

DES MATHÉMATIQUES ÉLÉMENTAIRES POUR DÉBUSQUER DES FRAUDES OU DES ERREURS EN ÉCONOMIE (OU AILLEURS)

Mots clés : Loi de BENFORD, 1^{er} chiffre significatif, logarithme, fraude, comptabilité

Résumé

Relevons au petit bonheur quelques nombres autour de nous : prix d'articles divers dans une publicité ou au hasard des rayons d'un magasin, valeurs numériques extraites d'articles de journaux, données numériques en géographie (populations de villes ou de pays, altitudes de montagnes, longueurs de fleuves, superficies de pays, ...), données économiques (PIB, chiffres extraits de comptabilités d'entreprises, cours de la bourse, ...).

Pour les données relevées, intéressons-nous à la proportion de chaque chiffre comme premier chiffre significatif (i.e. le premier non nul à gauche). *A priori*, on pourrait s'attendre à une répartition à peu près uniforme des chiffres.

Mais ô surprise : on constate expérimentalement que dans bien des ensembles de données il y a plus de 1 que de 2, de 2 que de 3, de 3 que de 4, etc.

Et ce n'est pas tout ! La proportion de chaque chiffre comme premier à gauche est assez souvent relativement stable et raisonnablement approchable par un logarithme (loi de BENFORD).

Encore plus fort : dans bien des ensembles de données, les 2èmes, 3èmes, ... chiffres significatifs ne se répartissent pas n'importe comment mais tendent très vite vers la répartition uniforme (conforme à notre intuition) lorsqu'on avance vers la droite (loi de BENFORD généralisée).

Ce résultat (de 1881), assez contre-intuitif, a été longtemps considéré comme une simple curiosité mais depuis une vingtaine d'années, il est largement exploité notamment pour débusquer des fraudes (erreurs ou falsifications) dans les comptabilités entre autres aux Etats-Unis et au Canada. Depuis peu, son utilisation tend à se répandre en Europe.

I – Petits pré-requis mathématiques

I.1 – Logarithme

Le logarithme (décimal) d'un nombre positif a , noté $\log a$, est la puissance à laquelle il faut élever 10 pour obtenir a .

$$10^{\log a} = a$$

Ainsi :

$$\log 1\,000 = 3 \quad \text{car} \quad 10^3 = 1\,000$$

$$\log 0,01 = -2 \quad \text{car} \quad 10^{-2} = 0,01$$

$$\log 2 \approx 0,301 \quad \text{car} \quad 10^{0,301} \approx 2$$

I.2 – Premier chiffre significatif

C'est le premier chiffre non nul le plus à gauche dans l'écriture d'un nombre.

$$\begin{array}{l} \text{Ainsi :} \quad 4\,853,746 \quad \rightarrow \quad 4 \\ \quad \quad \quad 0,03911 \quad \rightarrow \quad 3 \end{array}$$

I.3 – Notation scientifique

$$4\,853,746 = \underbrace{4,853746}_{\text{mantisse}} \cdot 10^3$$

$$0,03911 = \underbrace{3,911}_{\text{mantisse}} \cdot 10^{-2}$$

Le premier chiffre significatif d'un nombre est donc la partie entière de sa mantisse.

II – Idée générale

Livrons-nous à la petite expérience (*à priori* tout à fait saugrenue...) suivante. On choisit un ensemble de valeurs de l'un des types suivants, relevées au hasard :

- Prix relevés au hasard dans un magasin :
prix sur un assez long ticket de caisse (ou assemblage de plusieurs),
prix figurant sur une publicité, ...
- Nombres extraits de coupures de journaux
- Résultats des élections présidentielles françaises 2012
- Résultats sportifs (temps, distances, hauteurs, ...)
- Nombre d'habitants de toutes les communes françaises
- Altitudes de montagnes dans les Alpes
- Longueurs de fleuves dans le monde entier
- PIB de tous les pays du monde
- Cours de la bourse
- Nombres extraits de comptabilités d'entreprises
- ...

Puis on note le premier chiffre significatif des valeurs relevées.

Ensuite on dresse un tableau des fréquences d'apparitions de chacun des chiffres 1 à 9 comme premier chiffre significatif et on trace un histogramme.

Exemple Nombre d'habitants des 36 722 communes françaises au 1^{er} janvier 2009

Le site français officiel insee.fr/fr (de l'Institut National de la Statistique et des Etudes Economiques) fournit, dans la rubrique population, des tableaux reprenant commune par commune le nombre d'habitants au 1^{er} janvier de l'année de votre choix. Voici un extrait d'un tel tableau (honneur au département de la Moselle qui reçoit le congrès) avec les données les plus récentes, année 2009 :

57 - MOSELLE

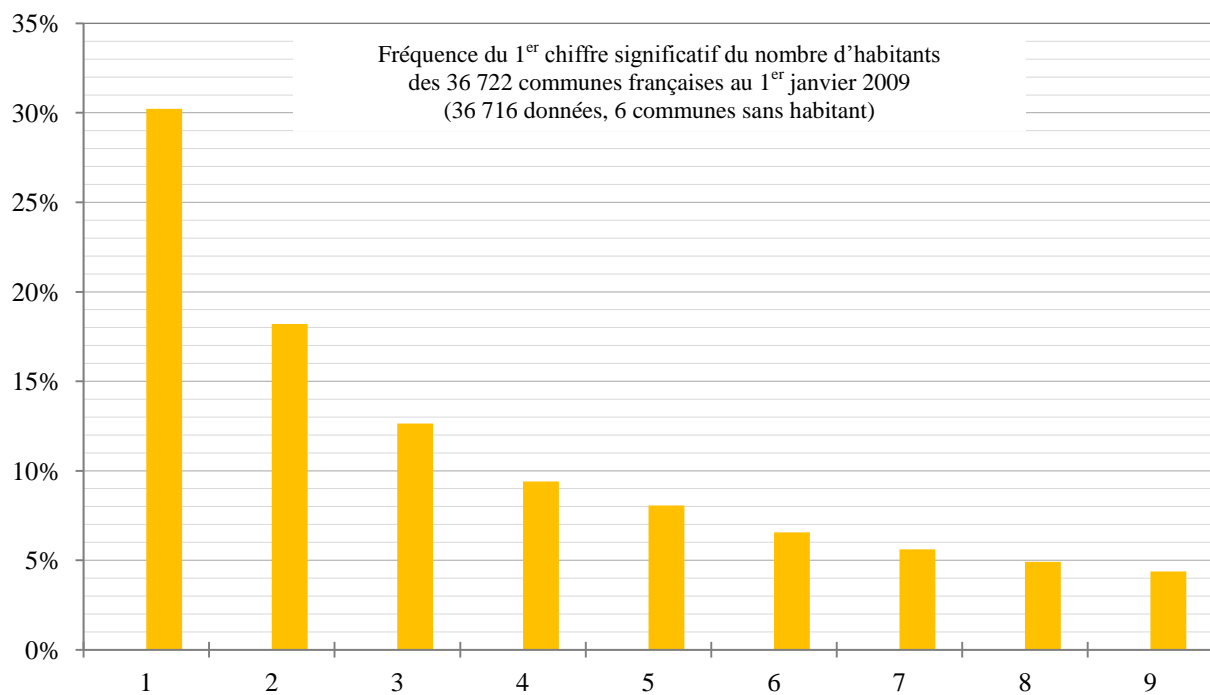
Tableau 2 - Populations légales des communes en vigueur à compter du 1^{er} janvier 2012 - date de référence statistique : 1^{er} janvier 2009

CODE			COMMUNES	Population totale	Population municipale	Population comptée à part
Arrondissement	Canton	Commune				
a	b	c	d	e = f + g	f	g
4	33	447	Marly	9 911	9 607	304
2	34	448	Marsal	281	268	13
4	23	449	Marsilly	500	487	13
2	07	451	Marthille	192	187	5
4	37	452	La Maxe	867	855	12
3	14	453	Maxstadt	296	295	1
4	33	454	Méclevues	1 189	1 169	20
1	03	455	Mégange	198	193	5
6	02	456	Meisenthal	751	734	17
1	04	457	Menskirch	166	160	6
7	31	459	Merschweiller	182	181	1
1	04	460	Merten	1 535	1 510	25
5	16	461	Métairies-Saint-Quirin	306	303	3
5	24	462	Metting	355	352	3
		463	METZ			
9	17	463	Metz-Ville 1er Canton	23 066	22 751	315
9	18	463	Metz-Ville 2e Canton	25 828	25 418	410
9	19	463	Metz-Ville 3e Canton	40 902	40 175	727
9	46	463	Metz-Ville 4e Canton	34 228	33 497	731
			TOTAL	124 024	121 841	2 183

Pour chaque commune, nous prenons note du 1^{er} chiffre significatif de son nombre d'habitants. Pour les communes ci-dessus, par exemple, nous relevons donc : 9, 2, 4, 1, 8, ...

