# Université de Nîmes

Laboratoire MIPA
Université de Nîmes, Site des Carmes
Place Gabriel Péri, 30021 Nîmes, France
http://mipa.unimes.fr


**The Benford phenomenon for random variables.**
**Discussion of Feller's way**

by


Michel Valadier


March 2012

# The Benford phenomenon for random variables. Discussion of Feller's way

Michel VALADIER

(draft version, March 9, 2012)

## 1 Introduction

F. Benford [B, 1938] (and earlier S. Newcomb [N, 1881]) observed that, in numerical data, when the numbers are written in base 10, very often the first digit, which is an integer between 1 and 9, takes the value 1 with a frequency much greater than 1/9 since close to[1] $\log 2 = 0.3010...$ More generally the *Benford phenomenon* would be that the first digit in base 10, let us denote it by $D$, follows the law:

$$\mathbf{P}(D = k) = \log\Big(\frac{k+1}{k}\Big) \quad (k \in \{1, ..., 9\})\,. \tag{1}$$

(Note that $\displaystyle\sum_{k=1}^{9} \log\Big(\frac{k+1}{k}\Big) = \log(10) - \log(1) = 1$.)

COMMENTS. The Benford phenomenon is not so intuitive. On the contrary: for example integers with 4 digits go from 1000 to 9999. "Random" would give the probability 1/9000 for each and the probability 1/9 for $\{D = 1\}$. For more see Raimi [R]. Not any data can satisfy the Benford phenomenon. As said by Scott et Fasli [SF] (just before Section 3) the height of men will give, essentially $D = 1$ if expressed in meters, and, if expressed in feet[2], $D$ will take mainly the values 4, 5 and 6. Academic example: if the law of $X$ is uniform on $[1, 2]$, $D = 1$ almost surely. For negative examples see Section 8.2.

This paper is devoted to the first digit of a random variable and not to the other digits or to dynamical systems. The litterature is tremendous:

[1] We denote by log the logarithm in base 10. The logarithm in base $e$ will be denoted by ln.

[2] The foot equals 0.3048 meter.

Hürlimann [Hu] gives till 2006, around 350 references; and Berger and Hill [BH3] quote around 600 papers (see also [Bee]). Maybe I missed some important results.

I will discuss some arguments starting from Feller [F, 1966] and rely specially on Engel-Leuenberger [EL, 2003], Dümbgen-Leuenberger [DL, 2008], Gauvrit-Delahaye [GD1, GD2, GD3, 2008–2009], Berger [Br, 2010]. Some "disaster" appears: see Section 7. Section 9 gives naive results. And I will not discuss papers relying on Fourier Analysis such as Pinkham [Pi, 1961], Good [Go, 1986], Boyle [Bo, 1994].

## 2 Preliminaries

Let $X$ be a random variable (briefly r.v.) with values in $\mathbb{R}_+^* = ]0, +\infty[$. Let us denote by $D(\omega)$ the first digit in base 10 of $X(\omega)$. It belongs to $\{1, ..., 9\}$. Let $n \in \mathbb{Z}$ and $k \in \{1, ..., 9\}$; when $X$ belongs to the interval $\left[10^n, 10^{n+1}\right[$,

$$D = k \text{ is equivalent to } X \in [k\,10^n, (k+1)10^n[\ .$$

We abbreviate $\{\omega\,;\, D(\omega) = k\}$ in $\{D = k\}$. The following covering is a partition (pairwise disjoint subsets)

$$\{D = k\} = \bigcup_{n \in \mathbb{Z}} \left\{X \in [k\,10^n, (k+1)10^n[\right\}.$$

Plenty of arguments are expressed with the r.v. $Y := \log(X)$, which takes its values in $\mathbb{R}$. With this r.v. the following partition holds

$$\{D = k\} = \bigcup_{n \in \mathbb{Z}} \left\{Y \in [\log(k) + n, \log(k+1) + n[\right\}. \tag{2}$$

Let $\mathcal{M}(y)$ denotes the *mantissa* of the real number $y$ defined by:

$$\text{if } n \in \mathbb{Z} \text{ and } y \in [n, n+1[\,, \quad \mathcal{M}(y) := y - n\,.$$

Thus $D = k$ is equivalent to (cf. (2))

$$\mathcal{M}(Y) \in [\log k, \log(k+1)[\ . \tag{3}$$

Assume $Y$ has the density $g$. Then $\mathcal{M}(Y)$ has the density (this is in [F, (8.3) p. 62], [Go, (5.7) p. 162], [DL, Section 2.1]),

$$[0, 1] \ni y \mapsto \sum_{n \in \mathbb{Z}} g(n + y) =: \bar{g}(y)\,, \tag{4}$$

(the point 1 should not be in the domain but later for expressing the total variation of $\bar{g}$ it will be useful). And let us denote by $\overline{G}$ the cumulative distribution function of $\mathcal{M}(Y)$:

$$\forall z \in [0,1], \quad \overline{G}(z) = \int_0^z \bar{g}(u)\,du\,.$$

We say that $g$ is *unimodal* if $g$ is non-decreasing till some abscissa, and then non-increasing.

