Les populations suivent-elles la loi des nombres anomaux ?

Frédéric SANDRON*

On appelle « premier chiffre significatif » d'un nombre son premier chiffre à gauche différent de zéro. Le premier chiffre significatif du nombre 325 est 3 et celui du nombre 0,8732 est 8. *A priori*, on pourrait penser que la distribution des premiers chiffres significatifs d'une série de nombres quelconques ou mesurant un phénomène quelconque est équiprobable. Or, il n'en est rien : dans une majorité de séries observées, le 1 apparaît plus souvent que le 2, qui lui-même apparaît plus souvent que le 3, etc. Le propos de cette note est d'illustrer cette loi dite de Benford avec les données relatives aux populations de l'ensemble des pays du monde et d'en montrer dans ce cas précis la logique sous-jacente liée au mode de croissance des populations.

I. La loi de Benford

En 1881, l'astronome mathématicien Simon Newcomb remarque que les tables de logarithmes de la bibliothèque de son établissement sont plus usées pour les premiers volumes que pour les suivants ; cela signifie que les consultations des nombres commençant par 1 ou 2 sont plus nombreuses que celles des nombres commençant par 8 ou 9. En 1938, l'ingénieur Benford refait la même constatation à partir des mêmes tables de logarithmes, dans l'ignorance des travaux de Newcomb. Dans son article[1], l'auteur compile de très nombreuses séries de données, allant des constantes physiques aux résultats de la ligue de base-ball en passant par des nombres divers relevés dans les journaux. La moyenne des apparitions des premiers chiffres significatifs suit une loi logarithmique déjà mise en évidence par Newcomb (Benford, 1938) :

$$F_d = \log_{10}(1 + 1/d)$$

— où F_d est la fréquence d'apparition du premier chiffre significatif d.

Le tableau 1 présente la fréquence d'apparition des premiers chiffres significatifs d'après la loi de Benford.

* Institut de recherche pour le développement, Laboratoire Population-Environnement, Université de Provence.

[1] Article intitulé « The law of anomalous numbers », que nous avons traduit au plus près par « La loi des nombres anomaux », c'est-à-dire irréguliers.

Population-F, 57(4-5), 2002, 761-768

TABLEAU 1.– FRÉQUENCE THÉORIQUE D'APPARITION DU PREMIER CHIFFRE SIGNIFICATIF SELON LA LOI DE BENFORD

Premier chiffre significatif	Fréquence théorique
1	0,301
2	0,176
3	0,125
4	0,097
5	0,079
6	0,067
7	0,058
8	0,051
9	0,046

On vérifie que la somme de ces probabilités est bien égale à 1 :

$$\sum_{i=1}^{i=9} F_i = \sum_{i=1}^{i=9} \log_{10}(1+1/i) = \sum_{i=1}^{i=9} \log_{10}[(i+1)/i] = \log_{10} \prod_{i=1}^{i=9} [(i+1)/i]$$

$$= \log_{10}\left(\frac{2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9}\right)$$

$$= \log_{10}(10) = 1$$

Il est en réalité difficile de généraliser cette loi; on trouve aussi des séries numériques qui ne la suivent pas. Mais Hill (1995) a mis en évidence que dans ce cas, les échantillons aléatoires tirés dans la réunion de leurs éléments suivaient la loi de Benford. Voyons ce qu'il en est pour la population des différents pays du monde contemporain.

II. La distribution de la population mondiale

L'Ined publie régulièrement une liste d'indicateurs sur les pays du monde. Nous utiliserons ici celle de 1997 qui renseigne sur 198 pays ou entités géopolitiques. La figure 1 permet de comparer la distribution théorique de la loi de Benford et celle réellement observée pour l'effectif de la population des différents pays. L'écart est extrêmement ténu et l'on peut affirmer que la distribution de la population des pays suit effectivement une loi de Benford.

Si l'on effectue le même type de calculs sur les superficies et sur les densités, les résultats sont de même nature (figures 2 et 3).

