Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p1

Benford or not Benford: new results on digits beyond the first

By Blondeau Da Silva

Published: 20 May 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 1-25
DOI: 10.1285/i20705948v15n1p1

Benford or not Benford: new results on digits beyond the first

Blondeau Da Silva*

Academy of Limoges
13 rue Francois Chénieux 87000 Limoges

Published: 20 May 2022

In this paper, we will see that the proportion of d as p^{th} digit, where $p > 1$ and $d \in [\![0, 9]\!]$, 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 the generalization of Benford's law to digits beyond the first one. These probability distributions fluctuate around theoretical values of the distribution of the p^{th} digit of Benford's law. Knowing beforehand the value of the upper bound can be a way to find a better adjusted law than Benford's one.

keywords: Benford's law, digits, experimental data.

1 Introduction

Benford's law is really amazing: 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; it rather follows a logarithmic distribution (see the recent books of Miller Miller (2015) and Berger and Hill Berger and Hill (2015)). Having been discovered by Newcomb in 1881 (Newcomb (1881)), this law was definitively brought to light by Benford in 1938 (Benford (1938)). He proposed the following probability distribution where the probability for d to be the first digit of a number is:

$$\log(1 + \frac{1}{d}) \ .$$

*Corresponding author: stephane.blondeau-da-silva@ac-limoges.fr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

It was quickly admitted that numerous empirical data sets follow Benford's law: economic data (Sehity et al. (2005)), social data (Golbeck (2015)), demographic data (Nigrini and Wood (1995); Leemis et al. (2000)), physical data (Knuth (1969); Burke and Kincanon (1991); Nigrini and Miller (2007); Alexopoulos and Leontsinis (2014)) or biological data (Costasa et al. (2008); Friar et al. (2012)) for instance; to such an extent that this law was used to detect possible frauds in lists of socio-economic data (Varian (1972); Nigrini (1999); Durtschi et al. (2004); Saville (2006); Tödter (2009); Rauch et al. (2011)) or in scientific publications (Alves et al. (2014)).

Nevertheless many discordant voices brought a significantly different message. By putting aside the distributions known to fully disobey Benford's law (Raimi (1976); Hill (1988); Tolle et al. (2000); Scott and Fasli (2001); Beer (2009); Deckert et al. (2011)), this law often appeared to be a good approximation of the reality, but no more than an approximation (Scott and Fasli (2001); Saville (2006); Deckert et al. (2011); Gauvrit and Delahaye (2011); Goodman (2016)). Goodman, for example, in Goodman (2016), discussed the necessity of introducing an error term. Even the 20 different domains, tested by Benford (in Benford (1938)), displayed large fluctuations around theoretical values. In Blondeau Da Silva (2019), Blondeau Da Silva, considering data as realizations of a homogeneous and expanded range of random variables following discrete uniform distributions, showed that, the proportion of each d as leading digit, $d \in [\![0, 9]\!]$, structurally fluctuates.

Benford's law can also be extended to digits beyond the first one: the probability for d, $d \in [\![0, 9]\!]$, to be the p^{th} digit of a number is equal to (see Hill (1995)):

$$\sum_{j=10^{p-2}}^{10^{p-1}-1} \log(1 + \frac{1}{10j + d}) \ .$$

Similarly to first digit case, these distributions have been observed in various areas (Geyer (2010); Alexopoulos and Leontsinis (2014); Alves et al. (2014)) and, in particular, have been used to detect frauds (Carslaw (1988); Thomas (1989); Mebane Jr (2006); Cho and Gaines (2007); Diekmann (2007); Joenssen (2013)). Once more, limits of such methods were also underlined (Mebane Jr (2006); Cho and Gaines (2007); Diekmann (2007)). Let us focus on a revealing example: in Saville (2006), Saville studied companies listed on the Johannesburg Stock Exchange; of the 17 companies known to have manipulated their accounts, none has passed Saville's test successfully, which is somewhat comforting. But, of the other 17 honest companies, 4 also failed the test, which is more troubling.

Building a very similar model to that described in Blondeau Da Silva (2019), the naturally occurring data will be considered as 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 strictly positive p-digits integers ($p > 1$) are equally likely to occur; note that we consider the most straightforward model where data is not the realization of a single random variable but of an expanded range of simple random variables.

Through this article we will demonstrate that the predominance of 0 over 1 (and of 1 over 2, and so on), as p^{th}, $(p > 1)$ digit 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, Benford's 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 p^{th} digit structurally fluctuates. For each $p > 1$, there is not a single law but numerous distinct laws that we will hereafter examine.

2 Notations and probability space

Let p and d be two strictly positive integers such that $p > 1$ and $d \in [\![0, 9]\!]$. Let m be a strictly positive integer such that $m \geq 10^{p-1}$. Let $\mathcal{U}\{10^{p-1}, m\}$ denote the discrete uniform distribution whereby integers between 10^{p-1} and m are equally likely to be observed.

Let n be a strictly positive integer such that $n \geq 10^{p-1}$. Let us consider the random experiment $\mathcal{E}_n$ of tossing two independent dice. The first one is a fair $(n + 1 - 10^{p-1})$-sided die showing $n + 1 - 10^{p-1}$ different numbers from 1 to $n + 1 - 10^{p-1}$. The number i rolled on it defines the number of faces on the second die. It thus shows i different numbers from 10^{p-1} to $i + 10^{p-1} - 1$.

Let us now define the probability space Ω_n as follows: $\Omega_n = \{(i, j) : i \in [\![1, n + 1 - 10^{p-1}]\!] \text{ and } j \in [\![10^{p-1}, i + 10^{p-1} - 1]\!]\}$. Our probability measure is denoted by P.

Let us denote by $D_{(n,p)}$ the random variable from Ω_n to $[\![0, 9]\!]$ that maps each element ω of Ω_n to the p^{th} digit of the second component of ω.

As our aim is to determine the probability that the p^{th} digit of the integer obtained thanks to the second throw is d, it can be considered with no consequence on our results that we first select an integer i equal to or less than n among at least p-digits integers (following the $\mathcal{U}\{10^{p-1}, n\}$ discrete uniform distribution); afterwards we select an other at least p-digits integer equal to or less than i (following the $\mathcal{U}\{10^{p-1}, i\}$ discrete uniform distribution).

3 Closely related approaches

In Janvresse and De La Rue (2004), Janvresse and De La Rue already provided a probabilistic explanation for the appearance of Benford's law in everyday-life numbers: they showed that it arised naturally when mixtures of uniform distributions were considered.

Furthermore, they established a connection with the theorem of Flehinger (Flehinger (1966)); *via* Cesaro-summation method, Flehinger recovered the probability distribution of the first digit. The cumulative average $P_N^2(1) = \frac{1}{N} \sum_{M=1}^{N} P_M^1(1)$ in Flehinger (1966) (where $P_N^1(1)$ is the proportion of the positive integers equal to or less than N, which have initial digit equal to or less than 1) is exactly $P(L_n = 1)$ in (Blondeau Da Silva, 2019, Proposition 2.1.) (*i.e.* the probability that the first digit of our second throw in the random experiment is 1, n here playing the same role as N in Flehinger (1966)). The other cumulative averages, for $A \in [\![2, 9]\!]$, $P_N^2(A)$ can also be linked to values of

$P(L_n = d)$, for $d \in [\![1, 9]\!]$ (A or d being the value of the considered first digit): indeed, $P_N^2(A) = \sum_{k=1}^{A} P(L_N = k)$. Fluctuations around Benford's values and the existence of an upperbound are common features of both Flehinger's and Blondeau Da Silva's first digit approaches.

Herzel (in Herzel (1956)) used urn's models identical or very close to the one described above. He gave three schemes: equal probability (as in our case) and probability linearly weighted: according to the size of the urn or according to the square of its size. None of these schemes produced a limit, but Herzel obtained numerical results close to Benford's (Blondeau Da Silva in Blondeau Da Silva (2019) confirmed these findings). Logan and Goudsmit (in Logan and Goudsmit (1978)) also used the same urn's model as ours, but found a limit by improperly neglecting a part of the calculation (see Raimi (1985) and Blondeau Da Silva (2019)).

All these articles deal mainly with the case where the studied digit is the first one. Henceforth we will study the case of other digits.

4 Proportion of d

Through the below proposition, we will express the value of $\mathrm{P}(D_{(n,p)} = d)$ *i.e.* the probability that the p^{th} digit of our second throw in the random experiment $\mathcal{E}_n$ is d.

Proposition 4.1. *Let k denote the integer such that:*

$$k = \max\{i \in \mathbb{N} : 10^{i+p} \leq n\}\,.$$

Let l denote the positive integer such that:

$$l = \lfloor \tfrac{n-(10^{p-1}+d)10^{k+1}}{10^{k+2}} \rfloor + 10^{p-2}\,.$$

The value of $\mathrm{P}(D_{(n,p)} = d)$ *is:*

$$\frac{1}{n+1-10^{p-1}}\Big(\sum_{i=0}^{k}\Big(\sum_{j=10^{p-2}}^{10^{p-1}-1}\sum_{b=(10j+d)10^i}^{(10j+(d+1))10^i-1}\frac{b-((9j+d)10^i+10^{p-2}-1)}{b+1-10^{p-1}}$$
$$+\sum_{j=10^{p-2}-1}^{10^{p-1}-1}\sum_{a=\max(10^{p+i-1},(10j+(d+1))10^i)}^{\min(10^{p+i}-1,(10(j+1)+d)10^i-1)}\frac{10^i(j+1)-10^{p-2}}{a+1-10^{p-1}}\Big)$$
$$+r_{(n,d,p)}\Big)\,,$$

where $r_{(n,d,p)}$ is, if the p^{th} digit of n is d:

$$\sum_{j=10^{p-2}}^{l}\sum_{b=(10j+d)10^{k+1}}^{\min(n,(10j+(d+1))10^{k+1}-1)}\frac{b-((9j+d)10^{k+1}+10^{p-2}-1)}{b+1-10^{p-1}}$$
$$+\sum_{j=10^{p-2}-1}^{l-1}\sum_{a=\max(10^{p+k},(10j+(d+1))10^{k+1})}^{(10(j+1)+d)10^{k+1}-1}\frac{10^{k+1}(j+1)-10^{p-2}}{a+1-10^{p-1}}\,,$$

or where $r_{(n,d,p)}$ is, if the p^{th} digit of n is all but d:

$$\sum_{j=10^{p-2}}^{l} \sum_{b=(10j+d)10^{k+1}}^{(10j+(d+1))10^{k+1}-1} \frac{b-((9j+d)10^{k+1}+10^{p-2}-1)}{b+1-10^{p-1}}$$
$$+ \sum_{j=10^{p-2}-1}^{l} \sum_{a=\max(10^{p+k},(10j+(d+1))10^{k+1})}^{\min(n,(10(j+1)+d)10^{k+1}-1)} \frac{10^{k+1}(j+1)-10^{p-2}}{a+1-10^{p-1}}\,.$$

