Math. & Sci. hum. / Mathematics and Social Sciences (46e année, n° 182, *2008*(2), p. 7–15)

POURQUOI LA LOI DE BENFORD N'EST PAS MYSTÉRIEUSE

Nicolas GAUVRIT[1], Jean-Paul DELAHAYE[2]

RÉSUMÉ – *La loi dite* de Benford *prévoit que le premier chiffre significatif d'un nombre tiré de manière aléatoire suit une loi logarithmique et non, comme on pourrait s'y attendre, une loi uniforme. Cette loi expérimentale a été démontrée mathématiquement pour diverses suites numériques, et a été vérifiée expérimentalement sur d'immenses corpus numériques. Sur ces données naturelles, la loi de Benford apparaît très souvent comme une bonne approximation de la réalité, mais il semble aussi qu'elle ne soit qu'une approximation.*

Nous proposons une nouvelle explication *de la loi de Benford, qui ne devrait pas, à notre avis, être considérée comme paradoxale mathématiquement. Nous énonçons un critère de régularité naturel sur une variable X et nous démontrons que, si ce critière est vérifié, alors X suit « à peu près » la loi de Benford.*

MOTS CLÉS – Biais d'équiprobabilité, Loi de Benford, Paradoxe.

SUMMARY – A new general explanation of Bendford's law

According to Benford's law*, the first digit of a random number does not follow a uniform distribution, as many people believe, but a logarithmic distribution. This law was at the begining purely experimental, but it is now established that it holds for various mathematical series and some natural data sets. Concerning data sets, Benford's law often appears as a good approximation of the reality, but as no more than an approximation.*

Our aim is to present a new explanation *for this law. We argue that it should not be considered as a mathematical paradox, but as a purely psychological paradox, a result of a cognitive bias. We express a general criterion of regularity on a random variable X and prove that, whenever X follow this criterion, X is approximately Benford.*

KEYWORDS – Benford's law, Equiprobability bias, Paradox.

D'abord remarquée par Newcomb [1881], la loi dite « de Benford » n'a connu son heure de gloire qu'à partir d'une nouvelle publication 57 ans plus tard [Benford, 1938]. Cette loi prévoit, dans sa version la plus faible, que le premier chiffre significatif d'un nombre tiré dans une série statistique à peu près quelconque ne suit absolument pas la loi uniforme sur $\{1, ..., 9\}$. Au contraire, d'après cette loi, le chiffre 1 est largement prépondérant, le chiffre 9 étant à l'inverse le moins fréquent.

[1]Equipe didactique des mathématiques (DIDIREM), EA 1547, Centre Chevaleret, Université Paris VII, 175 rue du Chevaleret 75013 Paris, adems@free.fr

[2]Laboratoire d'Informatique Fondamentale de Lille (LIFL), UMR USTL/CNRS 8022, Université des sciences et technologie de Lille, bâtiment M3, 59655 Villeneuve d'Ascq Cedex, jean-paul.delahaye@lifl.fr

La probabilité d'apparition d'un chiffre d en position de premier chiffre significatif (c'est-à-dire le chiffre non nul le plus à gauche dans l'écriture décimale du nombre) est $\log\left(1+\frac{1}{d}\right)$.

Largement considérée comme étonnante pour ne pas dire paradoxale, la loi de Benford (ou de Newcomb-Benford) a suscité depuis sa découverte un grand nombre de publications, qui cherchent essentiellement à répondre à deux questions :

(1) Pourquoi la plupart des données empiriques comme les constantes physiques (cf. [Knuth, 1969] ou [Burke, Kincanon, 1991]), certaines données économiques ou démographiques [Nigrini, Wood, 1995], vérifient-elles *approximativement* cette loi ?[3]

(2) Quelles conditions générales doit vérifier une variable aléatoire X pour suivre la loi de Benford ?