III. Une explication pour les effectifs de population

Mentionnons d'abord qu'il n'y a pas de biais d'arrondi pour les pays peu peuplés en faveur du chiffre significatif 1, les plus petits nombres de la table étant 0,02 (million).

L'explication est ailleurs. Pour la cerner, voyons quelle est la distribution des effectifs d'une population en croissance régulière, par exemple de 2 % l'an. Le tableau 2 montre qu'à taux de croissance constant, les effectifs de population dont le premier chiffre significatif est 1 sont plus nombreux que les autres et suivent la loi de Benford. Si l'on prolonge ces calculs à long terme, les fréquences observées pour les premiers chiffres significatifs suivent une parfaite loi de Benford, et ceci reste vrai avec d'autres taux de croissance.

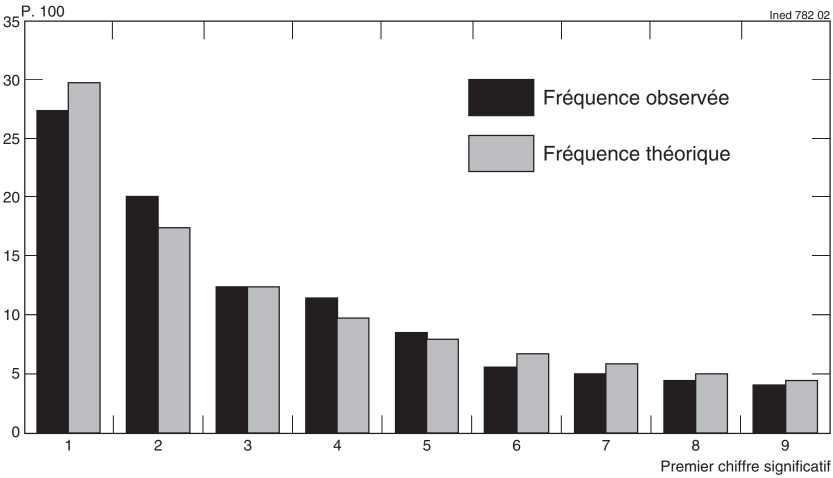

Figure 1. – Fréquence observée pour le premier chiffre significatif de l'effectif de la population de 198 pays en 1997 et fréquence théorique d'après la loi de Benford

Source : Ined 1997.

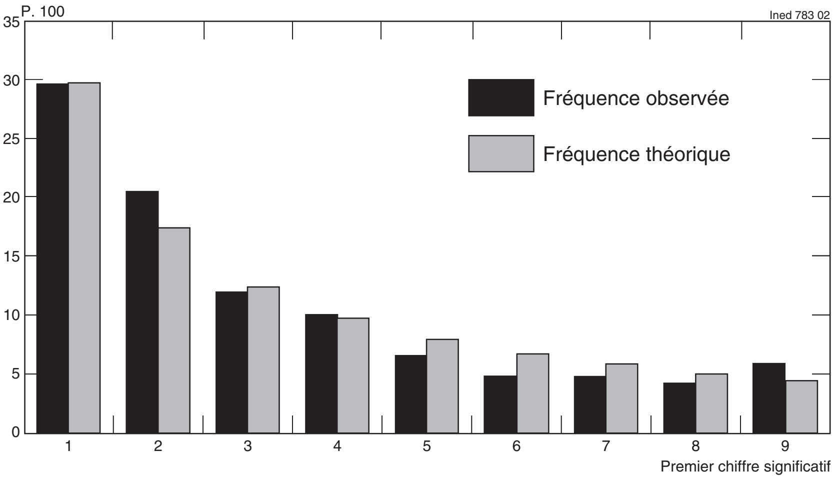

Figure 2. – Fréquence observée pour le premier chiffre significatif de la superficie de 198 pays en 1997 et fréquence théorique d'après la loi de Benford

Source : Ined 1997.

