Loi de Benford, relations de récurrence et suites équidistribuées. II

Autor(en): **Jolissaint, Paul**

Objekttyp: **Article**

Zeitschrift: **Elemente der Mathematik**

Band (Jahr): **64 (2009)**

PDF erstellt am: **26.04.2024**

Persistenter Link: https://doi.org/10.5169/seals-99147

Elem. Math. 64 (2009) 21 – 36
0013-6018/09/010021-16

Elemente der Mathematik

Loi de Benford, relations de récurrence et suites équidistribuées II

Paul Jolissaint

Paul Jolissaint received his doctoral degree from the University of Geneva in 1987. He is part time lecturer at the Institute of Mathematics of the University of Neuchâtel and part time professor of mathematics and physics at the Lycée cantonal de Porrentruy (Jura, CH). His main interests are operator algebras, group actions and ergodic theory.

1 Introduction

Depuis qu'elle a été présentée par F. Benford[1] en 1938, la loi qui régit la répartition quelque peu inattendue des premiers chiffres significatifs dans un grand nombre d'ensembles de données numériques a donné lieu à de nombreuses études. Cette loi affirme que, étant donné une base $b \geq 2$, pour chaque digit $d \in \{1, \ldots, b-1\}$, la proportion de valeurs dont le premier chiffre significatif est d vaut environ $\log_b(1 + 1/d)$. (Pour une définition plus précise et générale, voir la définition 2.1 ci-dessous et les remarques qui

1. Bien qu'elle ait été observée 57 ans plus tôt par S. Newcomb

Der wohl bekannteste Spezialfall des Benford-Gesetzes besagt, dass in einer zufälligen Folge (a_n) positiver Zahlen die Verteilung der führenden Ziffer im Dezimalsystem dem Gesetz $(\log_{10}(1+d^{-1}), d = 1, \ldots, 9)$ folgt. Eine Verallgemeinerung für Basen $b \geq 2$ lautet

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N : \text{Mantisse}_b(a_n) < t\}|}{N} = \log_b(t).$$

Inzwischen ist bekannt, dass ganze Familien von Folgen dem Benford-Gesetz für ein geeignetes b genügen. In einem früheren Aufsatz bewies der Autor, dass viele durch eine lineare Rekurrenzbeziehung definierte Folgen dazu gehören. Hier wird gezeigt, dass Folgen, die in gewissem Sinne äquivalent zu Folgen der Form $(n^\mu \xi^{Q(n)})$ mit $\mu \in \mathbb{R}, \xi \in (0,1) \cup (1,\infty)$ und einem nicht konstanten Polynom Q sind, ebenfalls dem Benford-Gesetz gehorchen. Bekannte Resultate über gleichverteilte Folgen werden benutzt und detaillierte Beweise derselben werden als Komfort für den Leser bereitgestellt.

la suivent.) On peut par exemple consulter les articles [3], [5] ainsi que [6] qui donnent de bons comptes-rendus de la problématique de la loi de Benford. Récemment, quelques articles ont été consacrés à la loi de Benford pour certaines suites: dans [1], les auteurs montrent que des suites réelles (x_n) définies par $x_{n+1} = T_n(x_n)$ où chaque T_n est une application convenable de $\mathbb{R}$ dans lui-même (ou d'une partie de $\mathbb{R}$ dans elle-même) satisfont la loi de Benford; dans [2], le résultat principal affirme que des suites obtenues par la méthode de Newton la satisfont également. Plus précisément, soit $f : I \to \mathbb{R}$ une fonction analytique réelle, non linéaire, et soit x^* un zéro de f. Pour x_0 dans un voisinage de x^*, considérons la suite $(x_n)_{n\geq 1}$ définie par $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$. Alors, pour presque tout x_0, les suites $(x_n - x^*)_{n\geq 1}$ et $(x_{n+1} - x_n)_{n\geq 1}$ satisfont la loi de Benford par rapport à toute base b.

Dans [7], nous avons considéré des suites de nombres positifs $(a_n)_{n\geq 0}$ qui satisfont une relation de récurrence de la forme $a_{n+q} = c_1 a_{n+q-1} + c_2 a_{n+q-2} + \ldots + c_q a_n$ où les $c_i \in \mathbb{R}$ et $c_q \neq 0$, et pour laquelle le polynôme caractéristique associé $p(x) = x^q - c_1 x^{q-1} - \ldots - c_q$ admet une racine distinguée $\xi > 1$, de multiplicité 1 et telle que $\xi > |\eta|$ pour toute autre racine η de $p(x)$. Nous avons alors montré que la suite $(a_n)_{n\geq 0}$ satisfait la loi de Benford par rapport à la base $b > 2$ si:

(a) $\log_b(\xi)$ est irrationnel,

(b) $\inf\left\{\frac{a_n}{\xi^n} \ : \ n \geq 1\right\} > 0$,

et que le cas échéant, il en est de même de toutes les sous-suites de (a_n) de la forme $(a_{Q(n)})$ pour tout polynôme $Q(x) \in \mathbb{Z}[x]$ tel que $Q(n) > 0$ pour tout $n > 0$. Cela généralise des cas déjà connus tels que la suite de Fibonacci ou la suite $(2^n)_{n\geq 1}$ par rapport à la base 10. La preuve repose sur le fait qu'une telle suite (a_n) s'exprime à l'aide des racines $\xi = \xi_1$, $\xi_2, \ldots, \xi_m$ de $p(x)$ et de leurs multiplicités respectives $\mu_1 = 1, \mu_2, \ldots, \mu_m$:

$$a_n = \alpha \xi^n + \sum_{j=2}^{m} \sum_{k=0}^{\mu_j - 1} \alpha_{j,k} n^k \xi_j^n$$

où α est un réel positif à cause de la condition (b), et où les $\alpha_{j,k}$ dépendent des conditions initiales $a_0, \ldots, a_{q-1}$. Quelques simulations numériques nous ont convaincu que l'hypothèse sur la multiplicité de la racine particulière ξ devait être superflue, et le but initial du présent article était d'établir ce fait. Nous allons démontrer le résultat général suivant (cas particulier du théorème 2.4):

Theorem 1.1 *Si $(a_n)_{n\geq 1}$ est une suite de nombres réels positifs tels qu'il existe $\alpha > 0$, $\mu \in \mathbb{R}$ et $\xi > 0$ tels que*

$$\lim_{x\to+\infty} \frac{a_n}{n^\mu \xi^n} = \alpha$$

et si $b \geq 2$ est un entier tel que $\log_b(\xi) \in \mathbb{R} \setminus \mathbb{Q}$, alors, pour tout polynôme non constant $Q(x)$ à coefficients dans $\mathbb{Z}$ tel que $Q(n) > 0$ pour tout $n > 0$, la suite $(a_{Q(n)})_{n\geq 1}$ satisfait la loi de Benford en base b.