Premier chiffre significatif du nombre d'habitants des 36 722 communes françaises au 1 ^{er} janvier 2009	effectif	fréquence
1	11 096	30,1 %
2	6 682	17,6 %
3	4 644	12,5 %

4	3 450	9,7 %
5	2 962	7,9%
6	2 411	6,7 %
7	2 062	5,8 %
8	1 801	5,1 %
9	1 608	4,6 %
Total	36 716	100 %



On constate qu'il y a à peu près un tiers de nombres commençant par 1, la moitié par 1 ou 2 et que « en gros » plus un chiffre est grand, moins il apparaît.

III – Loi de BENFORD

Ce phénomène plutôt inattendu s'observe assez souvent lorsqu'on analyse des données. Il a été relevé pour la première fois par l'astronome Simon NEWCOMB (1835-1909) en 1881.

Il avait constaté que les premières pages des tables de logarithmes étaient en général plus écornées, donc plus consultées, que les suivantes. Cela l'avait intrigué ! On comprendrait facilement cela pour un mauvais roman vite abandonné mais ici, qu'est-ce cela signifie ?

A l'intention des lecteurs les plus jeunes, rappelons qu'avant l'avènement des calculatrices et des PC, les tables de logarithmes rendaient de grands services pour effectuer des multiplications, des divisions et des calculs de racines. Et cette époque n'est pas si lointaine, les calculatrices n'ont vraiment été accessibles à tous qu'au cours des années 80 !

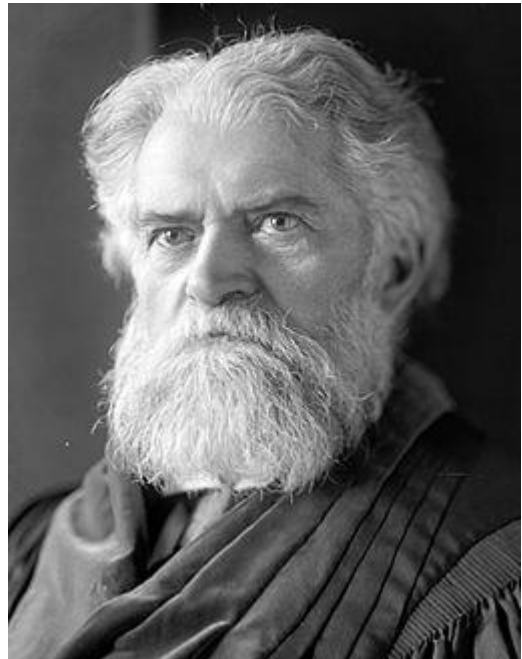
Or dans une table de logarithmes, on trouve des listes de nombres classés par mantisse croissante, par conséquent d'abord des pages de nombres commençant par 1, puis 2, etc.

Logiquement, NEWCOMB en tire la conclusion stupéfiante qu'on utilise plus de nombres avec des petites mantisses qu'avec des plus grandes ! Cela paraissait absurde et il a voulu comprendre pourquoi.

Après observation d'un grand ensemble de données (des milliers) d'origines diverses, il se rend compte que bien souvent, **la proportion de chaque chiffre c comme premier chiffre significatif des valeurs est relativement stable et peut être raisonnablement approchée**

par un log (et plus précisément par $\log \frac{c+1}{c}$).

A cette époque, personne n'a prêté la moindre attention à son idée et l'article qu'il publie dans l'*American Journal of Mathematics* est vite oublié !



Ignorant les travaux de NEWCOMB, 57 ans plus tard, en 1938, le physicien Frank BENFORD (1883-1948) fait la même constatation et ... devient célèbre !

Il rassemble plus de 20 000 données numériques de diverses provenances et énonce, sans le savoir, la même loi empirique que son prédécesseur. Plus tard, elle portera son nom mais, lui, publie un article dans lequel il l'appelle « **Loi des nombres anormaux** ».



Un ensemble de valeurs numériques suit la **loi de BENFORD** lorsque, pour chaque chiffre c (donc de 1 à 9),
la proportion de valeurs commençant par c vaut $\log \frac{c+1}{c}$.

Autrement dit, en terme de probabilités, on a

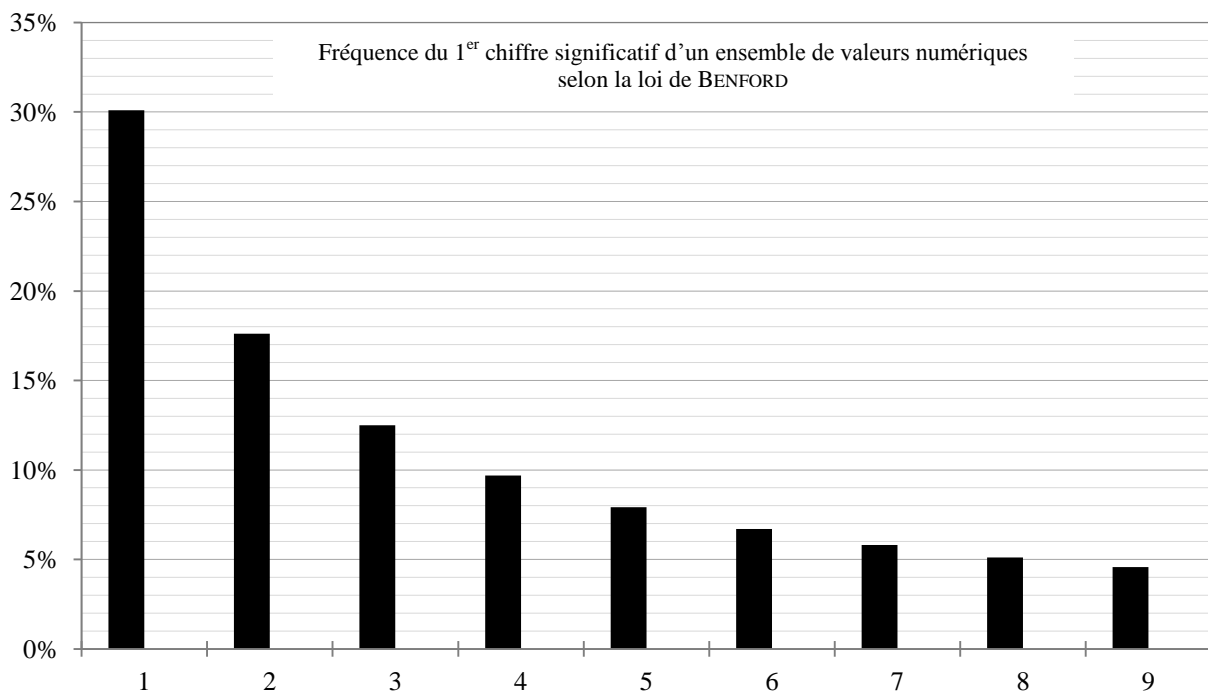
$$P(c \leq \text{mantissee} < c + 1) = \log \frac{c+1}{c}$$

Cette loi, très surprenante, est absolument empirique mais elle colle relativement bien à la réalité pour pas mal de situations.

