Math. & Sci. hum. / Mathematics and Social Sciences (47e année, n° 186, *2009*(2), p. 5–15)

LOI DE BENFORD GÉNÉRALE

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

RÉSUMÉ – *La loi dite « de Benford » s'applique à une variable X dont le logarithme a une partie fractionnaire uniforme. Il a été montré qu'elle s'applique approximativement à de nombreuses séries numériques réelles.*

Diverses explications ont été avancées, qui s'appuient sur certaines particularités des données utilisées, en lien avec les caractéristiques de la fonction log*. Une hypothèse bien plus élémentaire et générale a toutefois été proposée récemment, selon laquelle c'est le caractère régulier et étalé des données qui, seul, entraîne la loi de Benford.*

Si cette explication est bonne, la loi de Benford ne dépend pas fondamentalement de l'utilisation de la fonction log *. Dans cet article, nous testons la « loi de Benford générale » pour la fonction u selon laquelle la partie fractionnaire de $u(X)$ est uniforme. Des données réelles, des suites mathématiques, et des variables à densité, sont testées pour les fonctions* $\log \circ \log$*, racine, et* $x \longmapsto \pi x^2$. *Les résultats confirment que la fonction* log *n'a rien de particulier, et donnent quelques précisions sur l'intérêt et les propriétés des diverses variantes (quand on fixe u) de la loi de Benford générale.*

MOTS CLÉS – Densité modulo 1, Loi de Benford, Loi de Benford générale, mathématiques expérimentales

SUMMARY – General Benford Law

A variable X satisfies the "Benford law" if $log(X)$ has a uniform distribution modulo 1. This law approximately applies to many experimental or observational data sets.

Many theories have been put forward as explanations for this phenomenon, mostly based on the characteristics of the log function. An elementary new explanation has recently been published, based on the fact that any X whose distribution is "smooth" and "scattered" enough is Benford. Scatter and smoothness of usual data ensure that $log(X)$ is itself smooth and scattered, which in turn implies the Benford characteristic of X.

If this explanation is the good one, the Benford law should not depend on the log function itself. In this paper, we define and test a "General Benford Law" for a function u. X satisfies this law if $u(X)$ is uniform modulo 1. Statistical data, mathematical series and continuous variables are tested for functions $\log \circ \log$*,* $x \longmapsto \pi x^2$ *, and square root. The results suggest that the Benford law for function log is not pathological, and that other functions also apply to natural data. We discuss possible interests and properties of this general Benford law.*

KEYWORDS – Benford Law, Experimental mathematics, General Benford Law, Modulo 1 density

[1]Équipe de didactique des mathématiques (DIDIREM), EA 1547, Université Paris VII, centre Chevaleret, 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 technologies de Lille, Bât M3, 59655 Villeneuve d'Ascq cedex, jean-paul.delahaye@lifl.fr

1. INTRODUCTION

Dans sa forme la plus populaire, la loi de Benford (ou de Newcomb) prévoit que le premier chiffre significatif, i.e. le chiffre non nul le plus à gauche dans l'écriture décimale du nombre, d'une suite "aléatoire" de nombres suit une loi logarithmique et non pas uniforme, ce qui est contraire à l'intuition majoritaire. La probabilité que le premier chiffre significatif soit d est alors $\log\left(1+\frac{1}{d}\right)$. Cependant, les premières publications sur la question, celles de Newcomb [1881] et Benford [1938], proposaient une version plus exigente de la loi et qui implique la distribution logarithmique des premiers chiffres significatifs : si X est une variable aléatoire, on dira qu'elle suit la loi de Benford si la partie fractionnaire du logarithme en base 10 de X, $\{\log(X)\}$, suit une loi uniforme sur $[0,1[$.

Cette loi est approximativement vérifiée pour de nombreux ensembles de données naturelles comme les taux de la bourse, les surfaces de lacs ou les populations des pays [Nigrini et Wood, 1995]. Elle est également vérifiée pour certaines variables aléatoires ou suites mathématiques simples (cf. par exemple [Berger *et al.*, 2004 ; Jolissaint 2005 ; Posch, 2008]). Pour autant, l'idée répandue selon laquelle la loi de Benford est incoutournable reste très exagérée : une part non négligeable des données testées par les chercheurs ne suivent pas du tout la loi de Benford, et celles qui la respectent le font généralement de manière seulement approximative, quelle que soit la taille de l'ensemble de données [Scott et Fasli, 2001].