On voit donc que c'est le comportement asymptotique de la suite qui implique le fait qu'elle satisfait la loi de Benford.

À titre d'exemple, si $a_n > 0$ est de la forme

$$a_n = \sum_{j=1}^{m} P_j(Q(n))\xi_j^{Q(n)}$$

où Q est comme dans le théorème, $\xi_1 > |\xi_j|$ pour $2 \leq j \leq m$, $P_j(x) \in \mathbb{R}[x]$ pour tout j et $P_1(x) > 0$ pour tout $x > 0$ assez grand, alors la suite $(a_n)_{n\geq 1}$ satisfait la loi de Benford pour toute base b pour laquelle $\log_b(\xi_1)$ est irrationnel. C'est le cas de toute suite $(a_n)_{n\geq 1} \subset \mathbb{R}_+^*$ qui satisfait une relation de récurrence du type $a_{n+q} = c_1 a_{n+q-1} + c_2 a_{n+q-2} + \ldots + c_q a_n$ et dont le polynôme caractéristique $p(x) = x^q - c_1 x^{q-1} - \ldots - c_q = (x-\xi_1)^{\mu_1} \cdot \ldots \cdot (x-\xi_m)^{\mu_m}$ possède une racine $\xi = \xi_1 > 0$ telle que $\xi > |\xi_j|$ pour tout $2 \leq j \leq m$ et qui satisfait la condition (b) ci-dessus.

La preuve du théorème principal de [7] utilise le théorème de Weyl sur l'équidistributivité des suites $(P(n))_{n\geq 1}$ où P est un polynôme dont un coefficient au moins est irrationnel. Certains lecteurs de [7] ont exprimé le souhait de lire une preuve complète du théorème. C'est pourquoi nous donnons ici une preuve complète des résultats que nous utiliserons sur les suites équidistribuées. La plupart proviennent de [8].

Le paragraphe suivant rappelle les définitions importantes et contient l'énoncé du théorème principal, ainsi que sa preuve, conséquence de théorèmes d'équidistributivité tirés de [8]. Dans la section 3, nous établissons une version probabiliste de la loi de Benford pour les suites indexées par des nombres aléatoires, et les deux dernières sections sont consacrées aux démonstrations des critères d'équidistributivité utilisés ici.

2 Loi de Benford et suites équidistribuées

Dans tout l'article, b désigne un entier supérieur ou égal à 2. Comme la demi-droite réelle $(0, +\infty)$ admet la partition

$$(0, +\infty) = \bigsqcup_{k\in\mathbb{Z}} [b^k, b^{k+1}),$$

tout $x > 0$ s'écrit de façon unique

$$x = M_b(x) \cdot b^{e_b(x)}$$

où $M_b(x) \in [1, b)$ est la *mantisse* de x *en base* b et où $e_b(x) \in \mathbb{Z}$.

Par exemple, en base 10 (dix), on a $M_{10}(\pi) = \pi$, mais aussi $M_{10}(\frac{1}{\sqrt{2}}) = \frac{10}{\sqrt{2}} = 5\sqrt{2}$.

Pour tout nombre réel x, on note $\lfloor x \rfloor$ sa partie entière et $\langle x \rangle = x - \lfloor x \rfloor$ sa partie fractionnaire. La partie entière $\lfloor M_b(x) \rfloor \in \{1, \ldots, b-1\}$ est appelée le *premier chiffre significatif* de x (par rapport à la base b).

La définition ci-dessous, tirée de [1], précise la définition originale: on regarde la répartition des valeurs de la mantisse des nombres au lieu de celle du premier chiffre significatif (cf. [7], définition 1.1), qui, lui, correspond à la partie entière de la mantisse.

Definition 2.1 Soit $(a_n)_{n\geq 1} \subset (0, +\infty)$ et soit b comme ci-dessus. On dit que $(a_n)_{n\geq 1}$ *satisfait la loi de Benford en base b* si, pour tout $t \in [1,b)$, on a:

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N \ : \ M_b(a_n) < t\}|}{N} = \log_b(t).$$

On dit que $(a_n)_{n\geq 1}$ *satisfait la loi de Benford forte* si elle satisfait la loi de Benford en base b pour tout $b \geq 2$.

Remarques

(1) Dans [1], la limite dans la définition 2.1 est remplacée par

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N \ : \ M_b(a_n) \leq t\}|}{N} = \log_b(t)$$

pour tout $t \in [1,b)$. Or, les deux définitions sont équivalentes. Cela suit en effet de la continuité de la fonction $\log_b$ et des inégalité évidentes:

$$\begin{aligned} |\{n \leq N : M_b(a_n) < t\}| &\leq |\{n \leq N : M_b(a_n) \leq t\}| \\ &\leq |\{n \leq N : M_b(a_n) < t + \delta\}| \end{aligned}$$

pour tout $t \in [1,b)$, pour tout $\delta > 0$ tel que $t + \delta < b$ et pour tout entier positif N.

(2) La définition 2.1 implique celle de [7], par exemple, qui est classique: en effet, en notant $d(x) = \lfloor M_b(x) \rfloor$, on a $d(x) = d$ si et seulement si $d \leq M_b(x) < d + 1$. Ainsi, si la suite (a_n) satisfait la définition 2.1, on a pour tout $d \in \{1, \ldots, b-1\}$:

$$\begin{aligned} |\{n \leq N : d(a_n) = d\}| &= |\{n \leq N : d \leq M_b(a_n) < d + 1\}| \\ &= |\{n \leq N : M_b(a_n) < d + 1\}| \\ &\quad - |\{n \leq N : M_b(a_n) < d\}|, \end{aligned}$$

ce qui implique que

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N \ : \ d(a_n) = d\}|}{N} = \log_b(d+1) - \log_b(d) = \log_b\left(1 + \frac{1}{d}\right).$$

(3) Nous ne considérerons ici que des suites de nombres positifs. Néanmoins, on définit également la loi de Benford pour les suites réelles non nécessairement positives de la façon suivante (cf. [1]): $(a_n)_{n\geq 1} \subset \mathbb{R}$ satisfait la loi de Benford en base b si la suite des valeurs absolues $(|a_n|)_{n\geq 1}$ la satisfait.

