

La loi de Benford et les séries de Fourier

Mylène KOEHLIN David M.

Mémoire de maîtrise - 2003

Nous tenons tout d'abord à remercier Monsieur S. pour l'attrait des thèmes proposés, sa disponibilité, ses conseils constructifs et notre initiation à \LaTeX ainsi que Madame Q. pour nous avoir attribué ce sujet intéressant et actuel.

Un peu d'histoire...

Avant l'époque des calculatrices, les calculs se faisaient à la main, à l'aide de tables. Un jour de 1881, un astronome américain, Simon Newcomb, s'aperçut que les premières pages d'une table de logarithmes étaient plus usées que les autres. Se pouvait-il que les données recherchées dans cette table commençaient plus souvent par le chiffre "1" et que la distribution de leurs premiers chiffres significatifs ne soit pas uniforme ? Les résultats de son observation le conduisent à émettre la formule suivante :

$$\forall n = 1, \dots, 9 \quad \mathbf{P}(\text{1}^{\text{er}} \text{ chiffre significatif} = n) = \log \left(1 + \frac{1}{n} \right).$$

A l'époque, ceci ne convainquit personne.

Cinquante ans plus tard, vers 1938, un physicien américain, Frank Benford, redécouvrit les mêmes fréquences que celles résultant de l'application de la formule de Newcomb en répertoriant plus de 20 000 données sélectionnées dans des domaines aussi divers que les longueurs de plus de 300 fleuves, les recensements démographiques de plus de 3 000 régions, les masses atomiques des éléments chimiques, les cours de bourse, les constantes de la physique, etc. (cf. annexe A). Depuis, cette loi sur la fréquence d'apparition du premier chiffre significatif, s'appelle la loi de Benford.

Ce n'est, cependant, qu'en 1996 que Terence Hill démontra mathématiquement cette loi.

1 Introduction

La loi de Benford a été émise à la suite d'observations sur les données mathématiques provenant de phénomènes naturels. Traduire mathématiquement ces phénomènes, revient à répéter successivement une expérience avec les mêmes conditions initiales, c'est-à-dire de considérer des variables aléatoires indépendantes identiquement distribuées. Dans ce mémoire, nous supposerons de plus que ces variables sont à densité sur $L^2[0, 1]$.

Sous ces hypothèses, différentes démonstrations de la loi de Benford ont été établies depuis 1996. L'une d'entre elles considère l'approche par les coefficients de Fourier. Or, on ne peut utiliser cet outil que sur des fonctions à densité. Il nous suffit donc de travailler sur la mantisse de la variable aléatoire plutôt que sur son premier chiffre significatif. Nous obtiendrons donc un résultat bien plus "fin" que le résultat attendu.

Le préliminaire comporte les outils clés de notre mémoire. Tout d'abord, nous allons démontrer que si la mantisse suit la loi logarithmique, alors le premier chiffre significatif suit la loi de Benford. Ensuite, nous démontrerons un résultat pratique : si le logarithme d'une mantisse suit la loi uniforme sur $[0, 1]$, alors la mantisse suit la loi logarithmique.

Puis dans une seconde partie, après un rappel nécessaire sur les coefficients de Fourier, nous allons nous intéresser à la mantisse d'un produit de variables aléatoires. Notre but, dans un premier temps, est de démontrer la convergence de la loi d'une somme de variables aléatoires vers la loi uniforme sur $[0, 1]$. Celle-ci nous permettra, ensuite, grâce au travail préliminaire, de démontrer la convergence de la loi de la mantisse du produit vers la loi logarithmique. Le passage au modulo 1 est nécessaire pour définir les opérations sur les mantisses. En effet, nos variables aléatoires sont des mantisses définies sur $[1, 10[$, donc en prenant leur logarithme, nous sommes ramenés sur $[0, 1[$.

La partie suivante, qui traite du quotient et des puissances des variables aléatoires, est construite suivant la même stratégie que pour le produit. Dans la dernière partie, nous finirons par des exemples et des simulations.

2 Préliminaires

Soit $X = m_X \cdot 10^{n_X}$ où $1 \leq m_X < 10$ et $n_X \in \mathbb{Z}$.

On note $p_X = \lfloor m_X \rfloor$ la partie entière de m_X . Il s'agit donc du premier chiffre significatif de X .

Dans ce texte, on note \log le logarithme décimal et \ln le logarithme népérien.

Les convergences que nous cherchons à prouver dans ce mémoire sont des convergences en loi. Néanmoins, nous établirons des convergences en variation. En fait, nous aurons même un peu mieux car nous travaillons dans L^2 . Mais comme la convergence en variation implique la convergence au sens L^1 qui, elle-même, implique celle en loi, alors nous aurons le résultat attendu.

Proposition 1. *Si m_X suit la loi logarithmique, c'est-à-dire si sa fonction de répartition est égale à $\mathbf{P}(m_X \leq x) = \log x$, $\forall x \in [1, 10]$, alors : $\forall n \in \{1, \dots, 9\}$, $\mathbf{P}(p_X = n) = \log(n+1) - \log n$ (Loi de Benford).*

Démonstration.

$$\begin{aligned} \forall n \in \{1, \dots, 9\}, \quad \mathbf{P}(p_X = n) &= \mathbf{P}(n \leq m_X < n+1) \\ &= \mathbf{P}(m_X < n+1) - \mathbf{P}(m_X < n) \end{aligned}$$

Comme la fonction de répartition de m_X est continue,

$$\mathbf{P}(m_X < n) = \mathbf{P}(m_X \leq n) = \log n.$$

D'où :

$$\begin{aligned} \forall n \in \{1, \dots, 9\}, \quad \mathbf{P}(p_X = n) &= \mathbf{P}(m_X \leq n+1) - \mathbf{P}(m_X \leq n) \\ &= \log(n+1) - \log n \end{aligned}$$

□

Remarque 2. Dans la suite de ce mémoire, il nous suffira donc de démontrer que la loi de la mantisse de la multiplication de variables aléatoires indépendantes identiquement distribuées converge vers la loi logarithmique.

Proposition 3. *m_X suit la loi logarithmique de fonction de répartition $\mathbf{P}(m_X \leq x) = \log x$ $\forall x \in [1, 10]$, si et seulement si $\log m_X$ suit la loi uniforme sur $[0, 1]$.*

Démonstration. – Preuve de (\Rightarrow) :

$$\forall x \in [0, 1], \quad \mathbf{P}(\log m_X \leq x) = \mathbf{P}(\ln m_X \leq x \ln 10) \quad \text{car } \ln 10 > 0.$$

Puisque la fonction $x \mapsto 10^x$ est continue et strictement croissante :

$$\begin{aligned} \forall x \in [0, 1], \quad \mathbf{P}(\log m_X \leq x) &= \mathbf{P}(m_X \leq 10^x) \\ &= \log(10^x) \quad (\text{par hypothèse}) \\ &= x \end{aligned}$$

On reconnaît la fonction de répartition de la loi uniforme sur $[0, 1]$.
Donc $\log m_X$ suit la loi uniforme sur $[0, 1]$.

– Preuve de (\Leftarrow) :

Comme la fonction logarithmique est continue et strictement croissante, alors :

$$\begin{aligned} \mathbf{P}(m_X \leq y) &= \mathbf{P}(\log m_X \leq \log y) \\ &= \int_{]-\infty, \log y]} \mathbf{1}_{[0,1]}(x) \, dx \quad (\text{par hypothèse}) \\ &= \begin{cases} 0 & \text{si }]-\infty, \log y] \cap [0, 1] = \emptyset \\ \int_{[0, \log y]} dx & \text{si } 0 \leq \log y < 1 \\ \int_{[0,1]} dx & \text{si } \log y \geq 1 \end{cases} \end{aligned}$$

Or : $0 \leq \log y \leq 1 \Leftrightarrow 0 \leq \frac{\ln y}{\ln 10} \leq 1 \Leftrightarrow 0 \leq \ln y \leq \ln 10 \Leftrightarrow 1 \leq y \leq 10$.
D'où :

$$\mathbf{P}(m_X \leq y) = \begin{cases} 0 = \log 1 & \text{si } y \leq 1 \\ \log y & \text{si } 1 \leq y < 10 \\ 1 = \log 10 & \text{si } y \geq 10 \end{cases}$$

□

Remarque 4. Les théorèmes 7, 12, 14 aboutissent à une convergence vers la loi uniforme sur $[0, 1]$. Cette proposition nous permettra de revenir à notre problème, c'est à dire à une convergence vers la loi logarithmique sur $[1, 10]$.

3 Loi de Benford sur la multiplication de variables aléatoires

3.1 Résultats connus sur les coefficients de Fourier

Puisque nous sommes amenés à travailler modulo 1, les densités sont définies sur tout \mathbb{R} et sont périodiques de période 1. Les coefficients de Fourier

d'une fonction f sont alors définis par $\widehat{f}(n) = \int_0^1 f(t)e^{-2\pi int} dt, \forall n \in \mathbb{N}$.