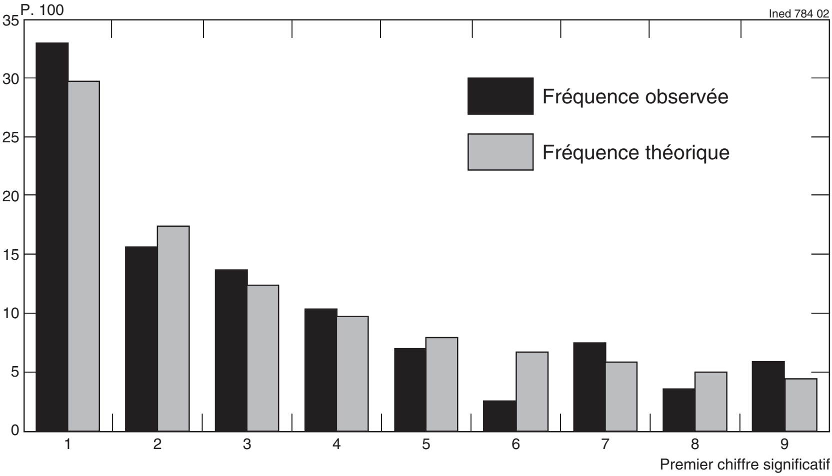

Figure 3.–Fréquence observée pour le premier chiffre significatif de la densité de population de 198 pays en 1997 et fréquence théorique d'après la loi de Benford
Source : Ined 1997.

Les démographes utilisent cette petite astuce qui consiste à diviser le nombre 70 par le taux de croissance (exprimé en %) d'une population pour en connaître le temps de doublement. On retrouve ce résultat de la manière suivante :

— la vitesse de croissance d'une population est définie par $\frac{P(T)}{P(0)} = e^{rT}$;

— avec $P(u)$ = effectif de la population au temps u;
— r = taux de croissance.

Le temps de doublement T_2 se déduit donc de $\frac{P(T_2)}{P(0)} = e^{rT_2} = 2$;

— d'où $rT_2 = \ln\left[\frac{P(T_2)}{P(0)}\right] = \ln 2$;

— soit $T_2 = \frac{\ln 2}{r} \approx \frac{0,693}{r} \approx \frac{0,70}{r} \approx \frac{70}{r'}$ où r' est le taux de croissance exprimé en %.

Pour passer d'un effectif de 100 à 200, il faut donc environ 35 périodes avec un taux de croissance de 2 %.

Pour passer de 200 à 300, il faut 21 périodes puisque :

$$T = \frac{1}{0,02}\ln\left[\frac{300}{200}\right] = \frac{\ln 1,5}{0,02}$$

Il en faut 15 pour passer de 300 à 400 :

$$T = \frac{1}{0,02} \ln\left[\frac{400}{300}\right] \approx \frac{\ln 1,333}{0,02}$$

et seulement 5 pour passer de 800 à 900 :

$$T = \frac{1}{0,02} \ln\left[\frac{900}{800}\right] = \frac{\ln 1,125}{0,02}$$

Le nombre de périodes nécessaires pour passer d'un effectif de population de la forme *(d)00* à *(d + 1)00* est donné par la formule :

$$T = \left(\frac{d+1}{d}\right) / r = \frac{1}{M} \log_{10}\left(\frac{d+1}{d}\right) / r = \frac{1}{M} F_d / r$$

— avec $\ln(x) = \frac{1}{M} \log_{10}(x)$

— où M est une constante

— et F_d la fréquence d'apparition du premier chiffre significatif d.

Cette formule est donc déterminée par la distribution de la fréquence des premiers chiffres significatifs F_d qui suit la loi de Benford. Sur longue période, le premier chiffre significatif de l'effectif de la population d'un pays quelconque est donc plus souvent le 1 que le 2, le 2 que le 3 et ainsi de suite jusqu'à 9.