Proof. Let us denote by $F_{(n,p)}$ the random variable from Ω_n to $[\![1, n+1-10^{p-1}]\!]$ 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 $(n+1-10^{p-1})$-sided die. For each $q \in [\![1, n+1-10^{p-1}]\!]$, we have:

$$\mathrm{P}(F_{(n,p)} = q) = \frac{1}{n+1-10^{p-1}}\,. \tag{1}$$

According to the law of total probability, we state:

$$\mathrm{P}(D_{(n,p)} = d) = \sum_{q=1}^{n+1-10^{p-1}} \mathrm{P}(D_{(n,p)} = d | F_{(n,p)} = q)\, \mathrm{P}(F_{(n,p)} = q)\,. \tag{2}$$

Thereupon two cases appear in determining the value, for $q \in [\![1, n+1-10^{p-1}]\!]$, of $\mathrm{P}(D_{(n,p)} = d | F_{(n,p)} = q)$. Let k_q be the integer such that $k_q = \max\{k \in \mathbb{N} : 10^{p+k} \leq q + 10^{p-1} - 1\}$ in both cases.

Let us study the first case where the p^{th} digit of $q + 10^{p-1} - 1$ is d. For all i in $[\![0, k_q]\!]$, there exist $9 \times 10^{p-2}$ sequences of 10^i consecutive integers from $(10j+d)10^i$ to $(10j+(d+1))10^i - 1$, where $j \in [\![10^{p-2}, 10^{p-1}-1]\!]$, whose p^{th} digit is d. The higher of these integers is $(10(10^{p-1}-1)+(d+1))10^{k_q} - 1$, the last $(p+k_q)$-digit number in this case. Thus, from 10^{p-1} to $10^{p+k_q} - 1$, the number of integers whose p^{th} digit is d is:

$$\sum_{i=0}^{k_q} \sum_{j=10^{p-2}}^{10^{p-1}-1} \sum_{(10j+d)10^i}^{(10j+(d+1))10^i-1} 1 = 9 \times 10^{p-2} \sum_{i=0}^{k_q} 10^i = 10^{p-2}(10^{k_q+1}-1)\,.$$

This equality still holds true for $k_q = -1$. Such types of sum would be considered null in the rest of the article. From 10^{p+k_q} to $q + 10^{p-1} - 1$, there exist t sequences of 10^{k_q+1} consecutive integers from $(10j+d)10^{k_q+1}$ to $(10j+(d+1))10^{k_q+1} - 1$, where $j \in [\![10^{p-2}, 10^{p-2}+t-1]\!]$, whose p^{th} digit is d. There also exist $q+10^{p-1}-1-(10(10^{p-2}+t)+d)10^{k_q+1}+1$ additional integers in this case between $(10(10^{p-2}+t)+d)10^{k_q+1}$ and $q + 10^{p-1} - 1$. Finally the total number of integers whose p^{th} digit is d is:

$$10^{p-2}(10^{k_q+1}-1) + t \times 10^{k_q+1} + q + 10^{p-1} - 1 - (10(10^{p-2}+t)+d)10^{k_q+1} + 1$$
$$\text{i.e.}\quad q + 10^{p-1} - 1 - \Big((9(10^{p-2}+t)+d)10^{k_q+1} - 1\Big)\,.$$

It may be inferred that:

$$\mathrm{P}(D_{(n,p)} = d | F_{(n,p)} = q) = \frac{q + 10^{p-1} - 1 - \Big(\big(9(10^{p-2} + t) + d\big) 10^{k_q+1} - 1 \Big)}{q}\,, \qquad (3)$$

the p^{th} digit of $q + 10^{p-1} - 1$ being here d.

In the second case, we consider the integers $q + 10^{p-1} - 1$ whose p^{th} digits are different from d. On the basis of the previous case, the total number of integers whose p^{th} digit is d is, where t is the number of sequences of consecutive integers lower than $q + 10^{p-1} - 1$:

$$\begin{aligned} &10^{p-2}(10^{k_q+1} - 1) + t \times 10^{k_q+1} \\ \text{i.e.}\quad &10^{k_q+1}(10^{p-2} + t) - 10^{p-2}\,. \end{aligned}$$

It can be concluded that:

$$\mathrm{P}(D_{(n,p)} = d | F_{(n,p)} = q) = \frac{10^{k_q+1}(10^{p-2} + t) - 10^{p-2}}{q}\,, \qquad (4)$$

the p^{th} digit of $q + 10^{p-1} - 1$ being here different from d.
Using equalities (1Proportion of $d$equation.4.1), (2Proportion of $d$equation.4.2), (3Proportion of $d$equation.4.3) and (4Proportion of $d$equation.4.4), we get our result. $\square$

For example, we get:

Examples 4.2. Let us first determine the value of $\mathrm{P}(D_{(10003,5)} = 2)$. The probability that the fifth digit of a randomly selected number in $[\![10000, 10000]\!]$ is 2 is $\frac{0}{1}$, those in $[\![10000, 10001]\!]$ is $\frac{0}{2}$, those in $[\![10000, 10002]\!]$ is $\frac{1}{3}$ and those in $[\![10000, 10003]\!]$ is $\frac{1}{4}$. Hence we have:

$$\mathrm{P}(D_{(10003,5)} = 2) = \frac{1}{4}\Big(\frac{0}{1} + \frac{0}{2} + \frac{1}{3} + \frac{1}{4}\Big) \approx 0.1458\,.$$

It is the second case of Proposition 4.1theorem.4.1, where $n = 10003$, $d = 2$, $p = 5$, $k = -1$ and $l = 1000$.
Let us now determine the value of $\mathrm{P}(D_{(1113,3)} = 1)$ (first case of Proposition 4.1theorem.4.1); in this case, we have $k = 0$ and $l = 11$.

$$\begin{aligned} \mathrm{P}(D_{(1113,3)} = 1) = \frac{1}{1014}\Big(&\sum_{j=10}^{99} \frac{j-9}{10j-98} + \sum_{j=10}^{98} \sum_{a=10j+2}^{10(j+1)} \frac{j-9}{a-99} + \sum_{a=992}^{999} \frac{90}{a-99} + \sum_{a=1000}^{1009} \frac{90}{a-99} \\ &+ \sum_{b=1010}^{1019} \frac{b-919}{b-99} + \sum_{a=1020}^{1109} \frac{100}{a-99} + \sum_{b=1110}^{1113} \frac{b-1009}{b-99}\Big) \\ &= \frac{1}{1014}\Big(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{11} + \frac{2}{12} + \frac{2}{13} + \ldots + \frac{89}{891} + \frac{90}{892} + \frac{90}{893} + \ldots + \frac{90}{910} \\ &+ \frac{91}{911} + \ldots + \frac{100}{920} + \frac{100}{921} + \ldots + \frac{100}{1010} + \frac{101}{1011} + \ldots + \frac{104}{1014}\Big) \\ &\approx 0.1028\,. \end{aligned}$$

Let us determine the value of $\mathrm{P}(D_{(212,2)} = 9)$ (second case of Proposition 4.1theorem.4.1); in this case, we have $k = 0$ and $l = 1$.

$$\begin{aligned}
\mathrm{P}(D_{(212,2)} = 9) &= \frac{1}{203}\Big(\frac{9}{10} + \sum_{j=1}^{8}\sum_{a=10(j+1)}^{10(j+1)+8}\frac{j}{a-9} + \sum_{a=100}^{189}\frac{9}{a-9} + \sum_{b=190}^{199}\frac{b-180}{b-9} + \sum_{a=200}^{212}\frac{19}{a-9}\Big) \\
&= \frac{1}{203}\Big(\frac{1}{10} + \frac{1}{11} + \ldots + \frac{1}{19} + \frac{2}{20} + \frac{2}{21} + \ldots + \frac{8}{89} + \frac{9}{90} + \frac{9}{91} + \ldots + \frac{9}{180} + \frac{10}{181} \\
&+ \frac{11}{182} + \ldots + \frac{19}{190} + \frac{19}{191} + \ldots + \frac{19}{203}\Big) \\
&\approx 0.0759\,.
\end{aligned}$$

5 Study of a particular subsequence

It is natural that we take a specific look at the values of n positioned one rank before the integers for which the number of digits has just increased.

