for which these minima are reached are correctly approximated by our results as illustrated in Table 6.

The second function we need to study is defined below:

Lemma 4.10. *The maximum M_d of the function g_d from $[\frac{1}{10d}, \frac{1}{d+1}]$ that maps x onto $(d+1)\beta_d x - \frac{10}{9}x\big(\ln(x) + \ln(d+1)\big)$ is reached when $x = \frac{10^{\frac{1}{9}\left(1-\frac{1}{d+1}\right)^d}}{d+1}$. Its value is $M_d = \frac{10^{\frac{10}{9}\left(1-\frac{1}{d+1}\right)^d}}{9(d+1)}$.*

Proof. $\forall x \in [\frac{1}{10d}, \frac{1}{d+1}]$, $g'_d(x) = (d+1)\beta_d - \frac{10}{9}(1+\ln(x)+\ln(d+1))$. By solving the equation $(d+1)\beta_d - \frac{10}{9}(1+\ln(x)+\ln(d+1)) = 0$, we obtain $x = \frac{10^{\frac{1}{9}\left(1-\frac{1}{d+1}\right)^d}}{d+1}$. Finally $g_d(\frac{10^{\frac{1}{9}\left(1-\frac{1}{d+1}\right)^d}}{d+1}) = \frac{10^{\frac{10}{9}\left(1-\frac{1}{d+1}\right)^d}}{9(d+1)}$. □

Remark 4.11. We note that:

$$
\begin{aligned}
g_d(\frac{1}{d+1}) &= \beta_d \qquad \text{and} \\
g_d(\frac{1}{10d}) &= (d+1)\beta_d \times \frac{1}{10d} - \frac{10}{9} \times \frac{1}{10d}\big(\ln(\frac{1}{10d}) + \ln(d+1)\big) \\
&= \frac{9 + \ln 10 + 9d\ln(1-\frac{1}{d+1})}{81d} - \frac{1}{9d}\big(\ln(\frac{1}{10d}) + \ln(d+1)\big) \\
&= \frac{9 + 10\ln 10 + 9(d+1)\ln(1-\frac{1}{d+1})}{81d} = \alpha_d\,.
\end{aligned}
$$

Using the approximation of Proposition 4.5, the definition of $(\widehat{P}_{(d,n)})_{n\in\mathbb{N}^*}$ (for $n \in [\![(d+1)\times 10^{k_i}, d\times 10^{k_d+1} - 1]\!]$) and the properties of g_d (Lemma 4.10), we can approximate the ranks of local maxima. Let i be a strictly positive integer. Let us denote by $M_{(d,i)}$ the rank of the local maximum of $P_{(d,n)}$ in $[\![(d+1)\times 10^i, (d+1)\times 10^{i+1} - 1]\!]$ and by $\widehat{M}_{(d,i)}$ the estimate of this rank. Let us gather in the following table the first values of those ranks and values of local maxima, for $d = 1$.

i	$\widehat{M}_{(1,i)}$	$M_{(1,i)}$	$P_{(1,M_{(1,i)})}$
1	31	27	0.402
2	310	304	0.364
3	3097	3090	0.359
4	30971	30963	0.359

Table 7: First four values of above defined ranks and associated values of local maxima of $P_{(1,n)}$. We round off values of $P_{(1,n)}$ to three significant digits and values of $\widehat{M}_{(1,i)}$ to unity. Note that $M_1 \approx 0.359$.

Indeed, for $n \in [\![20, 99]\!]$, our approximation of the rank $M_{(1,1)}$ for which the maximum is reached verifies: $\frac{10^1}{n_M} = \frac{5^{\frac{1}{9}}\times 0.5^{\frac{8}{9}}}{2}$, *i.e.* $\widehat{M}_{(1,1)} \approx 31$.

Both values of maxima (approximately 0.359 according to Lemma 4.10) and ranks for which these maxima are reached are correctly approximated by our results as illustrated in Table 7, for $d = 1$.

