Expérimentations sur la loi de Benford

Franck Jeannot - Août 2017 - C82 - v1.0

Abstract : Experimentations based on *Frank **Benford*** results (1937) on **anomalous numbers**.

Keywords : *Benford's law ; fraud detection ; Newcomb ; First-digit ; Significant-digit law ; Luatex ; Lua*

1 Loi de Benford

Loi de Benford simplifiée

La loi de **Benford** simplifiée établit, qu'en considérant un jeu de données, les premiers chiffres *significatifs* « *the first significant decimal digit* » de chaque valeur ont une fréquence de distribution, en base b selon :

$$\mathrm{Prob}(D_1 = d) = f(d) = \log_b\left(1 + \frac{1}{d}\right) , \quad d = 1, 2 \ldots, 9 \tag{1}$$

1.1 Implémentation d'exemple de la loi de Benford simplifiée

La loi (1) donne [1] [2] [3] la valeur asymptotique f de la fréquence d'apparition du premier chiffre *significatif* d'un nombre d d'un résultat de mesure exprimé dans une base **b**. On implémente de manière minimaliste la formule de Benford en **Luatex** et **Lua** :

```
\documentclass{article}
\usepackage{luacode}
\begin{luacode*}
        function benford (d)
        return ((math.log(1+(1/d)))  /math.log(10))
        end
\end{luacode*}
\begin{document}
        $benfordresu(1)$  = $\directlua{tex.sprint(benford(1))}$\\
        $benfordresu(2)$  = $\directlua{tex.sprint(benford(2))}$\\
        $benfordresu(3)$  = $\directlua{tex.sprint(benford(3))}$\\
\end{document}
```

1. https://fr.wikipedia.org/wiki/Loi_de_Benford
2. https://github.com/mancap314/mancap314.github.io/blob/master/benford.html
3. https://mancap314.github.io/benford.html

On obtient alors :

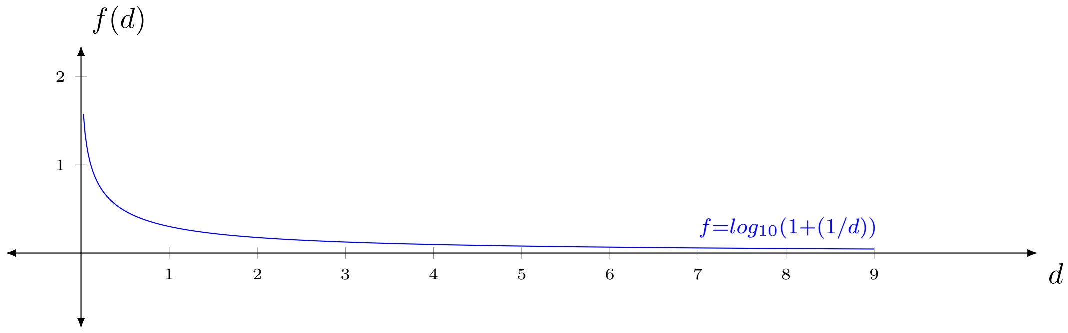
```
benfordresu(1) = 0.30102999566398
benfordresu(2) = 0.17609125905568
benfordresu(3) = 0.1249387366083
```

Tableau récapitulatif, on appelle D_1 le premier chiffre significatif dans les exemples ci-dessous :

d	Valeur dec. approx.	D_1	$f(d) =$	$benfordresu(D_1)\ approx.$
e^π	23.14	2	$\log_{10}(2/2+1/2)$	0.1760
π^{-1}	0.3183	3	$\log_{10}(3/3+1/3)$	0.1249
$\sqrt{2}$	1.414	1	$\log_{10}(2)$	0.3010

On remarque que la probabilité d'avoir un 3 comme premier chiffre significatif est de 12.49% et de 30.1% pour un 1.