Chercher la validité transversale de cette régularité longitudinale revient à effectuer un tirage aléatoire dans un ensemble de 198 séries suivant une loi de Benford, à condition que les deux hypothèses suivantes soient vérifiées, ce que l'on peut raisonnablement supposer :

Hypothèse 1 : les effectifs des populations des pays sont indépendants ;

Hypothèse 2 : le début de la transition démographique d'un pays donné est indépendante du premier chiffre significatif de l'effectif de sa population.

Conclusion

Cette explication de l'adéquation des séries de données de population à la loi de Benford ne doit pas occulter l'existence d'autres séries de données non longitudinales et obéissant néanmoins à la loi de Benford, comme celles des superficies des pays par exemple. Si certaines avancées effectuées grâce à la théorie des probabilités ont pu permettre de mieux comprendre les principaux aspects de cette loi, il n'en reste pas moins qu'une explication complètement satisfaisante reste à développer (Hill, 1999).

TABLEAU 2.– ÉVOLUTION DE L'EFFECTIF D'UNE POPULATION SUR 172 PÉRIODES AVEC UN TAUX DE CROISSANCE DE 2 % *

Périodes										
1- 36	37-56	57-71	72-82	83-91	92-99	100-106	107-111	112-117	118-152	153-172
100,00	203,99	303,12	407,95	507,24	606,20	710,26	815,86	900,78	1014,43	2028,74
102,00	208,07	309,18	416,11	517,39	618,32	724,46	832,18	918,80	1034,71	2069,31
104,04	212,23	315,36	424,44	527,73	630,69	738,95	848,83	937,17	1055,41	2110,70
106,12	216,47	321,67	432,93	538,29	643,30	753,73	865,80	955,92	1076,52	2152,91
108,24	220,80	328,10	441,58	549,05	656,17	768,81	883,12	975,03	1098,05	2195,97
110,41	225,22	334,67	450,42	560,03	669,29	784,18		994,53	1120,01	2239,89
112,62	229,72	341,36	459,42	571,24	682,68	799,87			1142,41	2284,69
114,87	234,32	348,19	468,61	582,66	696,33				1165,26	2330,38
117,17	239,01	355,15	477,98	594,31					1188,56	2376,99
119,51	243,79	362,25	487,54						1212,33	2424,53
121,90	248,66	369,50	497,29						1236,58	2473,02
124,34	253,63	376,89							1261,31	2522,48
126,82	258,71	384,43							1286,54	2572,93
129,36	263,88	392,11							1312,27	2624,39
131,95	269,16	399,96							1338,51	2676,88
134,59	274,54								1365,28	2730,42
137,28	280,03								1392,59	2785,02
140,02	285,63								1420,44	2840,72
142,82	291,35								1448,85	2897,54
145,68	297,17								1477,83	2955,49
148,59									1507,38	
151,57									1537,53	
154,60									1568,28	
157,69									1599,65	
160,84									1631,64	
164,06									1664,27	
167,34									1697,56	
170,69									1731,51	
174,10									1766,14	
177,58									1801,46	
181,14									1837,49	
184,76									1874,24	
188,45									1911,73	
192,22									1949,96	
196,07									1988,96	
199,99										

* Le tableau se lit séquentiellement en colonnes.

Remerciements : L'auteur remercie les rapporteurs anonymes, ainsi qu'Éva Lelièvre et Etienne van de Walle pour leurs précieux commentaires.

RÉFÉRENCES

BENFORD F., 1938, « The law of anomalous numbers », *Proceedings of the American Philosophical Society*, 78(4), p. 551-572.

INED, 1997, « Tous les pays du monde », *Population et Sociétés*, n° 326, 4 p.

HILL T., 1995, « A statistical derivation of the significant-digit law », *Statistical Science*, 10(4), p. 354-363.

HILL T., 1999, « Le premier chiffre significatif fait sa loi », *La Recherche*, n° 316, p. 72-75.

NEWCOMB S., 1881, « Note on the frequency of use of the different digits in natural numbers », *Am. J. Mathematics*, n° 4, p. 39-40.