To this end we will consider the sequence $\big(\mathrm{P}(D_{n,p} = d)\big)_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$. In the interests of simplifying notation, we will denote by $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$ this sequence. Let us study the subsequence $(P_{(d,\phi_{(d,p)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$ where $\phi_{(d,p)}$ is the function from $\mathbb{N}\setminus[\![0,p-1]\!]$ to $\mathbb{N}$ that maps n to $10^n - 1$. We get the below result:

Proposition 5.1. *The subsequence* $(P_{(d,\phi_{(d,p)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$ *converges to:*

$$10^{-1} + \frac{n_{(d,p)} + m_{(d,p)} - 9l_{(d,p)} - d\times k_{(d,p)}}{9\times 10^{p-1}} + \frac{1}{90}\ln\Big(\frac{10^{p-1}+d}{10^{p-1}}\Big) + \frac{1}{9}\ln\Big(\frac{10^p}{10^p - 10 + d + 1}\Big)\,,$$

where:

$$\begin{cases}
k_{(d,p)} = \sum_{j=10^{p-2}}^{10^{p-1}-1}\ln\big(\frac{10j+(d+1)}{10j+d}\big) \\
l_{(d,p)} = \sum_{j=10^{p-2}}^{10^{p-1}-1} j\ln\big(\frac{10j+(d+1)}{10j+d}\big) \\
m_{(d,p)} = \sum_{j=10^{p-2}}^{10^{p-1}-2}\ln\big(\frac{10(j+1)+d}{10j+(d+1)}\big) \\
n_{(d,p)} = \sum_{j=10^{p-2}}^{10^{p-1}-2} j\ln\big(\frac{10(j+1)+d}{10j+(d+1)}\big)\,.
\end{cases}$$

Proof. Let n be a positive integer such that $n \geq p$. According to Proposition 4.1theorem.4.1, we have $P_{(d,\phi_{(d,p)}(n),p)} = P_{(d,10^n-1,p)}$ *i.e.*, knowing that in this case $k = \max\{i\in\mathbb{N} : 10^{i+p} \leq 10^n - 1\} = n - p - 1$:

$$\begin{aligned}
\frac{1}{10^n - 10^{p-1}}\Big(\sum_{i=0}^{n-p}\Big(&\sum_{j=10^{p-2}}^{10^{p-1}-1}\sum_{b=(10j+d)10^i}^{(10j+(d+1))10^i-1}\frac{b-((9j+d)10^i + 10^{p-2}-1)}{b+1-10^{p-1}} \\
&+ \sum_{j=10^{p-2}-1}^{10^{p-1}-1}\sum_{a=\max(10^{p+i-1},(10j+(d+1))10^i)}^{\min(10^{p+i}-1,(10(j+1)+d)10^i-1)}\frac{10^i(j+1)-10^{p-2}}{a+1-10^{p-1}}\Big)\Big)\,.
\end{aligned}$$

Let us denote by $b_{(i,d,p)}$ the positive number:

$$\sum_{j=10^{p-2}}^{10^{p-1}-1}\sum_{b=(10j+d)10^i}^{(10j+(d+1))10^i-1}\frac{b-((9j+d)10^i+10^{p-2}-1)}{b+1-10^{p-1}}\,,$$

and by $a_{(i,d,p)}$ the positive number:

$$\sum_{j=10^{p-2}-1}^{10^{p-1}-1} \sum_{a=\max(10^{p+i-1},(10j+(d+1))10^i)}^{\min(10^{p+i}-1,(10(j+1)+d)10^i-1)} \frac{10^i(j+1)-10^{p-2}}{a+1-10^{p-1}}.$$

Thus we have:

$$P_{(d,\phi_{(d,p)}(n),p)} = \frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} \Big(b_{(i,d,p)} + a_{(i,d,p)} \Big).$$

Let us first find an appropriate lower bound of $P_{(d,\phi_{(d,p)}(n),p)}$. We have:

$$\begin{aligned} b_{(i,d,p)} &= \sum_{j=10^{p-2}}^{10^{p-1}-1} (10^i - \sum_{b=(10j+d)10^i}^{(10j+(d+1))10^i-1} \frac{(9j+d)10^i + 10^{p-2} - 10^{p-1}}{b+1-10^{p-1}}) \\ &= 9 \times 10^{p+i-2} - \sum_{j=10^{p-2}}^{10^{p-1}-1} ((9j+d)10^i + 10^{p-2} - 10^{p-1}) \sum_{b=(10j+d)10^i}^{(10j+(d+1))10^i-1} \frac{1}{b+1-10^{p-1}} \end{aligned}$$

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

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

Consequently, we obtain, for $i \geq 1$:

$$\begin{aligned} b_{(i,d,p)} &\geq 9 \times 10^{p+i-2} - \sum_{j=10^{p-2}}^{10^{p-1}-1} (9j+d)10^i \ln(\frac{(10j+(d+1))10^i - 10^{p-1}}{(10j+d)10^i - 10^{p-1}}) \\ &\geq 9 \times 10^{p+i-2} - \sum_{j=10^{p-2}}^{10^{p-1}-1} (9j+d)10^i \big(\ln(\frac{10j+(d+1)}{10j+d}) + \ln(1 + \frac{\frac{10^{p-1}}{10j+(d+1)}}{10^i(10j+d) - 10^{p-1}})) \\ &\geq 9 \times 10^{p+i-2} - d \times 10^i \sum_{j=10^{p-2}}^{10^{p-1}-1} \ln(\frac{10j+(d+1)}{10j+d}) - 9 \times 10^i \sum_{j=10^{p-2}}^{10^{p-1}-1} j \ln(\frac{10j+(d+1)}{10j+d}) \\ &\quad - \sum_{j=10^{p-2}}^{10^{p-1}-1} (9j+d)10^i \ln(1 + \frac{\frac{10^{p-1}}{10j+(d+1)}}{10^i(10j+d) - 10^{p-1}})). \end{aligned}$$

Let us denote by $k_{(d,p)}$ the positive number $\sum_{j=10^{p-2}}^{10^{p-1}-1} \ln(\frac{10j+(d+1)}{10j+d})$ and $l_{(d,p)}$ the positive number $\sum_{j=10^{p-2}}^{10^{p-1}-1} j \ln(\frac{10j+(d+1)}{10j+d})$. Knowing that for all $x \in]-1;+\infty[$, we have $\ln(1+x) \leq x$, we obtain:

$$\begin{aligned} b_{(i,d,p)} &\geq 9 \times 10^{p+i-2} - d \times 10^i k_{(d,p)} - 9 \times 10^i l_{(d,p)} - \sum_{j=10^{p-2}}^{10^{p-1}-1} (9j+d)10^i \frac{\frac{10^{p-1}}{10j+(d+1)}}{10^i(10j+d) - 10^{p-1}} \\ &\geq 9 \times 10^{p+i-2} - d \times 10^i k_{(d,p)} - 9 \times 10^i l_{(d,p)} - \sum_{j=10^{p-2}}^{10^{p-1}-1} 10^i \frac{10^{p-1}}{10^i \times 10^{p-1} - 10^{p-1}} \\ &\geq 9 \times 10^{p+i-2} - d \times 10^i k_{(d,p)} - 9 \times 10^i l_{(d,p)} - 9 \times 10^{p-2} \frac{10^i}{10^i - 1}. \end{aligned}$$

Similarly, we have thanks to inequalities (5Study of a particular subsequenceequation.5.5):

$$\begin{aligned}
a_{(i,d,p)} \geq & \sum_{j=10^{p-2}}^{10^{p-1}-2} (10^i(j+1) - 10^{p-2}) \ln(\frac{(10(j+1)+d)10^i + 1 - 10^{p-1}}{(10j+(d+1))10^i + 1 - 10^{p-1}}) \\
& + (10^{p-2+i} - 10^{p-2}) \ln(\frac{(10^{p-1}+d)10^i + 1 - 10^{p-1}}{10^{p+i-1} + 1 - 10^{p-1}}) \\
& + (10^{p-1+i} - 10^{p-2}) \ln(\frac{10^{p+i} + 1 - 10^{p-1}}{(10^p - 10 + d + 1)10^i + 1 - 10^{p-1}}) \\
\geq & \, 10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} j \ln(\frac{10(j+1)+d}{10j+(d+1)}) + 10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} j \ln(1 + \frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i + 1 - 10^{p-1}}) \\
& + (10^i - 10^{p-2})\Big(\sum_{j=10^{p-2}}^{10^{p-1}-2} \big(\ln(\frac{10(j+1)+d}{10j+(d+1)}) + \ln(1 + \frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i + 1 - 10^{p-1}}))\Big) \\
& + (10^{p-2+i} - 10^{p-2})\big(\ln(\frac{10^{p-1}+d}{10^{p-1}}) + \ln(1 + \frac{\frac{d(10^{p-1}-1)}{10^{p-1}+d}}{10^{p-1}10^i + 1 - 10^{p-1}})) \\
& + (10^{p-1+i} - 10^{p-2})(\ln(\frac{10^p}{10^p - 10 + d + 1}) + \ln(1 + \frac{\frac{(10^{p-1}-1)(10-d-1)}{10^p}}{(10^p - 10 + d + 1)10^i + 1 - 10^{p-1}})).
\end{aligned}$$

Let us denote by $m_{(d,p)}$ the positive number $\sum_{j=10^{p-2}}^{10^{p-1}-2} \ln(\frac{10(j+1)+d}{10j+(d+1)})$ and $n_{(d,p)}$ the positive number $\sum_{j=10^{p-2}}^{10^{p-1}-2} j \ln(\frac{10(j+1)+d}{10j+(d+1)})$:

$$\begin{aligned}
a_{(i,d,p)} \geq & \, 10^i n_{(d,p)} + (10^i - 10^{p-2})\Big(m_{(d,p)} + \sum_{j=10^{p-2}}^{10^{p-1}-2} \ln(1 + \frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i + 1 - 10^{p-1}})\Big) \\
& + (10^{p-2+i} - 10^{p-2}) \ln(\frac{10^{p-1}+d}{10^{p-1}}) + (10^{p-1+i} - 10^{p-2}) \ln(\frac{10^p}{10^p - 10 + d + 1}).
\end{aligned}$$