Lemme 5. *Si $f \in L^2[0, 1]$ est une densité de probabilité, alors :*

1. $\widehat{f}(0) = 1$
2. $\forall n \neq 0, |\widehat{f}(n)| < 1$
3. $\sup_{n \neq 0} |\widehat{f}(n)| < 1$.

Démonstration. 1. Puisque f est une densité de probabilité, alors :

$$\widehat{f}(0) = \int_0^1 f(t)e^0 dt = \int_0^1 f(t) dt = 1$$

2.

$$|\widehat{f}(n)| = \left| \int_0^1 f(t)e^{-2\pi int} dt \right| \leq \int_0^1 f(t)|e^{-2\pi int}| dt = 1$$

car $|e^{-2\pi int}| = 1$ et f est une densité de probabilité.

Démontrons maintenant l'inégalité stricte :

$\int_0^1 f(t)e^{-2\pi int} dt \in \mathbb{C}$, donc il existe $\rho \in \mathbb{R}_+$ et $\alpha \in \mathbb{R}$ tels que :

$$\int_0^1 f(t)e^{-2\pi int} dt = \rho e^{i\alpha}.$$

Ainsi, $\int_0^1 f(t)e^{-i(\alpha+2\pi nt)} dt = \rho \in \mathbb{R}_+$. D'où :

$$\begin{aligned} |\widehat{f}(n)| &= \left| \int_0^1 f(t)e^{-2i\pi nt} dt \right| = |\rho e^{i\alpha}| = \rho \\ &= \int_0^1 f(t)e^{-i(\alpha+2\pi nt)} dt \\ &= \int_0^1 f(t) \cos(\alpha + 2\pi nt) dt - i \int_0^1 f(t) \sin(\alpha + 2\pi nt) dt \\ &= \int_0^1 f(t) \cos(\alpha + 2\pi nt) dt \quad \text{car } \rho \in \mathbb{R}. \end{aligned}$$

Soit $E_n = \{t \in [0, 1] \mid \cos(\alpha + 2\pi nt) < 1\}$ et soit μ la mesure de densité f .

On a ainsi $\mu(E_n) = \int_{E_n} f(t) dt$.

Posons $E_{n,k} = \{t \in [0, 1] \mid \cos(\alpha + 2\pi nt) \leq 1 - \frac{1}{k}\}$.

Il est clair que $E_n = \bigcup_{k \geq 1} E_{n,k}$ et que $E_{n,k} \uparrow E_n$ ($k \rightarrow \infty$).

Ainsi, $\mu(E_{n,k}) \uparrow \mu(E_n)$ ($k \rightarrow \infty$).

- (a) Si $\mu(E_n) = \int_{E_n} f(t) dt > 0$.
Alors il existe $k_0 \in \mathbb{N}$ tel que $\mu(E_{n,k_0}) > 0$.

$$\begin{aligned}
|\widehat{f}(n)| &= \left| \int_{[0,1]} f(t) e^{-2i\pi nt} dt \right| \\
&= \int_{[0,1]} f(t) \cos(\alpha + 2\pi nt) dt \\
&= \int_{E_{n,k_0}} f(t) \cos(\alpha + 2\pi nt) dt + \int_{[0,1] \setminus E_{n,k_0}} f(t) \cos(\alpha + 2\pi nt) dt \\
&\leq \left(1 - \frac{1}{k_0}\right) \int_{E_{n,k_0}} f(t) dt + \int_{[0,1] \setminus E_{n,k_0}} f(t) dt \\
&= \int_{[0,1]} f(t) dt - \frac{1}{k_0} \int_{E_{n,k_0}} f(t) dt \\
&= 1 - \frac{1}{k_0} \int_{E_{n,k_0}} f(t) dt \\
&< 1 \quad \text{car } \mu(E_{n,k_0}) > 0.
\end{aligned}$$

- (b) Si $\mu(E_n) = \int_{E_n} f(t) dt = 0$.
Alors :

$$\cos(\alpha + 2\pi nX) = 0 \text{ p.s.} \quad (1)$$

Or $\alpha + 2\pi nx = 2k\pi \Leftrightarrow x = \frac{k}{n} - \frac{\alpha}{2\pi n}$.

Soit donc $A_n = \left\{ \frac{k}{n} - \frac{\alpha}{2\pi n} \mid k \in \mathbb{Z} \right\}$ pour tout $n \neq 0$.

Alors d'après (1), on a : $\mathbf{P}(X \in A_n) = 1$ p.s..

D'autre part, $\mathbf{P}(X \in A_n) = \mathbf{P}_X(A_n) = \mu(A_n) = \int_{A_n} f(t) dt = 0$ p.s. car A_n est un ensemble discret tandis que μ est une mesure à densité.

Nous avons une contradiction donc : $\forall n \neq 0, \mu(E_n) \neq 0$.

Ainsi, $\forall n \neq 0, |\widehat{f}(n)| < 1$.

3. Nous avons besoin de la formule de Parseval :

$$\int_0^1 f^2(t) dt = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2 \quad (2)$$

Nous savons que $\widehat{f}(-n) = \overline{\widehat{f}(n)}$. Ainsi, d'après (2) :

$$\int_0^1 f^2(t) dt = \sum_{n=0}^{+\infty} |\widehat{f}(n)|^2 + \sum_{n=0}^{+\infty} |\widehat{f}(n)|^2.$$

Or $\int_0^1 f^2(t) dt < +\infty$ car $f \in L^2[0, 1]$. Ainsi,

$$\sum_{n=0}^{+\infty} |\widehat{f}(n)|^2 + \sum_{n=0}^{+\infty} |\widehat{f}(n)|^2 < +\infty$$

et comme les deux séries sont à termes positifs, alors elles convergent toutes les deux et les termes généraux tendent vers 0.

D'où $\lim_{n \rightarrow \infty} |\widehat{f}(n)| = 0$.

En outre,

$$\lim_{n \rightarrow \infty} |\widehat{f}(n)| = 0 \Leftrightarrow \forall \epsilon > 0, \quad \exists N(\epsilon) \in \mathbb{N}^*, \quad \forall n > N(\epsilon) \quad |\widehat{f}(n)| < \epsilon$$

Soit $\epsilon < 1$ fixé. On sait alors qu'il y a au plus $N(\epsilon)$ valeurs de n telles que $\epsilon \leq |\widehat{f}(n)| < 1$ car d'après le 2., $|\widehat{f}(n)| < 1$ pour tout $n \neq 0$. $N(\epsilon)$ est fini, donc

$$\sup_{n \in \{1, \dots, N(\epsilon)\}} |\widehat{f}(n)| = \max_{n \in \{1, \dots, N(\epsilon)\}} |\widehat{f}(n)| < 1.$$

D'où

$$\sup_{n \neq 0} |\widehat{f}(n)| < 1.$$

□

Nous rappelons ici que nous travaillons modulo 1 et que les densités sont définies sur tout \mathbb{R} et sont périodiques de période 1. Ainsi,

$$\forall f \in L^2(\mathbb{R}), \quad \forall u \in \mathbb{R}, \quad \int_{-u}^{1-u} g(x) dx = \int_0^1 g(x) dx = 1.$$

Soient maintenant X et Y des variables aléatoires indépendantes distribuées sur $[0, 1]$ de densités respectives f et g appartenant à $L^2[0, 1]$. Si $Z = X + Y \pmod{1}$ a pour densité h , alors :

$$\forall t \in [0, 1], \quad h(t) = \int_0^1 f(u)g(t-u) du = f * g \quad (3)$$

Il faut remarquer que $g(t-u) = g(1+t-u)$ si $t-u < 0$ car g est définie sur \mathbb{R} et périodique de période 1. Il s'agit du produit de convolution de f et g sur $[0, 1]$.