Pour expliquer cette loi qui défie l'intuition, un très grand nombre d'hypothèses ont été avancées [Hürlimann, 2006]. Si l'on suppose par exemple que les données sont obtenues à partir de nombreux facteurs qui agissent de manière multiplicative, on obtient à la limite une loi de Benford [Boyle, 1994]. Ce résultat est un équivalent logarithmique de la loi des grands nombres. Si les données utilisées sont obtenues en mélangeant de manière *ad hoc* un ensemble de lois de probabilités, on obtient à nouveau une loi de Benford (cf. [Hill, 1996 ; Janvresse et De La Rue, 2004]). Toutes ces « explications » sont limitées, puisqu'elles ne s'appliquent qu'à un type de données très particulier. Si l'on veut les appliquer à grande échelle, il faut ajouter des hypothèses difficiles à formuler. La liste des populations des pays n'est en effet pas obtenue par mélange de diverses variables. Elle n'est pas non plus, *a priori*, le résultat d'un processus où de nombreux facteurs agissent multiplicativement.

Nous avons récemment proposé une explication à la fois beaucoup plus simple – la démonstration du théorème correspondant et l'idée intuitive qu'il formalise sont élémentaires – et bien plus générale : toute variable aléatoire X à densité, étalée et régulière, suit approximativement la loi de Benford [Gauvrit et Delahaye, 2008]. Il suffit que la densité de $\log(X)$ n'ait qu'un changement de monotonie et soit majorée par un nombre suffisamment petit pour que la loi de Benford soit approximativement vérifiée.

Plus précisément, le théorème est le suivant ($\{x\}$ désigne la partie fractionnaire de x) :

THÉORÈME 1. *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, $\{\log(X_n)\}$ converge en loi vers la loi uniforme sur $[0, 1[$.

La loi de Benford affirme l'uniformité de la partie fractionnaire de $\log(X)$, mais la fonction log peut être remplacée par une autre fonction u strictement croissante et C^1 quelconque, donnant ainsi une version plus large du théorème, dont la démonstration reste élémentaire :

THÉORÈME 2. *Soit X une variable aléatoire réelle à valeurs dans un intervalle I, de densité f. Soit u une application C^1 strictement croissante de I dans $\mathbb{R}$, telle que $\frac{f}{u'} : x \longmapsto \frac{f(x)}{u'(x)}$ vérifie les conditions suivantes : $\exists a > 0$ tel que (1) $\max\left(\frac{f}{u'}\right) = m = \frac{f}{u'}(a)$ et (2) $\frac{f}{u'}$ est croissante sur $]-\infty, a] \cap I$, puis décroissante sur $[a, +\infty[\cap I$. Dans ce cas, pour tout $z \in [0, 1[$,*

$$|P(\{u(X)\} < z) - z| < 2m.$$

En particulier, si (X_n) est une suite de telles variables, avec $\lim_{+\infty}(m_n) = 0$, la suite $\{u(X_n)\}$ converge en loi vers la loi uniforme sur $[0, 1[$ lorsque $n \longrightarrow \infty$.

La preuve est similaire à celle du cas général donnée dans [Gauvrit et Delahaye, 2008].

L'idée que ce théorème formalise est la suivante : si X est une variable étalée et régulière, et u une fonction C^1 strictement croissante, $u(X)$ est encore une variable régulière et étalée, pourvu que u soit « lisse » (dans un sens intuitif) et n'écrase pas trop les valeurs à l'infini. On dira qu'une variable aléatoire X ou un ensemble de données suit la loi de Benford générale pour u (ou « est u-Benford ») si $\{u(X)\}$ est uniforme. Si l'intuition selon laquelle la loi de Benford s'explique par les considérations d'étalement et de régularité est juste, la plupart des ensembles de données devraient suivre la loi de Benford générale pour certaines fonctions u.

C'est ce que nous testons dans la suite de l'article, sur des suites mathématiques, puis sur des données réelles et des variables continues d'utilisation courante. Les résultats positifs que nous obtenons attestent fortement que l'explication que nous proposons de la loi de Benford est correcte.