Let us gather in the following table the first values of those ranks and values of local maxima, for $d \in [\![2, 9]\!]$.

d	$\widehat{M}_{(d,4)}$	$M_{(d,4)}$	M_d	$P_{(d,M_{(d,4)})}$
2	52263	52258	0.213	0.213
3	73412	73409	0.151	0.151
4	94515	94515	0.118	0.118
5	115597	115600	0.096	0.096
6	136668	136673	0.081	0.081
7	157733	157741	0.070	0.070
8	178793	178803	0.062	0.062
9	199851	199863	0.056	0.056

Table 8: Values of above defined ranks and associated values of local maxima of $P_{(d,n)}$, for $n \in [\![(d+1)\times 10^4, d\times 10^5 - 1]\!]$. We round off values of $P_{(d,n)}$ and M_d to three significant digits and values of $\widehat{M}_{(d,4)}$ to unity.

Indeed, for $n \in [\![30000, 199999]\!]$, our approximation of the rank $M_{(2,4)}$ for which the maximum is reached verifies: $\frac{10^4}{\widehat{M}_{(2,4)}} = \frac{10^{\frac{1}{9}}\left(1-\frac{1}{2+1}\right)^2}{2+1} = \frac{4\times 10^{\frac{1}{9}}}{27}$, *i.e.* $\widehat{M}_{(2,4)} \approx 52263$.

Both values of maxima and ranks for which these maxima are reached are correctly approximated by our results as illustrated in Table 8.

When considering the random experiment defined in the beginning of the article, we have thus determined the values of the proportions of selected numbers whose leading digit is d and its bounds: these values seem to fluctuate between m_d and M_d.

5 Central values

From previous Figures, we notice that there exist fluctuations in the graph of $(P_{(d,n)})_{n\in\mathbb{N}^*}$. We can calculate over each "pseudo-cycle", *i.e.* for all $n \in [\![d\times 10^i, d\times 10^{i+1} - 1]\!]$ where $i \in \mathbb{N}$, the mean value $C_{(d,i)}$ of $P_{(d,n)}$. For example, we obtain:

Examples 5.1.

$$C_{(2,0)} = \frac{1}{18}\sum_{i=2}^{19} P_{(2,i)} \approx 0.197$$

$$C_{(5,1)} = \frac{1}{450}\sum_{i=50}^{499} P_{(5,i)} \approx 0.074$$

$$C_{(9,2)} = \frac{1}{8100}\sum_{i=900}^{8999} P_{(9,i)} \approx 0.043\,.$$

We will now consider the sequence $(C_{(d,n)})_{n\in\mathbb{N}}$ and will demonstrate that it converges. Before we go any further, let us prove the following lemma:

Lemma 5.2. *For all $(p,q) \in \mathbb{N}^2$, such that $4 < p < q$, we have:*

$$\frac{\ln\big(q(p-1)\big)\ln(\frac{q}{p-1})}{2} \leq \sum_{n=p}^{q} n\ln n \leq \frac{\ln\big((q+1)p\big)\ln(\frac{q+1}{p})}{2}\,.$$

Proof. The function from $[3;+\infty[$ to $\mathbb{R}$ that maps x onto $x\ln x$ is increasing on $[3;+\infty[$. Thus:

$$\int_{p-1}^{q} x\ln x\,\mathrm{d}x \leq \sum_{n=p}^{q} n\ln n \leq \int_{p}^{q+1} x\ln x\,\mathrm{d}x$$
$$\left[\frac{(\ln x)^2}{2}\right]_{p-1}^{q} \leq \sum_{n=p}^{q} n\ln n \leq \left[\frac{(\ln x)^2}{2}\right]_{p}^{q+1}.$$

The result follows. □

The below proposition can thereupon be stated:

Proposition 5.3.

$$C_{(d,n)} \underset{+\infty}{\sim} \frac{(18d(\alpha_d-1)-(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+18+2(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{162d}.$$

Proof. Let $\mathcal{E}$ be a real number such that $0<\mathcal{E}<1$.
For all $n\in\mathbb{N}^*$, we have:

$$\begin{aligned}
C_{(d,n)} &= \frac{1}{9d\times 10^n}\sum_{i=d\times 10^n}^{d\times 10^{n+1}-1} P_{(d,i)}\\
&= \frac{1}{9d\times 10^n}\Big(\sum_{i=d\times 10^n}^{\lfloor (d+\mathcal{E})\times 10^n\rfloor -1} P_{(d,i)} + \sum_{i=\lfloor (d+\mathcal{E})\times 10^n\rfloor}^{(d+1)\times 10^n-1} P_{(d,i)}\\
&\quad + \sum_{i=(d+1)\times 10^n}^{\lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor -1} P_{(d,i)} + \sum_{i=\lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor}^{d\times 10^{n+1}-1} P_{(d,i)}\Big).
\end{aligned}$$