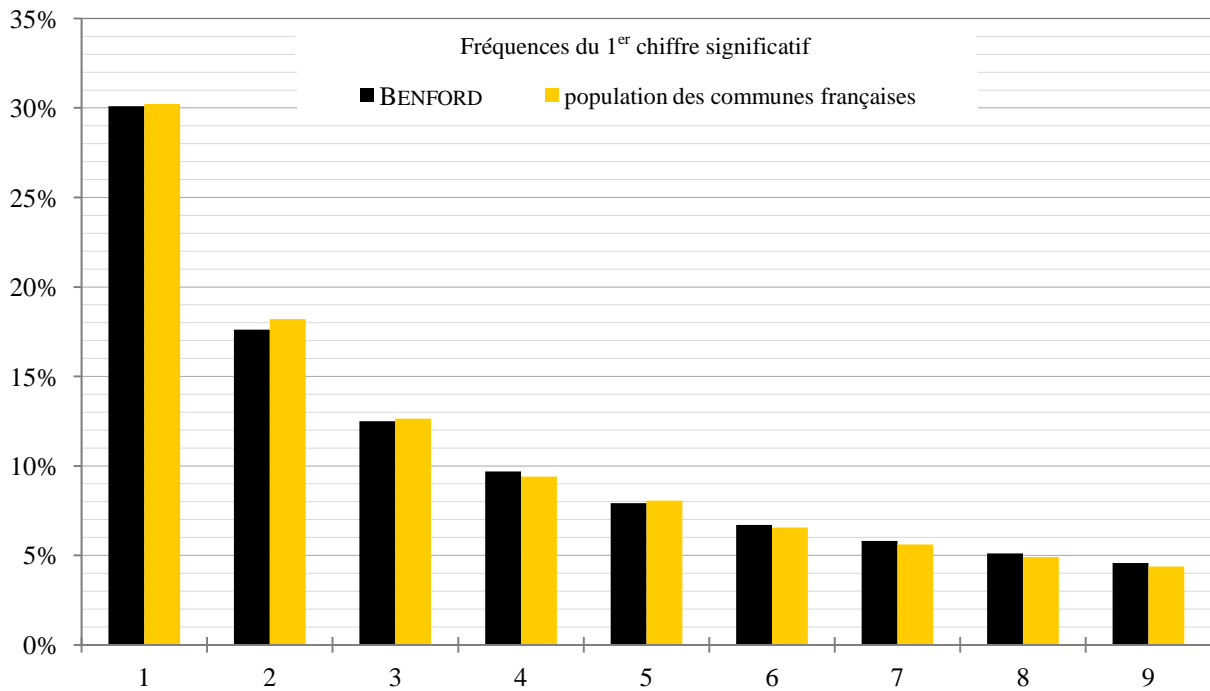
Elle a été longtemps considérée comme une aimable curiosité mathématique qui intrigue et fait un peu rêver à l'existence de lois encore inconnues régissant les mesures dans l'univers.

Il n'y a qu'une bonne vingtaine d'années que cette loi a trouvé des applications sous l'impulsion de l'économiste américain Mark NIGRINI.

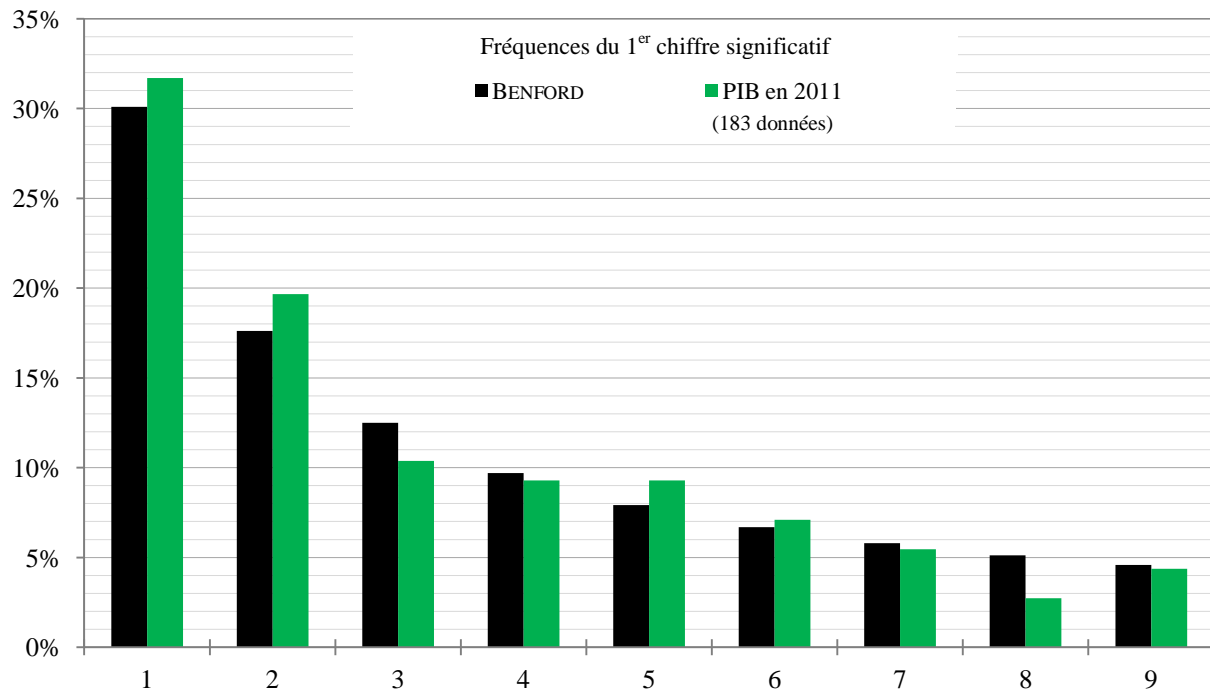
Premier chiffre significatif c	Fréquence théorique selon la loi de BENFORD $\log \frac{c+1}{c}$
1	$\log \frac{1+1}{1} = \log 2 = 0,301$
2	$\log \frac{2+1}{2} = \log \frac{3}{2} = 0,176$
3	$\log \frac{3+1}{3} = \log \frac{4}{3} = 0,125$
4	$\log \frac{4+1}{4} = \log \frac{5}{4} = 0,097$
5	$\log \frac{5+1}{5} = \log \frac{6}{5} = 0,079$
6	$\log \frac{6+1}{6} = \log \frac{7}{6} = 0,067$
7	$\log \frac{7+1}{7} = \log \frac{8}{7} = 0,058$
8	$\log \frac{8+1}{8} = \log \frac{9}{8} = 0,051$
9	$\log \frac{9+1}{9} = \log \frac{10}{9} = 0,046$
Total	1



De nombreux exemples de la vie quotidienne illustrent cette loi. Ainsi en va-t-il, comme nous l'avons vu plus haut, du nombre d'habitants des 36 722 communes françaises au 1^{er} janvier 2009.

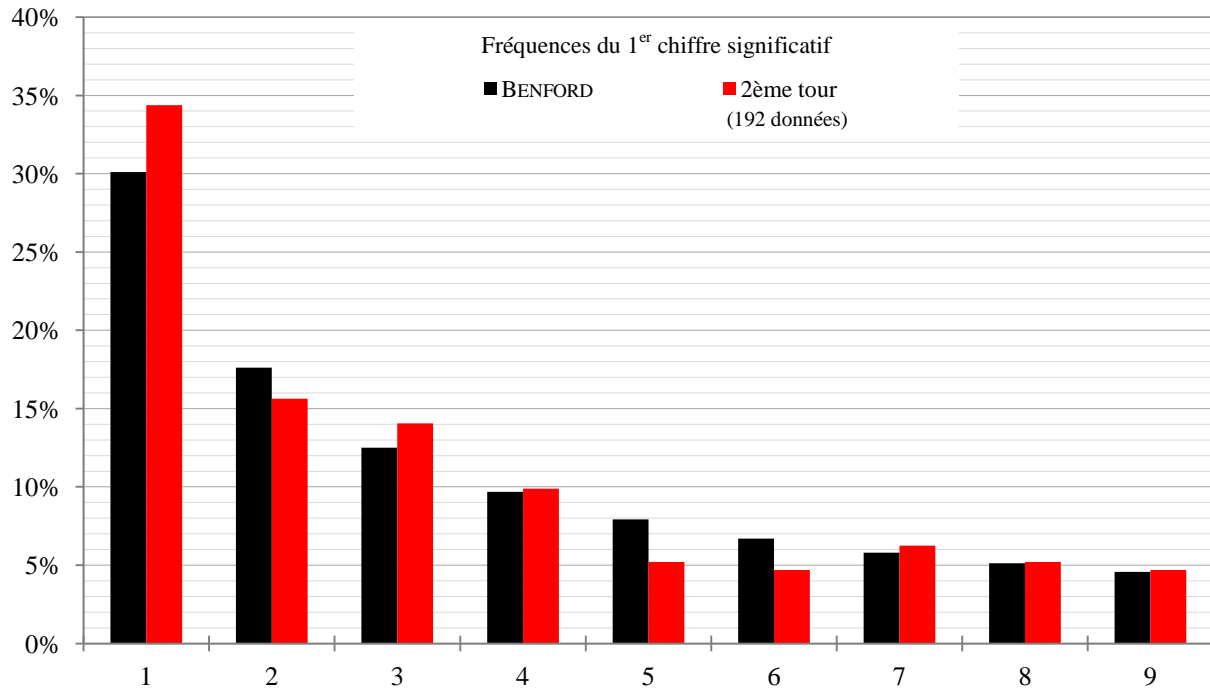


Le PIB pour 2011 et en dollars américains de l'ensemble de (presque) tous les pays du monde suit assez bien aussi la loi de BENFORD.