2. SUITES PARTICULIÈRES

Certaines suites mathématiques simples sont connues pour ne pas suivre la loi de Benford classique, par exemple la suite $(\pi n)_{n\in\mathbb{N}}$ (et plus généralement $(an)_{n\in\mathbb{N}}$ pour une constante a telle que $\log(a) \notin \mathbb{Q}$), la suite des nombres premiers [Caldwell, 2006], ou encore la suite $(\sqrt{n})_{n\in\mathbb{N}}$. À l'inverse, les suites $(n^n)_{n\in\mathbb{N}}$, $(n!)_{n\in\mathbb{N}}$ ou $(e^n)_{n\in\mathbb{N}}$ suivent la loi de Benford classique. Dans cette première partie, nous testons expérimentalement ces six suites, pour la loi de Benford générale associée à quatre fonctions u.

2.1. EXPÉRIENCES

Les fonctions utilisées sont les suivantes :

$$\begin{aligned} x &\longmapsto \log\left[\log\left(x\right)\right] \\ x &\longmapsto \log\left(x\right) \\ x &\longmapsto \sqrt{x} \\ x &\longmapsto \pi x^2 \end{aligned}$$

La première fonction, à croissance très lente en l'infini, écrase les valeurs. On s'attend à ce qu'elle donne de moins bons résultats que les autres. La seconde permet de tester la loi de Benford classique. Le coefficient π de la dernière fonction sert à compenser le fait que dans la plupart des données réelles, les valeurs sont entières, et donc les parties fractionnaires de x^2 nulles.

Les résultats de cette expérience sont donnés dans la Table 1.

v_n	$\log\circ\log\left(v_n\right)$	$\log\left(v_n\right)$	$\sqrt{v_n}$	πv_n^2
$\sqrt{n}$ $(N = 10\ 000)$	$68,90\ (0,000)$	$45,90\ (0,000)$	$4,94\ (0,000)$	$0,02\ (1,000)$
πn $(N = 10\ 000)$	$44,08\ (0,000)$	$26,05\ (0,000)$	$0,19\ (1,000)$	$0,80\ (0,544)$
p_n $(N = 10\ 000)$	$53,92\ (0,000)$	$22,01\ (0,000)$	$0,44\ (0,990)$	$0,69\ (0,719)$
e^n $(N = 1\ 000)$	$6,91\ (0,000)$	$0,76\ (1,000)$	$0,63\ (0,815)$	$0,79\ (0,560)$
$n!$ $(N = 1\ 000)^{(*)}$	$7,39\ (0,000)$	$0,58\ (0,887)$	$0,61\ (0,844)$	$0,90\ (0,387)$
n^n $(N = 1\ 000)^{(*)}$	$7,45\ (0,000)$	$0,80\ (0,543)$	$16,32\ (0,000)$	$0,74\ (0,646)$

TAB. 1. Résultats des tests de Kolmogorov-Smirnov appliqués à la partie fractionnaire de $u\left(v_n\right)$ pour quatre fonctions u (présentées en en-tête de colonne) et six suites présentées en en-tête de ligne. Chaque suite est testée en prenant pour n les valeurs de 1 à N. Les valeurs indiquées dans les cellules sont les z de Kolmogorov-Smirnov, qui mesurent l'écart à l'uniformité de la partie fractionnaire de $u\left(v_n\right)$. Les valeurs entre parenthèses sont les significations p associées

(*) Pour les suites (n^n) et $(n!)$ et la fonction $\log\circ\log$, la valeur $n = 1$ n'est pas prise en compte (car $\log\left(\log\left(u_1\right)\right)$ n'est pas définie).

Parmi les six suites utilisées, trois sont à croissance rapide (elles convergent vers l'infini plus vite que tout polynôme). Ce sont les trois suites qui vérifient la loi de Benford classique. Pour ces trois suites, des difficultés de calculs dues à la taille des nombres apparaissent rapidement, ce qui explique que nous n'utilisons ici que les 1000 premiers termes de ces suites.

Aucune des six suites considérées ne suit la loi de Benford générale pour la fonction $\log\circ\log$. Ceci provient, dans notre interprétation, du fait que $\log\circ\log$ écrase les valeurs à l'infini. Des suites à convergence encore plus rapide, comme $10^{\exp(n)}$ conviennent. En effet, $\log\circ\log\left(10^{\exp(n)}\right) = \log\left(e^n\right)$ dont la partie fractionnaire est uniforme, puisque (e^n) suit la loi de Benford classique.