For all $n\in\mathbb{N}^*$, let us consider $\widehat{C}_{(d,n)} = \frac{1}{9d\times 10^n}\sum_{i=d\times 10^n}^{d\times 10^{n+1}-1}\widehat{P}_{(d,i)}$. We know that:

$$\forall i\in\mathbb{N}^*,\quad |P_{(d,i)}-\widehat{P}_{(d,i)}| \leq |P_{(d,i)}| + |\widehat{P}_{(d,i)}| \leq 2\,.$$

There also exists an integer t such that for all $i\geq t$ and $i\in I_{\mathcal{E}}$, $|P_{(d,i)}-\widehat{P}_{(d,i)}|\leq\mathcal{E}$ (see Proposition 4.5).

Thus, for all $n\in\mathbb{N}^*$ such that $10^n\geq t$, we have:

$$\begin{aligned}
|C_{(d,n)}-\widehat{C}_{(d,n)}| &\leq \frac{1}{9d\times 10^n}\sum_{i=d\times 10^n}^{d\times 10^{n+1}-1}|P_{(d,i)}-\widehat{P}_{(d,i)}|\\
&\leq \frac{1}{9d\times 10^n}\Big(\sum_{i=d\times 10^n}^{\lfloor (d+\mathcal{E})\times 10^n\rfloor -1}|P_{(d,i)}-\widehat{P}_{(d,i)}| + \sum_{i=\lfloor (d+\mathcal{E})\times 10^n\rfloor}^{(d+1)\times 10^n-1}|P_{(d,i)}-\widehat{P}_{(d,i)}|\\
&\quad + \sum_{i=(d+1)\times 10^n}^{\lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor -1}|P_{(d,i)}-\widehat{P}_{(d,i)}| + \sum_{i=\lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor}^{d\times 10^{n+1}-1}|P_{(d,i)}-\widehat{P}_{(d,i)}|\Big)\\
&\leq \frac{1}{9d\times 10^n}\Big(2(\lfloor (d+\mathcal{E})\times 10^n\rfloor - d\times 10^n) + ((d+1)\times 10^n - \lfloor (d+\mathcal{E})\times 10^n\rfloor)\mathcal{E}\\
&\quad + 2(\lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor - (d+1)\times 10^n)\\
&\quad + (d\times 10^{n+1} - \lfloor ((d+1)+\mathcal{E})\times 10^n\rfloor)\mathcal{E}\Big)\\
&\leq \frac{1}{9d\times 10^n}\Big(2\mathcal{E}\times 10^n + 10^n\mathcal{E} + 2\mathcal{E}\times 10^n + 8\times 10^n\mathcal{E}\Big) \leq \frac{(5+(9d-1))\mathcal{E}}{9}\,.
\end{aligned}$$

Consequently $C_{(d,n)} \underset{+\infty}{\sim} \widehat{C}_{(d,n)}$. We will henceforth study $(\widehat{C}_{(d,n)})_{n\in\mathbb{N}}$. We have for all $n \in \mathbb{N}^*$:

$$\widehat{C}_{(d,n)} = \frac{1}{9d}\Big(\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\widehat{P}_{(d,i)}}{10^n} + \sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1} \frac{\widehat{P}_{(d,i)}}{10^n} \Big). \tag{11}$$

We consider the first term of this sum:

$$\begin{aligned}\frac{1}{9d}\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\widehat{P}_{(d,i)}}{10^n} &= \frac{1}{9d}\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \Big(\frac{d(\alpha_d-1)}{i} + \frac{1}{10^n} + \frac{9d-1}{9d}(\ln(\frac{10^n}{i}) + \ln d)\Big) \\ &= \frac{9d(\alpha_d-1) + (9d-1)n\ln(10) + (9d-1)\ln d}{81d} \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{1}{i} \\ &\quad - \frac{9d-1}{81d}\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i} + \frac{1}{9d}.\end{aligned}$$