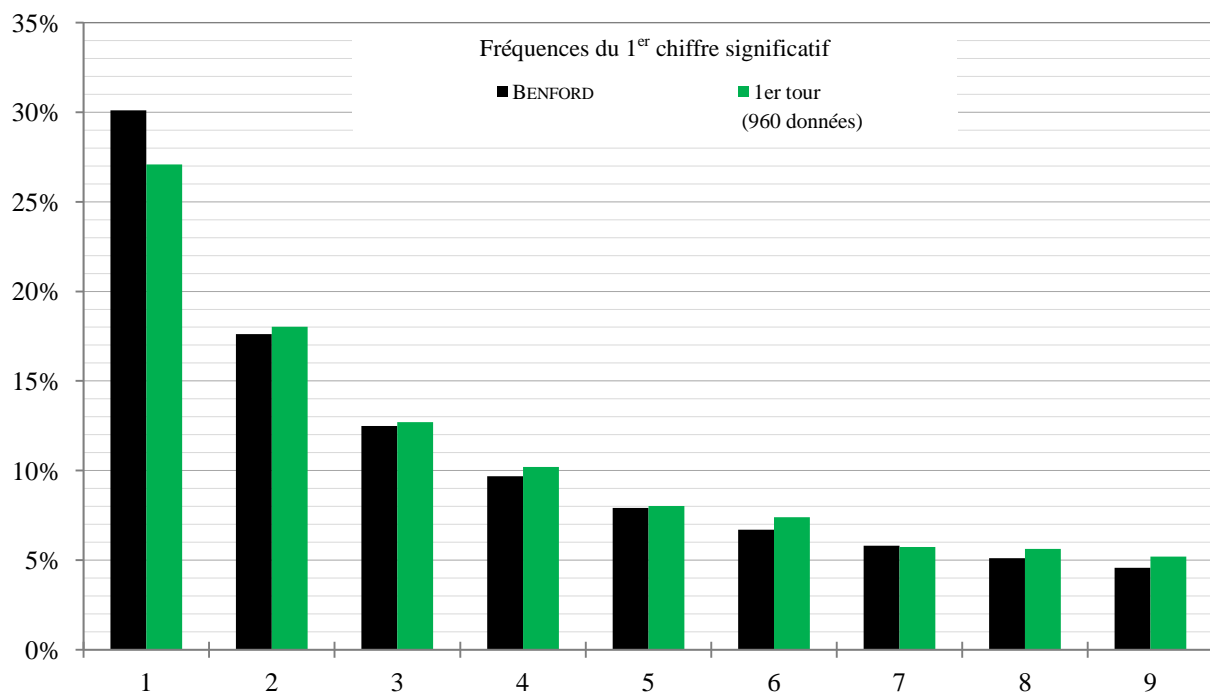


Intéressons-nous maintenant aux nombres de votes obtenus par les différents candidats aux dernières élections présidentielles françaises pour les 96 départements métropolitains.

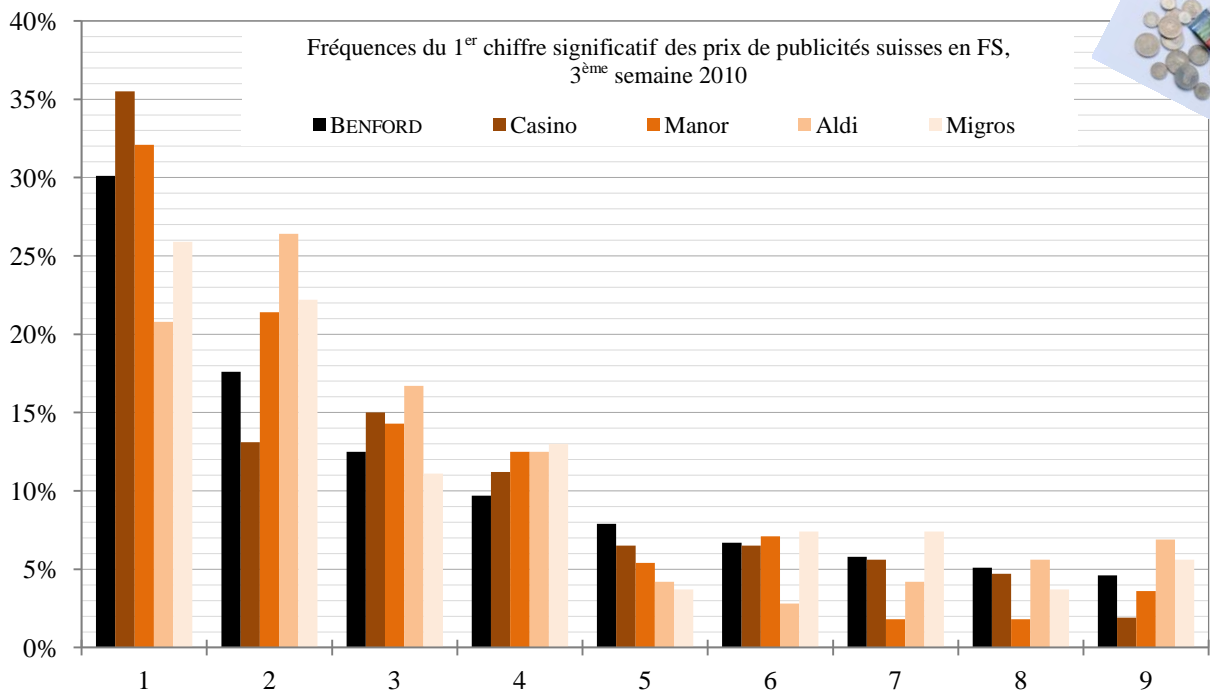
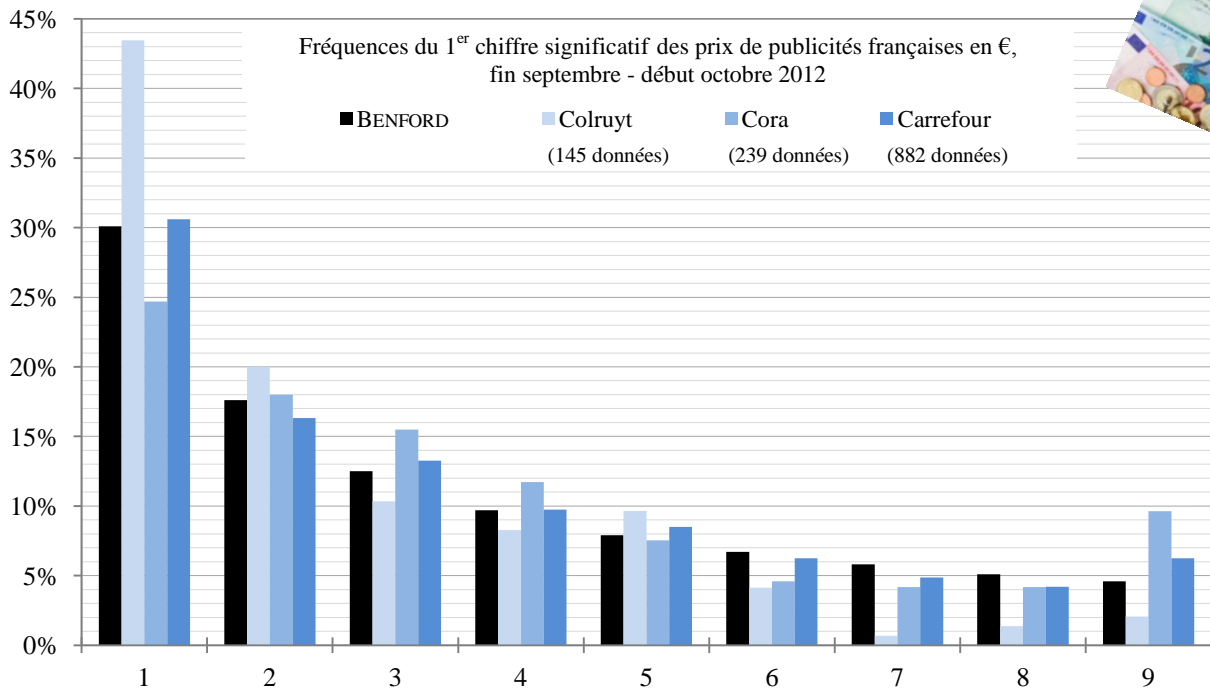
D'abord avec relativement peu de données en considérant le 2^{ème} tour, le 6 mai 2012 : il y a deux candidats Hollande et Sarkozy, donc $2 \times 96 = 192$ données.