Hence we have:

$$\begin{aligned}
P_{(d,\phi_{(d,p)}(n),p)} \geq & \frac{1}{10^n}\Big(a_{(0,d,p)} + b_{(0,d,p)} + \sum_{i=1}^{n-p} \big(9 \times 10^{p+i-2} - d \times 10^i k_{(d,p)} - 9 \times 10^i l_{(d,p)} \\
& + 10^i n_{(d,p)} + 10^i m_{(d,p)} + 10^{p-2+i} \ln(\frac{10^{p-1}+d}{10^{p-1}}) + 10^{p-1+i} \ln(\frac{10^p}{10^p - 10 + d + 1}) \\
& - 9 \times 10^{p-2}\frac{10}{9} - 10^{p-2}\big(m_{(d,p)} + \ln(\frac{10^{p-1}+d}{10^{p-1}}) + \ln(\frac{10^p}{10^p - 10 + d + 1})) \\
& + (10^i - 10^{p-2}) \sum_{j=10^{p-2}}^{10^{p-1}-2} \ln(1 + \frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i + 1 - 10^{p-1}}))\Big).
\end{aligned}$$

In light of the following equality $\sum_{i=1}^{n-p} 10^i = \frac{10^{n-p+1}-10}{9}$, we have:

$$\begin{aligned}
P_{(d,\phi_{(d,p)}(n),p)} \geq & \, 10^{-1} + \frac{10^{-p+1}(n_{(d,p)} + m_{(d,p)} - 9l_{(d,p)} - dk_{(d,p)})}{9} + \frac{10^{-1}}{9} \ln(\frac{10^{p-1}+d}{10^{p-1}}) \\
& + \frac{1}{9} \ln(\frac{10^p}{10^p - 10 + d + 1}) + \epsilon_{(d,n,p)},
\end{aligned}$$

where $\epsilon_{(d,n,p)}$ is:

$$\frac{a_{(0,d,p)}+b_{(0,d,p)}}{10^n}-\frac{10^{p-1}}{10^n}+\frac{dk_{(d,p)}+9l_{(d,p)}-n_{(d,p)}-m_{(d,p)}}{9\times 10^{n-1}}-\frac{10^{p-1}}{9\times 10^n}\ln(\frac{10^{p-1}+d}{10^{p-1}})$$
$$-\frac{10^p}{9\times 10^n}\ln(\frac{10^p}{10^p-10+d+1})-\frac{10^{p-1}(n-p)}{10^n}-\frac{10^{p-2}(n-p)}{10^n}\big(m_{(d,p)}+\ln(\frac{10^{p-1}+d}{10^{p-1}})$$
$$+\ln(\frac{10^p}{10^p-10+d+1})\big)+\frac{1}{10^n}\sum_{i=1}^{p-3}(10^i-10^{p-2})\sum_{j=10^{p-2}}^{10^{p-1}-2}\ln(1+\frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i+1-10^{p-1}}).$$

Knowing that for all $x\in]-1;+\infty[$, we have $\ln(1+x)\leq x$, we obtain, for all $i\in\{1,...,p-3\}$:

$$\sum_{j=10^{p-2}}^{10^{p-1}-2}\ln(1+\frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i+1-10^{p-1}})\leq\sum_{j=10^{p-2}}^{10^{p-1}-2}\frac{\frac{9\times(10^{p-1}-1)}{10(j+1)+d}}{(10j+d+1)10^i+1-10^{p-1}}$$
$$\leq 10^{p-1}\frac{\frac{10^p}{10^{p-1}}}{d+2}\leq 10^p$$

From the above upper bound and the definition of $\epsilon_{(d,n,p)}$, it may be deduced that $\lim_{n\to+\infty}\epsilon_{(d,n,p)}=0$.

Let us now find an appropriate upper bound of $P_{(d,\phi_{(d,p)}(n),p)}$. Thanks to inequalities (5Study of a particular subsequenceequation.5.5):

$$b_{(i,d,p)}\leq 9\times 10^{p+i-2}-\sum_{j=10^{p-2}}^{10^{p-1}-1}((9j+d)10^i+10^{p-2}-10^{p-1})$$
$$\ln(\frac{(10j+(d+1))10^i+1-10^{p-1}}{(10j+d)10^i+1-10^{p-1}})$$
$$\leq 9\times 10^{p+i-2}-\sum_{j=10^{p-2}}^{10^{p-1}-1}((9j+d)10^i+10^{p-2}-10^{p-1})\big(\ln(\frac{10j+(d+1)}{10j+d})$$
$$+\ln(1+\frac{\frac{10^{p-1}-1}{10j+(d+1)}}{10^i(10j+d)+1-10^{p-1}})\big)$$
$$\leq 9\times 10^{p+i-2}-d\times 10^ik_{(d,p)}-9\times 10^il_{(d,p)}+10^{p-1}k_{(d,p)}\,.$$

Similarly, we have thanks to inequalities (5Study of a particular subsequenceequation.5.5):

$$a_{(i,d,p)}\leq\sum_{j=10^{p-2}}^{10^{p-1}-2}10^i(j+1)\ln(\frac{(10(j+1)+d)10^i-10^{p-1}}{(10j+(d+1))10^i-10^{p-1}})$$
$$+10^{p-2+i}\ln(\frac{(10^{p-1}+d)10^i-10^{p-1}}{10^{p+i-1}-10^{p-1}})+10^{p-1+i}\ln(\frac{10^{p+i}-10^{p-1}}{(10^p-10+d+1)10^i-10^{p-1}})$$
$$\leq 10^in_{(d,p)}+10^i\sum_{j=10^{p-2}}^{10^{p-1}-2}j\ln(1+\frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i-10^{p-1}})$$
$$+10^i\Big(m_{(d,p)}+\sum_{j=10^{p-2}}^{10^{p-1}-2}\ln(1+\frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i-10^{p-1}})\Big)$$
$$+10^{p-2+i}\big(\ln(\frac{10^{p-1}+d}{10^{p-1}})+\ln(1+\frac{\frac{d\times 10^{p-1}}{10^{p-1}+d}}{10^{p-1}10^i-10^{p-1}})\big)$$
$$+10^{p-1+i}\big(\ln(\frac{10^p}{10^p-10+d+1})+\ln(1+\frac{\frac{10^{p-1}(10-d-1)}{10^p}}{(10^p-10+d+1)10^i-10^{p-1}})\big).$$

Hence we have:

$$\begin{aligned}
P_{(d,\phi_{(d,p)}(n),p)} \leq \frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} \Big(& 9 \times 10^{p+i-2} - d \times 10^i k_{(d,p)} - 9 \times 10^i l_{(d,p)} + 10^i m_{(d,p)} \\
& + 10^i n_{(d,p)} + 10^{p-2+i} \ln(\frac{10^{p-1}+d}{10^{p-1}}) + 10^{p-1+i} \ln(\frac{10^p}{10^p - 10 + d + 1}) \\
& + 10^{p-1} k_{(d,p)} + 10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} j \ln(1 + \frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i - 10^{p-1}}) \\
& + 10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} \ln(1 + \frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i - 10^{p-1}}) \\
& + 10^{p-2+i} \ln(1 + \frac{\frac{d\times 10^{p-1}}{10^{p-1}+d}}{10^{p-1}10^i - 10^{p-1}}) + 10^{p-1+i} \ln(1 + \frac{\frac{10^{p-1}(10-d-1)}{10^p}}{(10^p - 10 + d + 1)10^i - 10^{p-1}}) \Big).
\end{aligned}$$

In light of the following equality $\sum_{i=0}^{n-p} 10^i = \frac{10^{n-p+1}-1}{9}$, we have:

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 9 \times 10^{p+i-2}) = 10^{-1}$$

$$\lim_{n\to+\infty} (-\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} d \times 10^i k_{(d,p)}) = \frac{-dk_{(d,p)}}{9 \times 10^{p-1}}$$

$$\lim_{n\to+\infty} (-\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 9 \times 10^i l_{(d,p)}) = -l_{(d,p)} 10^{1-p}$$

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 10^i n_{(d,p)}) = \frac{m_{(d,p)}}{9 \times 10^{p-1}}$$

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 10^i n_{(d,p)}) = \frac{n_{(d,p)}}{9 \times 10^{p-1}}$$

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 10^{p-2+i} \ln(\frac{10^{p-1}+d}{10^{p-1}})) = \frac{1}{90} \ln(\frac{10^{p-1}+d}{10^{p-1}})$$

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 10^{p-1+i} \ln(\frac{10^p}{10^p - 10 + d + 1})) = \frac{1}{9} \ln(\frac{10^p}{10^p - 10 + d + 1}))$$

$$\lim_{n\to+\infty} (\frac{1}{10^n - 10^{p-1}} \sum_{i=0}^{n-p} 10^{p-1} k_{(d,p)}) = 0\,.$$

Knowing that for all $x \in]-1;+\infty[$, we have $\ln(1+x) \leq x$, we obtain, for $i \geq 1$:

$$10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} j \ln(1 + \frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i - 10^{p-1}}) \leq 10^{i+p-1} \frac{\frac{10^p}{10^{p-1}}}{10^{p-1}10^i - 10^{p-1}} = \frac{10^{i+1}}{10^i - 1} \leq \frac{100}{9}$$

$$10^i \sum_{j=10^{p-2}}^{10^{p-1}-2} \ln(1 + \frac{\frac{9\times 10^{p-1}}{10(j+1)+d}}{(10j+d+1)10^i - 10^{p-1}}) \leq 10^i \frac{\frac{10^p}{10^{p-1}}}{10^{p-1}10^i - 10^{p-1}} \leq \frac{100}{9 \times 10^{p-1}}$$

$$10^{p-2+i} \ln(1 + \frac{\frac{d\times 10^{p-1}}{10^{p-1}+d}}{10^{p-1}10^i - 10^{p-1}}) \leq 10^{p-2+i} \frac{\frac{d\times 10^{p-1}}{10^{p-1}+d}}{10^{p-1}10^i - 10^{p-1}} \leq \frac{10^{p-1+i}}{10^{p-1}10^i - 10^{p-1}} \leq \frac{10}{9}$$

$$10^{p-1+i} \ln(1 + \frac{\frac{10^{p-1}(10-d-1)}{10^p}}{(10^p - 10 + d + 1)10^i - 10^{p-1}}) \leq \frac{10^{p-1+i}}{10^{p-1}10^i - 10^{p-1}} \leq \frac{10}{9}\,.$$