1.2 Graphe de la loi de Benford simplifiée



1.3 Valeurs pour les entiers de 1 a 9 de la loi de Benford

On implemente pareillement le calcul d'une table des valeurs en Lua/Luatex, soit :

```
\documentclass{article}
\usepackage{luacode}
\begin{luacode*}
        function benfordtable ()
        for d = 1, 9, 1 do
        f = (math.log(1+(1/d)))  /math.log(10)
        tex.print(string.format('%1.0f & %1.9f \\\\', d, f))
        end
        end
\end{luacode*}
\newcommand{\benfordtable}{\luadirect{benfordtable()}}
\begin{document}
        \begin{tabular}{lcc}
                \hline
                & $d$ & $f(d)$ \\
                \hline
```

```
                \benfordtable
                \hline
        \end{tabular}
\end{document}
```

On obtient donc les valeurs suivantes :

```
d f(d)
1 0.301029996
2 0.176091259
3 0.124938737
4 0.096910013
5 0.079181246
6 0.066946790
7 0.057991947
8 0.051152522
9 0.045757491
```

1.4 Explications de la loi de Benford

Il n'existe pas de justifications ou explications parfaites sur ce type de distribution, cependant, selon Paul D. Scott et Maria Fasli (2001, *University of Essex*), ces caractéristiques sur certains jeux de données proviennent de ([3], trad. libre) « *la multiplication de facteurs aléatoires,..., ces multiplications tendant les jeux de données à se rapprocher de* ***distributions normales*** » [4]. Leurs **expérimentations ont aussi montré que peu de jeux de données étaient finalement conformes à cette loi**. Il convient de revoir l'analyse approfondie faite par Arnaud Berger [1].

2 Usage de la loi de Benford

Cette loi est utilisée en détection de fraude, pour des élections, pour des données financières,...etc. En cas de de grande différence entre les données de la distribution et celles données par la loi de Benford, on peut penser à une éventuelle manipulation des données (à confirmer). Cette loi fut premièrement découverte par l'astronome Américan Simon Newcomb puis redécouvert par Frank Benford en 1938 et prouvée par Knuth dans son célèbre livre « The Art of Computer Programming ».

3 Publication des travaux de Benford

On retrouve le résumé des travaux de Benford de 1937 [2] et le résumé de sa publication :

4. Citation exacte originale de [3] "The multiplication of random factors is the most plausible explanation for a data sets conformity to Benford's Law. Multiplying numbers together is equivalent to adding their logarithms. The central limit theorem states that the sum of independent random variates tends to a normal distribution as the number of variates is increased."

THE LAW OF ANOMALOUS NUMBERS

FRANK BENFORD

Physicist, Research Laboratory, General Electric Company, Schenectady, New York

(Introduced by Irving Langmuir)

(*Read April 22, 1937*)

Abstract

It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9. A compilation of some 20,000 first digits taken from widely divergent sources shows that there is a logarithmic distribution of first digits when the numbers are composed of four or more digits. An analysis of the numbers from different sources shows that the numbers taken from unrelated subjects, such as a group of newspaper items, show a much better agreement with a logarithmic distribution than do numbers from mathematical tabulations or other formal data. There is here the peculiar fact that numbers that individually are without relationship are, when considered in large groups, in good agreement with a distribution law—hence the name " Anomalous Numbers."

A further analysis of the data shows a strong tendency for bodies of numerical data to fall into geometric series. If the series is made up of numbers containing three or more digits the first digits form a logarithmic series. If the numbers contain only single digits the geometric relation still holds but the simple logarithmic relation no longer applies.

An equation is given showing the frequencies of first digits in the different orders of numbers 1 to 10, 10 to 100, etc.

The equation also gives the frequency of digits in the second, third $\cdots$ place of a multi-digit number, and it is shown that the same law applies to reciprocals.

There are many instances showing that the geometric series, or the logarithmic law, has long been recognized as a common phenomenon in factual literature and in the ordinary affairs of life. The wire gauge and drill gauge of the mechanic, the magnitude scale of the astronomer and the sensory response curves of the psychologist are all particular examples of a relationship that seems to extend to all human affairs. The Law of Anomalous Numbers is thus a general probability law of widespread application.

Références

[1] Berger, Arno, Hill, Theodore P. A basic theory of benford's law. *Probab. Surveys*, 8 :1–126, 2011.

[2] Frank Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4) :551–572, 1938.

[3] Paul D Scott and Maria Fasli. Benford's law : An empirical investigation and a novel explanation. *Unpublished manuscript*, 2001.