Lemme 6. *Sous les hypothèses précédentes, on a : $\widehat{h}(n) = \widehat{f}(n)\widehat{g}(n)$.*

Démonstration.

$$\begin{aligned}
\widehat{h}(n) &= \int_0^1 h(t) e^{-2\pi i n t} dt \\
&= \int_0^1 \left(\int_0^1 f(u) g(t-u) du \right) e^{-2\pi i n t} dt \quad \text{d'après (3)} \\
&= \int_0^1 \left(\int_0^1 f(u) g(t-u) e^{-2\pi i n (t-u)} e^{-2\pi i n u} du \right) dt \\
&= \int_0^1 f(u) e^{-2\pi i n u} \left(\int_0^1 g(t-u) e^{-2\pi i n (t-u)} dt \right) du
\end{aligned}$$

Par un changement de variable $x = t - u$, on a alors :

$$\widehat{h}(n) = \int_0^1 f(u) e^{-2\pi i n u} \left(\int_{-u}^{1-u} g(x) e^{-2\pi i n x} dx \right) du$$

$x \mapsto e^{-2\pi i n x}$ et g sont périodiques de période 1. D'où :

$$\begin{aligned}
\widehat{h}(n) &= \int_0^1 f(u) e^{-2\pi i n u} \left(\int_0^1 g(x) e^{-2\pi i n x} dx \right) du \\
&= \int_0^1 f(u) e^{-2\pi i n u} \widehat{g}(n) du \\
&= \widehat{f}(n) \widehat{g}(n)
\end{aligned}$$

□

3.2 Loi de la mantisse du produit

Théorème 7. *Soit $(X_n)_{n \geq 0}$ une famille de variables aléatoires indépendantes et identiquement distribuées sur $[0, 1]$ de densité $f \in L^2[0, 1]$ et soit $Z_k = \sum_{i=1}^k X_i \pmod{1}$. Alors la loi de Z_k converge vers la loi uniforme sur $[0, 1]$ quand k tend vers l'infini.*

Démonstration. Si Z_k a une densité h_k , alors par le lemme 6, nous avons :

$$\widehat{h}_k(n) = \widehat{f}(n)^k \tag{4}$$

$$\begin{aligned}
\|h_k - 1\|_2^2 &= \int_0^1 (h_k(t) - 1)^2 dt \\
&= \int_0^1 h_k^2(t) dt - 2 \int_0^1 h_k(t) dt + \int_0^1 1 dt
\end{aligned}$$

$$\begin{aligned}
\|h_k - 1\|_2^2 &= \int_0^1 (h_k(t) - 1)^2 dt \\
&= \sum_{n=-\infty}^{+\infty} |\widehat{h}_k(n)|^2 - 2 + 1 \quad \text{d'après (2)} \\
&= \sum_{n=-\infty}^{+\infty} |\widehat{h}_k(n)|^2 - 1 \\
&= \sum_{n=-\infty}^{+\infty} |\widehat{h}_k(n)|^2 - \widehat{h}_k(0) \\
&= \sum_{n \neq 0} |\widehat{h}_k(n)|^2 \\
&= \sum_{n \neq 0} |\widehat{f}(n)|^{2k} \quad \text{d'après (4)} \\
&= \sum_{n \neq 0} |\widehat{f}(n)|^2 |\widehat{f}(n)|^{2k-2} \\
&\leq \left(\sup_{n \neq 0} |\widehat{f}(n)| \right)^{2k-2} \left(\sum_{n \neq 0} |\widehat{f}(n)|^2 \right)
\end{aligned}$$

Puisque $\sup_{n \neq 0} |\widehat{f}(n)| < 1$ (cf. 3 du Lemme 5) et $\sum_{n \neq 0} |\widehat{f}(n)|^2 < \infty$ (cf. preuve du 3 du Lemme 5), alors :

$$\lim_{k \rightarrow +\infty} \left(\sup_{n \neq 0} |\widehat{f}(n)|^{2k-2} \right) = 0 \quad \text{et} \quad \lim_{k \rightarrow +\infty} \|h_k - 1\|_2^2 = 0$$

L'inégalité de Hölder :

$$\int_{\Omega} f(t)g(t) dt \leq \left(\int_{\Omega} f^2(t) dt \right)^{\frac{1}{2}} + \left(\int_{\Omega} g^2(t) dt \right)^{\frac{1}{2}},$$

nous donne avec $g(t) = 1$ quelque soit t de Ω :

$$\int_{\Omega} f(t) dt \leq \left(\int_{\Omega} f^2(t) dt \right)^{\frac{1}{2}}$$

c'est-à-dire :

$$\|f\|_1 \leq \|f\|_2.$$

Ainsi,

$$\lim_{k \rightarrow +\infty} \|h_k - 1\|_1^2 \leq \lim_{k \rightarrow +\infty} \|h_k - 1\|_2^2 = 0$$

La densité h_k tend donc dans L^1 vers 1 qui est la densité de la loi uniforme sur $[0, 1]$. □

Remarque 8. La somme modulo 1 de deux variables aléatoires sur $[0, 1]$ peut suivre la loi uniforme sans que les deux variables soient uniformément distribuées. (Détails dans les Annexes B et C)

Théorème clé 9. Soient $(X_i)_{i \in \mathbb{N}}$ une famille de variables aléatoires réelles identiquement distribuées et indépendantes sur $[0, 1]$ et $(m_{X_i})_{i \in \mathbb{N}}$ la famille de variables aléatoires réelles telle que $X_n = m_{X_n} 10^{n_{X_n}}$, $m_{X_n} \in [1, 10[$ et $n_{X_n} \in \mathbb{Z}$. Soit $Z_k = \prod_{i=1}^k X_i$, alors la loi de sa mantisse m_{Z_k} converge vers la loi logarithmique quand k tend vers l'infini.

Démonstration. Nous allons effectuer deux étapes pour conclure.

1. $\log X_n \equiv \log m_{X_n} \pmod{1}$ pour tout $n \in \mathbb{N}$.

Démonstration.

$$\begin{aligned} \log X_n = \log(m_{X_n} 10^{n_{X_n}}) &= \log m_{X_n} + \log(10^{n_{X_n}}) \\ &= \log m_{X_n} + n_{X_n} \log 10 \\ &= \log m_{X_n} + n_{X_n} \end{aligned}$$

Puisque $n_{X_n} \in \mathbb{Z}$ alors : $\log X_n \equiv \log m_{X_n} \pmod{1}$. □

2. La loi de $\log m_{Z_k} \pmod{1}$ converge vers la loi uniforme sur $[0, 1]$ quand k tend vers l'infini.

Démonstration. Par définition,

$$m_{xy} = \begin{cases} m_x m_y & \text{si } 1 \leq m_x m_y \leq 10 \\ \frac{m_x m_y}{10} & \text{si } 10 \leq m_x m_y \leq 100 \end{cases}$$

On établit facilement que $\log m_{xy} \equiv \log m_x + \log m_y \pmod{1}$, et par récurrence immédiate : $\sum_{i=1}^k \log m_{X_i} \equiv \log m_{\prod_{i=1}^k X_i} = \log m_{Z_k} \pmod{1}$.

En outre, on a, grâce au premier point :

$$\begin{aligned} \sum_{i=1}^k \log X_i &\equiv \sum_{i=1}^k \log m_{X_i} \pmod{1} \\ &\equiv \log m_{Z_k} \pmod{1} \end{aligned}$$

D'où $\log m_{Z_k} \equiv \sum_{i=1}^k \log m_{X_i} \pmod{1}$.

En outre, la famille de variables aléatoires $(\log m_{X_i})_{i \in \mathbb{N}}$ vérifie les hypothèses du théorème 7. Par conséquent, la loi de $\log m_{Z_k}$ converge vers la loi uniforme sur $[0, 1]$ quand k tend vers l'infini. □

On en déduit, grâce à la proposition 3, que la loi de m_{Z_k} converge vers la loi logarithmique sur $[1, 10]$ quand k tend vers l'infini. □

4 Loi de Benford sur le quotient et les puissances de variables aléatoires

Remarque 10. Les théorèmes qui suivent permettent de généraliser le théorème 7. Quant au lemme suivant, il servira à leur démonstration.

Lemme 11. *Soit X une variable aléatoire sur $[0, 1]$ de densité $f \in L^2[0, 1]$ et soit $Z_m = mX \pmod{1}$ tel que $m \in \mathbb{Z}$. Si h_m est la densité de Z_m , alors $\forall n \in \mathbb{Z}, \widehat{h_m}(n) = \widehat{f}(nm)$.*

Démonstration. Rappelons que $\widehat{h_m}(n) = \int_0^1 h_m(t) e^{-2\pi i n t} dt$. On cherche à exprimer la loi Z_m en fonction de celle de X . Le piège et la difficulté proviennent de l'opération $\pmod{1}$. Soit donc

$$\forall 0 \leq v < 1, \quad A_v = \{\omega \in \Omega \mid 0 \leq Z_m(\omega) \leq v\}.$$

La traduction du $\pmod{1}$ donne

$$A_v = \{\omega \in \Omega \mid \exists k \in \mathbb{N} \text{ tel que } k \leq mX(\omega) \leq k + v\}.$$

Le symbole d'existence se transforme en langage ensembliste en union.

$$A_v = \bigcup_{k=0}^{\infty} \{\omega \in \Omega \mid k \leq mX(\omega) \leq k + v\}$$

On note $A_k^v = \{\omega \in \Omega \mid k \leq mX(\omega) \leq k + v\}$. On se rappelle que $X \in [0, 1]$ et donc, pour $k = m, \dots, +\infty$, les ensembles A_k^v sont vides.