En marge de ces tentatives de résolution du paradoxe s'est posée une autre question : cette loi de Benford est-elle vraiment vérifiée ? Il apparaît que dans les faits bien des ensembles de données ne suivent pas du tout la loi de Benford [Scott, Fasli, 2001]. C'est le cas par exemple des nombres pseudo-aléatoires donnés par des humains [Hill, 1988]. Ces résultats ont déjà été utilisés pour repérer des fraudes, notamment en matière fiscale, mais aussi des données contrefaites dans des articles scientifiques. Mais surtout, ce qui est peut-être plus important, beaucoup de lois empiriques suivent *à peu près* la loi de Benford, en conservant vis-à-vis d'elle une différence significative que la multiplication des données ne résorbe pas.

Cette loi fut d'abord empirique, mais il est maintenant prouvé qu'elle est rigoureusement vraie pour certains types de données, comme les orbites de certains systèmes dynamiques [Berger *et al.*, 2004].

Hill, après avoir démontré que la loi de Benford était la seule loi possible si on impose l'invariance par changement d'échelle ou de base [1995(a)], a prouvé un théorème selon lequel une suite de valeurs obtenues en sélectionnant, selon certaines contraintes, différents échantillons dans différentes populations pour des variables diverses donne finalement une loi de Benford [1995(b)]. C'est un équivalent, au fond, du théorème central limite : un échantillonnage bien fait doit mener à une loi particulière. Boyle [1994] montre que la multiplication entre elles de variables indépendantes conduit à la loi de Benford. Autrement dit, la loi de Benford serait naturelle si les nombreux facteurs qui expliquent telle ou telle grandeur agissent *multiplicativement*.

Certaines suites numériques, comme (n^n) ou $n!$ [Posch, à paraître] suivent exactement la loi de Benford. C'est le cas également de la suite des (a^n), avec $\log_{10}(a) \in \mathbb{R}\backslash\mathbb{Q}$ et plus généralement de toute suite définie par une relation de récurrence polynomiale et dont le polynôme définitoire satisfait certaines conditions [Jolissaint, 2005].

[3]La terminologie varie d'un auteur à l'autre. On dit parfois que *la variable X suit une loi de Benford* pour signifier que le premier chiffre significatif de X suit cette loi logarithmique. Il arrive aussi qu'on dise alors que *le premier chiffre significatif* de X suit une loi de Benford. Enfin, il arrive que l'expression *X suit une loi de Benford* indique que la partie fractionnaire de $\log(X)$ suit une loi uniforme. Nous utiliserons dans la suite les trois expressions indifféremment, le contexte permettant toujours de comprendre de quelle notion il s'agit.

Deux raisons peuvent expliquer l'étonnement que suscite la loi de Benford.

La première est qu'elle est souvent présentée (à tort) comme une loi universelle, vraie pour tout ensemble de données empiriques ou mathématiques. De nombreuses données empiriques « aléatoires » ne la vérifient pourtant pas, et bien des séries ou des variables mathématiquement simples non plus. Les suites (n), (kn), des nombres premiers, sont dans ce cas.

Enfin, de nombreuses listes de données numériques suivent une loi proche de la loi de Benford, mais tout de même différente. Dans l'article initial de Benford [1938], par exemple, la moitié des listes considérées s'écartent significativement de la loi prévue.

La seconde raison est d'ordre psychologique : si nous considérons le *dernier chiffre* d'une série aléatoire suffisamment étalée et régulière de nombres entiers, nous nous attendons à trouver une loi uniforme, donc autant de 0 que de 1, que de 2, etc. Et plus généralement, si des réels sont choisis « aléatoirement », on s'attend également à ce que la partie fractionnaire des nombres suive une loi uniforme. En l'occurrence, cette attente paraît raisonnable, bien qu'il faille, si l'on veut être rigoureux, préciser ce qu'on entend par « série aléatoire de nombres » (la référence par défaut, la loi uniforme, ne pouvant être considérée ici). Cette « loi de la partie fractionnaire » ou du « chiffre des unités », que l'on attend uniforme, est donc une hypothèse rationnelle. Or, une approche superficielle qui s'appuie plus sur la forme que sur le fond, nous pousse à considérer la question du « dernier chiffre » comme parfaitement similaire à celle du « premier chiffre ».