Nous allons utiliser une caractérisation de la loi de Benford qui s'appuie sur la notion de suite équidistribuée; pour cela, rappelons la définition 1.1 de [8], qui, à l'origine, est due à H. Weyl [9]:

Definition 2.2 Soit $(a_n)_{n\geq 1}$ une suite réelle. On dit qu'elle est *équidistribuée modulo* 1 si, pour tous $0 \leq c < d \leq 1$, on a

$$\lim_{N\to\infty} \frac{|\{1 \leq n \leq N \ : \ c \leq \langle a_n \rangle < d\}|}{N} = d - c.$$

Le résultat suivant permet d'utiliser de nombreux résultats sur les suites équidistribuées. Il est dû à P. Diaconis (théorème 1, p. 74 de [4]), et est basé sur l'observation suivante: pour tout $x > 0$, on a $\log_b(x) = \log_b(M_b(x)) + e_b(x)$, et puisque $\log_b(M_b(x)) \in [0,1)$, il s'ensuit que $\log_b(M_b(x)) = \langle \log_b(x) \rangle$.

Proposition 2.3 *Si $(a_n)_{n\geq 1} \subset (0, +\infty)$ et si $b \geq 2$ est un entier, alors elle satisfait la loi de Benford en base b si et seulement si la suite des logarithmes $(\log_b(a_n))_{n\geq 1}$ est équidistribuée mod* 1.

Voici le résultat principal de l'article; il généralise le théorème 1.2 de [7].

Theorem 2.4 *Soient $\alpha > 0$, $\xi > 0$ et μ des nombres réels, et soit une fonction Q, définie sur $[1, +\infty)$, qui satisfait:*

(a) *il existe un entier $k \geq 1$ et un nombre réel $x_0 \geq 1$ tels que Q est k fois dérivable sur $(x_0, +\infty)$;*

(b) $\lim_{x\to\infty} Q^{(k)}(x)$ *existe et est un nombre rationnel non nul.*

Soit $(a_n)_{n\geq 1} \in (0, +\infty)$ une suite telle que

$$\lim_{n\to\infty} \frac{a_n}{n^\mu \xi^{Q(n)}} = \alpha.$$

Alors pour tout entier $b \geq 2$ tel que $\log_b(\xi) \in \mathbb{R} \setminus \mathbb{Q}$, la suite $(a_n)_{n\geq 1}$ satisfait la loi de Benford en base b. En particulier, si, pour tout entier positif m, ξ^m n'est pas entier, alors $(a_n)_{n\geq 1}$ satisfait la loi de Benford forte.

La preuve repose sur l'exercice 3.7, p. 31 de [8]; nous allons en donner une démonstration à la section 5.

Theorem 2.5 *Soient $k \geq 1$ un entier, $\theta \in \mathbb{R}$ un nombre irrationnel et $f : [1, +\infty) \to \mathbb{R}$ une fonction pour laquelle il existe un nombre $x_0 \geq 1$ tel que f soit k fois dérivable sur $(x_0, +\infty)$. Si $\lim_{x\to\infty} f^{(k)}(x) = \theta$ alors la suite $(f(n))_{n\geq 1}$ est équidistribuée mod* 1.

Preuve du théorème 2.4. On écrit pour tout $n \geq 1$:

$$a_n = n^\mu \xi^{Q(n)} \cdot \frac{a_n}{n^\mu \xi^{Q(n)}}$$

de sorte que $\log_b(a_n) = \mu \log_b(n) + Q(n) \log_b(\xi) + \eta_n$ avec $\eta_n = \log_b\left(\frac{a_n}{n^\mu \xi^{Q(n)}}\right) \to \log_b(\alpha)$ lorsque $n \to \infty$. Puisque la suite (η_n) est convergente, il suffit de démontrer que la suite $(\mu \log_b(n) + \log_b(\xi) Q(n))_{n\geq 1}$ est équidistribuée mod 1 (voir par exemple le lemme 3.1 de [7] ou le théorème 1.2 de [8]). Puisque $\log_b(\xi)$ est irrationnel et que $Q^{(k)}(x) \to q \in \mathbb{Q}^*$, on a, puisque $k \geq 1$,

$$\lim_{x\to\infty} \frac{d^k}{dx^k} \left(\mu \log_b(x) + \log_b(\xi) Q(x)\right)$$
$$= \lim_{x\to\infty} \left(\frac{\mu}{\log(b)} \cdot \frac{(-1)^{k-1}(k-1)!}{x^k} + \log_b(\xi) Q^{(k)}(x)\right) = q \log_b(\xi)$$

qui est irrationnel par hypothèse. Le théorème 2.5 s'applique donc. □

Remarques

(1) Soit $P(x) = a_k x^k + a_{k-1}x^{k-1} + \ldots + a_1 x + a_0$ un polynôme non constant à coefficients réels et tel que a_k soit irrationnel. Le théorème 2.5 implique immédiatement que la suite $(P(n))_{n\geq 1}$ est équidistribuée mod 1 (théorème de Weyl).

(2) Une classe typique de fonctions Q qui satisfont les hypothèses du théorème 2.4 est constituée des fonctions rationnelles $Q = P/R$ où P et R sont des polynômes à coefficients rationnels, $R(x) \neq 0$ pour tout $x \geq 1$, avec $\deg(P) > \deg(R)$ car, par division euclidienne de P par R, il existe deux polynômes q et r à coefficients rationnels tels que $P/R = q + r/R$ et $r = 0$ ou $\deg(r) < \deg(R)$. Si $k \geq 1$ est le degré de q, on a

$$\frac{d^k}{dx^k}\left(q(x) + \frac{r(x)}{R(x)}\right) = q^{(k)}(x) + \frac{r_k(x)}{R_k(x)}$$

pour des polynômes convenables r_k et R_k tels que $\deg(R_k) > \deg(r_k)$, et $q^{(k)}(x) =: q_k$ est un rationnel non nul. Enfin, si P et R sont comme ci-dessus, toute fonction de la forme $Q(x) = P(x)/R(x) + c_1 x^\beta + c_2 \log(x)$, où $\beta \in \mathbb{R}$, par exemple, satisfait encore les hypothèses du théorème.

Une preuve complète du théorème 2.5 sera présentée dans les deux derniers paragraphes où nous allons rappeler les critères de Weyl et le théorème des différences de Van der Corput.