Trois des six suites suivent la loi de Benford classique. Deux des trois suites qui ne suivent pas la loi de Benford classique suivent la loi de Benford générale pour la fonction racine. Une seule des suites log-Benford ne satisfait pas la loi de Benford généralisée pour la fonction racine, mais le cas de cette suite (n^n) est particulier : pour tout n pair, on a en effet $\sqrt{n^n}$ entier, si bien que la partie fractionnaire est nulle une fois sur deux. Si l'on ne tient compte que des nombres impairs et non carrés, la suite des parties fractionnaires n'est plus significativement différente de la loi uniforme $(z = 0,45, p = 0,987)$.

Enfin, les six suites suivent la loi de Benford générale pour la fonction $x \longmapsto \pi x^2$.

2.2. DISCUSSION

Dans la table 1, les colonnes vont de la fonction u à croissance la plus lente $\log \circ \log$, à la fonction à croissance la plus rapide $x \longmapsto \pi x^2$. Les lignes sont également rangées de la suite tendant la plus lentement vers ∞ à celle qui croît le plus rapidement.

Si l'on néglige le cas particulier de $\sqrt{n^n}$, on constate que la vitesse de convergence (vers ∞) de $u(v_n)$ détermine complètement le caractère u-Benford de (v_n) : si $(u(v_n))$ diverge plus vite que $\sqrt{n}$, (v_n) suit la loi de Benford pour u. Si $u(v_n)$ croît plus lentement que $\ln(n)$, (v_n) ne suit pas la loi de Benford. Remarquons d'abord que ce critère empirique ne saurait être vrai tout le temps. On peut en effet, sans modifier la partie fractionnaire de $u(v_n)$, imposer la vitesse de convergence voulue, en y ajoutant une suite à valeur dans $\mathbb{Z}$.

Néanmoins, les cas présentés nous poussent à considérer plus attentivement des suites de la forme $f(n)$ où f est une fonction de référence croissante concave tendant vers ∞. Nous cherchons une condition pour que $(\{f(n)\})_n$ soit, à la limite, uniforme. Une idée intuitive qui explique le phénomène observé est alors la suivante : il est possible de définir des « tranches » d'entiers de la forme $[f^{-1}(n), f^{-1}(n+1)-1[\cap\mathbb{N}$ où $f(n)$ conserve une partie entière constante. Une des raisons pour lesquelles $f(n)$ pourrait ne pas suivre la loi de Benford pour id est que ces « tranches », où $(\{f(n)\})$ est croissante, prennent un poids important dans la distribution observée, si bien que la distribution obtenue en ne conservant que les termes $f(0), ..., f(N)$ n'a pas de limite, parce qu'elle dépend de la place de N dans la « tranche » concernée. Il faudrait donc que la taille des tranches représente une partie relative de $f^{-1}(n)$ tendant vers 0, ce qu'on peut exprimer par :

$$\frac{f^{-1}(n) - f^{-1}(n+1)}{f^{-1}(n)} \underset{\infty}{\longrightarrow} 0.$$

Pourvu que la fonction f soit suffisamment régulière et en utilisant le théorème des accroissements finis, cela revient à poser l'hypothèse

$$\frac{(f^{-1})'(x)}{f^{-1}(x)} \underset{\infty}{\longrightarrow} 0,$$

soit encore

$$\left[\ln\left(f^{-1}(x)\right)\right]' \underset{\infty}{\longrightarrow} 0. \qquad \text{(Condition 1)}$$

Les fonctions définies par $f(x) = x^\alpha$ avec $\alpha > 0$ vérifient cette condition. On est donc amené à penser que toute suite de la forme $n^\alpha, \alpha > 0$ aura une partie fractionnaire uniforme à l'infini, sauf cas pathologique $(\alpha \in \mathbb{Q})$. Le cas $\alpha = \frac{1}{\pi}$ donne effectivement une suite de parties fractionnaires non significativement différente de la loi uniforme $(N = 1000,\ z = 1,331,\ p = 0,058)$. La fonction log ne vérifie pas la condition 1, et la suite $(\{\log(n)\})$ est effectivement différente de l'uniformité (puisque (n) ne suit pas la loi de Benford classique).