C'est justement là que l'intuition fait défaut. Car s'il y a bien un lien entre la loi de la partie fractionnaire et du premier chiffre significatif, il est bien moins direct qu'on ne le pense. On imagine volontiers que si X est une variable aléatoire suffisamment régulière, $\log(X)$ l'est aussi, et l'on s'attend donc à ce que la partie fractionnaire de X, *mais aussi celle de* $\log(X)$, suive une loi uniforme. Or, dans ce dernier cas, cette attente intuitive est précisément une version de la loi de Benford...

En réalité, l'attente d'une loi uniforme sur le premier ou le dernier chiffre d'un nombre est le résultat d'une illusion bien connue des psychologues : le biais d'équiprobabilité [Lecoutre, 1992]. Ce biais est une tendance humaine à considérer que le hasard implique nécessairement l'uniformité. Les humains considèrent spontanément que tout ce qui est aléatoire est uniforme. Le premier chiffre significatif, évidemment obtenu par hasard, devrait alors suivre une loi uniforme. La loi de Benford peut donc être comprise comme un paradoxe psychologique (en non mathématique), provenant d'un biais humain dans la perception du hasard, non d'une incohérence mathématique ou d'une particularité des séries de données utilisées.

Notre but est ici de montrer que des conditions générales et naturelles portant sur les variables impliquent que la loi de Benford est au moins approximativement vérifiée.

Il semble que seules des variables dont le logarithme est suffisamment étalé puisse correspondre à une loi de Benford. Nous donnons dans la première section une version de ce lien en termes d'approximation. Plus précisément, nous montrons que si $\log X$ suit une loi suffisamment régulière et étalée, alors X suit *approximativement* la loi de Benford.

Mais cela ne répond pas entièrement à la question du pourquoi tant que nous n'avons pas saisi ce que la régularité de $\log X$ a de naturel – même si cette régularité est assez intuitive. Dans la dernière section, nous nous attachons à donner une condition suffisante sur X pour que la partie fractionnaire de $\log X$ suive approximativement une loi uniforme sur $[0, 1[$, et donc que X suive approximativement la loi de Benford.

1. CONVERGENCE DE LA PARTIE FRACTIONNAIRE VERS LA LOI UNIFORME

On notera dans la suite $\lfloor y \rfloor$ la partie entière d'un réel y (plus grand entier inférieur ou égal à y) et $\{y\} = y - \lfloor y \rfloor$ sa partie fractionnaire. Le logarithme en base 10 sera noté log, la notation ln désignant le logarithme népérien. Le logarithme dans une autre base b sera noté $\log_b$. Id désigne l'application identité.

Tout nombre strictement positif x s'écrit (c'est la notation « scientifique ») :

$$x = 10^{\lfloor \log x \rfloor}.10^{\{\log x\}}.$$

Le nombre $10^{\{\log x\}-1}$ est appelée la *mantisse* de x. $\{\log x\}$ est la *mantisse logarithmique* de x.

Cette écriture permet un passage entre certaines caractéristiques des chiffres dans l'écriture de x en notation décimale, et d'autres propriétés portant sur $\{\log x\}$. Par exemple, le premier chiffre significatif de x est $\left\lfloor 10^{\{\log x\}} \right\rfloor$.

Un fait maintenant bien connu (cf. par exemple [Diaconis, 1977]) est que l'uniformité de la *mantisse logarithmique* $\{\log X\}$ d'une variable aléatoire réelle strictement positive X est équivalente à la loi de Benford sur X. En fait, c'est sous cette forme plus générale impliquant la mantisse logarithmique que la loi de Benford a la première fois été énoncée par Newcomb [1881][4].