3 Suites indexées par des nombres aléatoires

On considère une suite $(u(n))_{n\geq 1} \subset [0,1]$ qui est équidistribuée mod 1. Pour tout entier m assez grand, on considère également $N < m$ variables aléatoires indépendantes et identiquement distribuées (abrégées désormais *i.i.d.*) $X_{1,m}, \ldots, X_{N,m} : \Omega \to \{1, \ldots, m\}$ telles que, pour tout $1 \leq i \leq N$ et tout $1 \leq k \leq m$,

$$P(X_{i,m} = k) = \frac{1}{m}.$$

Par exemple, on peut prendre $\Omega = \{1, \ldots, m\}^N$ muni de la mesure de probabilité P uniforme:

$$P(\{(\omega_1, \ldots, \omega_N)\}) = \frac{1}{m^N} \quad \forall (\omega_1, \ldots, \omega_N) \in \Omega,$$

et $X_{n,m}$ est la projection sur la n-ième coordonnée. Cela constitue un modèle de tirage de N nombres aléatoires dans $\{1, \ldots, m\}$. On va démontrer:

Proposition 3.1 *Avec les hypothèses ci-dessus, pour tous $0 \leq c < d \leq 1$, la suite de variables aléatoires $(\frac{1}{N}\sum_{n=1}^{N} \chi_{[c,d)}(u(X_{n,m})))_{N,m}$ converge en moyenne vers $d - c$:*

$$\lim_{N\to\infty}\left(\lim_{m\to\infty} E\left(\left|\frac{1}{N}\sum_{n=1}^{N} \chi_{[c,d)}(u(X_{n,m})) - (d-c)\right|\right)\right) = 0.$$

Preuve. Posons $Z_{n,m} = \chi_{[c,d)}(u(X_{n,m}))$ pour tous n et m; ce sont également des variables aléatoires i.i.d. et $Z_{n,m}^2 = Z_{n,m}$ puisqu'elles sont à valeurs dans $\{0,1\}$. On a:

$$\begin{aligned} E(Z_{n,m}) &= \int_\Omega Z_{n,m}(\omega) dP(\omega) \\ &= \sum_{k=1}^{m} \int_{X_{n,m}=k} \chi_{[c,d)}(u(k)) dP(\omega) \\ &= \frac{1}{m} \sum_{k=1}^{m} \chi_{[c,d)}(u(k)) =: \mu_m(c,d). \end{aligned}$$

Par hypothèse, $\mu_m(c,d) \to d-c$ lorsque $m \to \infty$ puisque $(u(n))_{n\geq 1}$ est équidistribuée. Calculons encore la variance de $Z_{n,m}$:

$$\begin{aligned} \mathrm{Var}(Z_{n,m}) &= E((Z_{n,m} - E(Z_{n,m}))^2) \\ &= E(Z_{n,m}^2) - 2E(Z_{n,m}E(Z_{n,m})) + E(Z_{n,m})^2 \\ &= E(Z_{n,m}) - E(Z_{n,m})^2 = \mu_m(c,d) - \mu_m(c,d)^2 \end{aligned}$$

car $Z_{n,m}^2 = Z_{n,m}$. En particulier, comme la moyenne, elle est indépendante de n. On obtient alors, par l'indépendance des $Z_{n,m}$:

$$\mathrm{Var}\Big(\frac{1}{N}\sum_{n=0}^{N} Z_{n,m}\Big) = \frac{1}{N^2}\cdot \mathrm{Var}\Big(\sum_{n} Z_{n,m}\Big) = \frac{1}{N^2} N \cdot \mathrm{Var}(Z_{1,m}) = \frac{\mu_m(c,d) - \mu_m(c,d)^2}{N}$$

qui tend vers 0 lorsque N et m tendent vers ∞. Pour $N < m$ fixés, on a

$$\begin{aligned} & E\Big(\Big|\frac{1}{N}\sum_{n=1}^{N} Z_{n,m} - (d-c)\Big|\Big) \\ &\quad \leq E\Big(\Big|\frac{1}{N}\sum_{n=1}^{N} Z_{n,m} - \mu_m(c,d)\Big|\Big) + |\mu_m(c,d) - (d-c)| \\ &\quad \leq E\Big(\Big(\frac{1}{N}\sum_{n=1}^{N} Z_{n,m} - \mu_m(c,d)\Big)^2\Big)^{1/2} + |\mu_m(c,d) - (d-c)| \\ &\quad = \mathrm{Var}\Big(\frac{1}{N}\sum_{n=0}^{N} Z_{n,m}\Big)^{1/2} + |\mu_m(c,d) - (d-c)| \\ &\quad = \frac{\sqrt{\mu_m(c,d) - \mu_m(c,d)^2}}{\sqrt{N}} + |\mu_m(c,d) - (d-c)| \\ &\quad \leq \frac{1}{\sqrt{N}} + |\mu_m(c,d) - (d-c)| \end{aligned}$$

car $\mu_m(c,d) - \mu_m(c,d)^2 \leq \mu_m(c,d) \leq 1$ pour tout m.

Ainsi, si $\varepsilon > 0$ est fixé, il existe $M > 0$ tel que, pour tout $m \geq M$, on ait $|\mu_m(c,d) - (d-c)| < \varepsilon/2$. En prenant $N > 4/\varepsilon^2$, on obtient

$$E\left(\left|\frac{1}{N}\sum_{n=1}^{N} Z_{n,m} - (d-c)\right|\right) < \varepsilon. \qquad \square$$

Soit maintenant un entier $b \geq 2$ et soit $(a_n)_{n\geq 1} \subset \mathbb{R}_+^*$ une suite telle que la suite des parties fractionnaires $u(n) = \langle \log_b(a_n)\rangle$ soit équidistribuée mod 1. Par la proposition 2.3, la suite $(a_n)_{n\geq 1}$ satisfait la loi de Benford en base b. Si $(X_{n,m})_{1\leq n\leq N}$ est comme dans la proposition précédente, on a:

Proposition 3.2 *Avec les hypothèses ci-dessus, la suite de variables aléatoires $(a_{X_{n,m}})$ satisfait la loi de Benford en moyenne. Plus précisément, on a pour tout $t \in [1,b)$:*

$$\lim_{N\to\infty}\left(\lim_{m\to\infty} E\left(\left|\frac{|\{1\leq n\leq N \,:\, M_b(a_{X_{n,m}}) < t\}|}{N} - \log_b(t)\right|\right)\right) = 0.$$

En particulier, la proposition s'applique lorsque $a_n = \xi^n$ où ξ est un nombre réel supérieur à 1 tel que $\log_b(\xi)$ soit irrationnel. A titre d'exemple, le programme *Mathematica*© ci-dessous illustre le cas de $\xi = 2$ et de $b = 10$ avec $m = 10^7$ et $N = 10^5$. On constate que la loi de Benford est relativement bien vérifiée puisque 30 349 valeurs commencent par 1, 17 655 par 2, etc. et 4 648 par 9.