Voyons ce qui se passe lorsque le nombre de données augmente en considérant maintenant le 1^{er} tour, le 22 avril 2012 : 10 candidats, donc $10 \times 96 = 960$ données.



Et qu'en est-il de listes de prix dans des grands magasins ? Voici deux exemples.



Les exemples illustrés ici montrent que le fait que des listes de prix suivent assez bien la loi de BENFORD n'est pas propre à une monnaie particulière.

Le lecteur intéressé trouvera d'autres exemples dans le diaporama PowerPoint associé.

IV – Loi de BENFORD généralisée

La loi de BENFORD stipule que

$$P(c \leq \text{mantisse} < c+1) = \log \frac{c+1}{c} \quad \text{avec } c \text{ entier entre 1 et 9}$$

Loi de BENFORD généralisée

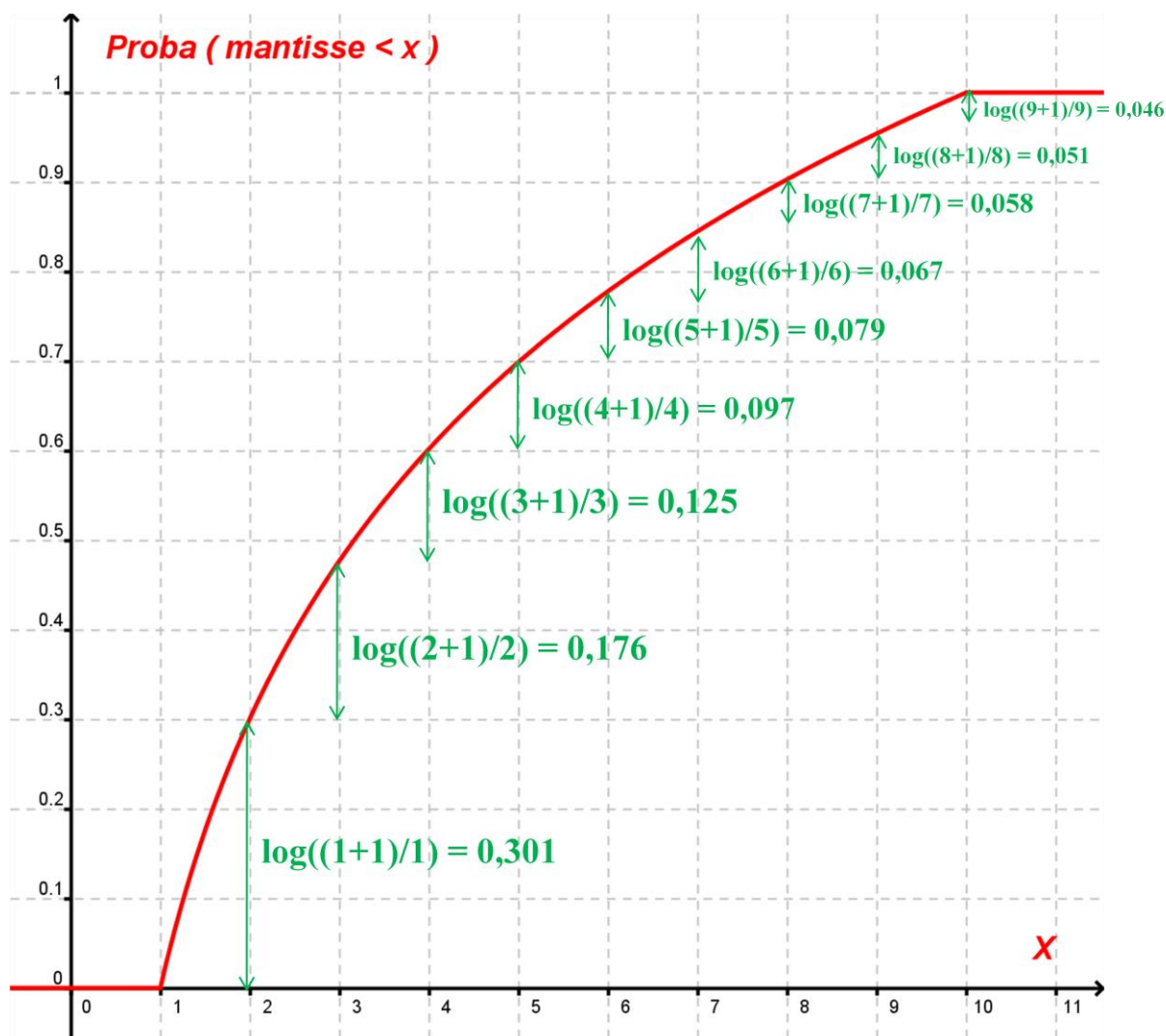
$$P(a \leq \text{mantisse} < b) = \log \frac{b}{a} \quad \text{avec } a \text{ et } b \text{ réels entre 1 et 10 et } a < b$$

En prenant $a = 1$ et $b = x$, il en découle

$$P(1 \leq \text{mantisse} < x) = \log \frac{x}{1} = \log x \quad \text{avec } x \text{ entier entre 1 et 10}$$

et donc

$$P(\text{mantisse} < x) = \log x$$



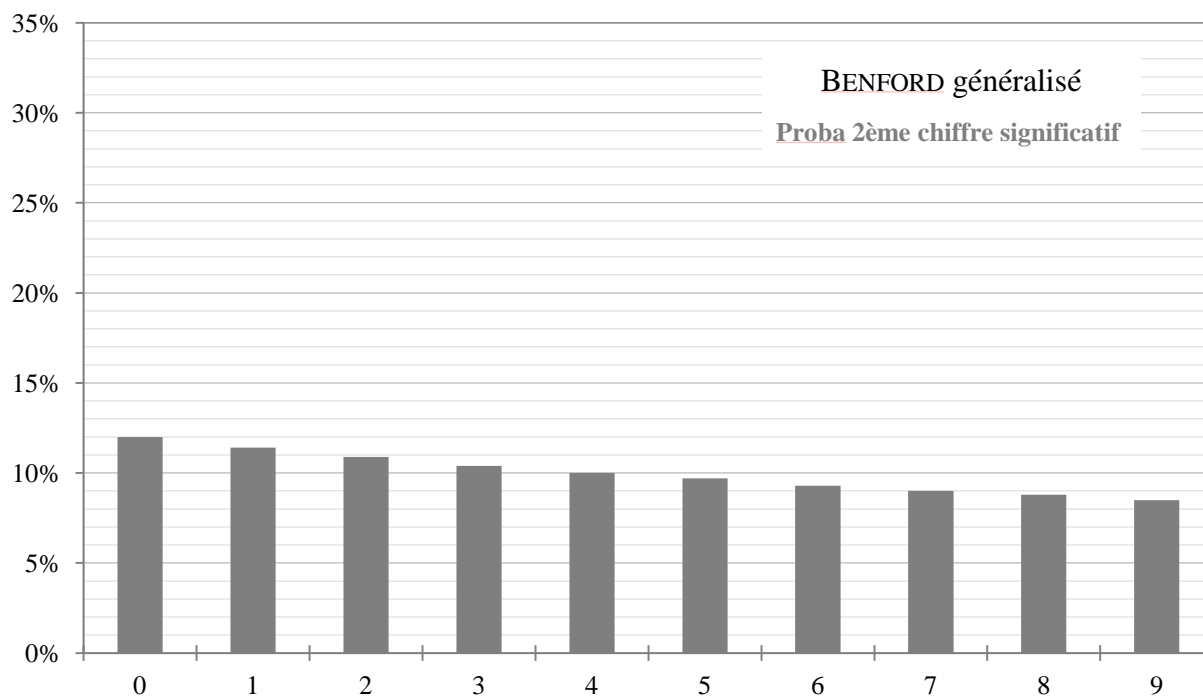
Calculons avec la loi de BENFORD généralisée les probabilités des chiffres 0 à 9 comme 2^{ème} chiffre significatif. Prenons par exemple le chiffre 3, la probabilité de 3 comme 2^{ème} chiffre significatif se calcule via le tableau suivant :