The end of this Section is not necessary to understand the remaining of the paper, but it corrects the impression caused by the seemingly non-smooth definition of the mantissa. Classically the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is identified to $[0,1[$. With this identification the canonical surjection $\varphi : \mathbb{R} \to \mathbb{T}$ coincides with the mantissa $\mathcal{M}$. A geometrical view is: use as for $\mathbb{T}$ the unit circle $\mathbb{U}$ via the identification

$$\mathbb{T} \sim [0,1[\ \ni t \mapsto e^{2\pi i t} \in \mathbb{U}$$

and as for $\mathbb{R}$ the helicoid $\mathbb{H}$ via the identification

$$\mathbb{R} \ni t \mapsto (e^{2\pi i t}, t) \in \mathbb{H} \subset \mathbb{U} \times \mathbb{R}\,.$$

Then $\varphi$ becomes the very smooth map

$$\mathbb{H} \ni (z,y) \mapsto z \in \mathbb{U}\,.$$

# 3 Position of the problem

Benford's phenomenon for the first digit is exactly satisfied[3] if

$$\forall k \in \{2,\ldots,9\}, \quad \overline{G}(\log k) = \log k\,.$$

As for approximation one can ask for

$$\forall k \in \{2,\ldots,9\}, \quad \overline{G}\,(\log k) - \log k \ \text{ is small}$$

or "$\bar{g}$ is close to the constant function $\mathbf{1}_{[0,1]}$".

When we will have found a good sufficient condition expressed with $g$ (that is with $Y$) the difficulty will be, after expressing it with $X = 10^Y$, to inventory which laws satisfy the sufficient condition. See Section 7.

[3] We will not use the following: $\forall z \in [0,1], \overline{G}(z) = z$ (equivalently $\forall z,\ \bar{g}(z) = 1$) in which case one commonly says $X$ satisfies "the *Benford law*"; more correct would be: $X$ modulo 10 (in the multiplicative group $\mathbb{R}_+^*$) obeys to the Benford law.

## 4 Poincaré's roulette problem

This title comes from Feller [F, Section 8 (b) p. 62][4]. If a ball is launched from a given zero point on a circle[5] of circumference 1, if the length path is $Y$, the final position of the ball will be $\mathcal{M}(Y)$.

When is the law of $\mathcal{M}(Y)$ close to the uniform distribution? Intuitively if one throws the ball with sufficient force and no special effort to get an integer number of revolutions or some other precise result, the uniform distribution will be approached.

Feller [F, p. 62] says that a valid assumption is "$g$ is sufficiently spread" (implying a small maximum). This is a bit fuzzy. As soon after Feller gives a precise and relevant hypothesis: the density $g$ of $Y = \log X$ is unimodal and has a small maximum. (Below we will quote Pinkham [Pi, 1961] who worked with a better hypothesis.) In my opinion there is a flaw in [F] about which I could not find any precise reference in the litterature. It is the following: the point $x_k$ defined just after (8.4), which is nothing else but $\mathcal{M}(x-a)+a+k$, is not necessarily on the left of $[a+k, a+k+1[$, so the assertion, just below (8.5), "For $k < 0$ the integrand is $\leq 0$" is not correct. Nevertheless Raimi [R, p. 533] quotes Feller.

Under the unimodality hypothesis Gauvrit and Delahaye in 2008 [GD1] gave a correct proof of

$$\forall z \in [0,1], \quad |\overline{G}(z) - z| \leq 2 \max g \,. \tag{5}$$

Their proof is in the line of Feller but they they do not quote him; in [GD2] they spoke of "scatter and regularity" which are surely not the good words. Dümbgen-Leuenberger in 2008 [DL, Th.1 and Cor.2] still starting from the same "spreadness idea" (they assume that the *total variation* of $g$ is small) give far more better bounds: see Section 6 below. Already in 1961 Pinkham [Pi, Corollary p. 1229] (quoted by Raimi [R, (8.12) p. 533]) using Fourier Analysis arguments obtained

$$\sup_{0 \leq z \leq 1} |\overline{G}(z) - z| \leq \mathrm{TV}(g)/6 \,.$$

[4] Note that Feller says at the end of his Section (b) that Poisson's formula could be used. And when he turns to Benford in Section (c) he quotes the name (not any paper) of Pinkham.

[5] Poincaré [Po] speaks of the roulette pages 11–13, pages 148–150, and of digits pages 313–320. Two figures in [Few] look a bit like the figure in [Po] page 149.

## 5 The proof of Gauvrit-Delahaye

Despite the existence of [Pi, 1961] and [DL, 2008] I reproduce the proof by Gauvrit and Delahaye because it is elementary and pleasant. We assume the density $g$ of $Y = \log X$ is unimodal and we denote by $M$ its maximum over $\mathbb{R}$ (maximum to be small).

*Proof.* The density $g$ is non-decreasing on $]-\infty, b]$ and non-increasing on $[b, +\infty[$. Let $M = g(b)$. Without loss of generality we can translate $g$ by an integer $n \in \mathbb{Z}$, so we may suppose[6] $b \in [0, 1]$. Let $z \in ]0, 1]$. We will prove the two following inequalities:

$$\overline{G}(z) \leq z + 2\,M \quad \text{and} \quad \overline{G}(z) \geq z - 2\,M.$$

The idea (not far from the idea in Feller's book) is that on left of $b$ the mean of $g$ on $[n, n+z]$ is less than the mean[7] of $g$ on $[n, n+1]$ and that on right of $b$ the mean of $g$ on $[n, n+z]$ is less than that of $g$ on $[n+z-1, n+z]$.

