Chiffres, fraudes et hasard

Paul Jolissaint

24 février 2016

Introduction

Un exemple typique de fraude (d'après M. Nigrini)

En 1993, *Wayne J. Nelson*, un employé du Trésor de l'état d'Arizona, est reconnu coupable d'avoir détourné près de 2 millions de dollars en versant à des personnes fictives 23 chèques dont voici la liste des montants :

Date d'émission	Montants en dollars
9 octobre 1992	1'927.48
	27'902.31
14 octobre 1992	86'241.90
	72'117.46
	81'321.75
	97'473.96
19 octobre 1992	93'249.11
	89'658.17
	87'776.89
	92'105.83
	79'949.16
	87'602.93

19 octobre 1992	96’879.27
	91’806.47
	84’991.67
	90’831.83
	93’766.67
	88’338.72
	94’639.49
	83’709.28
	96’412.21
	88’432.86
	71’552.16
Total	**1’878’687.58**

Indices de fraude :

- Le fraudeur a commencé par de petites valeurs, puis les montants et leur fréquence ont augmenté.
- Tous les montants étaient inférieurs à $100'000 : des montants supérieurs auraient sans doute fait l'objet de vérifications par un supérieur hiérarchique.
- Les chiffres significatifs sont trop grands : plus de 90 % admettent 7,8 ou 9 comme premier chiffre. De plus, les paires de premiers chiffres 87, 88, 93 et 96 ont été utilisées deux fois dans les 23 montants.

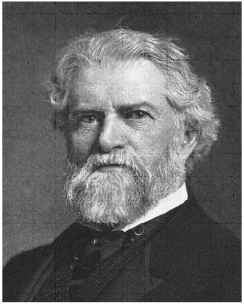
Constation empirique. Considérons un ensemble de valeurs de l'un des types suivants, relevées au hasard :

- valeurs boursières
- listes de prix d'articles divers
- valeurs numériques variées tirées de publications périodiques
- grandeurs géographiques (populations de villes, superficies des lacs d'un continent choisi, débit de rivières, etc.)
- Dans une certaine mesure : les numéros des maisons dans les rues (aux USA, par exemple).
- ...

Le *premier chiffre significatif* (p. ex. **2** est le chiffre significatif de **2**345.6, **7** est celui de 0.0**7**8) de ces valeurs est *plus souvent* 1 que 2, qui est *plus fréquent* que 3, etc., le chiffre le *moins fréquent* étant 9.

Cette constatation reste vraie même si l'on change les unités !

Le phénomène a été observé pour la première fois par l'astronome Simon Newcomb en 1881.



Simon Newcomb (1835-1909)

Il a constaté que les tables de logarithmes (utilisées alors pour effectuer les multiplications, divisions, extractions de racines) étaient *plus usées* dans les premières pages que dans les suivantes.

Il en a déduit que les utilisateurs manient plus souvent des nombres commençant par 1 que par 2, par 2 que par 3 etc.

Il a écrit un article sur le sujet dans *American Journal of Mathematics*, qui est passé inaperçu.

La loi a été redécouverte par le physicien Frank Benford en 1938 :

Frank Benford (1883-1948)

F. Benford a rassemblé 20’229 données numériques provenant de sources diverses et a appelée cette règle **loi des nombres anormaux**.

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit 1	2	3	4	5	6	7	8	9	Count
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average.......		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

L'article a été publié dans les *Proceedings of the American Philosophical Society*.

Il était situé à côté d'un article beaucoup consulté :

The Multiple Scattering of Electrons, de H. Bethe *et al.*.

Définitions

Rappel. Le ***logarithme*** (en base 10) d'un nombre $x > 0$ est la puissance à laquelle il faut élever 10 pour obtenir x :

$$10^{\log(x)} = x.$$

Exemples.

(1) $\log(10) = 1$ car $10^1 = 10$

(2) $\log(100) = 2$ car $10^2 = 100$

(3) $\log(0.01) = -2$ car $0.01 = 1/100 = 10^{-2}$

(4) $\log(\sqrt{10}) = 1/2$ car $\sqrt{10} = 10^{1/2}$

(5) $\log(\pi) = 0.49714\ldots$ (table ou calculette).