$$A_v = \bigcup_{k=0}^{m-1} \left\{ \omega \in \Omega \mid \frac{k}{m} \leq X(\omega) \leq \frac{k+v}{m} \right\}$$

Les ensembles A_k^v sont deux à deux disjoints pour $0 \leq v < 1$.

$$\mathbf{P}(A_v) = \mathbf{P}\left(\bigsqcup_{k=0}^{m-1} A_k^v\right) = \sum_{k=0}^{m-1} \mathbf{P}(A_k^v) = \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+v}{m}} f(t) dt$$

Par un changement de variable affine $\phi : t \mapsto \frac{vt+k}{m}$, on obtient :

$$y = \phi(t) \leftrightarrow y = \frac{mt - k}{v} \leftrightarrow t = \frac{vy + k}{m}$$

d'où $\phi^{-1}(y) = \frac{vy+k}{m}$ et $|\frac{d}{dy}(\phi^{-1})| = \frac{v}{m}$

$$\sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+v}{m}} f(t) dt = \int_0^1 \sum_{k=0}^{m-1} f\left(\frac{vy+k}{m}\right) \frac{v}{m} dy$$

Ainsi pour tout $v \in [0, 1[$,

$$\begin{aligned} \mathbf{P}(A_v) &= \int_0^v h_m(t) dt = \int_0^1 \sum_{k=0}^{m-1} f\left(\frac{vy+k}{m}\right) \frac{v}{m} dy \\ &= \int_0^v \frac{1}{m} \sum_{k=0}^{m-1} f\left(\frac{y+k}{m}\right) dy \end{aligned}$$

Posons pour $t \in [0, 1]$, $h_m(t) := \frac{1}{m} \sum_{k=0}^{m-1} f\left(\frac{t+k}{m}\right)$. C'est une fonction mesurable positive et dans $L^1[0, 1]$. Par convergence dominée avec $\lim_{n \rightarrow +\infty} v_n = 1$, on a

$$\lim_{n \rightarrow +\infty} \int_0^1 h_m(t) 1_{[0, v_n]} dt = \int_0^1 \lim_{n \rightarrow +\infty} \frac{1}{m} \sum_{k=0}^{m-1} f\left(\frac{t+k}{m}\right) 1_{[0, v_n]} dt$$

et après un calcul utilisant un changement de variable et puisque f est une densité sur $[0, 1]$, on obtient

$$\int_0^1 h_m(t) dt = 1$$

$\forall v \in [0, 1]$, $\mathbf{P}(0 \leq h_m(v) \leq v) = \int_0^v h_m(t) dt$ et h_m répond à toutes les propriétés d'une densité et donc la loi de Z_m a pour densité h_m . Ainsi, dans les coefficients de Fourier, on a :

$$\begin{aligned} \int_0^1 h_m(t) e^{-2\pi i n t} dt &= \int_0^1 \sum_{k=0}^{m-1} f\left(\frac{t+k}{m}\right) \frac{e^{-2\pi i n t}}{m} dt \\ &= \sum_{k=0}^{m-1} \int_0^1 f\left(\frac{k+t}{m}\right) \frac{e^{-2\pi i n t}}{m} dt \end{aligned}$$

On réalise un changement de variable $\phi : t \mapsto \frac{vt+k}{m}$ avec $\phi(t) = y$, on

obtient :

$$\begin{aligned}
\int_0^1 h_m(t) e^{-2\pi i n t} dt &= \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+1}{m}} f(y) e^{-2\pi i n (m y - k)} dy \\
&= \sum_{k=0}^{m-1} \int_{\frac{k}{m}}^{\frac{k+1}{m}} f(y) e^{-2\pi i n m y} dy, \quad \text{car } \forall k \in \mathbb{Z}, e^{-2\pi i n k} = 1 \\
&= \int_0^1 f(y) e^{-2\pi i n m y} dy \\
&= \widehat{f}(nm)
\end{aligned}$$

On a donc le résultat. □

4.1 Quotient de variables aléatoires

Théorème 12. Soient X_1, \dots, X_n des variables aléatoires sur $[0, 1]$ identiquement distribuées et mutuellement indépendantes de densité $f \in L^2[0, 1]$, et soit $Z_{j,n} = X_1 + \dots + X_j - X_{j+1} - \dots - X_n \pmod{1}$. Alors la loi de $Z_{j,n}$ converge uniformément par rapport à j vers la loi uniforme sur $[0, 1]$ lorsque n tend vers l'infini.

Démonstration. Soit $h_{j,n}$ la densité de $Z_{j,n}$. On a la même égalité que dans la démonstration du théorème 7 :

$$\|h_{j,n}(t) - 1\|_2^2 = \sum_{n \neq 0} \widehat{h_{j,n}}(n).$$

En utilisant le lemme 11, avec $m = -1$, les coefficients de Fourier de $-X_i$ sont $\widehat{f}(-n)$ et grâce à la propriété des coefficients du lemme 6, on a :

$$\begin{aligned}
\|h_{j,n}(t) - 1\|_2^2 &= \sum_{n \neq 0} |\widehat{f}(n)|^{2j} |\widehat{f}(-n)|^{2k-2j} \\
&= \sum_{n \neq 0} |\widehat{f}(n)|^{2k} \quad \text{car } |\widehat{f}(-n)| = |\overline{\widehat{f}(n)}| = |\widehat{f}(n)|.
\end{aligned}$$

Et donc de la même manière que dans la démonstration du théorème 7 :

$$\|h_{j,n}(t) - 1\|_2^2 \leq \left(\sup_{n \neq 0} |\widehat{f}(n)| \right)^{2k-2} \sum_{n \neq 0} |\widehat{f}(n)|^2$$

et comme $\sup_{n \neq 0} |\widehat{f}(n)| < 1$ et $\sum_{n \neq 0} |\widehat{f}(n)|^2 < +\infty$ (cf. preuve Théorème ??), alors $\lim_{k \rightarrow \infty} \|h_{j,n}(t) - 1\|_1^2 = 0$ □

Théorème clé 13. Soient $(X_i)_{i \in \mathbb{N}}$ une famille de variables aléatoires réelles identiquement distribuées et indépendantes sur $[0, 1]$ et $(m_{X_i})_{i \in \mathbb{N}}$ la famille de variables aléatoires réelles telle que $\forall i \in \mathbb{N}, X_i = m_{X_i} 10^{n_{X_i}}, m_{X_i} \in [1, 10[$ et $n_{X_i} \in \mathbb{Z}$. Soit $Z_{j,k} = \frac{\prod_{i=1}^j X_i}{\prod_{i=j+1}^k X_i}$, alors la loi de sa mantisse $m_{Z_{j,k}}$ converge uniformément par rapport à j vers la loi logarithmique quand k tend vers l'infini.

Démonstration. La démonstration est similaire au théorème clé 9 :

$$m_{x/y} = \begin{cases} \frac{m_x}{m_y} & \text{si } 1 \leq m_x m_y \leq 10 \\ \frac{10 \cdot m_x}{m_y} & \text{si } 0 \leq m_x m_y \leq 1 \end{cases}$$

On a facilement $\log m_{x/y} \equiv \log m_x - \log m_y \pmod{1}$, et par récurrence immédiate :

$$\sum_{i=1}^j \log m_{X_i} - \sum_{i=j+1}^k \log m_{X_i} \equiv \log m_{\frac{\prod_{i=1}^j X_i}{\prod_{i=j+1}^k X_i}} = \log m_{Z_{j,k}} \pmod{1}$$

D'où

$$\log m_{Z_{j,k}} \equiv \sum_{i=1}^k \log m_{X_i} - \sum_{i=j+1}^k \log m_{X_i} \pmod{1}.$$

En outre, la famille de variables aléatoires $(\log m_{X_i})_{i \in \mathbb{N}}$ vérifie les hypothèses du théorème 12. Par conséquent, la loi de $\log m_{Z_{j,k}}$ converge uniformément par rapport à j vers la loi uniforme sur $[0, 1]$ quand k tend vers l'infini.

On en déduit, grâce à la proposition 3, que la loi de $m_{Z_{j,k}}$ converge uniformément par rapport à j vers la loi logarithmique quand k tend vers l'infini. \square

4.2 Puissance de variables aléatoires

Théorème 14. Soit X une variable aléatoire sur $[0, 1]$ de densité $f \in L^2[0, 1]$, et soit $Z_m = mX \pmod{1}$ de densité h_m où $m \in \mathbb{Z}$. Alors la loi de Z_m converge vers la loi uniforme sur $[0, 1]$ quand m tend vers l'infini.