Thanks to $P_{(d,\phi_{(d,p)}(n),p)}$ upper bound and the above inequalities, the result follows. $\square$

Let us denote by $\alpha_{(d,p)}$ the limit of $(P_{(d,\phi_{(d,p)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$. Few values of $P_{(d,\phi_{(d,p)}(n),p)}$ are gathered in Tables 1Values of $P_{(d,\phi_{(d,2)}(n),2)}$ and $\alpha_{(d,2)}$, for $n \in [\![2,5]\!]$. These values are rounded to the nearest ten-thousandth.table.caption.3 and 2Values of $P_{(d,\phi_{(d,3)}(n),3)}$ and $\alpha_{(d,3)}$, for $n \in [\![3,5]\!]$. These values are rounded to the nearest ten-thousandth.table.caption.4.

d	$P_{(d,\phi_{(d,2)}(2),2)}$	$P_{(d,\phi_{(d,2)}(3),2)}$	$P_{(d,\phi_{(d,2)}(4),2)}$	$P_{(d,\phi_{(d,2)}(5),2)}$	$\alpha_{(d,2)}$
0	0.1330	0.1144	0.1123	0.1121	0.1121
1	0.1190	0.1103	0.1092	0.1091	0.1091
2	0.1107	0.1068	0.1063	0.1062	0.1062
3	0.1044	0.1037	0.1035	0.1035	0.1035
4	0.0991	0.1007	0.1009	0.1009	0.1009
5	0.0945	0.0979	0.0983	0.0984	0.0984
6	0.0903	0.0953	0.0958	0.0959	0.0959
7	0.0865	0.0927	0.0935	0.0936	0.0936
8	0.0829	0.0902	0.0912	0.0913	0.0913
9	0.0796	0.0879	0.0889	0.0891	0.0891

Table 1: Values of $P_{(d,\phi_{(d,2)}(n),2)}$ and $\alpha_{(d,2)}$, for $n \in [\![2,5]\!]$. These values are rounded to the nearest ten-thousandth.

d	$P_{(d,\phi_{(d,3)}(3),3)}$	$P_{(d,\phi_{(d,3)}(4),3)}$	$P_{(d,\phi_{(d,3)}(5),3)}$	$\alpha_{(d,3)}$
0	0.1045	0.1015	0.1012	0.1012
1	0.1028	0.1011	0.1009	0.1009
2	0.1017	0.1008	0.1007	0.1006
3	0.1008	0.1004	0.1004	0.1004
4	0.1000	0.1001	0.1001	0.1001
5	0.0993	0.0998	0.0999	0.0999
6	0.0986	0.0995	0.0996	0.0996
7	0.0980	0.0992	0.0993	0.0994
8	0.0974	0.0989	0.0991	0.0991
9	0.0968	0.0986	0.0988	0.0989

Table 2: Values of $P_{(d,\phi_{(d,3)}(n),3)}$ and $\alpha_{(d,3)}$, for $n \in [\![3,5]\!]$. These values are rounded to the nearest ten-thousandth.

6 Graphs of $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$

Let us plot graphs of sequences $(P_{(d,n,2)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$ for values of n from 10 to 1000 (Figure 1For $d \in [\![0,9]\!]$, graphs of $(P_{(d,n,2)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$.figure.caption.5). Then we plot graphs of $(P_{(d,n,3)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$, for $n \in [\![100,20000]\!]$ (Figure 2For $d \in [\![0,9]\!]$, graphs of $(P_{(d,n,3)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$. Note that points have not been all represented.figure.caption.5).

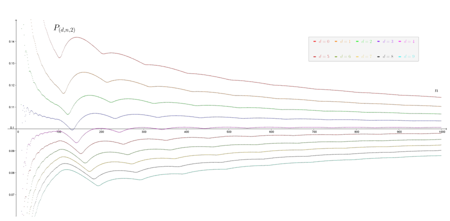

Figure 1: For $d \in \llbracket 0, 9 \rrbracket$, graphs of $(P_{(d,n,2)})_{n \in \mathbb{N} \setminus \llbracket 0, 10^{p-1}-1 \rrbracket}$.

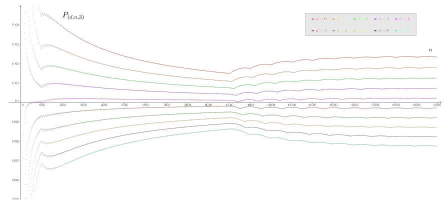

Figure 2: For $d \in \llbracket 0, 9 \rrbracket$, graphs of $(P_{(d,n,3)})_{n \in \mathbb{N} \setminus \llbracket 0, 10^{p-1}-1 \rrbracket}$. Note that points have not been all represented.

Let us plot two additional graphs of $P_{(d,n,2)}$ *versus* $\log(n)$ and $P_{(d,n,3)}$ *versus* $\log(n)$ for values of n from 10 to 2000000 (Figures 3For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6 and 4For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6).

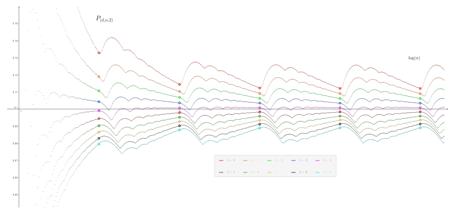

Figure 3: For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.

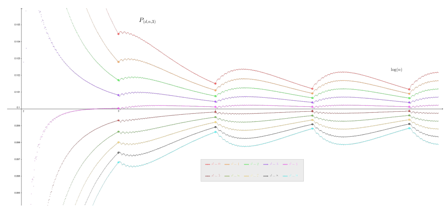

Figure 4: For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being represented by bigger plots.

Through Figures 3For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6 and 4For $d \in \llbracket 0, 9 \rrbracket$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6, the proportion of each d as p^{th} digit, for $d \in \llbracket 0, 9 \rrbracket$, seems to fluctuate and consequently not follow Benford's law. Each "pseudo cycle" seems to be composed of $9 \times 10^{p-2}$ short waves. Note that these observations were not obvious in view of Figures 1For $d \in \llbracket 0, 9 \rrbracket$, graphs of $(P_{(d,n,2)})_{n \in \mathbb{N} \setminus \llbracket 0, 10^{p-1}-1 \rrbracket}$.figure.caption.5 or 2For $d \in \llbracket 0, 9 \rrbracket$, graphs of $(P_{(d,n,3)})_{n \in \mathbb{N} \setminus \llbracket 0, 10^{p-1}-1 \rrbracket}$. Note that points have not been all represented.figure.caption.5.

Recall that, in the first digit case, similar cycles have already been highlighted (Herzel (1956); Flehinger (1966); Blondeau Da Silva (2019)).

We can also prove the following result:

Proposition 6.1. *For all $n \in \mathbb{N} \setminus [\![0, 10^{p-1}-1]\!]$ such that $n \geq 10^{p-1}+9$ and for all $(a,b) \in [\![0,9]\!]^2$ such that $a < b$, we have:*

$$P_{(a,n,p)} > P_{(b,n,p)} .$$

The relative position of graphs of $P_{(d,n,p)}$, for $d \in [\![0,9]\!]$, can be observed on Figures 1For $d \in [\![0,9]\!]$, graphs of $(P_{(d,n,2)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$.figure.caption.5, 2For $d \in [\![0,9]\!]$, graphs of $(P_{(d,n,3)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$. Note that points have not been all represented.figure.caption.5, 3For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6 and 4For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6.

Proof. $(a,b) \in [\![0,9]\!]^2$ such that $a < b$. For all $m \in [\![10^{p-1}, n]\!]$, let us denote by $\mathscr{E}_{(a,m)}$ the subset of $\mathbb{N}$ such that $\mathscr{E}_{(a,m)} = \{j \leq m : \text{the } p^{\text{th}} \text{ digit of } j \text{ is } a\}$.

For all $e \in \mathscr{E}_{(b,m)}$, we consider $e' = e - (b-a) \times 10^{dg-p}$ where dg is the number of digits of the integer e. It is clear that $e' \in \mathscr{E}_{(a,m)}$. Thus we get: $\left|E_{(a,m)}\right| \geq \left|E_{(b,m)}\right|$.

We also have $P_{(a,10^{p-1}+a,p)} = \frac{1}{a+1} > P_{(b,10^{p-1}+a,p)} = 0$. The result follows. □

Remark 6.2. For $n \in \mathbb{N} \setminus [\![0, 10^{p-1}-1]\!]$, we have, if $n < 10^{p-1}+d$, $P_{(d,n,p)} = 0$. Hence for all $n \in \mathbb{N} \setminus [\![0, 10^{p-1}-1]\!]$ and for all $(a,b) \in [\![0,9]\!]^2$ such that $a < b$, we have:

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

Let us henceforth provide the following equality:

Proposition 6.3.

$$P_{(d,n,p)} = \frac{1}{n+1-10^{p-1}} \Big(P_{(d,10^{k+p}-1,p)} \times (10^{k+p} - 10^{p-1}) + r_{(n,d,p)} \Big) ,$$

where:

$$k = \max\{i \in \mathbb{N} : 10^{i+p} \leq n\} .$$

Proof. The result comes directly from Proposition 4.1theorem.4.1. □

7 Study of $9 \times 10^{p-2}$ additional subsequences

To definitively bring to light the fact that the sequence $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$ does not converge, we will show that there exist additional subsequences that converge to limits different from those of $(P_{(d,\phi_{(d,p)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$.