Premier chiffre significatif i	$P(2^{\text{ème}} \text{ chiffre significatif} = 3 \mid 1^{\text{er}} \text{ chiffre significatif} = i)$ $= P(i,3 \leq \text{mantisse} < i,4) = \log \frac{i,4}{i,3}$
1	$\log \frac{1,4}{1,3} = 0,032$
2	$\log \frac{2,4}{2,3} = 0,018$
3	$\log \frac{3,4}{3,3} = 0,013$
4	$\log \frac{4,4}{4,3} = 0,010$
5	$\log \frac{5,4}{5,3} = 0,008$
6	$\log \frac{6,4}{6,3} = 0,007$
7	$\log \frac{7,4}{7,3} = 0,006$
8	$\log \frac{8,4}{8,3} = 0,005$
9	$\log \frac{9,4}{9,3} = 0,005$
Total $= P(2^{\text{ème}} \text{ chiffre significatif} = 3)$	0,104

En effectuant un calcul analogue pour chaque chiffre de 0 à 9, on trouve :

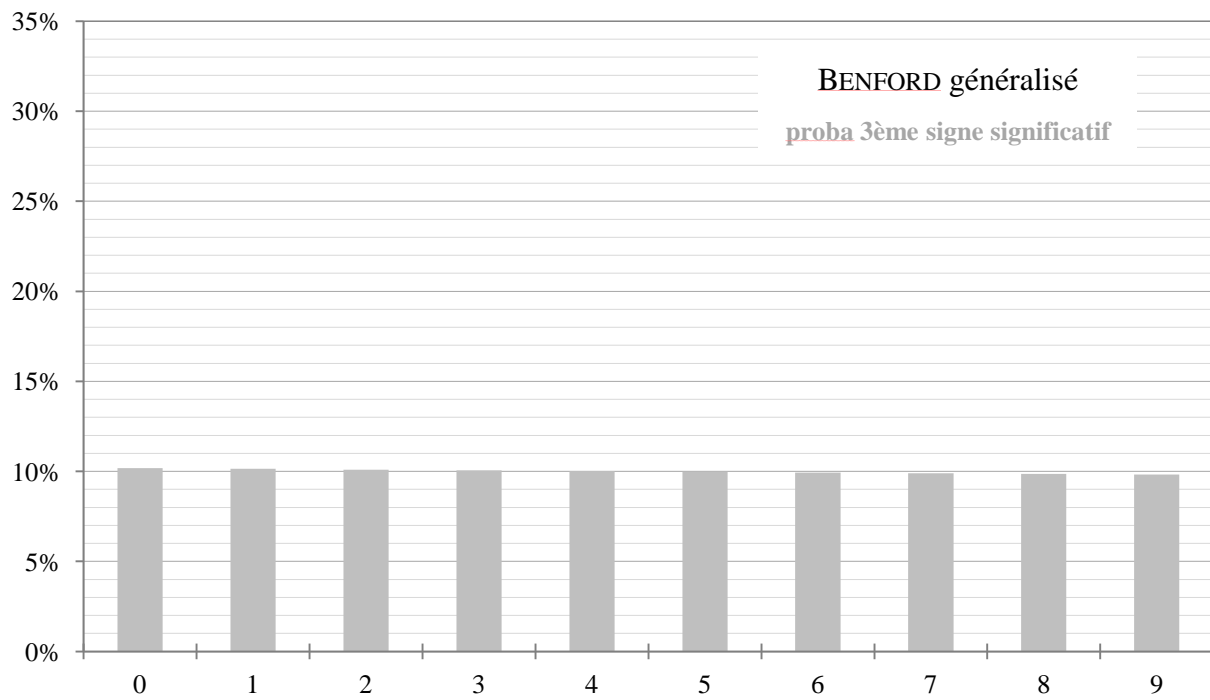
Deuxième chiffre significatif i	$P(2^{\text{ème}} \text{ chiffre significatif} = i)$
0	0,120
1	0,114
2	0,109
3	0,104
4	0,100
5	0,097
6	0,093
7	0,090
8	0,088
9	0,085

Total	1
-------	---



De la même manière, on calcule les probabilités des différents chiffres comme 3^{ème} chiffre significatif.

Troisième chiffre significatif i	$P(3^{\text{ème}} \text{ chiffre significatif} = i)$
0	0,1018
1	0,1014
2	0,1010
3	0,1006
4	0,1002
5	0,0998
6	0,0994
7	0,0990
8	0,0986
9	0,0983
Total	1



Avec la loi de BENFORD généralisée, plus un chiffre est loin à droite du 1^{er} chiffre significatif, plus il est distribué uniformément, plus il se distribue donc conformément à notre intuition...

V – Invariance par changement d'unités

Si une série de données suit la loi de BENFORD généralisée, alors cette loi est également suivie après un changement d'unités !

Ce résultat date de 1961 et a été démontré par l'américain Roger PINKHAM.

Les exemples précédents avec des prix montraient que la loi de BENFORD est à peu près suivie par des listes de prix même assez courtes (50 à 250 prix) ; qu'elles se réfèrent à des prix en FS ou en €. Ce n'est donc pas particulier à une unité.

Mais PINKHAM va plus loin et prouve que si on considère un ensemble de prix en € qui suivent à peu près la loi de BENFORD généralisée, ce sera toujours le cas si on les convertit en FS ou en \$!

De même, pour une liste de longueurs exprimées en km ou en miles.

Roger PINKHAM a même montré que **la loi de BENFORD est l'unique formulation pour obtenir une loi invariante par changement d'échelle.**

Un peu de poésie...

La loi de BENFORD a beaucoup fait rêver les scientifiques et en a amené plus d'un à se poser la question : existe-t-il dans la nature, une sorte de loi universelle régissant la proportion de chacun des chiffres 1 à 9 comme 1^{er} chiffre significatif ?

Si une telle loi existe, elle doit forcément être valable indépendamment des unités de mesure humaines et par conséquent, en tenant compte du résultat de PINKHAM, c'est forcément la loi de BENFORD...

Reste à savoir si une telle loi naturelle régit vraiment certaines mesures...

VI – Analyse des chiffres

Développée au début des années 90 par les travaux de l'économiste américain Mark NIGRINI, l'analyse des chiffres passe en revue les chiffres de données numériques pour lesquels on sait d'expérience qu'ils ont de bonnes chances de suivre raisonnablement certaines fréquences (comme par exemple celles prédites par la loi de BENFORD). Le but est de mettre en évidence des invraisemblances qui proviendraient de données pas très honnêtes, falsifiées ou carrément inventées.

Cela se pratique depuis une vingtaine d'années déjà au Canada et aux USA. Et récemment, l'analyse des chiffres s'est introduite en Europe.

APPLICATION 1:

Détection de fraudes (erreurs ou falsifications de données) dans les comptabilités !

Etats-Unis

Il y a déjà une vingtaine d'années, Mark NIGRINI, professeur de comptabilité au Texas, s'est convaincu, en étudiant des masses de données comptables, que la loi de BENFORD constitue un outil à ne pas négliger pour démasquer d'éventuelles fraudes. Il travaille en collaboration avec le fisc américain pour l'aider à démasquer des fraudeurs.

Constatations expérimentales :

- | |
|---|
| <ol style="list-style-type: none">1) Des données « honnêtes » suivent <i>assez souvent</i> la loi de BENFORD.2) Lorsqu'il y a fraude, les valeurs suivent <i>rarement</i> la loi de BENFORD. |
|---|

En 1993, cette méthode a permis à l'état d'Arizona de récupérer 2 millions de dollars détournés par un employé du trésor. Un test de comparaison avec la loi de BENFORD l'avait rendu suspect.

<p>Attention, l'éloignement à la loi de BENFORD peut amener une suspicion de fraude mais ce n'est en aucun cas une preuve, d'autant plus que des comptabilités tout à fait honnêtes peuvent s'en éloigner très fortement !</p>

Rien ne permet d'affirmer non plus que des données comptables qui suivent la loi de BENFORD soient nécessairement honnêtes !