Precisely: for any $n \leq -1$, since $g$ is non-decreasing on $]-\infty, 0]$, one has

$$\frac{1}{z}\int_n^{n+z} g(y)\,dy \leq \int_n^{n+1} g(y)\,dy$$

hence

$$\frac{1}{z}\sum_{n\in\mathbb{Z},\, n\leq -1}\int_n^{n+z} g(y)\,dy \leq \int_{-\infty}^{0} g(y)\,dy\,. \tag{6}$$

Similarly for any $n \geq 2$ thanks to the non-increasingness of $g$ on $[1+z, +\infty[$,

$$\frac{1}{z}\int_n^{n+z} g(y)\,dy \leq \int_{n+z-1}^{n+z} g(y)\,dy$$

hence

$$\frac{1}{z}\sum_{n\in\mathbb{Z},\, n\geq 2}\int_n^{n+z} g(y)\,dy \leq \int_{1+z}^{+\infty} g(y)\,dy\,. \tag{7}$$

Summing (6) and (7) gives

$$\frac{1}{z}\sum_{n\in\mathbb{Z},\, n\neq 0,\, n\neq 1}\int_n^{n+z} g(y)\,dy \leq \int_{-\infty}^{+\infty} g(y)\,dy = 1\,.$$

---

[6] This is the argument in [GD1]. Surely $b = 0$ is possible (here we have forgotten $D$ and the factor 10 relative to $X$) and maybe (5) could be improved of a factor 2.

[7] To prove $\frac{1}{z}\int_0^z g(y)\,dy \leq \int_0^1 g(y)\,dy$ when $g$ is non-decreasing on $[0, 1]$, express $\int_0^z g(y)\,dy$ as an integral over $[0, 1]$ by a linear change of variable.

On the left hand-side are lacking terms corresponding to $n = 0$ and $n = 1$. Each of them is bounded by

$$\frac{1}{z} \int_n^{n+z} g(y)\, dy \leq M$$

hence

$$\frac{1}{z}\, \overline{G}(z) \leq 1 + 2\, M$$

and

$$\overline{G}(z) \leq z + 2\, M\, .$$

Now we turn to

$$\overline{G}(z) \geq z - 2\, M\, .$$

On left of $b$ the mean of $g$ on $[n, n+z]$ is greater than the mean of $g$ on $[n+z-1, n+z]$. And on right of $b$ the mean of $g$ on $[n, n+z]$ is greater than the mean of $g$ on $[n, n+1]$. Thus for $n \leq -1$,

$$\frac{1}{z} \int_n^{n+z} g(y)\, dy \geq \int_{n+z-1}^{n+z} g(y)\, dy$$

and for $n \geq 1$,

$$\frac{1}{z} \int_n^{n+z} g(y)\, dy \geq \int_n^{n+1} g(y)\, dy$$

and summing

$$\begin{aligned} \frac{1}{z} \sum_{n \in \mathbb{Z},\, n \neq 0} \int_n^{n+z} g(y)\, dy &\geq \int_{-\infty}^{-1+z} g(y)\, dy + \int_1^{+\infty} g(y)\, dy \\ &= \int_{\mathbb{R}} g(y)\, dy - \int_{-1+z}^{1} g(y)\, dy\, . \end{aligned}$$

As the interval $[-1+z, 1]$ has length $\leq 2$, the last term has absolute value $\leq 2\, M$. $\square$

# 6 Some bounds expressed with total variation

Finite total variation encompasses unimodality. Precisely if $g$ is unimodal, its total variation is $2 \max g$. We will expose essentially some results by

Dümbgen-Leuenberger in 2008 [DL, Th.1 and Cor.2]. Thus several local minima and maxima are manageable[8].

Recall that $g$ is the density of $Y$ on $\mathbb{R}$. By the "stacking" operation, the density of $\mathcal{M}(Y)$ on $[0,1]$ is $\bar{g}(z)$ defined in (4). A classical notion is *total variation*. We assume that $g$ has a finite total variation which we define by[9]

$$\mathrm{TV}(g) := \sup\Big\{\sum_{i=1}^{m} |g(y_i) - g(y_{i-1})|\,;\, m \geq 1,\ -\infty < y_0 \leq \cdots \leq y_m < +\infty\Big\}\,.$$

If $g$ is unimodal, $g(y) \to 0$ when $|y| \to +\infty$, and $\mathrm{TV}(g) = 2\,\max_{\mathbb{R}} g$.

As for the total variation of $\bar{g}$ which is a function on the torus $\mathbb{T}$ identified to the half open interval $[0,1[$, one should consider

$$\sup\Big\{\sum_{i=1}^{m} |\bar{g}(z_i) - \bar{g}(z_{i-1})| + |\bar{g}(z_m) - \bar{g}(z_0)|\,;\, m \geq 1,\ 0 \leq z_0 < \cdots < z_m < 1\Big\}\,.$$

But considering $\bar{g}$ as defined on $[0,1]$ with[10] $\bar{g}(1) = \bar{g}(0)$ one can write

$$\mathrm{TV}(\bar{g}) = \sup\Big\{\sum_{i=1}^{m} |\bar{g}(z_i) - \bar{g}(z_{i-1})|\,;\, m \geq 1,\ 0 \leq z_0 \leq \cdots \leq z_m \leq 1\Big\}\,.$$

Now we observe that

$$\bar{g}(z) = \lim g_N(z) \quad \text{where} \quad g_N(z) = \sum_{n=-N}^{N} g(n+z)\,.$$

---

[8] In [GD1, GD2] the authors say that a finite number of bumps is possible. The proof could be tedious. Example 8.2.1 below shows that an infinite sequence of bumps may be bad.