Démonstration.

$$\|h_m(t) - 1\|_2^2 = \sum_{n \neq 0} |\widehat{h_m}(n)|^2 = \sum_{n \neq 0} |\widehat{f}(nm)|^2 \quad (\text{d'après le lemme 11})$$

Pour obtenir une majoration, il suffit de remarquer :

$$\begin{aligned} \sum_{|n| \geq m} |\widehat{f}(n)|^2 &= \sum_{n \neq 0} \sum_{r=0}^{m-1} |\widehat{f}(nm+r)|^2 \\ &= \sum_{r=1}^{m-1} \sum_{n \neq 0} |\widehat{f}(nm+r)|^2 + \sum_{n \neq 0} |\widehat{f}(nm)|^2 \end{aligned}$$

et donc $\sum_{n \neq 0} |\widehat{f}(nm)|^2 \leq \sum_{|n| \geq m} |\widehat{f}(n)|^2$. On obtient l'inégalité :

$$\|h_m(t) - 1\|_2^2 \leq \sum_{|n| \geq m} |\widehat{f}(n)|^2$$

$\|h_m(t) - 1\|_2^2$ est donc majoré par le reste d'une série convergente. D'où :

$$\lim_{m \rightarrow \infty} \|h_m(t) - 1\|_2^2 = 0.$$

La densité de h_k tend donc dans L^1 vers 1 qui est la densité de la loi uniforme sur $[0, 1]$. \square

Théorème clé 15. *Soit X une variable aléatoire sur $[0, 1]$ de densité $f \in L^2[0, 1]$ et soit $Z_k = X^k \pmod{1}$, $m \in \mathbb{Z}$. Alors la loi de sa mantisse m_{Z_k} converge vers la loi logarithmique sur $[1, 10]$ quand k tend vers l'infini.*

Démonstration. La démonstration est similaire aux théorèmes clés précédents.

$\exists i \in \mathbb{N} \quad 10^i \leq (m_X)^k \leq 10^{i+1}$. On a alors : $m_{Z_k} = \frac{(m_X)^k}{10^i}$. Ainsi :

$$\log m_{Z_k} = \log \frac{(m_X)^k}{10^i} = k \log m_X + \log 10^i \equiv k \log m_X \pmod{1}.$$

Donc : $\log m_{Z_k} \equiv k \log m_X \pmod{1}$.

En outre, la variable aléatoire $\log m_X$ vérifie les hypothèses du théorème 14. Par conséquent, la loi de $\log m_{Z_k}$ converge vers la loi uniforme sur $[0, 1]$ quand k tend vers l'infini.

On en déduit, grâce à la proposition 3, que la loi de m_Z converge vers la loi logarithmique sur $[1, 10]$ quand k tend vers l'infini. \square

Corollaire 16 (des Théorèmes clés 9, 13 et 15). *Soient X_1, \dots, X_k des variables aléatoires, $n_j \in \mathbb{Z}$ tel que $j = 1, \dots, k$. Si $Z_{k,j} = \frac{X_1^{n_1} \dots X_j^{n_j}}{X_{j+1}^{n_{j+1}} \dots X_k^{n_k}}$ est défini p.s alors la loi de sa mantisse $m_{Z_{k,j}}$ converge vers la loi logarithmique lorsque $\sum |n_j| \rightarrow +\infty$.*

5 Exemples et simulations

5.1 Vitesse de convergence théorique

Remarque 17. Dans les exemples suivants, nous étudions la vitesse de convergence dans $L^2[0, 1]$ qui est déjà très satisfaisante.

Exemple 1. Soit X une variable aléatoire de loi uniforme sur $[1, 10]$ et $Y = \log X \pmod{1}$. Y a donc pour densité $f(t) = \frac{1}{9} \ln 10 \cdot 10^t$, $0 \leq t \leq 1$. On a donc

$$\widehat{f}(n) = \int_0^1 \frac{\ln 10 \cdot 10^t}{9} e^{-2\pi i n t} dt = \frac{\ln 10}{9(\ln 10 - 2\pi i n)}.$$

Soient X_1, \dots, X_k des variables aléatoires indépendantes de même loi que Y . Si $Z_k = \sum_{i=1}^k X_i \pmod{1}$ alors ses coefficients de Fourier $\widehat{h}_k(n)$ sont égaux à $\left(\frac{\ln 10}{9(\ln 10 - 2\pi i n)}\right)^k$ (cf. Lemme 6).

Identiquement à la démonstration du théorème 7 :

$$\begin{aligned} \|h_k - 1\|_2^2 &= \sum_{n \neq 0} \left| \frac{\ln 10}{9(\ln 10 - 2\pi i n)} \right|^{2k} \\ &\leq 2 \sum_{n=1}^{+\infty} \left(\frac{\ln 10}{18\pi n} \right)^{2k} \\ &\leq 2 \left(\frac{\ln 10}{18\pi} \right)^{2k} \binom{2k}{2k-1} \end{aligned}$$

Il s'agit d'une vitesse de convergence exponentielle car $\frac{\ln 10}{18\pi} < 1$ et donc :

$$\|h_k - 1\|_2 = O\left(\left(\frac{\ln 10}{18\pi}\right)^k\right).$$

Le tableau ci-dessous nous informe sur la vitesse de convergence vers la loi logarithmique quand k variables suivant la loi uniforme sur $[0, 1]$ sont multipliées.

k	Majoration de $\ h_k - 1\ _2$
1	0.64602
2	0.69557
3	0.10110
4	0.00155
5	0.00024
10	0.000000025

Exemple 2. Soit X une variable aléatoire de loi uniforme sur $[0, 1]$ et $Y = \log X \pmod{1}$. Y a donc pour densité $f(t) = \ln 10 \cdot 10^t$, $0 \leq t \leq 1$. On a donc

$$\widehat{f}(n) = \int_0^1 \ln 10 \cdot 10^t e^{-2\pi i n t} dt = \frac{\ln 10}{\ln 10 - 2\pi i n}.$$

Soit $Z_m = mY \pmod{1}$ alors les coefficients de Fourier $\widehat{h}_m(n)$ sont égaux à $\frac{\ln 10}{\ln 10 - 2\pi i n m}$.

$$\begin{aligned} \|h_m - 1\|_2^2 &= \sum_{n \neq 0} \left| \frac{\ln 10}{\ln 10 - 2\pi i n m} \right|^2 \\ &\leq 2 \left(\frac{\ln 10}{2\pi m} \right)^2 \sum_{n=1}^{\infty} \frac{1}{n^2} \\ &\leq \frac{1}{12} \left(\frac{\ln 10}{m} \right)^2 \end{aligned}$$

On a ici une vitesse de convergence en $\frac{1}{m}$ car $\|h_m - 1\|_2 = O\left(\frac{1}{m}\right)$. Le tableau ci-dessous nous informe sur la vitesse de convergence vers la loi logarithmique quand une variable suivant la loi uniforme sur $[0, 1]$ est élevée à la puissance m .

m	Majoration de $\ h_m - 1\ _2$
1	0.44182
2	0.11045
3	0.04909
4	0.02761
5	0.01767
8	0.00690
10	0.00441
100	0.00004418

5.2 Simulations sous Maple

Proposition 18. m_X suit la loi logarithmique de fonction de répartition $\mathbf{P}(m_X \leq x) = \log x$ si et seulement si m_X a pour densité la fonction $x \mapsto \frac{1}{x \ln 10} \mathbf{1}_{[1, 10[}(x)$.

Démonstration. – Preuve de (\Rightarrow) :

$$\forall a, b \in \mathbb{R} \quad \mathbf{P}(a \leq m_X < b) = \int_{[1, 10[} f_{m_X}(x) \mathbf{1}_{[a, b[}(x) dx$$

Or, les fonctions $x \mapsto \frac{1}{x}$ et $x \mapsto \frac{1}{x \ln 10}$ sont Riemann intégrables sur $[1, 10[$ et donc :

$$\begin{aligned} \int_{[1,10[} \frac{1}{x \ln 10} \mathbf{1}_{[a,b[}(x) \, dx &= \frac{1}{\ln 10} \int_{[1,10[} \frac{1}{x} \mathbf{1}_{[a,b[}(x) \, dx \\ &= \frac{1}{\ln 10} \left[\ln x \right]_a^b \\ &= \frac{\ln b - \ln a}{\ln 10} \\ &= \log b - \log a \\ &= \mathbf{P}(a \leq m_X < b) \end{aligned}$$

D'où :

$$\int_{[1,10[} f_{m_X}(x) \mathbf{1}_{[a,b[}(x) \, dx = \int_{[1,10[} \frac{1}{x \ln 10} \mathbf{1}_{[a,b[}(x) \, dx$$

et donc, par identification, $f_{m_X}(x) = \frac{1}{x \ln 10} \mathbf{1}_{[1,10[}(x)$. Vérifions qu'il s'agit bien d'une densité :

$$\int_{\mathbb{R}} \frac{1}{x \ln 10} \mathbf{1}_{[1,10[}(x) \, dx = [\log x]_1^{10} = 1.$$

