

proposée dans l'encadré 3, ainsi qu'une analogie avec la croyance en l'équité d'une loterie dans l'encadré 2.

À côté de l'idée intuitive et un peu vague, il y a bien évidemment des formulations précises de conditions mathématiques qui assurent que l'écart à l'uniformité est faible, ou même qu'il est nul ; les lecteurs intéressés pourront les consulter dans les articles de la bibliographie.

L'origine de cette loi dont on va comprendre l'importance centrale est difficile à retracer. En 1912, Henri Poincaré a exprimé un principe équivalent sans énoncer de théorème précis. Il expliquait à propos de la troisième décimale des nombres trouvés dans une table de logarithmes qu'on observera « que les dix chiffres 0, 1, 2, 3, ..., 9 sont également répartis sur cette liste et par conséquent la probabilité pour que cette troisième décimale soit paire est égale à $1/2$ ». Pour lui, cet énoncé d'uniformité était évident ; il écrivit : « [...] un instinct invincible porte à le penser ».

L'idée de l'uniformisation des parties fractionnaires a été reprise par le mathématicien d'origine croate William Feller dans son célèbre traité de théorie des probabilités de 1966, mais comme on l'a noté depuis, en voulant formuler un énoncé mathématique, il s'est trompé. Sans connaître ni le texte de Poincaré ni celui de Feller, et cette fois avec un énoncé et une démonstration exacts, Nicolas Gauvrit et moi avons retrouvé l'idée en 2008, en même temps qu'un autre énoncé précis différent et correct était proposé indépendamment par les chercheurs allemands Lutz Dümbgen et Christoph Leuenberger. L'impossibilité de classer les résultats mathématiques pour les retrouver facilement comme on classe les mots d'un dictionnaire a pour conséquence que, souvent, des résultats qui ne se rattachent pas clairement à un domaine précis des mathématiques sont découverts indépendamment plusieurs fois.

ÉVIDENCE DE LA LOI DE BENFORD

Grâce à cette loi d'étalement à la fois intuitive et formalisable, on va obtenir une justification naturelle et simple de la loi de Benford. L'idée consiste simplement à appliquer un logarithme décimal à la loi précédente... et à réfléchir un peu.

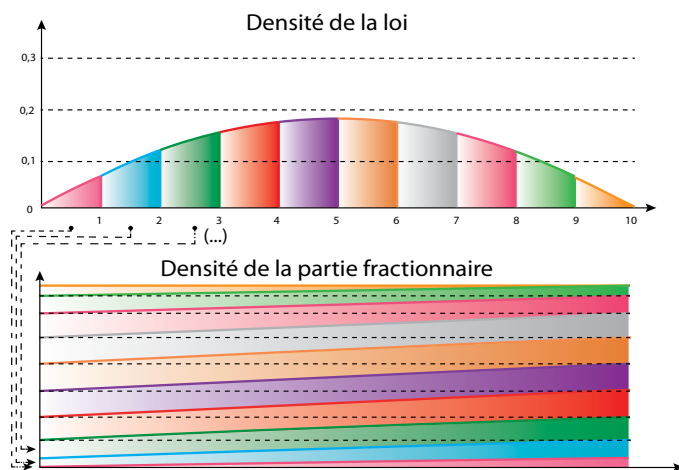
On reprend l'énoncé précédent et on l'applique non pas aux nombres r de la série considérée, mais à leur logarithme décimal, $\log_{10}(r)$. Si l'on choisit des nombres réels r au hasard sur une large plage couvrant plusieurs ordres de grandeur (par exemple entre 1 et 10^{20}), et que la loi qui indique la probabilité de tomber sur une des valeurs possibles est assez régulière et étalée, alors les parties fractionnaires des logarithmes décimaux des nombres, c'est-à-dire les $\{\log_{10}(r)\}$, seront, à peu de chose près, uniformément réparties entre 0 et 1.

3

LA LOI D'ÉTALEMENT DE LA PARTIE FRACTIONNAIRE

La partie fractionnaire $\{r\}$ d'un nombre r est ce qui est derrière la virgule. Par exemple, $\{52,98734\} = 0,98734$. La loi d'étalement uniforme de la partie fractionnaire permet de déduire la loi de Benford. En voici l'énoncé : si l'on choisit au hasard des nombres réels sur une plage large de plusieurs unités (par exemple entre 1 et 10), et que la loi donnant la probabilité de tomber sur une des valeurs possibles est assez régulière, alors la partie fractionnaire des nombres qu'on trouvera sera, à peu de chose près, uniformément

répartie entre 0 et 1. La figure ci-dessous, due à Nicolas Gauvrit, illustre l'idée intuitive de la loi d'étalement uniforme de la partie fractionnaire. Dans le graphique du bas, les tranches de la densité de départ sont superposées après réduction proportionnelle de taille, pour former la densité de la partie fractionnaire. Les pentes des tranches se compensent à peu près. Dans le cas d'une fonction linéaire sur chaque intervalle $[n, n+1]$, la compensation se fait parfaitement, et donc la partie fractionnaire est dans un tel cas parfaitement uniforme.



On ne le voit pas d'emblée, mais ce qu'on vient d'énoncer est la loi de Benford (ou plus exactement une loi plus puissante parfois dénommée « loi de Benford continue »). En effet, affirmer que c est le premier chiffre significatif du nombre r équivaut à énoncer que $\log_{10}(c) \leq \{\log_{10}(r)\} < \log_{10}(c+1)$, ce que l'on justifiera un peu plus loin.

Les parties fractionnaires des images par $\log_{10}$ des nombres r dont le premier chiffre significatif est c occupent donc dans l'intervalle $[0, 1]$ un intervalle de longueur $\log_{10}(c+1) - \log_{10}(c)$, ce qui signifie, si l'on admet l'uniforme répartition, que leur >