[9] Usually this formula is written with strict inequalities. It would give the same result (repetition of a value is useless). For a fine study of finite total variation functions in one variable, but for vector valued functions, see [M]. The total variation could be overestimated if one used "erratic values" of $g$. A non-erratic value at $y$ is a value between the two lateral limits which do exist, see for example [M, Prop.4.2 p. 11]. Note that variation is better adapted to cumulative functions than to densities!

[10] Here we can see the importance of a sharp definition of $\bar{g}$: for example if $g(y) = 2\,y$ on $[0,1]$ and 0 elsewhere, $\bar{g}(z) = 2\,z$ on $]0,1[$, but the downfall of 2 has to be added to the progressive increase of 2 to get the true value $\mathrm{TV}(\bar{g}) = 4$.

Then for any sequence $0 \le z_0 \le \cdots \le z_m \le 1$,

$$\begin{aligned}\sum_{i=1}^m |g_N(z_i) - g_N(z_{i-1})| &\le \sum_{i=1}^m \sum_{n=-N}^N |g(n+z_i) - g(n+z_{i-1})| \\ &= \sum_{n=-N}^N \sum_{i=1}^m |g(n+z_i) - g(n+z_{i-1})| \\ &\le \mathrm{TV}(g)\end{aligned}$$

hence the inequality (cf. the first assertion of [DL, Theorem 1] and [DL, formula (5) page 107]),

$$\mathrm{TV}(\bar g) \le \mathrm{TV}(g)\,.$$

Since

$$\sup \bar g \ge \int_0^1 \bar g(z)\,dz = 1 \ge \inf \bar g$$

one has

$$\mathrm{R}(g) := \sup_{z_1 \le z_2} |\bar g(z_2) - \bar g(z_1)| \ge \sup_z |\bar g(z) - 1|$$

Note that $|\bar g(z_2) - \bar g(z_1)| = \max([\bar g(z_2) - \bar g(z_1)]^+, [\bar g(z_2) - \bar g(z_1)]^-)$. Since $g$ is integrable on $\mathbb{R}$ it tends to 0 at infinity, and with the notation

$$\mathrm{TV}(g)^+ := \sup\Big\{\sum_{i=1}^m (g(y_i) - g(y_{i-1}))^+\,;\, m \ge 1,\ -\infty < y_0 \le \cdots \le y_m < +\infty\Big\}\,.$$

and the anologous with negative parts, one has $\mathrm{TV}^+(g) = \mathrm{TV}^-(g) = \mathrm{TV}(g)/2$. Hence (cf. [DL, Th.1])

$$\sup_{z\in[0,1]} |\bar g(z) - 1| \le \mathrm{R}(\bar g) \le \mathrm{TV}(g)/2\,.$$

Now we are going to prove [DL, Cor.2 p. 102]

$$\sup_{0\le z_1<z_2\le 1} |(\overline{G}(z_2) - \overline{G}(z_1) - (z_2 - z_1)| \le (z_2 - z_1)[1 - (z_2 - z_1)]\ \mathrm{TV}(g)/2\,.$$

Let $\delta := z_2 - z_1$. Then (we reproduce [DL, proof of Cor.2 p. 108])

$$\begin{aligned}
|(\overline{G}(z_2) - \overline{G}(z_1) - (z_2 - z_1)| &= \Big| \int_{z_1}^{z_2} \bar{g}(z)\, dz - \delta \int_{z_2-1}^{z_2} \bar{g}(z)\, dz \Big| \\
&= \Big| (1-\delta) \int_{z_1}^{z_2} \bar{g}(z)\, dz - \delta \int_{z_2-1}^{z_1} \bar{g}(z)\, dz \Big| \\
&= \Big| \delta(1-\delta) \int_0^1 \big[ \bar{g}(z_1 + \delta t) - \bar{g}(z_1 - (1-\delta)t) \big]\, dt \Big| \\
&\leq \delta(1-\delta) \int_0^1 \big| \bar{g}(z_1 + \delta t) - \bar{g}(z_1 - (1-\delta)t) \big|\, dt \\
&\leq \delta(1-\delta)\, \mathrm{R}(\bar{g})/2 \\
&\leq \delta(1-\delta)\, \mathrm{TV}(g)/2
\end{aligned}$$

which implies

$$\sup_{0 \leq z_1 < z_2 \leq 1} |(\overline{G}(z_2) - \overline{G}(z_1) - (z_2 - z_1)| \leq \mathrm{TV}(g)/8\,. \tag{8}$$

In 1961 Pinkham [Pi, bottom of page 1228] using Fourier Analysis arguments obtained

$$\sup_{0 \leq z \leq 1} |\overline{G}(z) - z| \leq \mathrm{TV}(g)/6\,.$$

All these results give better bounds than those of the foregoing Section. Indeed, if $g$ is unimodal, (5) gives

$$|(\overline{G}(z_2) - \overline{G}(z_1) - (z_2 - z_1)| \leq 4 \max_{\mathbb{R}} g = 2\ \mathrm{TV}(g)\,.$$

In their paper [DL] Dümbgen-Leuenberger give other fine bounds when $g$ admits derivatives.