3. DONNÉES CONCRÈTES ET VARIABLES ALÉATOIRES

Trois ensembles de données statistiques ont été testées : les valeurs du Dow Jones, à l'ouverture, mesurées le premier de chaque mois, d'octobre 1928 à novembre 2007 ; les aires en millions de km^2 des différents pays du monde ; enfin, les populations des différents pays du monde, estimées en 2008, en millions d'habitants. Les deux dernières séries proviennent du site de la CIA[3].

La Table 2 montre les résultats d'un test de Kolmogorov-Smirnov sur ces données.

	$\log\circ\log(v_n)$	$\log(v_n)$	$\sqrt{v_n}$	πv_n^2
Dow Jones ($N=950$)	$5,90\ (0,000)$	$5,20\ (0,000)$	$0,75\ (0,635)$	$0,44\ (0,992)$
Aire pays ($N=256$)	$1,94\ (0,001)$	$0,51\ (0,959)$	$0,89\ (0,404)$	$1,88\ (0,002)$
Populations ($N=242$)	$3,39\ (0,000)$	$0,79\ (0,568)$	$0,83\ (0,494)$	$0,423\ (1,994)$

Tab. 2. Résultats des tests de Kolmogorov-Smirnov appliqués à la partie fractionnaire de $u(v_n)$ pour quatre fonctions u (présentées en en-tête de colonne) et trois ensembles de données présentés en en-tête de ligne. "Dow Jones" est la valeur à l'ouverture, relevée en début de mois, d'octobre 1929 à novembre 2007. L'aire des pays est en millions de kilomètres carrées, et la population en millions d'habitants

La table confirme encore l'analyse proposée ici : la loi de Benford classique n'est pas plus souvent vérifiée que la loi correspondante pour la fonction racine ou la fonction $x \longmapsto \pi x^2$. On peut noter que la croissance en l'infini n'est pas le seul critère jouant un rôle, puisque pour la suite des aires des pays, la fonction la plus rapidement croissante, $x \longmapsto \pi x^2$, ne donne pas le résultat attendu. Les trois ensembles de données considérés suivent en revanche la loi de Benford générale pour la fonction racine.

Dans la version exprimée avec la partie fractionnaire, la loi de Benford s'applique de préférence à des variables à densité. Nous testons ici trois lois courantes : la loi uniforme sur $]0,k]$ $(k>0)$, la loi exponentielle de paramètre λ, dont la densité f est donnée par $f(x)=\lambda\exp(-\lambda x)$, et enfin la valeur absolue d'une variable normale centrée. Nous ne testons pas la fonction $\log\circ\log$, qui n'est pas définie partout sur $\mathbb{R}_+^*$, ce qui est problématique ici.

En ce qui concerne la loi uniforme, un résultat connu est qu'elle ne vérifie pas la loi de Benford classique, même à la limite. Pour $k=10^j-1$, on vérifie en effet sans peine que la loi du premier chiffre significatif est uniforme et non logarithmique.

Si X suit une loi uniforme sur $]0,k]$, la densité g de $\sqrt{X}$ est donnée par

$$g(x)=\frac{2x}{k},\ x\in\left]0,\sqrt{k}\right]$$

(et $g(x)=0$ sinon). Il s'agit donc d'une fonction croissante sur $]-\infty,\sqrt{k}]$ et décroissante sur $[\sqrt{k},+\infty[$, dont le maximum est $\frac{2}{\sqrt{k}}$, qui tend vers 0. Le

[3] http ://www.cia.gov/library/publications/the-world-factbook/docs/rankorderguide.html

Théorème 2 s'applique donc, montrant que $\left\{\sqrt{X}\right\}$ tend en loi vers la loi uniforme, lorsque k tend vers l'infini. Enfin, le Théorème 3 conclut que X suit, à la limite, une loi de Benford générale pour la fonction $x \longmapsto \pi x^2$.

THÉORÈME 3. *Si* X *suit une loi uniforme sur* $]0,k]$, *alors* $\{\pi X^2\}$ *converge en loi vers la loi uniforme sur* $[0,1[$ *lorsque* $k \longrightarrow \infty$.

Preuve. Soit X une variable aléatoire de loi uniforme sur $]0,k]$. La densité g de la variable $Y = \pi X^2$ est donnée par

$$g(x) = \frac{1}{2a\sqrt{x}},\ x \in]0,a^2]$$

en posant $a = k\sqrt{\pi}$.