Afrique du Sud

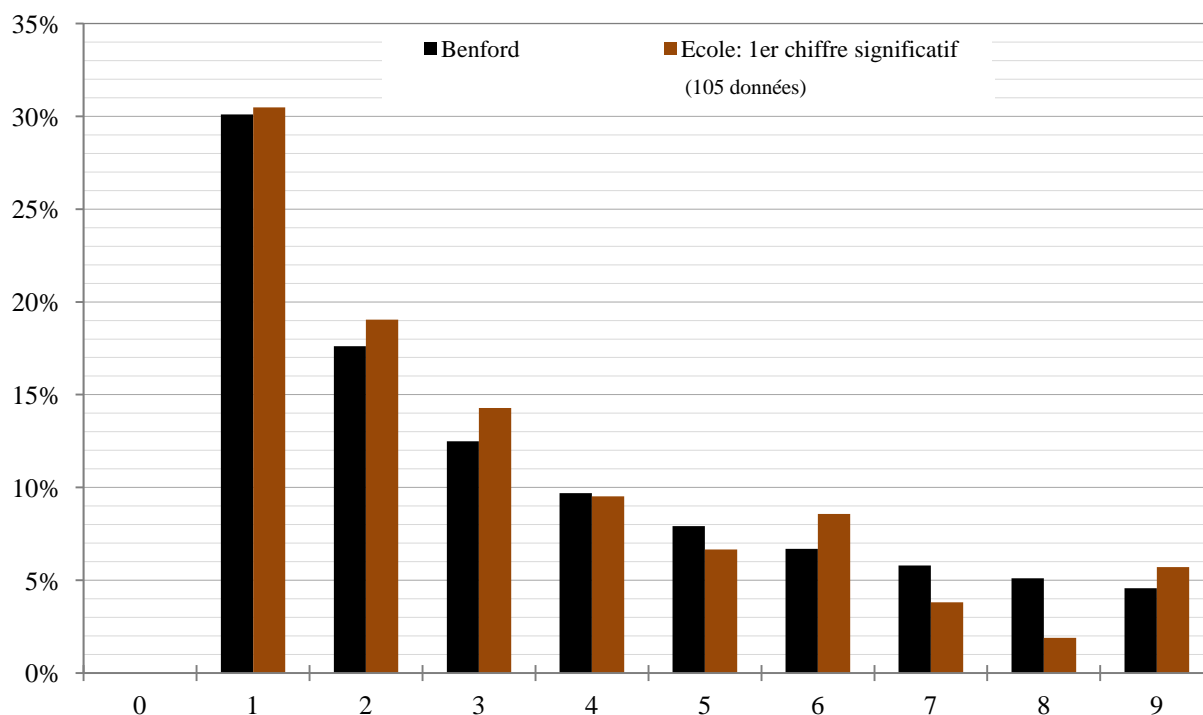
En 2006, une étude a été réalisée par A. SAVILLE de l'Université de Prétoria et a donné les résultats suivants :

- Un test statistique de compatibilité à la loi de BENFORD a été appliqué aux comptes (qu'on savait manipulés) de 17 entreprises et la loi de BENFORD n'était respectée dans aucun des 17 cas !
- Le même test a également été appliqué à 17 entreprises réputées honnêtes et là 4 comptabilités ne satisfaisaient pas la loi de BENFORD.

Tout cela justifie pleinement l'emploi d'un test de compatibilité à la loi de BENFORD pour démasquer d'éventuels fraudeurs mais nous rappelle aussi que la non-compatibilité n'est nullement une preuve de culpabilité.

Exemple 1er chiffre significatif des comptes d'une école neuchâteloise (Suisse) en 2011

Total des charges	44 494 631,89
Total des revenus	16 647 254,85 -
Charges/Revenus(-)	27 847 377,04
Charges de personnel	36 990 028,26
Personnel administratif et exploitation	3 233 951,65
Salaires occasionnels	52 000,15
Traitements du personnel enseignant	26 250 899,30
AVS, ALFA, chômage, accidents LAA	281 320,95
AVS, ALFA, AC, accidents LAA, autres	2 225 339,56
Caisse de pensions	352 946,20
Caisse de pensions, autres	2 895 846,05
Caisse de remplacement	245 123,91
Personnel temporaire facturés par tiers	1 377 819,14
Frais, formation et perfectionnement	65 152,25
Autres charges de personnel	9 629,10
Biens, services et marchandises	5 333 546,38
Imprimés et reliures	46 921,32
Matériel d'enseignement	139 446,26
Information	62 944,90
Machines, mobilier et équipement	907 771,69
Équipement informatique STS2	99 199,51
Eau, gaz, chauffage, électricité	586 367,15
Matières premières	197 246,59
⋮	⋮



Bien que les données soient peu nombreuses (car peu détaillées), bien moins nombreuses que lorsqu'on teste toute la comptabilité d'une entreprise, on voit déjà qu'il y a quelque chose de la loi de BENFORD.

Bien entendu, des données comptables ne suivent jamais exactement la loi de BENFORD. Alors comment décider, en pratique, si des données peuvent raisonnablement être considérées comme suivant la loi de BENFORD ou si elles s'en écartent significativement ?

Ici, la statistique nous vient en aide en offrant des techniques (les tests d'hypothèse) pour trancher cette question « objectivement ». Mais que faut-il comprendre par là ? Simplement que la réponse à la question ne dépend pas de la personne qui doit y répondre... Il ne manquerait plus que cela !

Test d'hypothèse

Un test d'hypothèse sert à trancher si des écarts entre des données relevées au hasard dans un document à contrôler et une loi théorique sont significatifs ou si les différences constatées sont imputables au seul hasard de l'échantillonnage.

Nous donnons ici une idée volontairement très générale et simpliste (pour être accessible aux non spécialistes) du déroulement d'un tel test. En effet, ce chapitre des statistiques est très vaste et assez complexe à saisir.

- 1) On écrit l'hypothèse qu'on souhaite tester, appelée hypothèse nulle et notée H_0 : « Les données suivent la loi théorique. ».
- 2) Un test d'hypothèse est toujours assorti d'un ou de plusieurs nombres qui quantifient le degré de fiabilité qu'on peut accorder aux résultats. En effet, puisqu'on travaille avec un échantillon de valeurs tirées au hasard dans les données, on ne peut jamais être certain à 100 % de la conclusion.

Le plus souvent, on se contente de choisir un nombre α (entre 0 et 1), appelé risque d'erreur de 1^{ère} espèce qui quantifie le risque de considérer à tort que les données s'écartent significativement de la loi théorique testée alors que ce ne serait pas vrai. En pratique, ce risque α est choisi avant de procéder au test d'hypothèse et on prend souvent $\alpha = 10 \%$, 5% , 2% , 1% ou 1% .

Revenons aux données extraites des comptes de l'école suisse considérée plus haut.

La question posée ici est :

Les écarts entre les données comptables et la loi de BENFORD sont-ils significatifs?

Nous allons donc tester l'hypothèse suivante :

Hypothèse nulle H_0 : « Les données suivent la loi de BENFORD. »

Reste à choisir un degré de fiabilité, par exemple $\alpha = 0,01$ (ou 0,02 ou 0,05 ou 0,10) qui quantifie le risque de rejeter l'hypothèse H_0 alors qu'elle est vraie, donc de conclure à tort que les données comptables s'écartent significativement de la loi de BENFORD.

Un test du khi-carré χ^2 permet de conclure qu'il n'y a pas de raison de rejeter l'hypothèse H_0 qui dit que les données suivent la loi de BENFORD (attention, cette conclusion est moins forte que de dire : les données suivent la loi de BENFORD...). Tous les détails concernant ce test peuvent être consultés dans le diaporama PowerPoint associé.

APPLICATION 2

Détection de fraudes plus générales

Une étude en psychologie expérimentale, réalisée à Zurich par A. DICKMANN, a montré que lorsqu'on demande à des gens d'imaginer des données, ils le font sans respecter la loi de BENFORD.

Au mieux, ceux qui connaissent la loi de BENFORD vont y penser pour le 1^{er} chiffre significatif, mais dès qu'on s'intéresse au 2^{ème}, la distribution devient ... à peu près n'importe quoi!

Cela valide l'idée de soupçonner des fraudes dans des listes de valeurs qui d'après des expériences précédentes devraient suivre approximativement la loi de BENFORD et s'en écartent manifestement ; ce peut par exemple être des données d'un inventaire ou des réponses à des questionnaires rentrés par des sondeurs ou des données d'enquêtes en sociologie ou ...

De nos jours sur Internet, on trouve des publicités pour des cours destinés à apprendre à détecter des fraudes dans des données numériques. Certaines de ces publicités citent explicitement la loi de BENFORD.

VII – Contre-exemples

Bien entendu, il existe beaucoup de séries de nombres qui ne suivent pas la loi de BENFORD (même approximativement) !