# 7 Return to $X$, the disaster

Now, what becomes an hypothesis concerning $g$ when expressed in term of $X$ or its density $f$? Recall that (this is change of variables in Integration)

$$g(y) = \ln(10)\, 10^y\, f(10^y) \quad \text{and} \quad f(x) = \frac{g(\log x)}{x\, \ln(10)}\,. \tag{9}$$

One cannot switch between the density of $X$ and the density of $Y$ only by changing[11] $x$ in $10^y$ or $y$ in $\log x$. Why? Because this would be possible if

[11] Despite the fact that $\mathbb{R} \ni y \mapsto 10^y \in \mathbb{R}_+^*$ is an isomorphism between ordered sets.

one had taken for density of $X$ the density of its law $\mathbf{P}_X$ with respect to the following Haar measure[12] on $\mathbb{R}_+^*$: the image (also called push-forward) of Lebesgue on $\mathbb{R}$ by $y \mapsto 10^y$. With respect to the Lebesgue measure this Haar measure has the density $x \mapsto [\ln(10)\, x]^{-1}$.

Obviously from (9), $\max g = \ln(10) \sup_{x \in \mathbb{R}_+^*} [x\, f(x)]$ and unimodality of $g$ is equivalent to unimodality of $x \mapsto x\, f(x)$.

Here the "disaster" occurs: even if $f$ is unimodal, $g$ may be not, see Section 8.1; and even if $\max f$ tends to 0 when a parameter converges to some value, the maximum of $x \mapsto x\, f(x)$ may not tend to 0. Despite the fact that log-normal laws (see 8.3.1) and Pareto laws (see 8.3.2) do the work, the uniform law on $[a, b]$ and the exponential law (see 8.2.2 and 8.2.3) exemplify the difficulty.

As allusively invoked above, classical usual laws described in textbooks are families depending on one or several parameters. The list is impressive, but the fact that the two most simple ones fail in exemplifying the Benford phenomenon calls for questioning. Surely the so many random variables which seem obey to Benford do not follow a classical "usual law" and the sentence "if the spread of the r.v.[13] is very large" (as in [F, p. 63 just after (8.6)]) is an unwise shortcut. See [Br, BH1] for more comments.

As already noticed by many authors, mixing of several data ([Hi1, Hi4, JR] and products [Bo] can give good laws.

I mention that Gauvrit and Delahaye [GD3, Th.2] say their more general result with a strictly increasing function in place of log applies to more situations.

# 8 Examples

## 8.1 Anoying examples

One could expect that the hypothesis "the density $g$ of $Y = \log X$ is unimodal with a small maximum" is usually encountered. Expressed with $X$, it means that $x \mapsto x\, f(x)$ is unimodal with a small maximum. This does not apply to the uniform law and to the exponential law: see below Section 8.2.

Let us give small examples showing the action of multiplication by $x$.

[12] I am indebted to J. Saint-Pierre [SP] for this idea of Haar measure. Note that as early as 1970 Hamming [Ha] used the measure with density $1/(\ln(10)\, x)$ on the interval $[10^{-1}, 1]$. See also Section 9.

[13] In Feller the r.v. is denoted $Y$ but it is the positive variable whose first digit is considered.

1) Let

$$f_0(x) = \begin{cases} x & \text{if } x \in ]0,1], \\ x^{-1}\,[1+(x-1)(x-2)/2] & \text{if } x \in [1,2], \\ 4\,x^{-3} & \text{if } x \in [2,+\infty[. \end{cases}$$

This is a positive integrable function, so it is, up to a multiplicative coefficient, a density. It is decreasing on $[1,2]$ because on this interval

$$f_0'(x) = \frac{1}{2} - \frac{2}{x^2} \leq 0$$

so $f_0$ is unimodal. But $x \mapsto x\,f_0(x)$ is no longer unimodal. It has two maxima, at $x=1$ and at $x=2$. $\square$

2) Let $f$ defined on $]0,+\infty[$ by $f(x)=0$ on $]0,1/2] \cup \bigcup_{n\geq 1}\{n-1/2\}$, $f(n)=1/n^2$ for all $n \geq 1$ and $f$ affine on all intervals $[n-1/2,n]$ and all intervals $[n,n+1/2]$. The graph of $f$ consists of a serie of bumps in form of isosceles triangles. The total variation is finite with value $2\sum_{n=1}^{\infty} 1/n^2$. As for $h(x) := x\,f(x)$ this function equals 0 at each $n-1/2$ and equals $1/n$ at each $n$. Since $\sum_{n=1}^{\infty} 1/n = +\infty$ the total variation of $h$ is infinite. $\square$

## 8.2 Negative examples

### 8.2.1 A kind of periodicity

If the law of $X$ is carried by the set

$$\bigcup_{n\in\mathbb{Z}} [0.9\,10^n, 10^n[$$

then $D \overset{\text{a.s.}}{=} 9$ (example inspired by [GD2, p.3]). And this in spite of, as soon as many intervals have $> 0$ probabilities, a large "scattering". This can be realized with a $C^\infty$ density taking strictly positive values on each open interval $]0.9\,10^n, 10^n[$.

### 8.2.2 Uniform law

The density is $f(x) = \frac{1}{b-a}\,\mathbf{1}_{[a,b]}(x)$ (with the parameters $a$ and $b$ satisfying $0 \leq a < b$). Surely $x \mapsto x\,f(x)$ is unimodal on $\mathbb{R}_+^*$ and the spread of $f$ is large when $b \to +\infty$. The maximum of $x\,f(x)$ is 1 if $a=0$. And, if $a>0$ the maximum is $\frac{b}{b-a}$ which decreases when $b$ increases to $+\infty$, but the limit is 1. So an inequality as (5) or (8) does not apply. For a finer study see [Br].