4 Les critères de Weyl

Dans ce paragraphe, on ne considère que des suites à valeurs dans l'intervalle $[0,1]$. Rappelons qu'une telle suite $(a_n)_{n\geq 1} \subset [0,1]$ est *équidistribuée mod* 1 si, pour tout intervalle $I \subset [0,1]$, on a

$$\lim_{N\to\infty} \frac{|\{1\leq n\leq N \,:\, a_n \in I\}|}{N} \to \ell(I)$$

où $\ell(I)$ désigne la longueur de I. Le critère suivant est dû à H. Weyl [9]; comme indiqué dans l'introduction, nous allons en donner une preuve détaillée.

Theorem 4.1 *Soit $(a_n)_{n\geq 1} \subset [0,1]$. Les trois conditions suivantes sont équivalentes:*

(1) *la suite $(a_n)_{n\geq 1}$ est équidistribuée;*

(2) *pour toute fonction continue f sur $[0,1]$, on a:*

$$S_N(f) := \frac{1}{N}\sum_{n=1}^{N} f(a_n) \to \int_0^1 f(x)dx$$

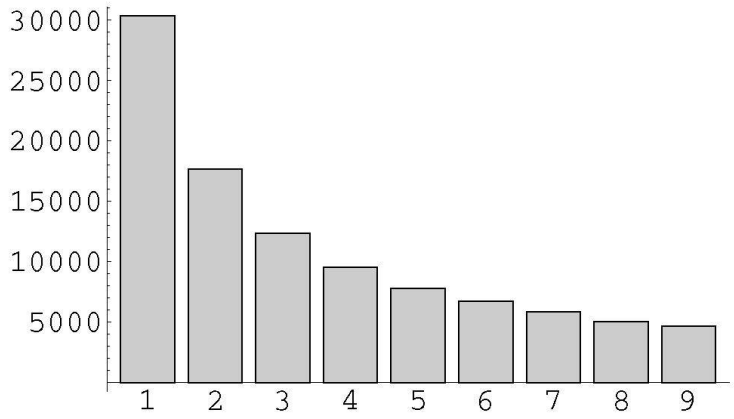
lorsque $N \to \infty$;

(3) *pour tout entier $k \neq 0$, on a:*

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

lorsque $N \to \infty$.

```
<< Statistics`DataManipulation`
<< Graphics`Graphics`
aleaben[n_, m_] := Module[{},
  val1 = Table[Random[Integer, {1, 10^n}], {10^m}];
  val2 = Table[N[FractionalPart[val1 * Log[10, 2]]]];
  freq = RangeCounts[val2, {0.301, 0.477, 0.602, 0.699,
                            0.778, 0.845, 0.903, 0.954}];
  BarChart[freq]]
aleaben[7, 5]
freq
```




```
{30349, 17655, 12387, 9534, 7810, 6712, 5845, 5060, 4648}
```

Fig. 1

Faisons quelques observations qui seront utiles dans la preuve du théorème ci-dessus. Fixons une suite $(a_n)_{n\geq 1} \subset [0,1]$. Pour une famille $\mathcal{F}$ de fonctions intégrables sur $[0,1]$, convenons que $\mathcal{F}$ satisfait la *condition (E)* par rapport à $(a_n)_{n\geq 1}$ si

$$S_N(f) \to \int_0^1 f(x)dx \quad \forall f \in \mathcal{F}$$

lorsque $N \to \infty$, où, rappelons-le, $S_N(f) = \frac{1}{N}\sum_{n=1}^{N} f(a_n)$.

Alors, on observe que les conditions (1), (2) et (3) sont des conditions (E) pour certaines familles de fonctions. En effet,

(a) l'assertion (1) est la condition (E) pour la famille $\mathcal{F}_1$ constituée des fonctions caractéristiques d'intervalles χ_I où I est un intervalle de $[0,1]$; en effet, $S_N(\chi_I) = \frac{1}{N}\sum_n \chi_I(a_n) = |\{1 \leq n \leq N \ : \ a_n \in I\}|/N$;

(b) (2) est la condition (E) pour la famille $\mathcal{F}_2$ de toutes les fonctions continues sur $[0,1]$;

(c) enfin, (3) est la condition (E) pour la famille des fonctions exponentielles complexes $\mathcal{F}_3 = \{e_k \ : \ e_k(x) = e^{2i\pi kx}\}$, car, pour $k \neq 0$, on a $\int_0^1 e_k(x)dx = 0$, et pour $k = 0$, on a $\int_0^1 e_0(x)dx = 1 = S_N(e_0)$ pour tout N. Remarquons que $\mathcal{F}_3 \subset \mathcal{F}_2$.

Le lemme suivant sera utile dans la preuve du théorème:

Lemma 4.2 *Soient $(f_l)_{l\geq 1}$ et f des fonctions intégrables sur $[0,1]$ telles que:*

(a) *la suite $(f_l)_{l\geq 1}$ satisfait la condition (E) par rapport à une suite $(a_n)_{n\geq 1} \subset [0,1]$;*

(b) *$f_l \to f$ uniformément sur $[0,1]$ lorsque $l \to \infty$.*

Alors la fonction f satisfait également la condition (E) par rapport à $(a_n)_{n\geq 1}$.

Preuve. Soit $\varepsilon > 0$ fixé. Il existe $l \geq 1$ tel que

$$\sup_{x\in[0,1]} |f_l(x) - f(x)| < \frac{\varepsilon}{3}.$$

Comme f_l satisfait (E), il existe N_0 tel que, pour tout $N \geq N_0$, on ait:

$$\left|S_N(f_l) - \int_0^1 f_l(x)\,dx\right| < \frac{\varepsilon}{3}.$$

Comme $|f_l - f| < \varepsilon/3$ sur $[0,1]$, on obtient en particulier

$$\begin{aligned} |S_N(f_l) - S_N(f)| &= \frac{1}{N}\left|\sum_{n=1}^{N} (f_l(a_n) - f(a_n))\right| \\ &\leq \frac{1}{N} \cdot N \cdot \sup_{x\in[0,1]} |f_l(x) - f(x)| < \frac{\varepsilon}{3} \end{aligned}$$

pour tout $N \geq N_0$. Enfin, si $N \geq N_0$,

$$\left|S_N(f) - \int_0^1 f\right| \leq |S_N(f) - S_N(f_l)| + \left|S_N(f_l) - \int_0^1 f_l\right| + \int_0^1 |f_l - f| < \varepsilon. \quad \square$$

Preuve du théorème 3.1. (1) $\Rightarrow$ (2): Si $f : [0,1] \to \mathbb{C}$ est continue, elle est limite uniforme de fonctions en escaliers, c'est-à-dire de la forme

$$g = \sum_{k=1}^{m} c_k \chi_{I_k}$$

où $I_k = [\frac{k-1}{m}, \frac{k}{m})$. Par hypothèse, chaque fonction g comme ci-dessus satisfait la condition (E), donc par le lemme, il en est de même de f.