Définition 1. (Élémentaire) Un ensemble de valeurs numériques suit la *loi de Benford* si, pour chaque chiffre $d \in \{1, \ldots, 9\}$ la proportion de valeurs qui commencent par d est

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

Explicitement :

Chiffre	Fréquence théorique		
1	$\log(2/1)$	=	0.301
2	$\log(3/2)$	=	0.176
3	$\log(4/3)$	=	0.125
4	$\log(5/4)$	=	0.097
5	$\log(6/5)$	=	0.079
6	$\log(7/6)$	=	0.067
7	$\log(8/7)$	=	0.058
8	$\log(9/8)$	=	0.051
9	$\log(10/9)$	=	0.046

Avant de donner la définition plus précise, introduisons deux notations :

- Soit $x > 0$; son ***significande*** est l'unique nombre $S(x) \in [1, 10)$ tel que

$$x = S(x) \cdot 10^k$$

pour un certain entier k. (Notation scientifique)

- Si $x > 0$, on note $D_1(x)$ le premier chiffre significatif de x ; c'est la partie entière de $S(x)$ et $D_1(x) \in \{1, \ldots, 9\}$. Plus généralement, pour $n \geq 2$, $D_n(x) \in \{0, 1, \ldots, 9\}$ désigne le n^{e} chiffre significatif de x.

Interprétons le fait que, pour un certain $x > 0$, on ait $D_1(x) = d$:

Cela signifie que $d \cdot 10^k \leq x < (d+1) \cdot 10^k$ pour un certain entier k.

Autrement dit,

$$d \leq S(x) < d + 1.$$

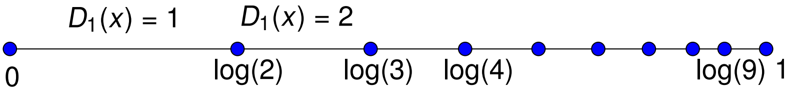
Ou encore :

$$\log(S(x)) \in [\log(d), \log(d+1)).$$

Or, $\log\left(\frac{d+1}{d}\right) = \log(d+1) - \log(d)$ est la **longueur** de l'intervalle

$$[\log(d), \log(d+1)).$$

Ainsi, dire qu'un ensemble de valeurs positives $\{x_1, \ldots, x_n\}$ satisfait la loi de Benford revient à dire que l'ensemble des valeurs $\{\log(S(x_1)), \ldots, \log(S(x_n))\}$ est **uniformément réparti dans l'intervalle** $[0, 1]$.



Cela mène à la définition générale suivante pour les suites $(x_n)_{n\geq 1} = (x_1, x_2, \ldots,)$:

Définition 2. Une suite (x_n) satisfait la ***loi de Benford*** si, pour tout $t \in [0, 1]$,

$$\lim_{N\to\infty} \frac{\#\{1 \leq n \leq N : \log(S(x_n)) < t\}}{N} = t.$$

Remarquons que, pour tout $x > 0$, $\log(S(x)) = \langle \log(x) \rangle$, où $\langle \cdot \rangle$ désigne la partie fractionnaire (appelée aussi **mantisse** du log).

Loi de Benford pour les 2 premiers chiffres

Utilisée couramment dans les applications.

Si $d_1 = 1, \ldots, 9$ et $d_2 = 0, 1, \ldots, 9$, alors la proportion de valeurs x telles que $D_1(x) = d_1$ et $D_2(x) = d_2$ est

$$\log\left(\frac{10 \cdot d_1 + d_2 + 1}{10 \cdot d_1 + d_2}\right) = \log\left(\frac{d_1 d_2 + 1}{d_1 d_2}\right).$$

Par exemple, dans la loi de Benford, la probabilité qu'un nombre admette 1 puis 0 comme premiers chiffres significatifs est

$$\log(11/10) = \log(1.1) = 0.041.$$

La probabilité que les premiers chiffres soient 3 puis 4 est

$$\log(35/34) = 0.0125.$$

Critère de H. Weyl

Définition 3. Une suite $(a_n) \subset \mathbb{R}$ est **uniformément distribuée modulo 1** si, pour tout $t \in [0, 1]$,

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : \langle a_n \rangle < t\}}{N} = t \left(= \int_0^t 1 \cdot dx \right).$$

Théorème de Weyl. *La suite (a_n) est u.d. $\mod 1$ si et seulement si, pour tout $k \in \mathbb{Z}$, $k \neq 0$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2k\pi i a_n} = 0.$$

Exemple. Soit $a > 0$ un nombre irrationnel. Alors la suite (na) est u.d. mod 1. En effet, fixons $k \neq 0$. On a

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N} e^{2k\pi i n a} &= \frac{1}{N}\sum_{n=1}^{N} (e^{2k\pi i a})^n \\ &= e^{2k\pi i a}\frac{1-e^{2k\pi i N a}}{N(1-e^{2k\pi i a})} \to_{N\to\infty} 0.\end{aligned}$$