### 8.2.3 Exponential law

The density is $f(x) = \lambda\, e^{-\lambda x}$ ($\lambda \in\, ]0, +\infty[$ is the parameter). One could naively expect a good Bendford approximation when $\lambda \to 0$. Derivating $h(x) := x\, f(x)$ one proves easily that the function $x\, f(x)$ is unimodal; and its maximum attained at $x = 1/\lambda$ has the value $1/e$ (particular case of (10) below). This maximum does not tends to 0 as $\lambda \to 0$. So an inequality as (5) does not apply, moreover $4\, \ln(10)\, e^{-1} = 3.388...$ is a very huge value. Here (8) gives

$$\leq 2\, \ln(10)[\max_{x>0} x\, f(x)]/8 = \ln(10)\, e^{-1}/4 = 0.211...$$

Engel and Leuenberger [EL] study the exact formula coming from (3)

$$\mathbf{P}(D = k) = \sum_{n=-\infty}^{+\infty} e^{-\lambda\, k 10^n} \left(1 - e^{-\lambda 10^n}\right).$$

They prove that Bendorf is almost satisfied with a periodical dependance on $\log \lambda$ and small gaps. But the error does not tend to 0 as $\lambda \to 0$.

## 8.3 Positive examples

### 8.3.1 Log-normal laws

When $X = \exp(Y_0)$ with $Y_0$ of law $\mathcal{N}(\mu, \sigma^2)$, one has $Y = \log X = Y_0/\ln(10)$. Recall that the density of $Y_0$ is

$$y \mapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

So the required properties of $g$ are clearly verified. An inequality as (5) applies: when $\sigma$ tends to infinity the Benford approximation is good.

### 8.3.2 Pareto laws

The Pareto law of type 1 (cf. [GD2, p.7]) depends on two parameters $\alpha$ and $x_0$ both in $\mathbb{R}_+^*$ and has the density

$$f(x) = \frac{\alpha\, x_0^\alpha}{x^{\alpha+1}}\, \mathbf{1}_{[x_0, \infty[}(x).$$

There $x \mapsto x\, f(x) = \dfrac{\alpha\, x_0^\alpha}{x^\alpha}\, \mathbf{1}_{[x_0,\infty[}(x)$ is non-decreasing on $]-\infty, x_0]$ (identically null on $]-\infty, x_0[$) and non-increasing on $[x_0, +\infty[$. The maximum reached at $a = x_0$ equals $m = a\,f(a) = \alpha$. Thanks to an inequality as (5), when $\alpha$ tends to 0 the probabilities $\mathbf{P}(D = k)$ converge to the values of Benford (1). Note that $X$ has no mean as soon as $\alpha \leq 1$ which indicates a large "scattering".

Pareto laws of type 2 are treated in [GD2].

### 8.3.3 Gamma laws

Given the parameters $p$ and $\lambda$ each in $]0, +\infty[$, the density is

$$f(x) = \frac{\lambda}{\Gamma(p)}\, e^{-\lambda x}\, (\lambda x)^{p-1}\,.$$

When $p = 1$ one recovers the exponential law which provides a rather negative example (see 8.2.3 above). Let $h(x) := x f(x) = \dfrac{\lambda^p}{\Gamma(p)}\, e^{-\lambda x}\, x^p$. As

$$h'(x) = \frac{\lambda^p}{\Gamma(p)}\, e^{-\lambda x}\, x^{p-1}[-\lambda x + p]$$

$h$ is increasing on $]0, p/\lambda]$ and decreasing on $[p/\lambda, +\infty[$. Its maximum is

$$h(p/\lambda) = \frac{1}{\Gamma(p)}\, e^{-1}\, p^p \tag{10}$$

This tends to 0 when $p \to 0$. Is this useful?

### 8.3.4 Fréchet and Weibull laws

These laws are used in the extreme value theory (Fisher-Tippett results).

The Fréchet law on $]0, +\infty[$ has for cumulative function ($\alpha$ is a parameter belonging to $]0, +\infty[$)

$$\mathbf{P}(X < x) = \exp(-x^{-\alpha}) \quad (\text{with } x > 0)\,.$$

Here $f(x) = \alpha\, x^{-\alpha-1} \exp(-x^{-\alpha})$ and with $h(x) := x\, f(x)$,

$$h'(x) = \alpha^2\, x^{-\alpha-1} \exp(-x^{-\alpha})\, [-1 + x^{-\alpha}]\,.$$

So unimodality holds with the maximum at $x = e^{1/\alpha}$

$$h(e^{1/\alpha}) = \alpha\, e^{-1} \exp(-e^{-1}) = (0.2546...)\, \alpha\,.$$

Hence the maximum tends to 0 as $\alpha \to 0$.

The Weibull law is, when considered on $]0,+\infty[$ and not $]-\infty,0[$, the law of cumulative function ($\beta$ is a parameter belonging to $]0,+\infty[$)

$$1-\exp(-x^\beta) \quad (\text{with } x>0)\,.$$

Here $f(x)=\beta\, x^\beta \exp(-x^\beta)$ and with $h(x):=x\,f(x)$,

$$h'(x)=\beta^2\, x^{\beta-1} \exp(-x^\beta)\,[1-x^\beta]\,.$$

So unimodality holds with the maximum at $x=1$

$$h(1)=\beta\, e^{-1}=(0.367879...)\,\beta\,.$$

Hence the maximum tends to 0 as $\beta\to 0$.