The proofs of Lemma 4.6 and Proposition 3.1 allow us to state:

$$\begin{aligned}\ln(\frac{(d+1)\times 10^n}{d\times 10^n}) \le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{1}{i} \le \ln(\frac{(d+1)\times 10^n - 1}{d\times 10^n - 1}) \\ \ln(\frac{d+1}{d}) \le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{1}{i} \le \ln(\frac{d+1}{d}) + \ln(1 + \frac{\frac{1}{d(d+1)}}{10^n - \frac{1}{d}}),\end{aligned}$$

i.e. $\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{1}{i} = \ln(\frac{d+1}{d}) + \underset{+\infty}{O}(\frac{1}{10^n})$. Then thanks to Lemma 5.2:

$$\begin{aligned}\frac{\ln\big(((d+1)\times 10^n - 1)(d\times 10^n - 1)\big)\ln(\frac{(d+1)\times 10^n-1}{d\times 10^n-1})}{2} &\le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i} \\ \frac{\ln(d(d+1)\times 10^{2n} - (2d+1)\times 10^n + 1)(\ln(\frac{d+1}{d}) + \ln(1 + \frac{\frac{1}{d(d+1)}}{10^n - \frac{1}{d}}))}{2} &\le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i}.\end{aligned}$$

We have:

$$\begin{aligned}\ln(d(d+1)\times 10^{2n} - (2d+1)\times 10^n + 1) &= \ln(d(d+1)\times 10^{2n}) + \ln(1 - \frac{(2d+1) - \frac{1}{10^n}}{d(d+1)\times 10^n}) \\ &= \ln(d(d+1)) + 2n\ln(10) + \underset{+\infty}{O}(\frac{2d+1}{d(d+1)\times 10^n}).\end{aligned}$$

We obtain:

$$\begin{aligned}\frac{\big(\ln(d(d+1)) + 2n\ln(10) + \underset{+\infty}{O}(\frac{1}{10^n})\big)\big(\ln(\frac{d+1}{d}) + \underset{+\infty}{O}(\frac{1}{10^n})\big)}{2} &\le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i} \\ \frac{(\ln(d(d+1)) + 2n\ln(10))\ln(\frac{d+1}{d}) + \underset{+\infty}{o}(1)}{2} &\le \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i}.\end{aligned}$$

Thanks to Lemma 5.2 we also have:

$$\begin{aligned}\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i} &\le \frac{\ln\big(d(d+1)\times 10^{2n}\big)\ln(\frac{(d+1)\times 10^n}{d\times 10^n})}{2} \\ \sum_{i=d\times 10^n}^{(d+1)\times 10^n-1} \frac{\ln i}{i} &\le \frac{(\ln(d(d+1)) + 2n\ln(10))\ln(\frac{d+1}{d})}{2}.\end{aligned}$$

Finally the first term of equality (11) verifies:

$$\frac{1}{9d}\sum_{i=d\times 10^n}^{(d+1)\times 10^n-1}\frac{\widehat{P}_{(d,i)}}{10^n} \underset{+\infty}{\sim} \frac{(9d(\alpha_d-1)+(9d-1)n\ln 10+(9d-1)\ln d)\ln(\frac{d+1}{d})}{81d}$$
$$-\frac{(9d-1)(\ln(d(d+1))+2n\ln(10))\ln(\frac{d+1}{d})}{2\times 81d}+\frac{1}{9d}$$
$$\underset{+\infty}{\sim}\frac{(18d(\alpha_d-1)-(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+18}{162d}.$$

We consider henceforth the second term of the equality (11):

$$\frac{1}{9d}\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n}=\frac{1}{9d}\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\Big((d+1)\beta_d\frac{1}{i}-\frac{10}{9d}\big(\ln(d+1)+\ln(\frac{10^n}{i})\big)\Big)$$
$$=\frac{9(d+1)\beta_d-10\ln(d+1)-10\ln(10)n}{81d}\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{1}{i}$$
$$+\frac{10}{81d}\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\ln i}{i}.$$