Voici une première famille d'exemples :

Théorème 1. *Soit* $x_n = r^n$ *une suite géométrique. Alors elle satisfait la LB si et seulement si* $\log(r) \notin \mathbb{Q}$. *Plus généralement, soit* $P(n)$ *un polynôme non constant à coefficients entiers tel que* $P(n) > 0$ *pour tout* n *assez grand. Alors la sous-suite* $(r^{P(n)})$ *de* (r^n) *satisfait la loi de Benford.*

Plus généralement, les conclusions sont vraies pour les suites (x_n) telles que

$$\lim_{n\to\infty} \frac{x_n}{r^n}$$

existe et est positive.

Cela s'applique à la suite de Fibonacci : $F_0 = 0$, $F_1 = 1$ et $F_{n+2} = F_{n+1} + F_n$, $n \geq 0$. On sait que

$$F_n = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}}$$

où $\varphi = \frac{1+\sqrt{5}}{2}$, car $\log(\varphi) \notin \mathbb{Q}$.

La preuve du théorème 1 utilise le critère de H. Weyl appliqué à la suite $\log(r^n) = n\log(r)$ pour la première partie et une généralisation due à Weyl pour la seconde.

Corollaire. *Si un ensemble de valeurs, classées par ordre croissant ou décroissant, suit (approximativement) une progression géométrique de raison r telle que $\log(r) \notin \mathbb{Q}$, alors elle satisfait (approximativement) la loi de Benford.*

Exemple : La suite des puissances de 2

C'est la suite (2^n), c'est-à-dire $2^1 = 2$, $2^2 = 2 \cdot 2 = 4$, $2^3 = 8$, $2^4 = 16 \ldots$. Elle satisfait la LB car $\log(2) \notin \mathbb{Q}$.
Écrivons les 20 premiers termes :

2, **4**, **8**, **1**6, **3**2, **6**4, **1**28, **2**56, **5**12, **1**024, **2**048, **4**096, **8**192, **1**6384, **3**2768, **6**5536, **1**31072, **2**62144, **5**24288, **1**048576.

Regardons ensuite les premiers chiffres des 40 premières valeurs de la suite :

$$2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,$$
$$2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,$$
$$2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1,$$
$$2,\ 4,\ 8,\ 1,\ 3,\ 6,\ 1,\ 2,\ 5,\ 1.$$

Continuons un peu :

$$2,\ 4,\ 8,\ 1,\ 3.$$

Israeli Gel'fand : Existe-t-il une puissance de 2 commençant par 7 ? par 9 ? Si oui, quelles sont les fréquences des chiffres significatifs ?

Il faut attendre jusqu'à $2^{46} = 70'368'744'177'664$ pour obtenir le premier 7,

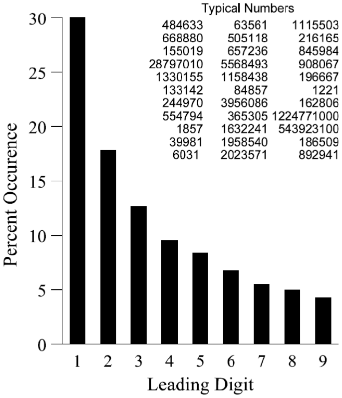
et $2^{53} = 9'007'199'254'740'992$ pour la première valeur commençant par 9.

Exemples

(1) Exemples qui ne satisfont *pas* la loi de Benford :

1. L'ensemble des numéros de téléphone d'une région donnée ; des ensembles de numéros de cartes de crédit, ou des ensembles de codes de sécurité ou de contrôle (code ISBN), numéros AVS.
2. Les ensembles qui suivent une loi normale : les valeurs sont $\pm$ concentrées autour de la moyenne. Par exemple, l'ensemble des tailles des adultes (quelle que soit l'unité choisie !).
3. Tout ensemble de nombres (pseudo-)aléatoires : un tel ensemble satisfait une répartition uniforme.