– *Preuve de (\Leftarrow) :*

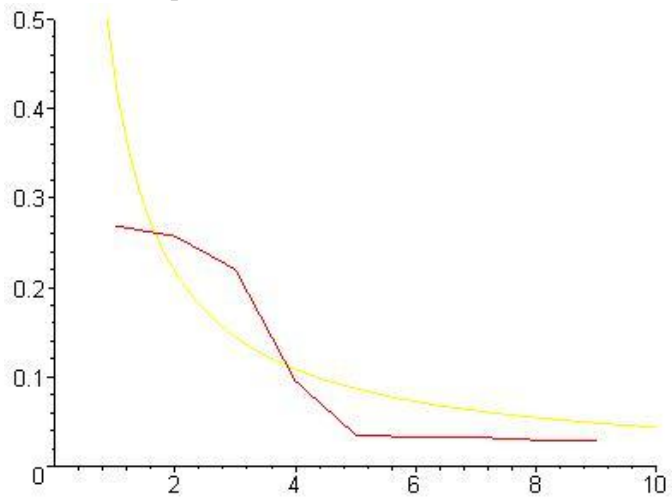
$$\mathbf{P}(m_X \leq x) = \int_{[1,x[} \frac{1}{t \ln 10} \, dt = [\log t]_1^x = \log x$$

□

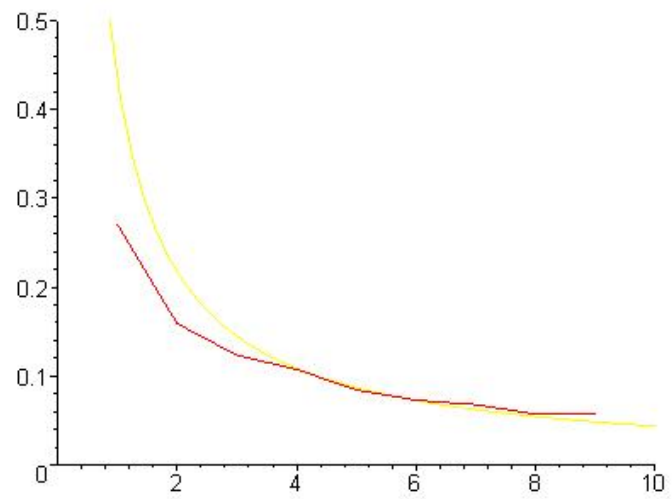
Ainsi, pour voir si la mantisse d'une variable aléatoire suit la loi logarithmique, il suffit de comparer sa densité à celle de la fonction $f : x \mapsto \frac{1}{x \ln 10}$ sur $[1, 10]$.

Nous avons, grâce au logiciel Maple, effectué plusieurs simulations (cf. Annexe D). Les variables aléatoires utilisées sont de densité $g(t) = \frac{10^t \ln 10}{9}$. Sur chacun des graphiques, on pourra comparer la distribution de la mantisse à la fonction f .

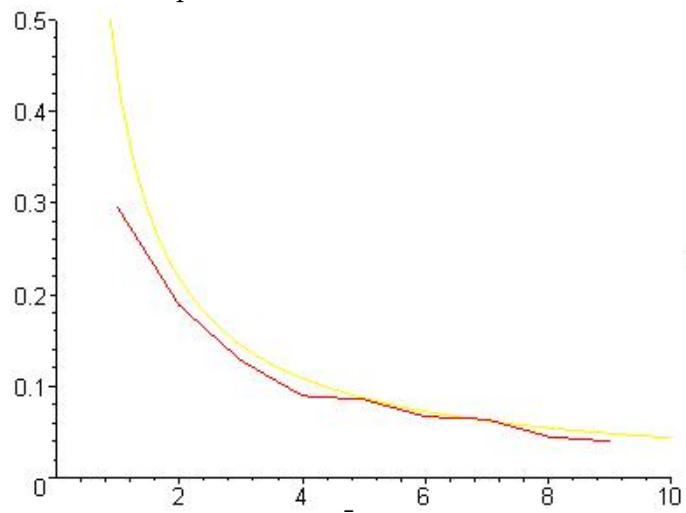
3000 réalisations du produit de 2 variables aléatoires de densité g :



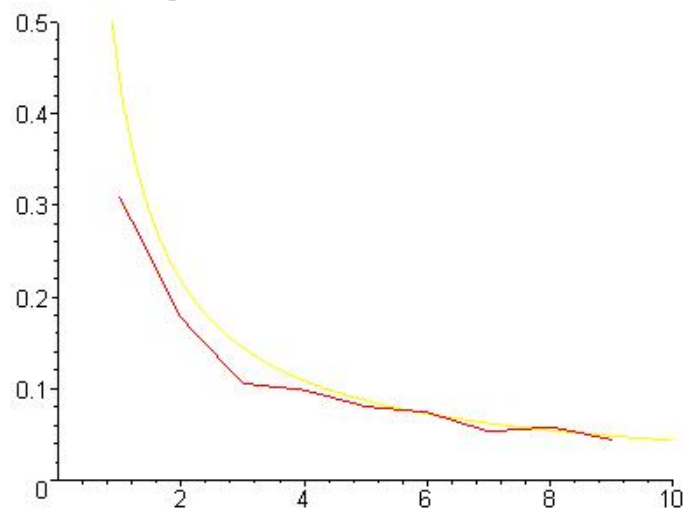
3000 réalisations du produit de 4 variables aléatoires de densité g :



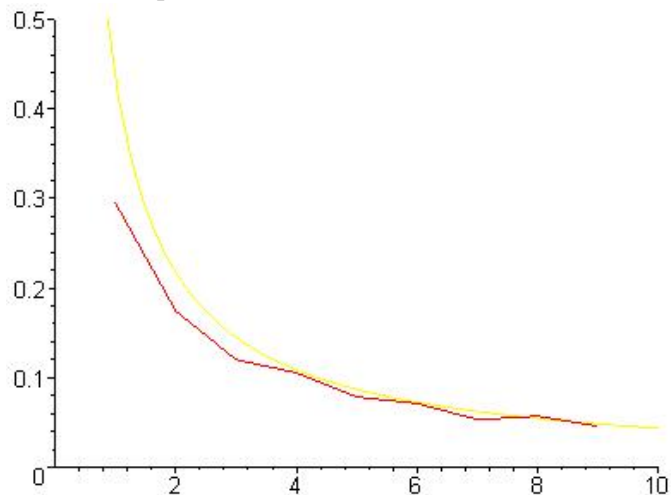
3000 réalisations du produit de 8 variables aléatoires de densité g :



3000 réalisations du produit de 15 variables aléatoires de densité g :



3000 réalisations du produit de 100 variables aléatoires de densité g :



Ainsi, nous voyons bien que plus on multiplie de variables aléatoires, et plus la loi de la mantisse du produit s'approche de la loi logarithmique, c'est à dire que le premier chiffre significatif suit la loi de Benford.

5.3 Simulation sous C++

Nous avons trouvé sur Internet, un programme qui utilise la liste de la taille de tous les fichiers contenus sur le disque dur d'un ordinateur et qui calcule la fréquence du premier chiffre significatif. Nous nous apercevons que ces données suivent aussi la loi de Benford. Ce programme est disponible sur le site http://plvw.free.fr/scm_benford.html. Le code se trouve dans l'annexe E.

6 Conclusion

Ainsi, nous avons mis en évidence le fait que si l'on multiplie, divise ou élève à une certaine puissance un grand nombre de variables aléatoires indépendantes, identiquement distribuées sur $[0, 1]$, alors la loi de la mantisse de la variable aléatoire ainsi obtenue approche la loi logarithmique, ce qui implique que son premier chiffre significatif suit la loi de Benford.

C'est un résultat étrange qui contredit toute forme d'intuition. C'est d'ailleurs pour cela, que cet outil simple et puissant se révèle être un sérieux atout, par exemple, dans le domaine des fraudes fiscales. En effet, la loi de Benford n'est vérifiée que sur des valeurs ayant une signification physique. Donc si les nombres composant un budget ne suivent pas cette loi, c'est que ces valeurs ne représentent rien de réel et sont choisies aléatoirement. Il suffit ensuite de déclencher un contrôle fiscal pour s'en assurer...

Bien que cette loi soit récente, des applications très diverses ont déjà été trouvées. Celles-ci ne cesseront, sans doute pas de se diversifier et de se multiplier, tant cette loi est proche du monde qui nous entoure !