La plupart des lois de probabilité continues rencontrées en mathématiques, mais aussi dans la pratique, notamment au travers des statistiques, ont une allure *régulière*. Intuitivement, elles présentent généralement l'allure d'une gaussienne déformée[5]. En particulier, et c'est une caractéristique qui nous intéresse en droite ligne, leurs densités admettent un unique maximum, atteint en un point a, et ne présentent qu'une seule inversion de monotonie.

Cette caractéristique n'est pas réservée aux données brutes : les lois dérivés, par exemple les lois de Student, de Fisher, etc., présentent aussi cette allure.

Cependant, ces lois, même très régulières, ne satisfont pas toujours à la propriété de Benford (pensez par exemple à la loi uniforme sur $[1, 2]$). Ce qui semble important,

[4]La loi de Benford peut aussi être généralisée à une base b quelconque : une variable X suit la loi de Benford en base b si $\{\log_b X\}$ est uniforme.

[5]Dépendantes d'un grand nombre de facteurs, ces lois tombent plus ou moins sous le coup du théorème central limite, ce qui explique leur allure « en cloche » déformée, si souvent constatée en sciences humaines.

c'est l'*étalement* de la densité (au sens de la variance par exemple). C'est ce que notent Scott et Fasli [2001] sur le cas particulier des lois log-normales.

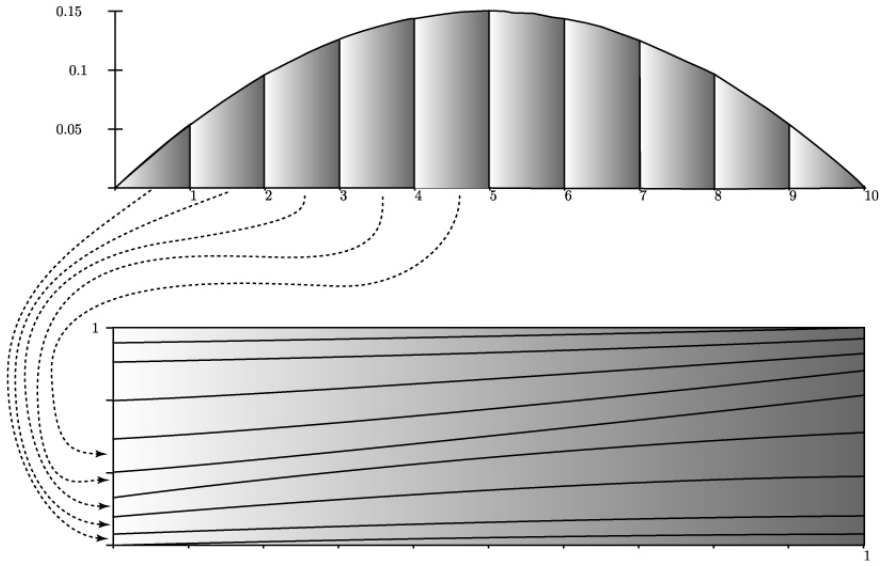

FIGURE 1. Illustration de l'idée intuitive selon laquelle la densité « régulière » d'une variable X donne une partie fractionnaire presque uniforme. Les « tranches » de la densité de départ sont superposées pour former la densité de la loi de X. Les pentes des tranches se compensent en partie (parfaitement si la densité de départ est affine sur tout segment $[n, n+1]$)

Considérons maintenant une variable Y « régulière » et « étalée » (ces notions seront formalisées plus loin dans l'énoncé du théorème). Une idée intuitive est que la partie fractionnaire d'une telle variable devrait suivre une loi presque uniforme. La régularité, associée à l'étalement, laisse en effet présager que sur chaque intervalle $[n, n+1[$, Y est plus ou moins affine, et que les écarts à l'uniformité se compensent en partie (cf. Figure 1). La somme des restrictions de la densité de Y à ces divers intervalles (après translation) forme la loi de $\{Y\}$, dès lors à peu près uniforme. C'est l'idée informelle qui mène à l'énoncé du théorème suivant.