The proofs of Lemma 4.7 and Proposition 3.2 allow us to state:

$$\ln(\frac{d\times 10^{n+1}}{(d+1)\times 10^n})\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{1}{i}\le\ln(\frac{d\times 10^{n+1}-1}{(d+1)\times 10^n-1})$$
$$\ln(\frac{10d}{d+1})\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{1}{i}\le\ln(\frac{10d}{d+1})+\ln(1+\frac{\frac{1}{d+1}-\frac{1}{10i}}{10^n-\frac{1}{d+1}}),$$

i.e. $\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{1}{i}=\ln(\frac{10d}{d+1})+\underset{+\infty}{O}(\frac{1}{10^n})$. Then thanks to Lemma 5.2:

$$\frac{\ln\big((d\times 10^{n+1}-1)((d+1)\times 10^n-1)\big)\ln(\frac{d\times 10^{n+1}-1}{(d+1)\times 10^n-1})}{2}\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\ln i}{i}$$
$$\frac{\ln(10d(d+1)\times 10^{2n}-(11d+1)\times 10^n+1)(\ln(\frac{10d}{d+1})+\ln(1+\frac{\frac{1}{d+1}-\frac{1}{10d}}{10^n-\frac{1}{d+1}}))}{2}\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\ln i}{i}.$$

We have:

$$\ln(10d(d+1)10^{2n}-(11d+1)10^n+1)=\ln(10d(d+1)10^{2n})+\ln(1-\frac{11d+1-\frac{1}{10^n}}{10d(d+1)10^{n+1}})$$
$$=\ln(10d(d+1))+2n\ln(10)+\underset{+\infty}{O}(\frac{11d+1}{10d(d+1)\times 10^{n+1}}).$$

We obtain:

$$\frac{\big(\ln(10d(d+1))+2n\ln(10)+\underset{+\infty}{O}(\frac{1}{10^n})\big)\big(\ln(\frac{10d}{d+1})+\underset{+\infty}{O}(\frac{1}{10^n})\big)}{2}\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\ln i}{i}$$
$$\frac{(\ln(10d(d+1))+2n\ln 10)\ln(\frac{10d}{d+1})+\underset{+\infty}{o}(1)}{2}\le\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\ln i}{i}.$$

Thanks to Lemma 5.2 we also have:

$$\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1} \frac{\ln i}{i} \leq \frac{\ln(d(d+1)\times 10^{2n+1})\ln(\frac{10d}{d+1})}{2}$$
$$\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1} \frac{\ln i}{i} \leq \frac{(\ln(10d(d+1))+2n\ln(10))\ln(\frac{10d}{d+1})}{2}\,.$$

Finally the second term of equality (11) verifies:

$$\begin{aligned}\frac{1}{9d}\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n} &\underset{+\infty}{\sim} \frac{(9(d+1)\beta_d - 10\ln(d+1) - 10n\ln(10))\ln(\frac{10d}{d+1})}{81d}\\ &\quad + \frac{5(\ln(10d(d+1))+2n\ln(10))\ln(\frac{10d}{d+1})}{81d}\\ &\underset{+\infty}{\sim} \frac{(9(d+1)\beta_d + 5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{81d}\,.\end{aligned}$$

Hence:

$$\begin{aligned}\widehat{C}_{(1,n)} &\underset{+\infty}{\sim} \frac{(18d(\alpha_d-1)-(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+18}{162d} + \frac{(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{81d}\\ &\underset{+\infty}{\sim} \frac{(18d(\alpha_d-1)-(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+18+2(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{162d}\,.\end{aligned}$$

The result follows. □

Let us denote by C_d the limit of $(C_{(d,n)})_{n\in\mathbb{N}}$.

Note that our first choice of "pseudo-cycle" give more weight at the values of proportion situated at the end of intervals $[\![d\times 10^i, d\times 10^{i+1}-1]\!]$ where $i\in\mathbb{N}$. We can also have defined the sequence $(\tilde{C}_{(d,n)})_{n\in\mathbb{N}}$ as follows:

$$\tilde{C}_{(d,n)} = \frac{1}{9(d+1)\times 10^n}\sum_{i=(d+1)\times 10^n}^{(d+1)\times 10^{n+1}-1} P_{(1,i)}\,.$$

If so, we would state that the limit $\tilde{C}_d$ of the sequence $(\tilde{C}_{(d,n)})_{n\in\mathbb{N}}$ is:

$$\frac{(90d(\alpha_d-1)-5(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+90+(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{81(d+1)}\,.$$

Proof. $\tilde{C}_{(d,n)} \underset{+\infty}{\sim} \widehat{C}_{(d,n)}$ where, for all $n\in\mathbb{N}^*$:

$$\widehat{C}_{(d,n)} = \frac{1}{9(d+1)}\Big(\sum_{i=(d+1)\times 10^n}^{d\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n} + \sum_{i=d\times 10^{n+1}}^{(d+1)\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n}\Big)\,. \tag{12}$$