Annexe A

Fréquences théoriques d'apparition du premier chiffre significatif selon la loi de Benford								
1	2	3	4	5	6	7	8	9
30.10%	17.61%	12.49%	9.69%	7.92%	6.69%	5.80%	5.12%	4.58%

Fréquences d'apparition (en %) du premier chiffre significatif tirées de l'observation de diverses données										
Nb. de données	Domaine	Premier chiffre significatif								
		1	2	3	4	5	6	7	8	9
335	Long. de fleuves	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1
3259	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2
104	Constantes	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6
100	Journaux	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0
1389	Chaleurs	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1
703	Pressions	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7
690	Immobilier	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6
1800	Mes. molaires	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2
159	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9
91	Mes. atomiques	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5
5000	n^{-1}, \sqrt{n}	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9
560	Plans	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6
707	Volts Ray. X	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8
741	Prix	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1
342	Adresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0
900	$n^1, n^2, \dots, n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5
418	Décès	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1
Moyenne		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7

Annexe B

Soient X, Y, Z des variables aléatoires sur $[0, 1]$ de densités respectives $f, g, h \in L^2[0, 1]$ telles que X et Y sont indépendantes et $Z = X + Y \pmod{1}$. Si $\forall n \neq 0, \widehat{f}(n)\widehat{g}(n) = 0$, alors Z suit la loi uniforme. En effet, dans ce cas :

$$\|h - 1\|_2^2 = \sum_{n \neq 0} |\widehat{h}(n)|^2 = \sum_{n \neq 0} |\widehat{f}(n)\widehat{g}(n)|^2 = 0.$$

Exemple 3. Soient $f(t) = 1 + \sin(2\pi t)$ et $g(t) = 1 + \sin(4\pi t)$.

1. Soit $I_n = \widehat{f}(n) \quad \forall n \neq 0$

$$I_n = \int_0^1 (1 + \sin(2\pi t)) e^{-2\pi i n t} dt = \int_0^1 \sin(2\pi t) e^{-2\pi i n t} dt$$

Intégration par parties :

$$\begin{aligned} u &= \sin(2\pi t) & u' &= 2\pi \cos(2\pi t) \\ v' &= e^{-2\pi i n t} & v &= -\frac{1}{2\pi i n} e^{-2\pi i n t} \end{aligned}$$

D'où :

$$\begin{aligned} I_n &= \left[\sin(2\pi t) \left(-\frac{1}{2\pi i n} e^{-2\pi i n t} \right) \right]_0^1 + \frac{1}{i n} \int_0^1 \cos(2\pi t) e^{-2\pi i n t} dt \\ &= \frac{1}{i n} \int_0^1 \cos(2\pi t) e^{-2\pi i n t} dt \end{aligned}$$

Intégration par parties :

$$\begin{aligned} u &= \cos(2\pi t) & u' &= -2\pi \sin(2\pi t) \\ v' &= e^{-2\pi i n t} & v &= -\frac{1}{2\pi i n} e^{-2\pi i n t} \end{aligned}$$

D'où :

$$\begin{aligned} I_n &= \left[\cos(2\pi t) \left(-\frac{1}{2\pi i n} e^{-2\pi i n t} \right) \right]_0^1 + \frac{1}{n^2} \int_0^1 \sin(2\pi t) e^{-2\pi i n t} dt \\ &= \frac{1}{n^2} \int_0^1 \sin(2\pi t) e^{-2\pi i n t} dt \\ &= \frac{1}{n^2} I_n \end{aligned}$$

Ainsi, $I_n(1 - \frac{1}{n^2}) = 0$ et donc :

$$\forall n \notin \{-1, 0, 1\}, \quad I_n = \widehat{f}(n) = 0 \quad (5)$$

2. Soit $J_n = \widehat{g}(n) \quad \forall n \neq 0$

$$J_n = \int_0^1 (1 + \sin(4\pi t)) e^{-2\pi i n t} dt = \int_0^1 \sin(4\pi t) e^{-2\pi i n t} dt$$

Intégration par parties :

$$\begin{aligned} u &= \sin(4\pi t) & u' &= 4\pi \cos(4\pi t) \\ v' &= e^{-2\pi i n t} & v &= -\frac{1}{2\pi i n} e^{-2\pi i n t} \end{aligned}$$

D'où :

$$\begin{aligned} J_n &= \left[\sin(4\pi t) \left(-\frac{1}{2\pi i n} e^{-2\pi i n t} \right) \right]_0^1 + \frac{2}{i n} \int_0^1 \cos(4\pi t) e^{-2\pi i n t} dt \\ &= \frac{2}{i n} \int_0^1 \cos(4\pi t) e^{-2\pi i n t} dt \end{aligned}$$

Intégration par parties :

$$\begin{aligned} u &= \cos(4\pi t) & u' &= -4\pi \sin(4\pi t) \\ v' &= e^{-2\pi i n t} & v &= -\frac{1}{2\pi i n} e^{-2\pi i n t} \end{aligned}$$

D'où :

$$\begin{aligned} J_n &= \left[\cos(4\pi t) \left(-\frac{1}{2\pi i n} e^{-2\pi i n t} \right) \right]_0^1 + \frac{4}{n^2} \int_0^1 \sin(4\pi t) e^{-2\pi i n t} dt \\ &= \frac{4}{n^2} \int_0^1 \sin(4\pi t) e^{-2\pi i n t} dt \\ &= \frac{4}{n^2} J_n \end{aligned}$$

Ainsi, $J_n(1 - \frac{4}{n^2}) = 0$ et donc :

$$\forall n \notin \{-2, 0, 2\}, \quad J_n = \widehat{g}(n) = 0 \quad (6)$$

Conclusion : (5) et (6) $\Rightarrow \forall n \neq 0, \quad \widehat{f}(n)\widehat{g}(n) = 0$.

On peut alors affirmer que si $Z = X + Y \pmod{1}$ où X et Y sont de densités respectives $f(t) = 1 + \sin(2\pi t)$ et $g(t) = 1 + \sin(4\pi t)$ sur $[0, 1]$, alors Z suit une loi uniforme sur $[0, 1]$.

Annexe C

Soient X, Y, Z de densités respectives $f, g, h \in L^2[0,1]$ telles que $\widehat{f}(n)\widehat{g}(n) = \widehat{h}(n)$ pour tout $n \in \mathbb{N}$.

Puisque $\forall n \neq 0 \quad |\widehat{g}(n)| < 1$ (cf. 2 du Lemme 5), alors $|\widehat{h}(n)| = |\widehat{f}(n)\widehat{g}(n)| \leq |\widehat{f}(n)|$.

1. S'il existe $n \neq 0$ tel que l'on ait l'égalité stricte, alors

$$\|h - 1\|_2^2 = \sum_{n \neq 0} |\widehat{h}(n)|^2 < \sum_{n \neq 0} |\widehat{f}(n)|^2 = \|f - 1\|_2^2$$

et donc si X suit une loi uniforme, alors Z aussi.

2. S'il y a égalité pour tout n , alors $|\widehat{f}(n)| = |\widehat{h}(n)|$ et donc :

$$|\widehat{f}(n)\widehat{g}(n)| = |\widehat{f}(n)| \tag{7}$$

Si $\widehat{f}(n) \neq 0$ alors (7) $\Rightarrow \widehat{g}(n) = 1$, ce qui est impossible d'après le 2 du Lemme 5.

D'où $\widehat{f}(n) = \widehat{h}(n) = 0 \quad \forall n \neq 0$. Ainsi, $\|h - 1\|_2^2 = \|f - 1\|_2^2 = \sum_{n \neq 0} |\widehat{h}(n)|^2 = 0$. X et Z suivent alors la loi uniforme.

Conclusion : Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes telles qu'au moins un des X_i suive la loi uniforme sur $[0, 1]$, alors $Z_n := \sum_{i=1}^n X_i \pmod{1}$ aussi.

Exemple 4. (Le résultat est analogue modulo 6) On jette deux dés dont l'un est équilibré, l'autre est truqué de façon inconnue. On note X, Y les nombres de points indiqués respectivement par le dé équilibré et le dé truqué.

La variable aléatoire X suit ainsi la loi uniforme sur $\{1, \dots, 6\}$, tandis que la loi de Y nous est inconnue : on sait seulement que l'ensemble des valeurs possibles de Y est $\{1, \dots, 6\}$, mais on ignore les valeurs des $\mathbb{P}(Y = j)$ pour $j \in \{1, \dots, 6\}$.

On suppose que le trucage n'affecte pas l'indépendance des deux dés.

On note R la variable aléatoire égale au représentant dans $\{0, \dots, 5\}$ de la classe d'équivalence de $X + Y \pmod{6}$.

1. $\forall r \in \{0, \dots, 5\}, \quad \forall j \in \{1, \dots, 6\}, \quad \exists i \in \{1, \dots, 6\}, \quad i + j \equiv r \pmod{6}$.