La fonction de répartition G de Y vérifie donc

$$G(x) = \frac{\sqrt{x}}{a},\ x \in]0,a^2].$$

Soit maintenant $\delta \in]0,1[$. Soit P_δ la probabilité que $\{Y\} < \delta$. On a :

$$\sum_{j=0}^{\lfloor a^2-\delta \rfloor} G(j+\delta) - G(j) \le P_\delta \le \sum_{j=0}^{\lfloor a^2-\delta \rfloor +1} G(j+\delta) - G(j)$$

$$\frac{1}{a}\sum_{j=0}^{\lfloor a^2-\delta \rfloor} \sqrt{j+\delta} - \sqrt{j} \le P_\delta \le \frac{1}{a}\sum_{j=0}^{\lfloor a^2-\delta \rfloor +1} \sqrt{j+\delta} - \sqrt{j}$$

où $\lfloor a^2-\delta \rfloor$ désigne la partie entière de $a^2-\delta$. Par concavité de la fonction racine, on peut écrire

$$\sqrt{j+\delta} - \sqrt{j} \ge \frac{\delta}{2\sqrt{j+\delta}}$$

et, pour $j > 0$,

$$\sqrt{j+\delta} - \sqrt{j} \le \frac{\delta}{2\sqrt{j}}.$$

On en déduit

$$\frac{\delta}{2a}\sum_{j=0}^{\lfloor a^2-\delta \rfloor} \frac{1}{\sqrt{j+\delta}} \ \le\ P_\delta \le \frac{1}{a}\left[\sqrt{\delta} + \sum_{1}^{\lfloor a^2-\delta \rfloor +1} \frac{\delta}{2\sqrt{j}}\right]$$

$$\frac{\delta}{2a}\sum_{j=0}^{\lfloor a^2-\delta \rfloor} \frac{1}{\sqrt{j+\delta}} \ \le\ P_\delta \le \frac{\sqrt{\delta}}{a} + \frac{\delta}{2a}\sum_{1}^{\lfloor a^2-\delta \rfloor +1} \frac{1}{\sqrt{j}}$$

La décroissance de la fonction $x \longmapsto \frac{1}{\sqrt{x}}$ permet encore d'écrire

$$\sum_{j=0}^{\lfloor a^2-\delta \rfloor} \frac{1}{\sqrt{j+\delta}} \ge \int_{\delta}^{\lfloor a^2-\delta \rfloor +1+\delta} \frac{1}{\sqrt{t}}dt \ge 2\left[\sqrt{\lfloor a^2-\delta \rfloor + 1 + \delta} - \sqrt{\delta}\right]$$

et

$$\sum_{1}^{\lfloor a^2-\delta \rfloor+1} \frac{1}{\sqrt{j}} \leq \int_0^{\lfloor a^2-\delta \rfloor+1} \frac{1}{\sqrt{t}} dt \leq 2\left[\sqrt{\lfloor a^2-\delta \rfloor+1}\right].$$

Au final, on obtient l'encadrement

$$\frac{\delta}{a}\left[\sqrt{\lfloor a^2-\delta \rfloor+1+\delta}-\sqrt{\delta}\right] \leq P_\delta \leq \frac{\delta}{a}\left[\sqrt{\lfloor a^2-\delta \rfloor+1}\right].$$

Par conséquent, pour tout δ fixé, $\lim_{a\longrightarrow\infty}(P_\delta)=\delta$, et $\{\pi X^2\}$ converge donc en loi vers la loi uniforme sur $[0,1[$. □

En ce qui concerne la loi exponentielle de paramètre λ, elle vérifie bien, à la limite (c'est-à-dire pour $\lambda \longrightarrow 0$), la loi de Benford classique. Une preuve de ce résultat, est donnée par Engel et Leuenberger [2003][4].

Si X suit une loi exponentielle de paramètre $\lambda > 0$, sa racine $\sqrt{X}$ a pour densité :

$$x \longmapsto 2\lambda x \exp\left(-\lambda x^2\right),$$

fonction croissante sur $\left]0, \frac{1}{2\lambda}\right]$, puis décroissante, et de maximum $\exp\left(-\frac{1}{4\lambda}\right)$ dont la limite en 0 est 0. Ainsi, $\left\{\sqrt{X}\right\}$ converge en loi vers la loi uniforme lorsque $\lambda \longrightarrow 0$, en application du Théorème 2. Le théorème suivant permet de conclure quant au cas de la fonction $x \longmapsto \pi x^2$.