The first term of this sum is similar to the second term of the associated sum in the proof of Proposition 5.3. The second one verifies:

$$\begin{aligned}\frac{1}{9(d+1)}\sum_{i=d\times 10^{n+1}}^{(d+1)\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n} &= \frac{90d(\alpha_d-1)+10(9d-1)((n+1)\ln(10)+\ln d)}{81(d+1)}\sum_{i=d\times 10^{n+1}}^{(d+1)\times 10^{n+1}-1}\frac{1}{i}\\ &\quad - \frac{10(9d-1)}{81(d+1)}\sum_{i=d\times 10^{n+1}}^{(d+1)\times 10^{n+1}-1}\frac{\ln i}{i} + \frac{10}{9(d+1)}\,.\end{aligned}$$

Finally the second term of equality (12) is:

$$\frac{1}{9(d+1)}\sum_{i=d\times 10^{n+1}}^{(d+1)\times 10^{n+1}-1}\frac{\widehat{P}_{(d,i)}}{10^n} \underset{+\infty}{\sim} \frac{(90d(\alpha_d-1)+10(9d-1)((n+1)\ln 10+\ln d))\ln(\frac{d+1}{d})}{81(d+1)}$$
$$-\frac{5(9d-1)(\ln(d(d+1))+2(n+1)\ln(10))\ln(\frac{d+1}{d})}{81(d+1)}+\frac{10}{9(d+1)}$$
$$\underset{+\infty}{\sim}\frac{(90d(\alpha_d-1)-5(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+90}{81(d+1)}.$$

Hence:

$$\widehat{C}_{(d,n)} \underset{+\infty}{\sim} \frac{(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{81(d+1)}+\frac{(90d(\alpha_d-1)-5(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+90}{81(d+1)}$$
$$\underset{+\infty}{\sim}\frac{(90d(\alpha_d-1)-5(9d-1)\ln(\frac{d+1}{d}))\ln(\frac{d+1}{d})+90+(9(d+1)\beta_d+5\ln(\frac{10d}{d+1}))\ln(\frac{10d}{d+1})}{81(d+1)}.$$

The result follows. □

Once more, means values over both "pseudo-cycles" are very close to the theoric value highlighted by Benford: $\log(1+\frac{1}{d})$ ([2]). Table 9 below gathers the whole values:

d	$\tilde{C}_d$	$\log(1+\frac{1}{d})$	C_d
1	0.281	0.301	0.301
2	0.160	0.176	0.191
3	0.113	0.125	0.139
4	0.088	0.097	0.109
5	0.072	0.079	0.090
6	0.061	0.067	0.077
7	0.053	0.058	0.067
8	0.047	0.051	0.059
9	0.042	0.046	0.053

Table 9: Values of C_d, $\tilde{C}_d$ and Benford's Law probabilities, these values being rounded to the nearest thousandth.

Indeed, according to Hill ([10]), it is absolutely normal. In a way, it can be considered as an equivalent to the central limit theorem ([8]).

Conclusion

To conclude, through our model, we have seen that the proportion of d as leading digit, $d \in [\![1,9]\!]$, in certain naturally occurring collections of data is more likely to follow a law whose probability distribution is $(d, P_{(d,n)})_{d\in[\![1,9]\!]}$, where n is the smaller integer upper bound of the physical, biological or economical quantities considered, rather than Benford's Law. These probability distributions fluctuate around Benford's value as can be seen in the literature (see [12], [5], [14] or [7] for example) in accordance with our model. Knowing beforehand the value of the upper bound n can be a way to find a better adjusted law than Benford's one.

The results of the article would have been the same in terms of fluctuations of the proportion of $d \in [\![1, 9]\!]$ as leading digit, of limits of subsequences, or of results on central values, if our discrete uniform distributions uniformly randomly selected were lower bounded by a positive integer different from 0: first terms in proportion formulas become rapidly negligible. Through our model we understand that the predominance of 1 as first digit (followed by those of 2 and so on) is all but surprising in experimental data: it is only due to the fact that, in the lexicographical order, 1 appears before 2, 2 appears before 3, *etc.*

However the limits of our model rest on the assumption that the random variables used to obtain our data are not the same and follow discrete uniform distributions that are uniformly randomly selected. In certain naturally occurring collections of data it cannot conceivably be justified. Studying the cases where the random variables follow other distributions (and not necessarily randomly selected) sketch some avenues for future research on the subject.