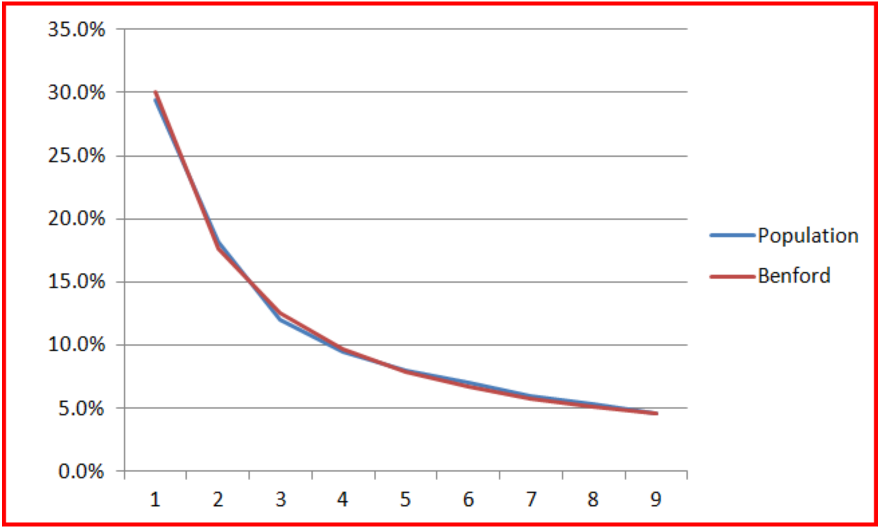
(2) Revenus imposables de 14'414 compagnies américaines (d'après S. W. Smith, 2007) :



(3) Populations des villes américaines (recensement de juillet 2009) : 19’509 villes recensées, de *Abbeville (Alabama)* à *Yoder (Wyoming)*

(3) Populations des villes américaines (recensement de juillet 2009) : 19’509 villes recensées, de *Abbeville (Alabama)* à *Yoder (Wyoming)*

Chiffre	**Fréquence observée**	**Fréquence théorique**
1	0.294	0.301
2	0.181	0.176
3	0.120	0.125
4	0.094	0.097
5	0.079	0.079
6	0.070	0.067
7	0.060	0.058
8	0.053	0.051
9	0.046	0.046


35.0%
30.0%
25.0%
20.0%
15.0%
10.0%
5.0%
0.0%
1
2
3
4
5
6
7
8
9
Population
Benford

(4) Le nombre d'articles publiés par année sur la loi de Benford depuis 1938 (*sic*).

(6) (Diaconis) La suite $(n!)$ satisfait la LB, mais pas la suite des nombres premiers $p_1 = 2, p_2 = 3, \ldots$.

(7) (Deligny, J) Toutefois, si ℓ est un entier positif tel que $\log(\ell) \notin \mathbb{Q}$, alors la sous-suite (p_{ℓ^n}) de la suite des nombres premiers satisfait la LB.

(8) (Berger et Hill) Soit $f : I \to \mathbb{R}$ une fonction suffisamment régulière, et soit $x^* \in I$ un zéro de f.

Soit $(x_n)_{n \geq 0}$ une suite obtenue par la formule de Newton :

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Alors pour presque toute valeur initiale x_0 proche de x^*, les suites $(|x_n - x^*|)$ et $(|x_{n+1} - x_n|)$ satisfont la LB.

Applications et autres exemples

(1) Détection de fraudes (erreurs ou falsifications de données) dans les comptabilités, audits :

Mark Nigrini (www.nigrini.com) a amassé dès le début des années 1990 un grand nombre de preuves empiriques qui justifient l'usage de la loi de Benford comme indicateur de fraude. Si la fraude est délibérée, les données suivent rarement la loi de Benford.

(2) Traitement d'images (Pérez-González et al.) : l'application à la transformée en cosinus discrète d'images aide à déterminer si elles contiennent des messages cachés (stéganographie). La méthode n'est pas plus fiable que d'autres, mais plus simple.

(3) Phénomènes naturels :

- Un ensemble de données d'un système naturel peut suivre la LB dans un certain état, mais pas dans un autre.
- Tremblements de terre (Sambridge et al.) : Avant un événement, les valeurs relevées par les sismographes ne satisfont pas la LB (bruit) ; mais elles la satisfont au début d'un séisme.
- Les intervalles de temps entre deux tremblements de terre consécutifs semblent suivre la loi de Benford : en 2012, 19'451 événements recensés (un événement toutes les 27 minutes) !

- Les demi-vies des éléments radioactifs dans les désintégrations α et β.
- Les énergies de transition dans les couches électroniques.
- Les distributions statistiques de Boltzmann-Gibbs et de Fermi-Dirac.
- Les tailles des fichiers dans les disques durs.