THÉORÈME 4. *Si X suit une loi exponentielle de paramètre $\lambda > 0$ (de densité f : $x \longmapsto \lambda\exp(-\lambda x)$), alors la partie fractionnaire de $Y = \pi X^2$ converge en loi vers la loi uniforme lorsque $\lambda \longrightarrow 0$.*

Preuve. Soit X suivant une loi exponentielle de paramètre λ. La variable $Y = \pi X^2$ a une densité g donnée par

$$g(x) = \frac{\mu}{2\sqrt{x}}\exp\left(-\mu\sqrt{x}\right),\ x \geq 0$$

où $\mu = \frac{\lambda}{\sqrt{\pi}}$. La fonction de répartition G de Y vérifie donc, pour tout $x \geq 0$

$$G(x) = 1 - e^{-\mu\sqrt{x}}.$$

Notons P_δ la probabilité que $\{Y\} < \delta$, pour $\delta \in]0,1[$.

On a

$$P_\delta = \sum_{j=0}^{\infty}\left[e^{-\mu\sqrt{j}} - e^{-\mu\sqrt{j+\delta}}\right]$$

[4]Dans un article récent [Gauvrit et Delahaye, 2008], nous donnions, suite à une erreur de notre part, la loi exponentielle comme exemple d'application du Théorème 1. En réalité, ce Théorème ne s'applique pas au cas de la loi exponentielle (le maximum $m(\lambda)$ ne tendant pas vers 0), bien que le résultat final soit correct.

Par convexité de $x \longmapsto \exp\left(-\mu\sqrt{x}\right)$, on a

$$\delta\frac{\mu}{2\sqrt{j+\delta}}e^{-\mu\sqrt{j+\delta}} \leq e^{-\mu\sqrt{j}} - e^{-\mu\sqrt{j+\delta}}$$

pour tout $j \geq 0$, et

$$e^{-\mu\sqrt{j}} - e^{-\mu\sqrt{j+\delta}} \leq \delta\frac{\mu}{2\sqrt{j}}e^{-\mu\sqrt{j}}$$

pour tout $j > 0$. On en déduit l'encadrement

$$\delta\sum_{j=0}^{\infty}\frac{\mu}{2\sqrt{j+\delta}}e^{-\mu\sqrt{j+\delta}} \leq P_\delta \leq 1 - e^{-\mu\sqrt{\delta}} + \delta\sum_{j=1}^{\infty}\frac{\mu}{2\sqrt{j}}e^{-\mu\sqrt{j}}.$$

Par décroissance de la fonction $x \longmapsto \frac{1}{\sqrt{x}}\exp\left(-\mu\sqrt{x}\right)$, on a

$$\begin{aligned}\delta\sum_{j=0}^{\infty}\frac{\mu}{2\sqrt{j+\delta}}e^{-\mu\sqrt{j+\delta}} &\geq \delta\int_{\sqrt{\delta}}^{\infty}\frac{\mu}{2\sqrt{t}}e^{-\mu\sqrt{t}}dt \\ &\geq \delta\left[-e^{-\mu\sqrt{t}}\right]_{\sqrt{\delta}}^{\infty} \\ &= \delta e^{-\mu\sqrt{\delta}}\end{aligned}$$

et

$$\begin{aligned}1 - e^{-\mu\sqrt{\delta}} + \delta\sum_{j=1}^{\infty}\frac{\mu}{2\sqrt{j}}e^{-\mu\sqrt{j}} &\leq 1 - e^{-\mu\sqrt{\delta}} + \delta\int_0^{\infty}\frac{\mu}{2\sqrt{t}}e^{-\mu\sqrt{t}}dt \\ &\leq 1 - e^{-\mu\sqrt{\delta}} + \delta\end{aligned}$$

Les deux expressions tendent vers δ lorsque $\mu \longrightarrow 0$, si bien que $P_\delta \longrightarrow \delta$. Le théorème est prouvé. $\square$

Pour la valeur absolue de la loi de Gauss centrée d'écart type 10 000, nous avons testé un échantillon de 2 000 valeurs selon le même principe que les données concrètes. Les résultats sont donnés dans la Table 3.