References

[1] T. W. Beer. Terminal digit preference: beware of benford's law. *Journal of Clinical Pathology*, 62(2):192, 2009.

[2] F. Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78:127–131, 1938.

[3] Bunimovich L. Hill T. Berger, A. One-dimensional dynamical systems and benford's law. *Transactions of the American Mathematical Society*, 357(1):197–219, 2004.

[4] J. Boyle. An application of fourier series to the most significant digit problem. *America Mathematical Monthly*, 101:879–886, 1994.

[5] J. Burke and E. Kincanon. Benford's law and physical constants: the distribution of initial digits. *American Journal of Physics*, 59:952, 1991.

[6] A. Diekmann. Not the first digit! using benford's law to detect fraudulent scientific data. *Journal of Applied Statistics*, 34(3):321–329, 2007.

[7] J. L. Friar, T. Goldman, and J. Pérez-Mercader. Genome sizes and the Benford distribution. *PLOS ONE*, 7(5), 2012.

[8] N. Gauvrit and J.-P. Delahaye. Pourquoi la loi de benford n'est pas mystérieuse. *Mathématiques et sciences humaines*, 182(2):7–15, 2008.

[9] T. Hill. Base-invariance implies benford's law. *Proceedings of the American Mathematical Society*, 123:887–895, 1995.

[10] T. Hill. A statistical derivation of the significant-digit law. *Statistical Science*, 10(4):354–363, 1995.

[11] P. Jolissaint. Loi de benford, relations de récurrence et suites équidistribuées. *Elemente der Mathematik*, 60(1):10–18, 2005.

[12] D. Knuth. *The Art of Computer Programming 2.* Addison-Wesley, New-York, 1969.

[13] R. Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4:39–40, 1881.

[14] M. Nigrini and W. Wood. Assessing the integrity of tabulated demographic data. 1995. Preprint.

[15] R. S. Pinkham. On the distribution of first significant digits. *The Annals of Mathematical Statistics*, 32(4):1223–1230, 1961.

[16] R. A. Raimi. The first digit problem. *American Mathematical Monthly*, 83(7):521–538, 1976.

[17] P. B. Sarkar. An observation on the significant digits of binomial coefficients and factorials. *Sankhya B*, 35:362–364, 1973.

[18] G. Van Rossum. *Python tutorial*, volume Technical Report CS-R9526. 1995. Centrum voor Wiskunde en Informatica (CWI).

[19] H. Varian. Benford's law (letters to the editor). *The American Statistician*, 26(3):62–65, 1972.

[20] L. C. Washington. Benford's law for fibonacci and lucas numbers. *The Fibonacci Quarterly*, 19(2):175–177, 1981.

Appendix: Python script

Using Propositions 2.1, we can determine the terms of $(P_{(d,n)})_{n\in\mathbb{N}^*}$, for $d \in [\![1, 9]\!]$. To this end, we have created a script with the Python programming language (Python Software Foundation, Python Language Reference, version 3.4. available at `http://www.python.org`, see [18]). The implemented function *expvalProp* has two parameters: the rank n of the wanted term of the sequence and the value ld of the considered leading digit. Here is the used algorithm:

```
def expvalProp(n,ld):
    if(ld>n):
        return(0)
    else:
        k=0
        while(ld*10**k<=n):
            k=k+1
        u=1;v=ld-1;S=0;T=0;
        for i in range(0,k-1):
            for b in range(ld*10**i,(ld+1)*10**i):
                T=T+(b-v)/b
            for a in range((ld+1)*10**i,ld*10**(i+1)):
                S=S+u/a
            u=u+10**(i+1)
            v=v*10+8
        if (n<(ld+1)*10**(k-1)):
            for b in range(ld*10**(k-1),n+1):
                T=T+(b-v)/b
        else:
```

```
    for b in range(ld*10**(k-1),(ld+1)*10**(k-1)):
        T=T+(b-v)/b
    for a in range((ld+1)*10**(k-1),n+1):
        S=S+u/a
return((S+T)/n)
```