(2) $\Rightarrow$ (3): banal car toute fonction e_k est continue.

(3) $\Rightarrow$ (1): La condition (3) implique que tout polynôme trigonométrique

$$p(x) = \sum_{k=-m}^{q} c_k e^{2i\pi kx}$$

satisfait la condition (E). Or, par le théorème de Weierstrass, toute fonction continue périodique de période 1 est limite uniforme sur $[0,1]$ de tels polynômes. Donc toute fonction continue périodique de période 1 satisfait également (E). C'est le cas pour toute fonction affine par morceaux f comme dans les figures ci-dessous:

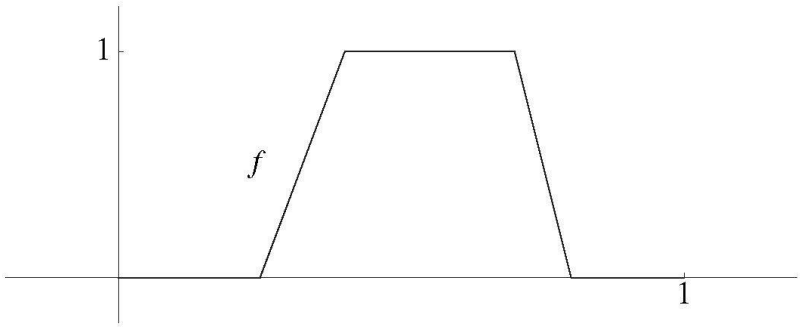


Fig. 2

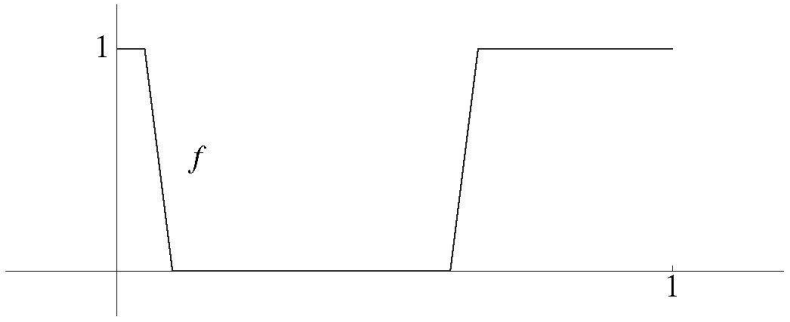


Fig. 3

Fixons alors un intervalle $I \subset [0,1]$ et $\varepsilon > 0$. Choisissons deux fonctions continues affines par morceaux f et g telles que $f(0) = f(1)$, $g(0) = g(1)$, et enfin

$$0 \leq g \leq \chi_I \leq f \leq 1 \quad \text{et} \quad \int_0^1 (f(x) - g(x))dx < \frac{\varepsilon}{5}.$$

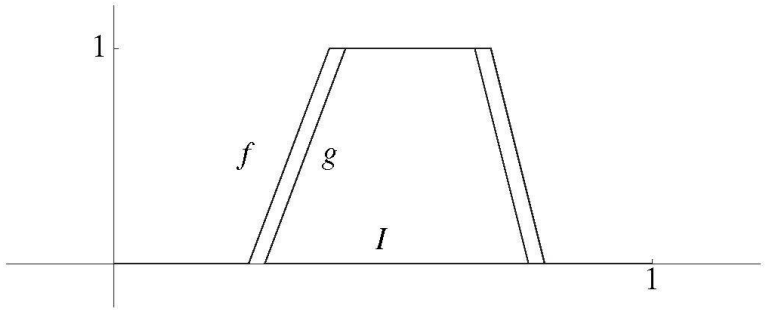


Fig. 4

On a:

$$S_N(g) \leq S_N(\chi_I) \leq S_N(f) \quad \forall N \quad \text{et} \quad \int_0^1 g(x)\,dx \leq \ell(I) \leq \int_0^1 f(x)\,dx.$$

On choisit alors N_0 assez grand pour que

$$\Big|\int_0^1 f(x)\,dx - S_N(f)\Big|, \Big|\int_0^1 g(x)\,dx - S_N(g)\Big| < \frac{\varepsilon}{5} \quad \forall N \geq N_0.$$

On obtient pour tout $N \geq N_0$:

$$\begin{aligned} 0 &\leq S_N(f) - S_N(\chi_I) \leq S_N(f) - S_N(g) \\ &\leq \Big|S_N(f) - \int_0^1 f(x)\,dx\Big| + \int_0^1 (f(x) - g(x))dx + \Big|\int_0^1 g(x)\,dx - S_N(g)\Big| < \frac{3\varepsilon}{5}. \end{aligned}$$

Par suite, on a pour $N \geq N_0$:

$$\begin{aligned} |\ell(I) - S_N(\chi_I)| &\leq \int_0^1 (f(x) - \chi_I(x))dx + \Big|\int_0^1 f(x)\,dx - S_N(f)\Big| + S_N(f) - S_N(\chi_I) \\ &< 2 \cdot \frac{\varepsilon}{5} + \frac{3\varepsilon}{5} = \varepsilon. \end{aligned}$$

Cela démontre le théorème. □

Rappelons un exemple typique de suite équidistribuée mod 1:

Exemple 4.3 Soit $\theta \in \mathbb{R}$ un nombre irrationnel. Alors la suite $(n\theta)_{n\geq 1}$ est équidistribuée mod 1.

Preuve. Il suffit de montrer que la suite des parties fractionnaires $(a_n)_{n\geq 1} := (\langle n\theta \rangle)_{n\geq 1}$ satisfait la condition (3) du théorème. Fixons donc un entier $k \neq 0$ et posons $z = e^{2\pi i k\theta}$. Observons que $z \neq 1$ quel que soit $k \neq 0$ puisque θ est irrationnel.
Comme $e^{2\pi i k a_n} = e^{2\pi i k n\theta} = z^n$ pour tout n, on a:

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i k a_n} = \frac{z}{N}\sum_{n=0}^{N-1} z^n = \frac{1}{N}\frac{z^{N+1} - z}{z - 1} \to 0$$

lorsque $N \to \infty$. □

5 Une preuve du théorème 2.5

Ce dernier paragraphe est consacré à la preuve complète du théorème 2.5. L'un des ingrédients est le *théorème des différences de Van der Corput*, cf. [8], Theorem 3.1, dont nous reproduisons la preuve.