(4) Psychologie ; on constate deux types de comportement :

- Productions de nombres sans signification : ne satisfont ni la LB ni la loi uniforme, mais on constate un phénomène de *priming* qui signifie que si on demande de produire un nombre de d chiffres, alors le digit d est le plus fréquent en tant que chiffre significatif.
- Productions de valeurs ayant un sens (populations,. . .) : les résultats s'approchent de la LB pour le premier chiffre significatif, mais pas pour les deux premiers.

Simulations du hasard

Les êtres humains simulent **mal** le hasard :

Choix d'une couleur :

Rouge ou **bleu**.

Choix d'un arbre :

Le **chêne**.

Choix d'un chiffre entre 1 et 9 :

Le **7**.

Choix de "pile ou face" :

P à 55% chez les francophones et **F** chez les anglophones (même pourcentage).

Raison probable : nous appelons ce jeu *pile ou face*,

ils l'appellent *head and tail*.

Exemple d'expérience

Un groupe de personnes est réparti en deux sous-groupes A et B de même taille, à l'aveugle pour l'expérimentateur.

Chaque personne du groupe A lance 100 fois une pièce de monnaie et note la séquence des résultats.

Chaque personne du groupe B *simule* 100 lancers de pièce de monnaie et note la séquence inventée.

L'expérimentateur doit déterminer à quel groupe appartient chaque séquence.

Sur 100 lancers consécutifs d'une pièce de monnaie, la probabilité d'obtenir au moins une séquence *PPPPPP* ou *FFFFFF* est d'environ

$$81\ \%;$$

la probabilité d'obtenir au moins une séquence *PPPPP* ou *FFFFF* est d'environ 97 % ...

Si on demande à des humains de simuler 100 lancers, il est très peu probable que l'on obtienne de telles séquences.

Théorèmes

Théorème 2 (T. Hill, 1996) : *Si des ensembles de données sont pris au hasard et si l'on en extrait des échantillons aléatoirement, alors les fréquences des chiffres significatifs suivent la loi de Benford.*

N.B. Les ensembles en question peuvent ne pas satisfaire la loi de Benford individuellement (nombres aléatoires, listes de numéros de téléphones, listes de tailles d'êtres humains, ...) ; le théorème affirme que si on tire au hasard des valeurs de chacun de ces ensembles, alors la liste ainsi obtenue satisfait la loi de Benford.

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
	Average.......	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
	Probable Error	±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

Théorème 3 (J. Boyle, 1994) : *Si on prend un ensemble de valeurs numériques et qu'on effectue des multiplications et des divisions entre elles un grand nombre de fois, alors l'ensemble final satisfait la loi de Benford.*

Cela fournit une explication plausible du fait que de nombreuses comptabilités satisfont cette loi.

Définition 4. Soit $\mathbb{P}$ une mesure de probabilité sur $\mathbb{R}^+$. Elle est de **Benford** si

$$\mathbb{P}(S(x) \leq t) = \log(t) = \int_1^t \frac{ds}{s\ln(10)} \quad (t \in [0, 10)).$$

Théorème 4. (Hill) *La probabilité $\mathbb{P}$ ci-dessus est la seule probabilité sur $\mathbb{R}^+$ invariante au sens suivant : $\mathbb{P}(aA) = \mathbb{P}(A)$ pour tout $a > 0$ et pour tout borélien A de la forme*

$$A = \bigcup_{k\in\mathbb{Z}} 10^k B \quad (B \subset [1, 10)).$$

De manière équivalente, pour tout $m \geq 1$, pour tout $1 \leq d_1 \leq 9$ et tous $0 \leq d_j \leq 9$ ($j = 2, \ldots, m$), et pour tout $a > 0$

$$\mathbb{P}(D_j(ax) = d_j, j = 1, \ldots, m) = \mathbb{P}(D_j(x) = d_j, j = 1, \ldots, m).$$

Théorème 5 (N. Gauvrit, J.-P. Delahaye, 2009) *Toute distribution de Pareto suit à peu près la loi de Benford.*

Les valeurs X d'une telle distribution satisfont :

$$P(X > x) = \left(\frac{x}{x_{\min}}\right)^{-k}$$

pour $x \geq x_{\min}$, où $k > 0$ et $x_{\min} > 0$ sont des paramètres.

Cette loi est appelée aussi *principe des 20-80* :

- **Fiscalité** : 20% des contribuables paient le 80% des impôts ;
- **Contrôle de gestion** : 20% des indicateurs fournissent le 80% des informations ;
- **Consommation** : 20% des êtres humains consomment le 80% de la nourriture ;
- **Service après vente** : 20% des clients sont responsables de 80% des réclamations !

Merci !