	$\log(X)$	$\sqrt{X}$	πX^2
Uniforme	NON	OUI	OUI
Exponentielle	OUI	OUI	OUI
Gauss	$14,49$ $(0,000)$	$0,647$ $(0,797)$	$28,726$ $(0,000)$

TAB. 3. Le tableau indique si les lois uniformes sur $]0,k]$ (avec $k \longrightarrow +\infty$), exponentielles de paramètre λ ($\lambda \longrightarrow 0$) suivent ou non les lois de Benford généralisées pour log, racine, et $x \longmapsto \pi x^2$. Pour les variables uniformes et exponentielles, c'est l'adéquation à la limite ($\lambda \longrightarrow 0,\ k \longrightarrow \infty$) qui est indiquée. Pour la variable gaussienne, un test de Kolmogorov-Smirnov a été effectué sur 2 000 tirages. On donne dans les cellules les valeurs du z ainsi que les significations entre parenthèses. Les valeurs suggèrent que cette variable suit une loi de Benford pour la fonction racine, mais ne suit pas la loi de Benford classique, ni la loi de Benford pour $x \longmapsto \pi x^2$

On voit encore une fois que la fonction racine donne plus souvent des résultats positifs que la fonction log. Ces derniers résultats confirment encore nos hypothèses, à savoir que la fonction log n'est pas particulière, et qu'il ne faut pas chercher une explication de la loi de Benford qui dépende d'elle.

4. CONCLUSION

Parmi les explications avancées pour résoudre le paradoxe de la loi de Benford, celle présentée par Gauvrit et Delahaye [2008] est à la fois la plus élémentaire et la plus générale. Selon cette hypothèse, si bon nombre de suites vérifient la loi de Benford classique, c'est parce que la fonction log est régulière et à croissance suffisamment rapide en l'infini d'une part, et parce que les données usuelles sont régulières et étalées d'autre part. Ces critères sont suffisants, selon les auteurs. Si le Théorème 1 prouve que cette explication est rationnelle, il restait à prouver qu'elle ne dépendait pas de manière profonde de la fonction log intervenant dans la loi de Benford classique. Nous venons de voir que c'est le cas, et que d'autres fonctions conviennent tout à fait. Ainsi, la loi de Benford n'est particulière ni en termes de données (des données sans structures fortes, mais simplement étalées régulièrement, suivent cette loi), ni en terme de fonction log[5].

Il ressort même de nos tests que la fonction racine semblerait être un meilleur candidat que la fonction log et, la *digital analysis* qui applique un test de Benford aux données fiscales pour détecter les fraudes, pourrait être remplacée par l'analyse de l'adéquation à la loi de Benford générale pour la fonction racine (à condition toutefois que cette fonction racine reste discriminante).

L'intuition première était que toute fonction régulière convient si elle n'écrase pas trop les valeurs, autrement dit, croît suffisamment vite en l'infini. L'échec constant de la fonction $\log \circ \log$ semble confirmer cette idée. Les suites que nous avons étudiées laissent penser que la vitesse de croissante de $u(v_n)$ est alors déterminante, sans être le seul critère à prendre en compte, pour savoir si (v_n) est u-Benford. Dans le cas des variables continues, le comportement en 0 de la densité semble avoir une importance.

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.

CALDWELL C., "Does Benford apply to prime numbers ?", 2006, http ://primes.utm.edu/notes/faq/BenfordsLaw.html

ENGEL H.-A., LEUENBERGER C., "Benford's law for exponential random variables", *Statistics and Probability Letters* 63, 2003, p. 361-365.

GAUVRIT N., DELAHAYE J.-P., « Pourquoi la loi de Benford n'est pas mystérieuse », *Mathématiques et Sciences humaines/Mathematics and Social Science* 182, 2008, p. 7-15.

[5] Il reste que la fonction log seule permet une interprétation en termes de premier chiffre significatif, qui explique le caractère pregnant de la loi de Benford classique.

HILL T., "A statistical derivation of the Significant-Digit Law", *Statistical Science* 10(4), 1996, p. 354-363.

HÜRLIMANN W., "Benford's law from 1881 to 2006", http ://arxiv.org/abs/math.ST/0607168

JANVRESSE É., DE LA RUE T., "From uniform distributions to Benford law", *Journal of Applied Probability* 41, 2004, p. 1203-1210.

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

NEWCOMB S., "Note on the frequency of use of the different digits in natural number", *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*, 11 (1), 2008, p. 1-19.

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