Theorem 5.1 *(Van der Corput) Soit $(x_n)_{n\geq 1}$ une suite de nombres réels telle que, pour tout entier positif h, la suite des différences $(x_{n+h} - x_n)_{n\geq 1}$ soit équidistribuée mod 1. Alors la suite $(x_n)_{n\geq 1}$ est équidistribuée mod 1.*

Preuve. Elle repose essentiellement sur l'inégalité suivante que nous démontrons d'abord (*inégalité fondamentale de van der Corput*, [8], Lemma 3.1): soient $u_1,u_2,\ldots,u_N$ des nombres complexes et $1 \le H \le N$ un entier; alors

$$(\star)\ H^2\Big|\sum_{n=1}^{N} u_n\Big|^2 \le H(N+H-1)\sum_{n=1}^{N}|u_n|^2+2(N+H-1)\sum_{h=1}^{H-1}(H-h)\Re\Big(\sum_{n=1}^{N-h} u_n\bar{u}_{n+h}\Big).$$

On pose $u_n = 0$ pour $n \le 0$ ou $n > N$. On observe que

$$(\star\star)\qquad \sum_{p=1}^{N+H-1}\sum_{h=0}^{H-1} u_{p-h} = H\sum_{n=1}^{N} u_n.$$

En effet, $(\star\star)$ se démontre par récurrence sur $H \ge 1$: c'est banal pour $H = 1$, et si c'est vrai pour H, on a

$$\sum_{p=1}^{N+H}\sum_{h=0}^{H} u_{p-h} = \sum_{p=1}^{N+H-1}\Big(\sum_{h=0}^{H-1} u_{p-h} + u_{p-H}\Big) + \sum_{h=0}^{H} u_{N+H-h}$$
$$= H\sum_{n=1}^{N} u_n + \sum_{p=H+1}^{N+H-1} u_{p-H} + u_N = H\sum_{n=1}^{N} u_n + \sum_{q=1}^{N-1} u_q + u_N = (H+1)\sum_{n=1}^{N} u_n.$$

Cela achève la démonstration de $(\star\star)$.

Démontrons alors $(\star)$. En appliquant l'inégalité de Cauchy-Schwarz à $(\star\star)$, on obtient

$$H^2\Big|\sum_{n=1}^{N} u_n\Big|^2 \le (N+H-1)\sum_{p=1}^{N+H-1}\Big|\sum_{h=0}^{H-1} u_{p-h}\Big|^2$$
$$= (N+H-1)\sum_{p=1}^{N+H-1}\Big(\sum_{r=0}^{H-1} u_{p-r}\Big)\Big(\sum_{s=0}^{H-1}\bar{u}_{p-s}\Big)$$
$$= (N+H-1)\sum_{p=1}^{N+H-1}\sum_{h=0}^{H-1}|u_{p-h}|^2 + 2(N+H-1)\Re\Big(\sum_{p=1}^{N+H-1}\sum_{r,s=0,s<r}^{H-1} u_{p-r}\bar{u}_{p-s}\Big)$$
$$=: (N+H-1)(\Sigma_1 + 2\Re(\Sigma_2)).$$

Par $(\star\star)$, $\Sigma_1 = H\sum_{n=1}^{N}|u_n|^2$. Enfin, tout élément $u_{p-r}\bar{u}_{p-s}$ apparaissant dans Σ_2 est de la forme $u_n\bar{u}_{n+h}$ pour une paire $(n,h) \in \{1,\ldots,N\}\times\{1,\ldots,H-1\}$. Soit alors $A = \{(p,r,s) : 1 \le p \le N+H-1, 0 \le r,s \le H-1, s < r, p-r \le N\}$ et soit $\phi : A \to \{1,\ldots,N\}\times\{1,\ldots,H-1\}$ l'application définie par $\phi(p,r,s) = (p-r,r-s)$. Pour (n,h) fixé, on a $\phi^{-1}(n,h) = \{(n+h+s,h+s,s) : 0 \le s \le H-h-1\}$ car $r = h+s \le H-1$. Ainsi, pour toute paire (n,h), il y a exactement $H-h$ triples $(p,r,s) \in A$ qui satisfont $u_{p-r}\bar{u}_{p-s} = u_n\bar{u}_{n+h}$. On obtient donc $\Sigma_2 = \sum_{h=1}^{H-1}(H-h)\sum_{n=1}^{N-h} u_n\bar{u}_{n+h}$ puisque $u_k = 0$ pour $k > N$. Cela démontre $(\star)$, que nous allons utiliser pour démontrer le théorème 5.1. Soit une suite $(x_n)_{n\ge1}$ qui satisfait les hypothèses du théorème: pour tout entier $h \ge 1$, la

suite $(x_{n+h} - x_n)$ est équidistribuée mod 1. Pour montrer que (x_n) l'est aussi, nous allons utiliser le critère (3) du théorème de Weyl. On fixe donc un entier non nul k, et on pose $u_n = e^{2i\pi k x_n}$. En appliquant ($\star$), on a

$$H^2\Big|\sum_{n=1}^{N} u_n\Big|^2 \leq H(N+H-1)N + 2(N+H-1)\sum_{h=1}^{H-1}(H-h)\Re\Big(\sum_{n=1}^{N-h} u_n\bar{u}_{n+h}\Big)$$
$$\leq H(N+H-1)N + 2(N+H-1)\sum_{h=1}^{H-1}(H-h)\Big|\sum_{n=1}^{N-h} e^{2i\pi k(x_n - x_{n+h})}\Big|$$

pour $1 \leq H \leq N$. En divisant par H^2N^2, on a

$$\Big|\frac{1}{N}\sum_{n=1}^{N} u_n\Big|^2 \leq \frac{N+H-1}{HN}$$
$$+ 2\sum_{h=1}^{H-1}\frac{(N+H-1)(H-h)(N-h)}{H^2N^2}\Big|\frac{1}{N-h}\sum_{n=1}^{N-h} e^{2i\pi k(x_n - x_{n+h})}\Big|.$$

Si $\varepsilon > 0$ est fixé, on choisit d'abord H assez grand pour que $1/H < \varepsilon/3$. Il existe alors $N_0 \geq H$ assez grand pour que, pour tout $N > N_0$, on ait $\frac{N+H-1}{HN} < 2\varepsilon/3$ et

$$2\sum_{h=1}^{H-1}\frac{(N+H-1)(H-h)(N-h)}{H^2N^2}\Big|\frac{1}{N-h}\sum_{n=1}^{N-h} e^{2i\pi k(x_n - x_{n+h})}\Big| < \varepsilon/3$$

puisque chaque suite $(x_{n+h} - x_n)$ est équidistribuée mod 1. Cela achève la démonstration du théorème de Van der Corput. □

La proposition ci-dessous est le premier pas dans la preuve du théorème 2.5; c'est le théorème 3.3 de [8].