THÉORÈME 1. *Soit Y une variable aléatoire réelle continue de densité f telle que (1) f admet un maximum M atteint en un point a_0, (2) f est croissante sur $]-\infty, a_0]$ et décroissante sur $[a_0, +\infty[$. Soit x un point de $[0, 1[$. Alors*

$$|P(\{Y\} < x) - x| < 2M.$$

En particulier, si une suite (Y_n) de variables aléatoires réelles satisfaisant ces hypothèses vérifie en outre $M_n \underset{n\to\infty}{\longrightarrow} 0$, alors $\{Y_n\}$ converge en loi vers la loi uniforme sur $[0, 1[$ quand $n \longrightarrow \infty$.

DÉMONSTRATION 1. On peut supposer sans nuire à la généralité de la chose que $a_0 \in [0; 1[$. Fixons $x \in]0; 1[$ (le cas $x = 0$ est évident). Notons $I_{n,x} = [n, n+x[$.

Pour tout $n \leq -1$, on a par croissance de f

$$\frac{1}{x}\int_{I_{n,x}} f(t)dt \leq \int_{n}^{n+1} f(t)dt$$

on tire de cette inégalité

$$\frac{1}{x}\sum_{n\leq -1}\int_{I_{n,x}} f(t)dt \leq \int_{-\infty}^{0} f(t)dt.$$

Pour tout $n \geq 2$, on a par décroissance de f,

$$\frac{1}{x}\int_{I_{n,x}} f(t)dt \leq \int_{n-1+x}^{n+x} f(t)dt$$

d'où

$$\frac{1}{x}\sum_{n\geq 2}\int_{I_{n,x}} f(t)dt \leq \int_{1+x}^{+\infty} f(t)dt.$$

D'autre part,

$$\int_{I_{0,x}} f \leq xM$$

et

$$\int_{I_{1,x}} f \leq xM,$$

si bien que

$$\frac{1}{x}\sum_{n\in\mathbb{Z}}\int_{I_{n,x}} f \leq \int_{-\infty}^{\infty} f + 2M.$$

On montre de manière analogue

$$\frac{1}{x}\sum_{n\in\mathbb{Z}}\int_{I_{n,x}} f \geq \int_{-\infty}^{\infty} f - 2M.$$

Comme $\sum_{\mathbb{Z}}\int_{I_{n,x}} f = P\left(\{Y\} < x\right)$, que $x < 1$ et que $\int_{-\infty}^{\infty} f = 1$, le théorème est démontré. ■

L'encadrement ainsi déterminé est grossier, mais largement suffisant pour nos besoins.

Remarquons qu'on peut obtenir le même type d'inégalité pour des densités admettant plus d'un changement de monotonie, si bien que le théorème reste valable pour des lois régulières bimodales par exemple.

2. MAJORATION DE L'ÉCART À LA LOI DE BENFORD

Soit maintenant X une variable aléatoire réelle strictement positive. Lorsque X suit certaines conditions, le résultat précédent s'applique à $\log(X)$, montrant ainsi que $\{\log(X)\}$ est « proche » d'une loi uniforme, donc que X suit à peu près la loi de Benford. C'est ce que montre le théorème suivant.

THÉORÈME 2. *Soit X une variable aléatoire réelle strictement positive de densité f telle que $Id.f : x \longmapsto xf(x)$ vérifie les conditions suivantes : $\exists a > 0$ tel que (1) $\max(Id.f) = m = a.f(a)$ et (2) $Id.f$ est croissante sur $]0, a]$, puis décroissante sur $[a, +\infty[$. Dans ce cas, pour tout $z \in]0, 1]$,*

$$|P(\{\log X\} < z) - z| < 2\ln(10)\, m$$

En particulier, si X_n est une suite de variables aléatoires de densité f_n satisfaisant ces conditions et telles que $m_n = \max(Id.f_n)$ tende vers 0, $\{\log(X_n)\}$ tend vers la loi uniforme sur $[0, 1[$ en loi.