For $i \in [\![10^{p-2}, 10^{p-1}-1]\!]$, let us in this way study the $9 \times 10^{p-2}$ subsequences $(P_{(d,\psi_{(d,p,i)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$ where $\psi_{(d,p,i)}$ is the function from $\mathbb{N} \setminus [\![0,p-1]\!]$ to $\mathbb{N}$ that maps n to $(10i + (d+1))10^{n-p+1} - 1$. We get the below result:

Proposition 7.1. *$i \in [\![10^{p-2}, 10^{p-1} - 1]\!]$.*
The subsequence $(P_{(d,\psi_{(d,p,i)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$ converges to:

$$\frac{\alpha_{(d,p)}10^{p-1} + i + 1 - 10^{p-2} - k_{(d,p,i)}d - 9l_{(d,p,i)} + m_{(d,p,i)} + n_{(d,p,i)} + 10^{p-2}\ln(\frac{10^{p-1}+d}{10^{p-1}})}{10i + d + 1},$$

where:

$$\begin{cases} k_{(d,p,i)} = \sum_{j=10^{p-2}}^{i} \ln(\frac{10j+(d+1)}{10j+d}) \\ l_{(d,p,i)} = \sum_{j=10^{p-2}}^{i} j\ln(\frac{10j+(d+1)}{10j+d}) \\ m_{(d,p,i)} = \sum_{j=10^{p-2}}^{i-1} \ln(\frac{10(j+1)+d}{10j+(d+1)}) \\ n_{(d,p,i)} = \sum_{j=10^{p-2}}^{i-1} j\ln(\frac{10(j+1)+d}{10j+(d+1)}) \,. \end{cases}$$

Proof. $i \in [\![10^{p-2}, 10^{p-1} - 1]\!]$. Thanks to Proposition 6.3theorem.6.3, we have, for $n \in \mathbb{N} \setminus [\![0, p-1]\!]$:

$$P_{(d,\psi_{(d,p,i)}(n),p)} = \frac{1}{\big(10i + (d+1)\big)10^{n-p+1} - 10^{p-1}}\Big(P_{(d,10^n-1,p)} \times (10^n - 10^{p-1}) \\ + r_{(\psi_{(d,p,i)}(n),d,p)}\Big)\,.$$

The first term of $r_{(\psi_{(d,p,i)}(n),d,p)}$ can be simplified as follows:

$$\sum_{j=10^{p-2}}^{i} \sum_{b=(10j+d)10^{n-p+1}}^{(10j+(d+1))10^{n-p+1}-1} \Big(1 - \frac{(9j+d)10^{n-p+1} + 10^{p-2} - 10^{p-1}}{b + 1 - 10^{p-1}}\Big)$$
$$= 10^{n-p+1}(i - 10^{p-2} + 1) - \sum_{j=10^{p-2}}^{i} \big((9j+d)10^{n-p+1} + 10^{p-2} - 10^{p-1}\big)$$
$$\sum_{b=(10j+d)10^{n-p+1}}^{(10j+(d+1))10^{n-p+1}-1} \frac{1}{b + 1 - 10^{p-1}}$$
$$\underset{n\to+\infty}{\sim} 10^{n-p+1}(i - 10^{p-2} + 1) - \sum_{j=10^{p-2}}^{i} (9j+d)10^{n-p+1}\ln(\frac{10j+(d+1)}{10j+d})\,,$$

thanks to inequalities (5Study of a particular subsequenceequation.5.5).

The second term of $r_{(\psi_{(d,p,i)}(n),d,p)}$ can be simplified as follows:

$$\sum_{j=10^{p-2}-1}^{i-1} \sum_{a=\max(10^n,(10j+(d+1))10^{n-p+1})}^{(10(j+1)+d)10^{n-p+1}-1} \frac{10^{n-p+1}(j+1)-10^{p-2}}{a+1-10^{p-1}}$$

$$= \left(10^{n-p+1}10^{p-2}-10^{p-2}\right) \sum_{a=10^n}^{(10^{p-1}+d)10^{n-p+1}-1} \frac{1}{a+1-10^{p-1}}$$

$$+ \left(10^{n-p+1}(j+1)-10^{p-2}\right) \sum_{j=10^{p-2}}^{i-1} \sum_{a=(10j+(d+1))10^{n-p+1}}^{(10(j+1)+d)10^{n-p+1}-1} \frac{1}{a+1-10^{p-1}}$$

$$\underset{n\to+\infty}{\sim} 10^{n-1}\ln\left(\frac{10^{p-1}+d}{10^{p-1}}\right) + \sum_{j=10^{p-2}}^{i-1} 10^{n-p+1}(j+1)\ln\left(\frac{10(j+1)+d}{10j+(d+1)}\right),$$

thanks to inequalities (5Study of a particular subsequenceequation.5.5).
Knowing that $P_{(d,10^n-1,p)} \underset{n\to+\infty}{\sim} \alpha_{(d,p)}$ (see Proposition 5.1theorem.5.1), the result follows. □

Let us denote by $\alpha_{(d,p,i)}$ the limit of $(P_{(d,\psi_{(d,p,i)}(n),p)})_{n\in\mathbb{N}\setminus[\![0,p-1]\!]}$. Few values of $P_{(d,\psi_{(d,p,i)}(n),p)}$ are gathered in Tables 3Values of $P_{(d,\psi_{(d,2,7)}(n),2)}$ and $\alpha_{(d,2,7)}$, for $n \in [\![2,5]\!]$ and $i = 7$. These values are rounded to the nearest ten-thousandth.table.caption.7 and 4Values of $P_{(d,\psi_{(d,3,23)}(n),3)}$ and $\alpha_{(d,3,23)}$, for $n \in [\![3,5]\!]$ and $i = 23$. These values are rounded to the nearest ten-thousandth.table.caption.8.

d	$P_{(d,\psi_{(d,2,7)}(2),2)}$	$P_{(d,\psi_{(d,2,7)}(3),2)}$	$P_{(d,\psi_{(d,2,7)}(4),2)}$	$P_{(d,\psi_{(d,2,7)}(5),2)}$	$\alpha_{(d,2,7)}$
0	0.1182	0.1152	0.1148	0.1148	0.1148
1	0.1127	0.1111	0.1109	0.1109	0.1109
2	0.1082	0.1074	0.1073	0.1073	0.1073
3	0.1042	0.1040	0.1040	0.1040	0.1039
4	0.1006	0.1008	0.1008	0.1008	0.1008
5	0.0973	0.0978	0.0979	0.0979	0.0979
6	0.0942	0.0950	0.0951	0.0951	0.0951
7	0.0913	0.0923	0.0924	0.0925	0.0925
8	0.0886	0.0898	0.0899	0.0900	0.0900
9	0.0860	0.0874	0.0876	0.0876	0.0876

Table 3: Values of $P_{(d,\psi_{(d,2,7)}(n),2)}$ and $\alpha_{(d,2,7)}$, for $n \in [\![2,5]\!]$ and $i = 7$. These values are rounded to the nearest ten-thousandth.

As a result, the sequence $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$ does not converge. The $9 \times 10^{p-2}$ convergent subsequences confirm the remarks raised by Figures 3For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6 and 4For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being

d	$P_{(d,\psi_{(d,3,23)}(3),3)}$	$P_{(d,\psi_{(d,3,23)}(4),3)}$	$P_{(d,\psi_{(d,3,23)}(5),3)}$	$\alpha_{(d,3,23)}$
0	0.1037	0.1023	0.1022	0.1021
1	0.1026	0.1018	0.1017	0.1017
2	0.1017	0.1012	0.1012	0.1012
3	0.1009	0.1007	0.1007	0.1007
4	0.1007	0.1002	0.1002	0.1002
5	0.0995	0.0997	0.0997	0.0997
6	0.0988	0.0992	0.0993	0.0993
7	0.0982	0.0987	0.0988	0.0988
8	0.0976	0.0983	0.0983	0.0983
9	0.0969	0.0978	0.0979	0.0979

Table 4: Values of $P_{(d,\psi_{(d,3,23)}(n),3)}$ and $\alpha_{(d,3,23)}$, for $n \in [\![3,5]\!]$ and $i = 23$. These values are rounded to the nearest ten-thousandth.

represented by bigger plots.figure.caption.6 about the existence of "pseudo cycles" in the graph of $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$.

8 Central values

From Figures 3For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,2)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first five values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6 and 4For $d \in [\![0,9]\!]$, graphs of $P_{(d,n,3)}$ *versus* $\log(n)$. Note that points have not been all ploted. The first four values of the above defined subsequence, for each d, being represented by bigger plots.figure.caption.6, we notice that there exist fluctuations in the graph of $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$. We define $C_{(d,p)}$ as follows:

Definition 8.1.

$$C_{(d,p)} = \frac{1}{9 \times 10^{p-2}} \sum_{i=10^{p-2}}^{10^{p-1}-1} \alpha_{(d,p,i)} \, .$$

Figure 5Graph of $P_{(0,n,2)}$ *versus* $\log(n)$. Note that points have not been all represented. Lines whose equation is $y = \alpha_{(0,2,i)}$, for $i \in [\![1,9]\!]$, have also been ploted. Note that those of equations $y = \alpha_{(0,2,1)}$ and $y = \alpha_{(0,2,7)}$ are almost coincident. We have $C_{(0,2)} \approx 0.1170$.figure.caption.9 below shows the different values of $\alpha_{(0,2,i)}$, for $i \in [\![1,9]\!]$ and also the values of $P_{(0,n,2)}$ *versus* $\log(n)$ for $n \in [\![10,2000000]\!]$:

These means values are very close to the theoric value highlighted in Hill (1995) as can be seen in below tables (Tables 5Values of $C_{(d,p)}$ and probabilities associated to the second digit (Hill (1995)), for $p = 2$. These values are rounded to the nearest ten-thousandth.table.caption.10 and 6Values of $C_{(d,p)}$ and probabilities associated to the third digit (Hill (1995)). These values are rounded to the nearest ten-thousandth.table.caption.11, where $p = 2$ and $p = 3$, respectively).