- 1) Si vous construisez une série de nombres avec un générateur de nombres aléatoires installé sur votre ordinateur, vous devriez en principe avoir une répartition uniforme de leurs 1^{ers} chiffres significatifs. Attention, de tels générateurs ne sont jamais tout à fait aléatoires car leur fonctionnement est déterministe, c'est pourquoi on devrait plutôt parler de nombres pseudo-aléatoires. En tout cas, vous n'obtiendrez pas une série de nombres obéissant à la loi de BENFORD !
- 2) De même, la loi de BENFORD ne vous sera d'aucune aide pour augmenter vos chances de gagner à une loterie ! En effet, le tirage est censé être aléatoire et chaque chiffre devrait donc avoir une probabilité de 0,1 de sortir en premier.
- 3) Relevez les tailles des adultes que vous croisez. Là non plus, pas de loi de BENFORD !
- 4) Pas plus d'ailleurs que dans la liste des numéros de téléphone figurant sur votre agenda car beaucoup de numéros sont probablement de votre région...
- 5) Les numéros des maisons dans une rue donnée ne suivent pas non plus la loi de BENFORD.

Et vous pouvez imaginer plein d'autres exemples.

VIII – Exemples « mathématiques »

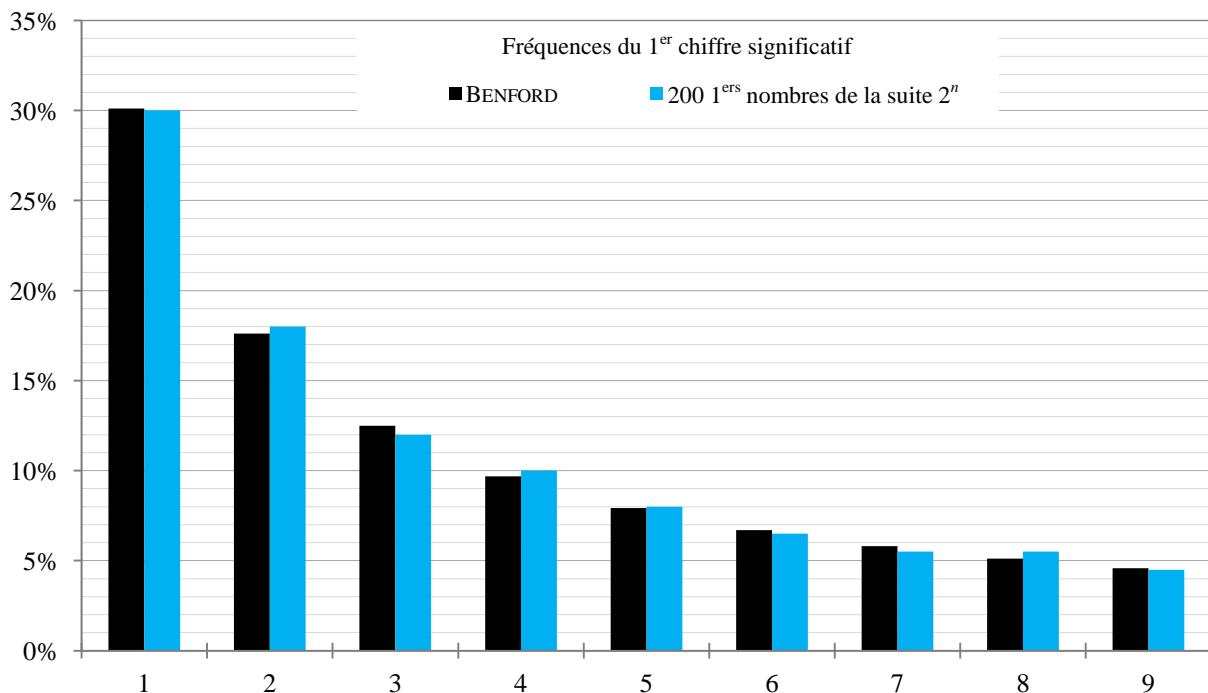
Des mathématiciens ont prouvé que certaines suites de nombres satisfont exactement la loi de BENFORD.

Ainsi, Vladimir ARNOLD et André AVEZ ont démontré que la suite 2^n satisfait la loi de BENFORD. La démonstration n'a rien d'évident.

Et un coup d'œil rapide au début de la liste suivante fait d'abord penser que la suite des 1^{ers} chiffres significatifs est périodique !

n	2^n	1 ^{er} chiffre significatif de 2^n	1 ^{ere} apparition de ce chiffre comme 1 ^{er} chiffre significatif	Remarques
1	2	2	2	Ces 10 1 ^{ers} chiffres significatifs se reproduisent 4 fois de suite !
2	4	4	4	
3	8	8	8	
4	16	1	1	
5	32	3	3	
6	64	6	6	
7	128	1		
8	256	2		
9	512	5	5	
10	1024	1		
11	2048	2		Est-on parti pour une suite périodique de période 10 ?
12	4096	4		
13	8192	8		
14	16384	1		
15	32768	3		
16	65536	6		
17	131072	1		
18	262144	2		
19	524288	5		
20	1048576	1		
21	2097152	2		Oui ? Non ? Peut-être ?
22	4194304	4		
23	8388608	8		
24	16777216	1		
25	33554432	3		
26	67108864	6		
27	134217728	1		
28	268435456	2		
29	536870912	5		
30	1073741824	1		
31	2147483648	2		Le suspense est insoutenable !
32	4294967296	4		
33	8589934592	8		
34	17179869184	1		
35	34359738368	3		
36	68719476736	6		
37	1.37439E+11	1		
38	2.74878E+11	2		
39	5.49756E+11	5		
40	1.09951E+12	1		
41	2.19902E+12	2		Non !!! L'illusion de périodicité s'envole avec $n = 46 \dots$ et l'apparition du 1 ^{er} chiffre 7.
42	4.39805E+12	4		
43	8.79609E+12	8		
44	1.75922E+13	1		
45	3.51844E+13	3		
46	7.03687E+13	7	7	
47	1.40737E+14	1		
48	2.81475E+14	2		
49	5.6295E+14	5		
50	1.1259E+15	1		
51	2.2518E+15	2		On ne l'attendait presque plus ce 1 ^{er} chiffre 9 ...
52	4.5036E+15	4		
53	9.0072E+15	9	9	
54	1.80144E+16	1		
55	3.60288E+16	3		
56	7.20576E+16	7		
57	1.44115E+17	1		
58	2.8823E+17	2		
59	5.76461E+17	5		
60	1.15292E+18	1		
⋮				

Mais on constate qu'au plus l'exposant n grandit, au plus les fréquences d'apparitions des chiffres 1 à 9 comme 1^{ers} chiffres significatifs de la suite 2^n se rapprochent des fréquences de la loi de BENFORD, comme le montre le graphique suivant.



Et ARNOLD et AVEZ ont démontré qu'asymptotiquement, la suite 2^n satisfait à la loi de BENFORD.

Le mathématicien suisse Paul JOLISSAINT a démontré que la célèbre suite de FIBONACCI (1, 1, 2, 3, 5, 8, 13, 21, 34, ...), déjà connue pour plein de propriétés sympathiques ou amusantes, suit elle aussi asymptotiquement la loi de BENFORD !

IX – Pourquoi des suites numériques issues du monde réel se conforment-elles raisonnablement à la loi de BENFORD ?

Cette curiosité fait rêver et plusieurs tentatives d'explications ont été formulées avec plus ou moins de succès.

Ainsi des suites de nombres s'étalant sur plusieurs ordres de grandeur et de manière assez régulière s'approcheraient relativement bien de la loi de BENFORD.

Comme le suggère J.-P. DELAHAYE, peut-être qu'un jour, quelque principe général qui nous échappe encore aujourd'hui, amènera une explication. Voilà un beau sujet de réflexion pour ceux d'entre vous qui souhaitent prolonger...

Pour en savoir plus

- [1] ARNOLD V. et AVEZ A., *Ergodics problems of classical mechanics*, Benjamin, 1968
- [2] BENFORD F., *The law of anomalous numbers*, Proceedings of the American Philosophical Society, 78 (4), 1938, pp 551-772
- [3] DELAHAYE J.-P., *L'étonnante loi de Benford*, Pour la Science, n° 351, 2007, pp 90-95
- [4] JOLISSAINT P., *Loi de Benford, relations de récurrence et suites équidistribuées*, Elemente der Mathematik, 60, 2005, pp 10-18
et www.institut-jurassien.ch/docs/Benford.pdf
- [5] NEWCOMB S., *Note of the frequency of the use of the digits in natural numbers*, American Journal of Mathematics, 1881, 4 pp 39-40

Françoise Valette-Duchêne est professeur à l'ESNE, Ecole Supérieure du canton de Neuchâtel, Suisse

francoise.duchene@rpn.ch