DÉMONSTRATION 2. Il s'agit d'une application du théorème 1. Appelons F la fonction de répartition de X et G celle de $\log X$, de densité g. On a

$$\begin{aligned} G(x) &= P(\log X < x) \\ &= P(X < 10^x) \\ &= F(10^x). \end{aligned}$$

d'où

$$g(x) = \ln(10)\, 10^x f(10^x).$$

g a donc un maximum, $\ln(10)\, m$, et est croissante sur $]-\infty, \log a]$, et décroissante sur $[\log a, +\infty[$. ■

Partant d'une variable X de densité f « ressemblant » à une gaussienne, croissante sur $]-\infty, b]$, on obtient une fonction $Id.f$ également croissante sur cet intervalle. La condition de décroissance à partir d'un certain a (*a priori* différent de b) est donc une condition portant essentiellement sur la décroissance de f à l'infini. Notons encore une fois qu'un théorème équivalent s'applique pour des fonctions du type $Id.f$ ayant un nombre de changements de monotonie fini et inférieur à une borne préalablement fixée.

Exemple. Soit X de loi exponentielle de paramètre λ, de densité donnée par

$$f(x) = \frac{e^{-\lambda x}}{\lambda}, \quad \forall x \geq 0.$$

La fonction $x \longmapsto xf(x)$ est croissante jusqu'à $\frac{1}{\lambda}$, puis décroissante. Son maximum vaut $\frac{e^{-1}}{\lambda}$. Ainsi, pour tout $z \in]0, 1]$,

$$|P(\{\log X\} < z) - z| \ < \ 2\ln(10)\, e^{-1}\frac{1}{\lambda} \ < \ \frac{1.7}{\lambda}.$$

En particulier, si $\lambda \longrightarrow \infty$, $\{\log X\} \longrightarrow U[0, 1[$ en loi. En particulier, le premier chiffre significatif de X suit à la limite une loi logarithmique, conformément à la loi de Benford.

Remarque. Certaines variables qui paraissent assez régulières et étalées ne vérifient pourtant pas la loi de Benford. Considérons par exemple X une variable uniforme sur $]0, u]$. Sa densité est $f = \frac{1}{u}\chi_{]0,u]}$, ($\chi_A$ désigne la fonction caractéristique de A) et on a donc bien $Id.f$ croissante jusqu'à u, puis décroissante. La fonction $Id.f$ atteint son maximum $m = 1$ en u. On a donc $2\ln(10)\, m = 2\ln 10 > 1$. Par conséquent, on ne peut rien déduire de notre théorème, même en faisant tendre u vers l'infini. Et il se trouve que la loi uniforme ne s'approche pas de la loi de Benford quand u tend vers l'infini (par exemple, pour $u = 10^n$, le premier chiffre significatif suit une loi uniforme et non logarithmique).

3. DISCUSSION CONCLUSIVE

Des données assez variées, dispersées, soit qu'elles proviennent de sources différentes comme dans le théorème de Hill, soit qu'elles proviennent de mesures physiques disparates, comme la taille des rivières et des fleuves, donnent assez généralement des variables étalées et régulières.

Nous avons vu que, sous des conditions raisonnables de régularité portant sur X, X suit *au moins approximativement* la loi de Newcomb-Benford. Il n'y a bien entendu aucune raison pour que X suive exactement cette loi en général. En tout état de cause, les théorèmes que nous avons démontrés ne permettent jamais de conclure qu'une variable donnée suit exactement la loi de Benford[6]. Cela permet de voir que la loi de Newcomb-Benford n'est au fond rien d'autre qu'une conséquence mathématique de la régularité de X. Et surtout, cela permet d'expliquer pourquoi bien des tables de données brutes fournissent des données suivant *à peu près* la loi de Benford (mais pas toujours parfaitement).