# 9 Two exact results

There exist in the litterature a lot of exact results, some relying on "scale invariance" see [Hi1, Hi2, Hi3], other relying on mixing of laws, see [JR]. I will give personal results (except the second part of Theorem 2) written when I was completely naive with the Benford phenomenon and being unaware of [Ha] and [BH2].

The next Theorem has no application (or I would be surprised!). It shows that the rough hypothesis "$\mathcal{M}(\log X)$ follows the uniform law on $[0,1]$" admits sufficient conditions. The second hypothesis comes from the caption of the figure in [GD1, GD2].

**Theorem 1** *Let $X$ be a random variable ($X>0$) and $Y:=\log(X)$. Suppose that $Y$ has the density $g$. Suppose either that $g$ is countably a step function, constant (equality Lebesgue a.e.) on each interval $[n,n+1]$ ($n\in\mathbb{Z}$), either that $g$ is continuous on $\mathbb{R}$ and affine on each interval $[n,n+1]$ ($n\in\mathbb{Z}$). Then $D$ follows the Benford law* (1).

*Proof.* 1) Let $\gamma_n$ be the value of $g$ on $[n,n+1]$. The serie (4) gives

$$\sum_{n\in\mathbb{Z}} g(n+y)=\sum_{n\in\mathbb{Z}}\gamma_n=1$$

which proves that $\mathcal{M}(\log X)$ follows the uniform law on $[0,1]$.

2) The integral of $g$ on $[n,n+1]$ is the aera of a trapezoid and it amounts to $\frac{1}{2}\big(g(n)+g(n+1)\big)$. As the sum is 1, it holds $\sum_{n=-\infty}^{+\infty} g(n)=1$. Above $[0,1]$ the function

$$y\mapsto \sum_{n=-N}^{N} g(n+y)$$

is affine and equals $\sum_{n=-N}^{N} g(n)$ at 0 and equals $\sum_{n=-N}^{N} g(n+1)$ at 1. All this converge to 1. For the incredulous reader if any: the affinity entails

$$\begin{aligned}
\sum_{n=-N}^{N} g(n+y) &= y \sum_{n=-N}^{N} g(n+1) + (1-y) \sum_{n=-N}^{N} g(n) \\
&= \sum_{n=-N}^{N} g(n) + y\,\big(g(N+1) - g(-N)\big) \\
&\to 1
\end{aligned}$$

when $N \to \infty$. $\square$

REMARKS. The hypotheses do not assert unimodality of $g$. Without the continuity of $g$ the second part does not hold: take $g(y) = 2\,y$ on $[0,1]$ and 0 elsewhere.

The next exact result is no longer realistic. It as already been obtained by Hamming [Ha, Section IV p. 1615] (quoted in [R, p. 535]). See also [BH2, Part 1 of Theorem 6.3]. One could imagine collecting data (richness, level of a river, etc.) in several places and several countries where the units are not the same. All this would be listed together.

The idea leading to a mathematical result is: multiply a given r.v. $X_0$ which models our physical quantity (at least in one precise unit) by a random coefficient belonging to $[1, 10]$, which gives $X$ (and as for the law of $X$ a mixing of the laws of the homothetic r.v. of $X_0$). Changing the unit of several times a factor 10 or 1/10 would not change the first digit in base 10. We assume that the coefficient obeys the Haar measure[14] of the multiplicative group $(\mathbb{R}_+^*, *)$ restricted to $[1, 10]$ (more precisely the image of Lebesgue measure by $u \mapsto 10^u$).

**Theorem 2** *Let $X_0$ be a random variable ($X > 0$) defined on* $(\Omega, \mathcal{F}, \mathbf{P}_0)$ *which has a density. The Lebesgue measure on* $[0,1]$ *is denoted by* $\mathbf{\Lambda}$. *Let $X$ be the r.v. on $\Omega \times [0,1]$ equipped with the probability measure* $\mathbf{P} := \mathbf{P}_0 \otimes \mathbf{\Lambda}$ *defined as*

$$X(\omega, u) = 10^u\, X_0(\omega)$$

*Then $D$ obeys to the Benford law* (1).

[14] See a foregoing footnote in Section 2.

*Proof.* Let $Y_0 := \log(X_0)$. For $k \in \{1, ..., 9\}$ one has

$$\begin{aligned} D(\omega, u) = k &\iff X(\omega, u) \in \bigcup_{n \in \mathbb{Z}} [k\, 10^n, (k+1)10^n[ \\ &\iff u + Y_0(\omega) \in \bigcup_{n \in \mathbb{Z}} [n + \log k, n + \log(k+1)[\ . \end{aligned}$$

The above unions are disjoint, hence we have to sum the terms

$$\mathbf{P}\Big(\big\{(\omega, u)\,;\ u + Y_0(\omega) \in [n + \log k, n + \log(k+1)[\big\}\Big)\,. \tag{11}$$

Let $g$ denotes the density of $Y_0$. The term (11) can be expressed by successive integration (firstly with respect to $u$ and then to $y$)

$$\begin{aligned} &\int_{-\infty}^{+\infty} \mathbf{\Lambda}([n + \log k - y, n + \log(k+1) - y] \cap [0,1])\, g(y)\, dy \\ =&\int_{-\infty}^{+\infty} \mathbf{\Lambda}([\log k - y, \log(k+1) - y] \cap [-n, -n+1])\, g(y)\, dy\,. \end{aligned}$$