Proposition 5.2 *Soit $(x_n)_{n\geq 1} \subset \mathbb{R}$ une suite telle que la suite des différences $\Delta x_n := x_{n+1} - x_n$ tende vers une limite $\theta \in \mathbb{R} \setminus \mathbb{Q}$. Alors $(x_n)_{n\geq 1}$ est équidistribuée mod* 1.

Preuve. On fixe un entier non nul h et on va montrer (critère de H. Weyl) que

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

Soit q un entier positif. Par hypothèse, il existe un entier $M > 0$ assez grand tel que $|\Delta x_N - \theta| \leq \frac{1}{q^2}$ pour tout $N \geq M$. Par suite, on a pour tout $N \geq M$:

$$|x_N - x_M - (N-M)\theta| = |(x_N - x_{N-1}) + (x_{N-1} - x_{N-2}) + \ldots +$$
$$+ (x_{M+1} - x_M) - (N-M)\theta|$$
$$= \Big|\sum_{j=M}^{N-1}(\Delta x_j - \theta)\Big| \leq \frac{N-M}{q^2}.$$

Par le théorème des accroissements finis, on a pour tous $u,v \in \mathbb{R}$, $|e^{2i\pi u} - e^{2i\pi v}| \leq 2\pi|u-v|$, donc

$$\begin{aligned}|e^{2i\pi hx_N} - e^{2i\pi h(x_M+(N-M)\theta)}| &\leq 2\pi|h||x_N - x_M - (N-M)\theta| \\ &\leq \frac{2\pi|h|(N-M)}{q^2}\end{aligned}$$

et ainsi

$$\begin{aligned}\Big|\sum_{n=M}^{M+q-1} e^{2i\pi hx_n}\Big| &= \Big|\sum_{n=M}^{M+q-1} (e^{2i\pi hx_n} - e^{2i\pi h(x_M+(n-M)\theta)} + \\ &\quad + \sum_{n=M}^{M+q-1} e^{2i\pi h(x_M+(n-M)\theta)}\Big| \\ &\leq \Big|\sum_{n=M}^{M+q-1} e^{2i\pi hx_M}\cdot(e^{2i\pi h\theta})^{n-M}\Big| + \frac{2\pi|h|}{q^2}\sum_{n=M}^{M+q-1}(n-M).\end{aligned}$$

Or, on a

$$\Big|\sum_{n=M}^{M+q-1} e^{2i\pi hx_M}\cdot(e^{2i\pi h\theta})^{n-M}\Big| = \Big|\sum_{n=M}^{M+q-1}(e^{2i\pi h\theta})^{n-M}\Big| \leq \frac{2}{|e^{2i\pi h\theta}-1|} = \frac{1}{\sin(\pi|h|\theta)}$$

et

$$\sum_{n=M}^{M+q-1}(n-M) = \sum_{k=0}^{q-1} k \leq \frac{q^2}{2}.$$

Par conséquent, on obtient

$$\Big|\sum_{n=M}^{M+q-1} e^{2i\pi hx_n}\Big| \leq \frac{1}{\sin(\pi|h|\theta)} + \frac{2\pi|h|}{q^2}\cdot\frac{q^2}{2} =: K.$$

Ainsi, on a pour tout entier positif H:

$$\Big|\sum_{n=M}^{M+Hq-1} e^{2i\pi hx_n}\Big| \leq H\cdot K,$$

et finalement, si $N \geq M$,

$$\Big|\frac{1}{N}\sum_{n=1}^{N} e^{2i\pi hx_n}\Big| \leq \frac{M-1}{N} + \frac{N-M}{qN}K + \frac{q}{N}$$

qui peut être rendu arbitrairement proche de K/q, et comme q peut être aussi grand qu'on veut, on a le résultat. □

Preuve du théorème 2.5. Rappelons que $f(x)$ est définie pour $x \geq 1$, qu'elle est k fois dérivable pour x assez grand et que

$$\lim_{x\to\infty} f^{(k)}(x) = \theta \in \mathbb{R} \setminus \mathbb{Q}.$$

On procède par récurrence sur k. Pour $k = 1$, on a $\Delta f(n) = f(n+1) - f(n) = f'(\alpha_n)$ pour $n < \alpha_n < n+1$ convenable. Ainsi, $\Delta f(n) \to \theta$ lorsque $n \to \infty$. On applique alors la proposition précédente.

Supposons le théorème vrai pour k et soit une fonction f telle que

$$\lim_{x\to\infty} f^{(k+1)}(x) = \theta \in \mathbb{R} \setminus \mathbb{Q}.$$

Par le théorème de Van der Corput, il suffit de montrer que pour tout entier $h \geq 1$ fixé, la suite $(f(n+h) - f(n))_{n\geq 1}$ est équidistribuée mod 1. Posons donc, pour h fixé, $g_h(x) := f(x+h) - f(x)$. On a

$$g_h^{(k)}(x) = f^{(k)}(x+h) - f^{(k)}(x) = f^{(k+1)}(\xi_{x,h})$$

pour $x < \xi_{x,h} < x + h$ convenable, et ainsi

$$\lim_{x\to\infty} g_h^{(k)}(x) = \lim_{x\to\infty} = f^{(k+1)}(\xi_{x,h}) = \theta.$$

Par hypothèse de récurrence, la suite $(g_h(n))_{n\geq 1}$ est équidistribuée mod 1, ce qui achève la preuve du théorème. □

Références

[1] Berger, A.; Bunimovich, L.A.; Hill, T.P.: One-dimensional dynamical systems and Benford's law. *Trans. Amer. Math. Soc.* 357 (2005), 197–219.

[2] Berger, A.; Hill, T.P.: Newton's method obeys Benford's law. *Amer. Math. Monthly* 114 (2007), 588–601.

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

[4] Diaconis, P.: The distribution of leading digits and uniform distribution mod 1. *Ann. Probab.* 5 (1977), 72–81.

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

[6] Hungerbühler, N.: Benfords Gesetz über führende Ziffern. EducETH, March 2007, `www.educ.ethz.ch`

[7] Jolissaint, P.: Loi de Benford, relations de récurrence et suites équidistribuées. *Elem. Math.* 60 (2005), 10–18.

[8] Kuipers, L.; Niederreiter, H.: *Uniform distribution of sequences*. Dover Publications, Inc., Mineola, New York 2006.

[9] Weyl, H.: Über die Gleichverteilung von Zahlen mod 1. *Math. Ann.* 77 (1916), 313–352.

Paul Jolissaint
Institut de Mathématiques
Université de Neuchâtel
Rue Emile-Argand 11
CH–2000 Neuchâtel
e-mail: `paul.jolissaint@unine.ch`