Notons également que les majorations que nous avons choisies sont très grossières, et qu'il est sans doute possible d'affiner un peu notre résultat et de donner ainsi un critère plus large de « ressemblance » à la loi de Benford.

Comme nous le suggérions au début de l'article, le paradoxe de Benford est psychologique et non mathématique : ce qui est étonnant n'est pas qu'on observe à peu près la loi de Benford, mais qu'on s'*attende* à trouver autre chose (une loi uniforme sur les chiffres). Considérer la loi de Benford comme paradoxale est une position du même acabit que celle qui consisterait à trouver paradoxal le théorème de la limite centrale, qui fait émerger comme limite non une loi uniforme, mais une gaussienne – et à échaffauder des explications de cette loi.

Notons enfin que ce que nous venons de faire avec le logarithme en base 10 est évidemment transposable en n'importe quelle base, mais aussi que d'autres fonctions que log sont parfaitement envisageables. De même qu'on a une loi de Benford, on peut imaginer une loi du carré ou de la racine qui, sous certaines contraintes de régularité, affirmeraient que $\left\{\sqrt{X}\right\}$, ou $\{X^2\}$, sont « proches » de la loi uniforme. Bien entendu, l'intérêt de la loi de Benford restera à part, du fait que l'uniformité

[6]Si l'on comprend ces théorèmes comme des théorèmes limites, la loi de Benford devrait émerger à l'infini. L'application finie de ces théorèmes est une explication possible du fait que de nombreuses lois sont proches de celle de Benford sans coïncider avec elle.

de $\{\log X\}$ se traduit de manière très parlante et donne des conclusions sur la répartition des premiers chiffres significatifs.

BIBLIOGRAPHIE

BENFORD F., "The law of anomalous numbers", *Proceedings of the American Philosophical Society* 78, 1938, p. 127-131.

BERGER A., BUNIMOVICH L., HILL T., "One-dimensional dynamical systems and Benford's law", *Transactions of the American Mathematical Society* 357(1), 2004, p. 197-219.

BOYLE J., "An application of Fourier series to the most significant digit problem", *America Mathematical Monthly* 101, 1994, p. 879-886.

BURKE J., KINCANON E., "Benford's law and physical constants: he distribution of initial digits", *American Journal of Physics* 59, 1991, p. 952.

DIACONIS P., "The distribution of leading digits and uniform distribution mod", *Annals of Probability* 5, 1977, p. 72-81.

HILL T., "Random-number guessing and the first-digit phenomenon", *Psychological Reports* 62, 1988, p. 967-971.

HILL T., "Base-invariance implies Benford's law", *Proceedings of the American Mathematical Society* 123, 1995(a), p. 887-895.

HILL T., "A statistical derivation of the Significant-Digit Law", *Statistical Science* 10(4), 1995(b), p. 354-363.

JOLISSAINT P., « Loi de Benford, relations de récurrence et suites équidistribuées », *Elemente der Mathematik* 60(1), 2005, p. 10-18. http://ww.jura.ch/ijsla/Benford.pdf

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

LECOUTRE M., "Cognitive models and problem spaces in 'purely random' situations", *Educational Studies in Mathematics* 23, 1992, p. 557-568.

NEWCOMB S., "Note on the frequency of use of the different digits in natural numbers", *American Journal of Mathematics* 4, 1881, p. 39-40.

NIGRINI M., WOOD W., "Assessing the integrity of tabulated demographic data", preprint, University of Cincinnati and St. Mary's University, 1995.

POSCH P.N., "A survey of sequences and distribution functions satisfying the first-digit-law", *Journal of Statistics & Management Systems*, [à paraître].

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. http://citeseer.ist.psu.edu/709593.html