We furthermore note, thanks to Table 5Values of $C_{(d,p)}$ and probabilities associated to the second digit (Hill (1995)), for $p = 2$. These values are rounded to the nearest

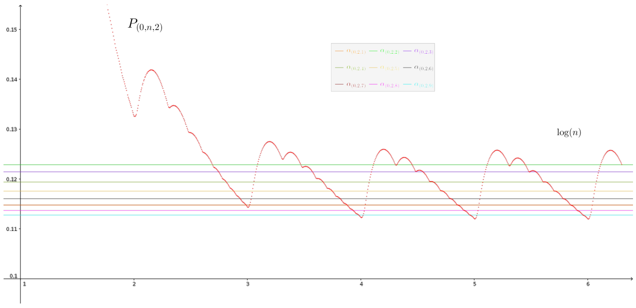

Figure 5: Graph of $P_{(0,n,2)}$ *versus* $\log(n)$. Note that points have not been all represented. Lines whose equation is $y = \alpha_{(0,2,i)}$, for $i \in [\![1, 9]\!]$, have also been ploted. Note that those of equations $y = \alpha_{(0,2,1)}$ and $y = \alpha_{(0,2,7)}$ are almost coincident. We have $C_{(0,2)} \approx 0.1170$.

ten-thousandth.table.caption.10, that $C_{(0,2)}$ slightly underestimates $\sum_{j=1}^{9} \log(1 + \frac{1}{10j})$ as can be infered from Figure 5Graph of $P_{(0,n,2)}$ *versus* $\log(n)$. Note that points have not been all represented. Lines whose equation is $y = \alpha_{(0,2,i)}$, for $i \in [\![1, 9]\!]$, have also been ploted. Note that those of equations $y = \alpha_{(0,2,1)}$ and $y = \alpha_{(0,2,7)}$ are almost coincident. We have $C_{(0,2)} \approx 0.1170$.figure.caption.9.

It can be added that Definition 8.1theorem.8.1 comes close to:

$$P^3_{10^{p-1}}(A) = \sum_{M=1}^{10^{p-1}} P^2_M(A) \ ,$$

where A is the first digit considered, defined in Flehinger (1966). These cumulative averages allowed Flehinger to approximate ever more finely Benford's first digit values.

9 Applications

Among the different domains studied by Benford (Benford (1938)), some could be well adapted to our model: sizes of populations or street addresses for example (see Blondeau Da Silva (2019) for a detailed explanation). In Janvresse and De La Rue (2004), Janvresse and De La Rue advised precisely to use their own similar model in the case of street addresses or when considering the first-page numbers of articles in a bibliography.

Indeed 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.

The below proposition could help us to determine whether a data set is likely to verify these conditions or not. Actually our model induces the following distribution of positive integers in our datasets:

Proposition 9.1. *Let us keep the previous notations. We denote by $X_{(n,p)}$ the random variable from Ω_n to $\mathbb{N}$ that maps each element ω of Ω_n to the second component of ω.*

d	$C_{(d,2)}$	$\sum_{j=1}^{9}\log(1+\frac{1}{10j+d})$
0	0.1170	0.1197
1	0.1122	0.1139
2	0.1079	0.1088
3	0.1039	0.1043
4	0.1001	0.1003
5	0.0967	0.0967
6	0.0935	0.0934
7	0.0905	0.0904
8	0.0878	0.0876
9	0.0851	0.0850

Table 5: Values of $C_{(d,p)}$ and probabilities associated to the second digit (Hill (1995)), for $p=2$. These values are rounded to the nearest ten-thousandth.

We have:

$$\forall k \in [\![10^{p-1};n]\!], \qquad \mathrm{P}(X_{(n,p)}=k)=\frac{1}{n-10^{p-1}+1}\sum_{i=k+1-10^{p-1}}^{n+1-10^{p-1}}\frac{1}{i}\ .$$

Proof. Let us recall that $F_{(n,p)}$ is the random variable from Ω_n to $[\![1,n+1-10^{p-1}]\!]$ that maps each element ω of Ω_n to the first component of ω. According to the law of total probability, we state, for $k\in[\![10^{p-1};n]\!]$:

$$\begin{aligned}
\mathrm{P}(X_{(n,p)}=k) &= \sum_{i=1}^{n+1-10^{p-1}} \mathrm{P}(X_{(n,p)}=k|F_{(n,p)}=i)\,\mathrm{P}(F_{(n,p)}=i)\\
&= \sum_{i=1}^{n+1-10^{p-1}} \mathrm{P}(X_{(n,p)}=k|F_{(n,p)}=i)\times\frac{1}{n+1-10^{p-1}}\\
&= \frac{1}{n+1-10^{p-1}}\sum_{i=k+1-10^{p-1}}^{n+1-10^{p-1}} \mathrm{P}(X_{(n,p)}=k|F_{(n,p)}=i)\\
&= \frac{1}{n+1-10^{p-1}}\sum_{i=k+1-10^{p-1}}^{n+1-10^{p-1}} \frac{1}{i}\,.
\end{aligned}$$

□

This result is still valid for $p=1$.

Hence, in order to conform as closely as possible to our model, the studied database must have a distribution similar to that described in Proposition 9.1theorem.9.1. Figures

d	$C_{(d,3)}$	$\sum_{j=10}^{99} \log(1 + \frac{1}{10j+d})$
0	0.1016	0.1018
1	0.1013	0.1014
2	0.1009	0.1010
3	0.1005	0.1006
4	0.1002	0.1002
5	0.0998	0.0998
6	0.0994	0.0994
7	0.0991	0.0990
8	0.0987	0.0986
9	0.0984	0.0983

Table 6: Values of $C_{(d,p)}$ and probabilities associated to the third digit (Hill (1995)). These values are rounded to the nearest ten-thousandth.

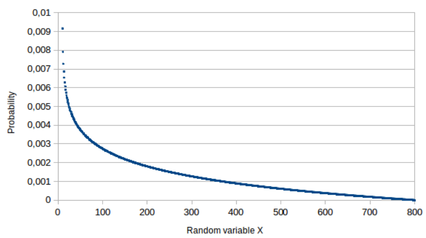

Figure 6: The second digit case ($p = 2$), the value of n being 800.

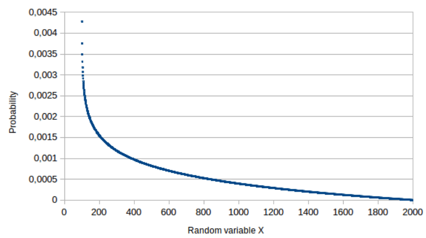

Figure 7: The third digit case ($p = 3$), the value of n being 2000.

6The second digit case ($p = 2$), the value of n being 800.figure.caption.12 and 7The third digit case ($p = 3$), the value of n being 2000.figure.caption.12 provide two examples:

Let's take an example:

Example 9.2. The addresses of 42975 French educational establishments are available on the site `www.data.gouv.fr` (data.gouv.fr (2017)), which is an open platform for French public data. Among these establishments, we limit ourselves to the study of those whose street numbers are strictly less than 200: they represent more than 95% of them.

Their distribution (Figure 8The distribution of the street numbers (less than 200) of French educational establishments.figure.caption.13) looks similar to the one described by our model (Proposition 9.1theorem.9.1). In Table 7Values of frequency of d as second digit in the database, Benford's theoretical values and $P_{(d,199,2)}$ values, for $d \in [\![0, 9]\!]$. These values are rounded to the nearest ten-thousandth.table.caption.14 below, the frequencies of each $d \in [\![0, 9]\!]$, regarding the second digit of address numbers, are listed.

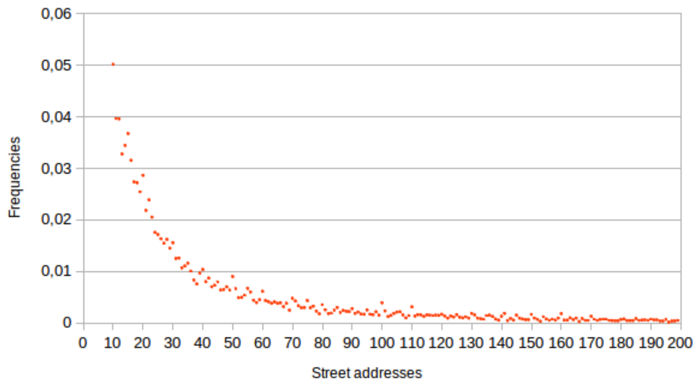

Figure 8: The distribution of the street numbers (less than 200) of French educational establishments.

d	Frequency in the database	$P_{(d,199,2)}$	Benford's values
0	0.1503	0.1349	0.1197
1	0.1187	0.1230	0.1139
2	0.1137	0.1143	0.1088
3	0.0980	0.1069	0.1043
4	0.0968	0.1003	0.1003
5	0.1024	0.0944	0.0967
6	0.0889	0.0889	0.0934
7	0.0792	0.0838	0.0904
8	0.0786	0.0790	0.0876
9	0.0736	0.0745	0.0850

Table 7: Values of frequency of d as second digit in the database, Benford's theoretical values and $P_{(d,199,2)}$ values, for $d \in [\![0,9]\!]$. These values are rounded to the nearest ten-thousandth.

It can be seen that the values for $d = 0$ and $d \in \{6,7,8,9\}$ are much better with our model. This can be quantified with Pearson's chi-squared test of goodness of fit (Pearson (1900)). The chi-squared test statistic is approximately 85.9 regarding our model and approximately 300.8 considering Benford's values. The critical value from the chi-squared distribution with 9 degrees of freedom and 95% confidence level is approximately 16.92. The observed frequencies distributions in both cases differ from the theoretical distributions, but the results of our model are undeniably better.

However, limits of our example lie in the fact that (i) the distribution of street numbers (Figure 8The distribution of the street numbers (less than 200) of French educational establishments.figure.caption.13) is far from being perfectly identical to the one described in Proposition 9.1theorem.9.1 (indeed the curve tends more slowly towards the abscissa axis and some street numbers are over-represented, as some round numbers: 50, 60, 70, 75, 80, 100, 110 *etc.*); (ii) the value 199 is rather arbitrary, it enables avoiding part of the pitfalls described above in (i) but deviates us from the model (indeed beyond 200 the curve continues to tend slowly toward the axis, contrary to our model, and the weight of the over-represented numbers becomes overly large).