But

$$\begin{aligned} \sum_{n \in \mathbb{Z}} \mathbf{\Lambda}([\log k - y, \log(k+1) - y] \cap [-n, -n+1]) \\ =&\mathbf{\Lambda}([\log k - y, \log(k+1) - y]) \\ =&\log(k+1) - \log k\,. \end{aligned}$$

As $\displaystyle\int_{-\infty}^{+\infty} g(y)\, dy = 1$ this proves

$$\sum_{n=-\infty}^{\infty} \mathbf{P}\Big(\big\{(\omega, u)\,;\ Y_0(\omega) \in [n - u + \log k, n - u + \log(k+1)[\big\}\Big) = \log\Big(\frac{k+1}{k}\Big) \quad \square$$

Comment. The hypothesis that the unit could be random and obey to a Haar measure is debatable. As said by Pinkham [Pi], there is a ratio 10 between the decimeter and the meter but as for volumes one gets the ratio of 10 between 100 dm$^3$ and one m$^3$ (and not between one dm$^3$ and one m$^3$). And usual units are certainly numerous but in a finite number: cf. meters and feet (argument of [SF] quoted above in Section 1).

## References

[A good part of the references are loadable on the Net. I give some links.]

[Bee] Beebe N.H.F., *A bibliography of publications about Benford's law, Heap's law, and Zipf's law.* On the Net.

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

[Br] Berger A., *Large spread does not imply Benford's Law*, Draft version (2010) 5 pages.

[BH1] Berger A., Hill, T.P., *Benford's law strikes back: no simple explanation in sight for mathematical gem*, The Mathematical Intelligencer **33** (2011) 85–91.

[BH2] Berger A., Hill, T.P., *A basic theory of Benford's Law*, Probability Surveys **8** (2011) 1–126.

[BH3] Berger A., Hill, T.P., *Benford Online Bibliography.* On:

http://www.benfordonline.net/

[Bo] Boyle J., *An application of Fourier series to the most significant digit problem*, Amer. Math. Monthly **101** (1994) 879–886.

[D] Delahaye, J.-P., *L'étonnante loi de Benford*, Pour la Science **351** (2007), 90–95.

[DL] Dümbgen, L., Leuenberger, C., *Explicit bounds for the approximation error in Benford's law*, Electron. Commun. Probab. **13** (2008), 99–112. On the Net, tenth paper on

http://ecp.ejpecp.org/issue/view/33

[EL] Engel, H.-A., Leuenberger, C., *Benford's law for exponential random variables*, Statist. Probab. Lett. **63** (2003), 361–365.

[F] Feller, W., *An Introduction to Probability Theory and Its Applications vol.2*, J. Wiley, New York, 1966 (second edition 1971).

[Few] Fewster, R.M., *A simple explanation of Benford's law*, American Statistician **63** (2009) 20–25.

[GD1] Gauvrit, N., Delahaye, J.-P., *Pourquoi la loi de Benford n'est pas mystérieuse*, Math. Sci. Hum. Math. Soc. Sci. **182** (2008), 7–15.

[GD2] Gauvrit, N., Delahaye, J.-P., *Scatter and regularity imply Benford's Law... and more*, arXiv0910.1359. On:

http://arxiv.org/PS_cache/arxiv/pdf/0910/0910.1359v1.pdf

[GD3] Gauvrit, N., Delahaye, J.-P., *Loi de Benford générale*, Math. Sci. Hum. Math. Soc. Sci. **186** (2009), 5–15.

[Go] Good, I.J., *Some statistical applications of Poisson's work*, Statistical Science **1** (1986), 157–170.

[Ha] Hamming, R.W., *On the distribution of numbers*, Bell System Tech. J. **49** (1970) 1609–1625.

[Hi1] Hill, T.P., *A statistical derivation of the significant-digit law*, Statist. Sci. **10** (1995) 354–363.

[Hi2] Hill, T.P., *The significant-digit phenomenon*, Amer. Math. Monthly **102** (1995) 322–327.

[Hi3] Hill, T.P., *Base-invariance implies Benford's law*, Proc. Amer. Math. Soc. **123** (1995) 887–895.

[Hi4] Hill, T.P., *Le premier chiffre significatif fait sa loi*, La Recherche **316** (1999) 72–75.

[Hu] Hürlimann, W., *Benford's law from 1881 to 2006: a bibliography.*

[JR] Janvresse, É., de la Rue, T., *From uniform distributions to Benford's law*, J. Appl. Probab. **41** (2004), 1203–1210.

[M] Moreau, J.-J., *Bounded variation in time*, in Topics in nonsmooth mechanics, 1–74, Birkhäuser, Basel, 1988.

[N] Newcomb, S., *Note on the Frequency of Use of the Different Digits in Natural Numbers*, Amer. J. Math. **4** (1881) 39–40.

[Pi] Pinkham, R.S., *On the distribution of first significant digits*, Annals Math. Stat. **32** (1961) 1223–1230.

[Po] Poincaré, H., *Calcul des probabilités*, Gauthier-Villars, Paris, 1912.

[R] Raimi, R.A., *The first digit problem*, Amer. Math. Monthly **83** (1976) 521–538.

[SP] Saint-Pierre, J., Personal mail in date of November 9, 2011.

[SF] Scott, P.D., Fasli, M., *Benford's Law: an empirical investigation and a novel explanation*, CSM Technical Report 349, Department of Computer Science, University of Essex, 2001.

[T] Tao, T., Blog, July 3, 2009

http://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/