CURIOSITE

La loi de Benford

Ph. Paturel,

Résumé : *La loi de Benford fait partie de l'intéressante famille des curiosités mathématiques et autres paradoxes numériques.*

La loi de Benford s'énonce sans formalisme, de la façon suivante : "dans une liste de nombres mesurant des grandeurs physiques quelconques, on trouvera plus de nombres commençant par le chiffre '1' que de nombres commençant par le chiffre '2'. De la même façon, un premier chiffre '2' sera plus fréquent qu'un premier chiffre '3' et ainsi de suite". Intrigué ? Bien sûr ! On s'attendrait à une fréquence de 1/10, identique pour chaque chiffre ou plutôt 1/9 si on ne compte pas le zéro comme premier chiffre possible. Et pourtant ...

Les fréquences réellement constatées sont en fait les suivantes :

chiffre 1 : 30.1 %
chiffre 2 : 17.6 %
chiffre 3 : 12.5 %
chiffre 4 : 9.7 %
chiffre 5 : 7.9 %
chiffre 6 : 6.7 %
chiffre 7 : 5.8 %
chiffre 8 : 5.1 %
chiffre 9 : 4.6 %

Il y a donc pratiquement une chance sur deux (47.7 %) que le premier chiffre soit 1 ou 2 et presque deux chances sur trois (60.2 %) que le premier chiffre soit 1, 2 ou 3.

N'hésitez pas à faire un essai, en inscrivant par exemple sur une feuille de papier, la longueur d'une sélection de fleuves quelconques, la surface de lacs, le nombre d'habitants de n'importe quelles villes, les prix des articles sur votre dernier ticket de caisse de grande surface, la taille des fichiers sur votre ordinateur, etc. Ce travail a déjà été fait de nombreuses fois et il confirme bien que la fréquence moyenne du premier chiffre des nombres ainsi relevés n'est pas uniforme.

Ce phénomène possède d'intéressantes propriétés. La fréquence du premier chiffre n'est pas modifiée si l'on change les unités de mesure des nombres. Ainsi, si la longueur des fleuves de l'exemple précédent est exprimée en mètres, inch, miles ou multiple de la longueur de votre bras, ou si les prix sont convertis en pesos colombiens, les fréquences restent inchangées. Ainsi, les schtroumpfs qui mesurent les surfaces en unité schtroumpf doivent également rencontrer la même loi. Cette loi possède un caractère universel et invariant par changement d'échelle.

Dans le même ordre d'idée, la conversion des nombres dans une autre base que la base de 10 (base 16, base 8, etc.) maintiendra une différence de fréquence en faveur des petits chiffres. Enfin, les anomalies de fréquences portent également sur les chiffres suivants avec une intensité décroissante au fur et à mesure que l'on avance "vers la droite" de l'écriture des nombres.

Attention, certaines familles de nombres ne conviennent pas, comme on le comprendra plus loin. Les numéros de téléphones, les tirages du loto, les âges, les codes postaux ou les plaques d'immatriculation sont insensibles à cette loi. Ils ne mesurent pas exactement une grandeur physique, ou bien souffrent d'un biais statistiques artificiel.

Cette bizarrerie mathématique a été découverte fortuitement en 1881 par le mathématicien et astronome Simon Newcomb. Il a constaté en effet que les tables de logarithmes utilisées pour les calculs numériques étaient plus usées sur les pages des nombres commençant par les petits chiffres. En d'autres termes, il s'effectuait plus de calculs sur des données commençant par de petits chiffres. Il proposa une formule mathématique donnant la probabilité d'occurrence d'un premier chiffre, loi à base de logarithme (encore !) et publia un article qui passa inaperçu dans la communauté scientifique, peut-être en raison de sa production par ailleurs abondante ou de l'absence de démonstration du phénomène. La formule proposée était correcte ; la probabilité P(D) que le premier chiffre soit "D" est :

$$P(D)= \log_{10}(1+1/D).$$

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 analysant plus de 20 000 données prises dans des domaines très divers. Il confirma la validité de la loi de Simon Newcomb mais avec plus d'écho dans la communauté scientifique, notamment grâce au volume exceptionnel de données étudiées ! Cette loi porte aujourd'hui son nom : la Loi de Benford

Par la suite, les travaux de Roger Pinkham en 1961 ont montré que la formulation mathématique de la loi de Benford était l'unique solution permettant d'obtenir une loi universelle et invariante par changement d'échelle.

L'explication "réelle" du phénomène est venue encore plus tard avec les travaux du mathématicien Theodore Hill en 1996. Certaines grandeurs suivent des distributions de probabilités identifiées : gaussiennes ou courbes en cloches, lois exponentielles, lois uniformes, etc. D'autre part, des données collectées dans des domaines très divers suivront une distribution de probabilité résultant d'un mélange de ces distributions plus simples. De la même façon, des phénomènes macroscopiques du monde réel, sont issus d'une multitude de phénomènes élémentaires en interaction. Leur distribution de probabilité suivra donc une sorte de mélange des distributions de probabilité de chaque phénomène : ce mélange, ou encore cette "distribution des distributions" peut se calculer mathématiquement et c'est la loi de Benford.

Depuis les années 1990, des applications concrètes de la loi de Benford voient le jour, la faisant sortir du simple domaine des curiosités mathématiques. Il a été montré (M.Nigrini, 1992) que les données financières des entreprises suivent cette loi. L'application d'un test statistique classique, le test du chi-2, constitue donc un outil simple et rapide pouvant fournir des présomptions de fraude en indiquant si une série de données n'est pas assez ... naturelle. En effet, un individu qui invente ou modifie des chiffres artificiellement ne suit inconsciemment pas une telle loi. Il ne crée pas assez de nombre commençant par les petits chiffres et crée trop de nombres commençant par 5 ou 6. Les données d'essais cliniques pharmaceutiques, sont également de bonnes candidates pour ce test. La loi de Benford peut aussi participer à la validation de modèles et simulation numériques (démographie, météorologie, etc.). D'autres applications suivront certainement.

Nous venons de voir les principales facettes de cette étonnante loi de Benford. Elle est amusante par son contenu, intéressante par son histoire, prometteuse par ses applications, et vient nous rappeler s'il en est encore besoin, la grande subtilité du monde réel.

■

Effectivement, mais c'est une autre loi, qui d'ailleurs s'applique également aux tables de logarithmes quand on essaye de les lire comme un roman.

Essayez de tester la loi de Benford avec les mesures les plus diverses que vous pourrez trouver et envoyez-nous vos conclusions. Nous les publierons.
NdlR