Démonstration. $R \equiv X + Y \pmod{6}$

$\forall r \in \{0, \dots, 5\}, \quad \forall j \in \{1, \dots, 6\}, \quad \exists i \in \{1, \dots, 6\}, \quad i + j \equiv r \pmod{6}$.

Montrons que i est unique.

Supposons qu'il existe $i, i' \in \{1, \dots, 6\}$, $i + j \equiv i' + j \equiv r \pmod{6}$.

Alors $i - i' \equiv 0 \pmod{6}$.

Supposons que $i \geq i'$. Alors : $\exists k \in \mathbb{N} \quad i - i' = 6k$. Si $k \geq 1$, alors $i = 6k + i' > 6$. Ce qui est absurde car $i \in \{1, \dots, 6\}$.

D'où $k = 0$ et donc $i = i'$. \square

2. L'événement $\{R = r\}$ est réunion de 6 événements deux à deux disjoints du type $\{X = i, Y = j\}$.

Démonstration.

$$\begin{aligned} \{R = r\} &= \{X + Y \equiv r \pmod{6}\} \\ &= \bigcup_{j \in \{1, \dots, 6\}} \{X \equiv r - j \pmod{6}, Y \equiv j \pmod{6}\} \end{aligned}$$

Il est clair que les 6 événements sont deux à deux disjoints. \square

3. $\forall r \in \{0, \dots, 5\}$, $\mathbb{P}(R = r) = \frac{1}{6}$ et donc R suit la loi uniforme sur $\{0, \dots, 5\}$ et ceci quelle que soit la loi de Y , c'est-à-dire quel que soit le trucage du deuxième dé.

Démonstration.

$$\mathbf{P}(R = r) = \mathbf{P}\left(\bigcup_{j \in \{1, \dots, 6\}} \{X \equiv r - j \pmod{6}, Y = j\}\right)$$

Puisque les événements sont deux à deux disjoints :

$$\mathbf{P}(R = r) = \sum_{j \in \{1, \dots, 6\}} \mathbf{P}(X \equiv r - j \pmod{6}, Y = j)$$

Puisque X et Y sont indépendants :

$$\begin{aligned} \mathbf{P}(R = r) &= \sum_{j \in \{1, \dots, 6\}} \mathbf{P}(X \equiv r - j \pmod{6}) \mathbf{P}(Y = j) \\ &= \sum_{j \in \{1, \dots, 6\}} \frac{1}{6} \mathbf{P}(Y = j) \\ &= \frac{1}{6} \sum_{j \in \{1, \dots, 6\}} \mathbf{P}(Y = j) = \frac{1}{6} \times 1 = \frac{1}{6} \end{aligned}$$

\square

Annexe D

```
> restart;
> with(stats);
  [anova, describe, fit, importdata, random, statevalf,
  statplots, transform]

> with(random);

  [beta, binomiald, cauchy, chisquare, discreteuniform,
  empirical, exponential, fratio, gamma, laplaced, logistic,
  lognormal, negativebinomial, normald, poisson, studentst,
  uniform, weibull]
```

Simulation de variables aléatoires.
Le programme simule la variable aléatoire de densité
 $f(t) = ((\ln 10) * 10^{-t}) / 9$ k fois.
Nous avons calculé à la main la fonction inverse.

```
simulX := proc(k)
local i, K, L;
  L := [uniform[0, 1](k)];
  K := [];
  for i to k do K := [op(K), log(9*L[i])] end do;
  K
end proc
```

Donne le produit de k valeurs de la variable aléatoire de densité
 $f(t) = ((\ln 10) * 10^{-t}) / 9$.

```
Zk := proc(k) product(simulX(k)[i], i = 1 .. k) end proc
```

Extrait la mantisse du nombre n.

```
extractmantisse := proc(n)
local l, M, s;
  l := abs(n);
  while floor(l) = 0 do l := 10*l end do;
  M := length(floor(l));
  s := l*10^(-M + 1);
```

```

s
end proc

```

Nous construisons une liste contenant k réalisations de la mantisse du produit de n variables aléatoires de densité $f(t) = ((\ln 10) * 10^{-t})/9$.

En fait, on construit une liste de la variables aléatoire qui nous intéresse.

```

liste := proc(k, n)
local L, i;
  L := [];
  for i to k do L := [op(L), extractmantisse(Zk(n))] end do
  ;
  L
end proc

```

Des petits graphiques

On prend une liste L de la variable aléatoire m_z et le nombre de paquets decoupage de l'intervalle $[0,1]$.

```

paquet := proc(L, n)
local i, j, M;
  M := [seq(0, i = 1 .. 9*n)];
  for j to nops(L) do
    i := 0;
    while not ((n + i)/n < L[j] and L[j] < (n + i + 1)/n)
    do i := i + 1
    end do;
    M[i + 1] := M[i + 1] + 1
  end do;
  [seq(M[i]/nops(L), i = 1 .. 9*n)]
end proc

```

```

> with(plots);
[animate, animate3d, animatecurve, arrow, changecoords, complexplot,
  complexplot3d, conformal, conformal3d, contourplot,
  contourplot3d, coordplot, coordplot3d, cylinderplot,
  densityplot, display, display3d, fieldplot, fieldplot3d,
  gradplot, gradplot3d, implicitplot, implicitplot3d, inequal,
  listcontplot, listcontplot3d, listdensityplot, listplot,

```

```

listplot3d, loglogplot, logplot, matrixplot, odeplot, pareto,
pointplot, pointplot3d, polarplot, polygonplot, polygonplot3d,
polyhedra_supported, polyhedraplot, replot, rootlocus,
semilogplot, setoptions, setoptions3d, spacecurve,
sparsematrixplot, sphereplot, surfdata, textplot, textplot3d,
tubepplot]

```

Forme le graphe des paquets que l'on a construit.

```

dessin := proc(L, n)
local i, U;
  U := [seq((n + i)/n, i = 1 .. 9*n)];
  plot([seq([U[i], L[i]], i = 1 .. nops(L))])
end proc

benford := proc()
local i, S;
  for i to 10 do S[i] := log(i + 1) - log(i) end do;
  [seq(evalf(S[i]), i = 1 .. 10)]
end proc

benford2 := proc()
local i, S;
  for i to 9 do S[i] := 1/(ln(10)*i) end do;
  [seq(evalf(S[i]), i = 1 .. 9)]
end proc

> C:=benford();

C := [.6931471806, .4054651084, .287682072, .223143551, .182321557,
      .154150680, .133531393, .117783035, .105360516, .095310180]

> B:=benford2();

B := [.4342944819, .2171472410, .1447648273, .1085736205,
      .08685889638, .07238241366, .06204206886, .05428681024,
      .04825494243]

dessin1 := proc(L) plot([seq([i, L[i]], i = 1 .. 9)]) end proc

```

Annexe E

```
// Programme Benford.exe (c) 2002 Pascal Levy//
// Pour compiler sous Windows avec GCC, copiez-collez le contenu
// de cette fenêtre dans un fichier nommé benford.cpp, puis
// tapez en ligne de commande ://// gcc -o benford.exe benford.cpp

#include <windows.h>
#include <stdio.h>#
include <string.h>
static int iFileNr;
static int iProb[10];
static void Recurse(char * _psz){ WIN32_FIND_DATA wfd;
HANDLE h; char sz[260]; int k;

printf("Exploration de : %s\n",_psz); strcpy(sz,_psz);
if (sz[strlen(sz)-1]!='\\') strcat(sz,"\\"); strcat(sz,"*.");
h=FindFirstFile(sz,&wfd); if (h!=INVALID_HANDLE_VALUE) { do
{
if (wfd.cFileName[0]!='.' && wfd.cFileName[0]!='?') {
if (wfd.dwFileAttributes&FILE_ATTRIBUTE_DIRECTORY)
{
strcpy(sz,_psz);
if (sz[strlen(sz)-1]!='\\') strcat(sz,"\\");
strcat(sz,wfd.cFileName); Recurse(sz);
} else {
k=wfd.nFileSizeLow; while (k/10) k/=10;
if (k) {
iProb[k%10]++; iFileNr++;
} } }
while (FindNextFile(h,&wfd)); FindClose(h); }}void main(){
double d; int i; iFileNr=0; for (i=0; i<10; i++) iProb[i]=0;
Recurse("C:\\"); for (i=1; i<10; i++) {
d=(double) iProb[i]*100.0/(double) iFileNr;
printf("%d : %.2lf\n",i,d); }}
```

Références

- [1] http://www.webridge.fr/Accueil/loi_de_benford.htm
- [2] BENFORD, F. (1938), The law of anomalous numbers. *Proceedings of the American Philosophical Society*, **78**, No 4, 551–572.
- [3] http://plvw.free.fr/scm_benford.html
- [4] *Revue Française de Comptabilité*, No 321, avril 2000.