Finally, Benford's law, as previously noted, is not as effective as expected in determining with certainty whether a dataset is fraudulent or not. Our probability distributions could obtain better results allowing a certain flexibility in the frequencies of occurrence of digits around Benford's values: some data sets falsely judged fraudulent could thus be declared "normal" and *vice versa*.

10 Conclusion

To conclude, through our model, we have seen that the proportion of d as p^{th} digit, $d \in [\![0, 9]\!]$, in certain naturally occurring collections of data is more likely to follow a law whose probability distribution is $(d, P_{(d,n,p)})_{d\in[\![0,9]\!]}$, where n is the smaller integer upper bound of the physical, biological or economical quantities considered, rather than the generalized Benford's law. 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 $d \in [\![0, 9]\!]$ as p^{th} 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 10^{p-1}: first terms in proportion formulas become rapidly negligible. Through our model we understand that the predominance of 0 as p^{th} digit (followed by those of 1 and so on) is all but surprising in experimental data: it is only due to the fact that, in the lexicographical order, 0 appears before 1, 1 appears before 2, *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.

Appendix: Python script

Using Proposition 4.1theorem.4.1, we can determine the terms of $(P_{(d,n,p)})_{n\in\mathbb{N}\setminus[\![0,10^{p-1}-1]\!]}$, for $d \in [\![0, 9]\!]$. 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 Van Rossum (1995)). The implemented function *expvalProp* has three parameters: the rank n of the wanted term of the sequence, the position p of the considered digit and the value d of this digit. Here is the used algorithm:

```
def expvalProp(n,d,p):
```

```
k=-1;
while(10**(k+p+1)¡=n):
    k=k+1
l=math.floor((n-(10**(p-1)+d)*10**(k+1))/10**(k+2))+10**(p-2);S=0;T=0;
if (k!=-1):
    for i in range(0,k+1):
        for j in range(10**(p-2),10**(p-1)):
            for b in range((10*j+d)*10**i,(10*j+(d+1))*10**i):
                T=T+(b-((9*j+d)*10**i+10**(p-2)-1))/(b+1-10**(p-1))
        for j in range(10**(p-2)-1,10**(p-1)):
            for a in range(max(10**(p+i-1),(10*j+(d+1))*10**i),min(10**(p+i),(10*
                                                                    (j+1)+d)*10**i)):
                S=S+((j+1)*10**i-10**(p-2))/(a+1-10**(p-1))
if ((math.floor(n/10**(k+1))-10*math.floor(n/10**(k+2)))==d):
    for j in range(10**(p-2),l+1):
        for b in range((10*j+d)*10**(k+1),min(n,(10*j+(d+1))*10**(k+1)-1)+1):
            T=T+(b-((9*j+d)*10**(k+1)+10**(p-2)-1))/(b+1-10**(p-1))
    for j in range(10**(p-2)-1,l):
        for a in range(max(10**(p+k),(10*j+(d+1))*10**(k+1)),(10*(j+1)+d)*10**
                                                                            (k+1)):
            S=S+((j+1)*10**(k+1)-10**(p-2))/(a+1-10**(p-1))
else:
    for j in range(10**(p-2),l+1):
        for b in range((10*j+d)*10**(k+1),(10*j+(d+1))*10**(k+1)):
            T=T+(b-((9*j+d)*10**(k+1)+10**(p-2)-1))/(b+1-10**(p-1))
    for j in range(10**(p-2)-1,l+1):
        for a in range(max(10**(p+k),(10*j+(d+1))*10**(k+1)),min(n,(10*(j+1)+d)
                                                                    *10**(k+1)-1)+1):
            S=S+((j+1)*10**(k+1)-10**(p-2))/(a+1-10**(p-1))
return((S+T)/(n+1-10**(p-1)))
```

References

Alexopoulos, T. and Leontsinis, S. (2014). Benford's Law in astronomy. *Journal of Astrophysics and Astronomy*, 35(4):639–648.

Alves, A. D., Yanasse, H. H., and Soma, N. Y. (2014). Benford's Law and articles of scientific journals: comparison of JCR and Scopus data. *Scientometrics*, 98:173–184.

Beer, T. W. (2009). Terminal digit preference: beware of Benford's Law. *Journal of Clinical Pathology*, 62(2):192.

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

Berger, A. and Hill, T. (2015). *An Introduction to Benford's Law*. Princeton University Press, Princeton, NJ.

Blondeau Da Silva, S. (2019). Benford or not Benford: a systematic but not always well-founded use of an elegant law in experimental fields. *Communications in Mathematics and Statistics*. https://doi.org/10.1007/s40304-018-00172-1.

Burke, J. and Kincanon, E. (1991). Benford's Law and physical constants: the distribution of initial digits. *American Journal of Physics*, 59:952.

Carslaw, C. (1988). Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review*, 63(2):321–327.

Cho, W. K. T. and Gaines, B. J. (2007). Breaking the (Benford) Law: Statistical fraud detection in campaign finance. *The American Statistician*, 61(3):218–223.

Costasa, E., Lopez-Rodasa, V., Torob, F., and Flores-Moya, A. (2008). The number of cells in colonies of the cyanobacterium microcystis aeruginosa satisfies Benford's Law. *Aquatic Botany*, 89(3):341–343.

data.gouv.fr (2017). Adresse et géolocalisation des établissements d'enseignement du premier et second degrés. https://www.data.gouv.fr/fr/datasets/adresse-et-geolocalisation-des-etablissements-denseignement-du-premier-et-second-degres-1/.

Deckert, J., Myagkov, M., and Ordeshook, P. (2011). Benford's Law and the detection of election fraud. *Political Analysis*, 19:245–268.

Diekmann, A. (2007). Not the first digit! Using Benford's Law to detect fraudulent scientific data. *Journal of Applied Statistics*, 34(3):321–329.

Durtschi, C., Hillison, W., and Pacini, C. (2004). The effective use of benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, V:17–34.

Flehinger, B. J. (1966). On the probability that a random integer has initial digit a. *American Mathematical Monthly*, 73(10):1056–1061.

Friar, J. L., Goldman, T., and Pérez-Mercader, J. (2012). Genome sizes and the Benford distribution. *Plos One*, 7(5).

Gauvrit, N. and Delahaye, J.-P. (2011). *Scatter and Regularity Imply Benford's Law... and More*, pages 53–69. World Scientific.

Geyer, D. (2010). Detecting fraud in financial data sets. *Journal of Business and Economics Research*, 8(7):75–83.

Golbeck, J. (2015). Benford's Law applies to online social networks. *Plos One*, 10(8).

Goodman, W. (2016). The promises and pitfalls of Benford's Law. *Significance*, 13(3):38–41.

Herzel, A. (1956). Sulla distribuzione delle cifre iniziali dei numeri statistic. In *Atti XV e XVI Riunione sci., Roma*, pages 205–228.

Hill, T. (1988). Random-number guessing and the first digit phenomenon. *Psychological Reports*, 62(3):967–971.

Hill, T. (1995). The significant-digit phenomenon. *The American Mathematical Monthly*, 102(4):322–327.

Janvresse, E. and De La Rue, T. (2004). From uniform distributions to Benford's Law. *Journal of Applied Probability*, 41(4):1203–1210.

Joenssen, D. W. (2013). Two digit testing for benford's law. In *Proceedings of the ISI World Statistics Congress, 59th Session in Hong Kong.*

Knuth, D. (1969). *The Art of Computer Programming 2.* Addison-Wesley, New-York.

Leemis, L., Schmeiser, B., and Evans, D. (2000). Survival distributions satisfying Benford's Law. *The American Statistician*, 54(4):236–241.

Logan, J. L. and Goudsmit, S. A. (1978). The first digit phenomenon. In *Proceedings of the American Philosophical Society*, volume 122, pages 193–197.

Mebane Jr, W. R. (2006). Election forensics: The second-digit benford's law test and recent american presidential elections. In *Proceedings of the Election Fraud Conference*.

Miller, S. J., editor (2015). *Benford's Law: Theory and Applications*. Princeton University Press, Princeton, NJ.

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

Nigrini, M. (1999). I've got your number : How a mathematical phenomenon can help CPAs uncover fraud and other irregularities. *Journal of Accountancy*.

Nigrini, M. and Miller, S. (2007). Benford's Law applied to hydrology data-results and relevance to other geophysical data. *Mathematical Geology*, 39(5):469–490.

Nigrini, M. and Wood, W. (1995). Assessing the integrity of tabulated demographic data. Preprint.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, 50(302):157–175.

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

Raimi, R. A. (1985). The first digit phenomenon again. In *Proceedings of the American Philosophical Society*, volume 129, pages 211–219.

Rauch, B., Göttsche, M., Brälher, G., and Engel, S. (2011). Fact and fiction in EU-governmental economic data. *German Economic Review*, 12(3):243–255.

Saville, A. (2006). Using Benford's Law to detect data error and fraud: An examination of companies listed on the Johannesburg Stock Exchange. *South African Journal of Economic and Management Sciences*, 9(3):341–354.

Scott, P. D. and Fasli, M. (2001). *Benford's Law: an empirical investigation and a novel explanation*. CSM Technical Report 349, University of Essex. https://cswww.essex.ac.uk/technical-reports/2001/CSM-349.pdf.

Sehity, T., Hoelz, E., and Kirchler, E. (2005). Price developments after a nominal shock: Benford's Law and psychological pricing after the euro introduction. *International Journal of Research in Marketing*, 22(4):471–480.

Tödter, K. (2009). Benford's Law as an indicator of fraud in economics. *German Economic Review*, 10:339–351.

Thomas, J. K. (1989). Unusual patterns in reported earnings. *Accounting Review*, 64(4):773–787.

Tolle, C., Budzien, J., and Laviolette, R. (2000). Do dynamical systems follow Benford's Law? *Chaos*, 10(2).

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

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