Tabeillon, Inn et loi de Benford

Denis Odiet
Collège de Delémont

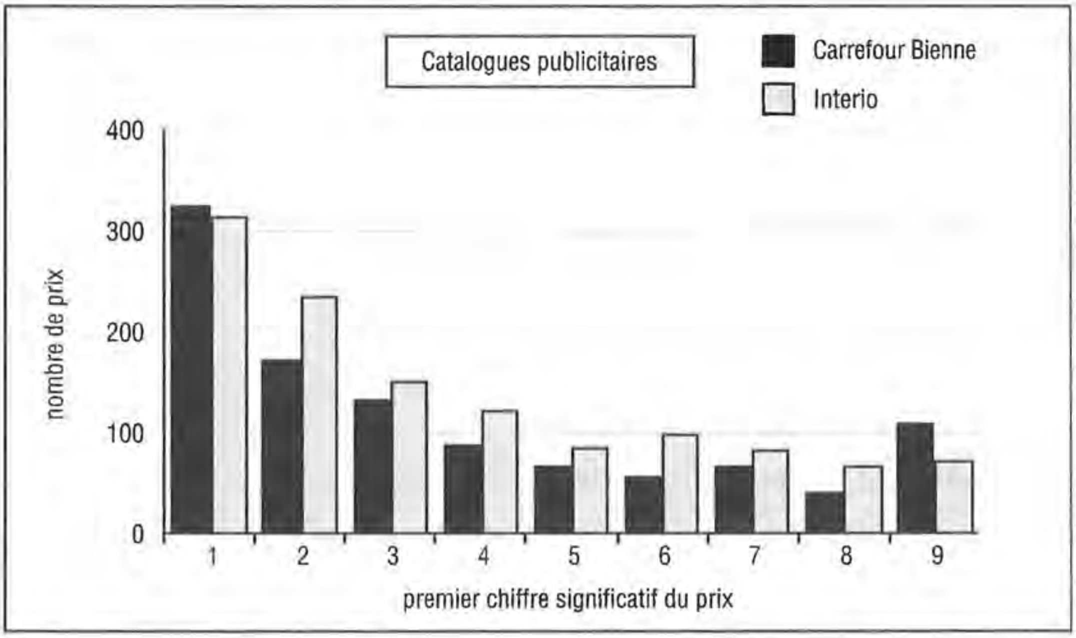
Le bucolique ruisseau du Tabeillon prend sa source sur les contreforts du plateau des Franches-Montagnes, parcourt une jolie combe en direction de la vallée de Delémont pour se jeter dans la Sorne. Sa longueur est de 14 km.

Familière des cruciverbistes, l'Inn s'écoule, quant à elle, sur une longueur de 510 km. Si l'on considère tous les cours d'eau prenant leur source sur territoire helvétique, on pourrait raisonnablement penser qu'il y en a autant dont le premier chiffre[1] du nombre exprimant leur longueur est 1 que de ceux dont le premier chiffre est 5. Or, il n'en est rien, plus que probablement. Bien que n'ayant procédé à aucune vérification, je pense pouvoir affirmer que le Tabeillon – premier chiffre significatif égal à 1 – appartient à un ensemble de cardinal nettement plus grand que celui dont l'Inn – premier chiffre significatif égal à 5 – est élément.

Afin de justifier cette affirmation, voyons ce qui se passe pour d'autres séries de nombres, pris dans des domaines variés de la vie de tous les jours.

Commençons par un premier domaine, celui des prix indiqués dans différents catalogues publicitaires.

Il s'agit donc de relever tous les prix qui y figurent, puis de compter ceux dont le premier chiffre significatif est égal à 1, respectivement 2, 3, ... 8 ou 9. Le travail a été réalisé pour les 1219 prix du catalogue de la maison « Interio » et les 1051 prix de quatre dépliants publicitaires du supermarché « Carrefour » de Bienne. Voici les résultats :



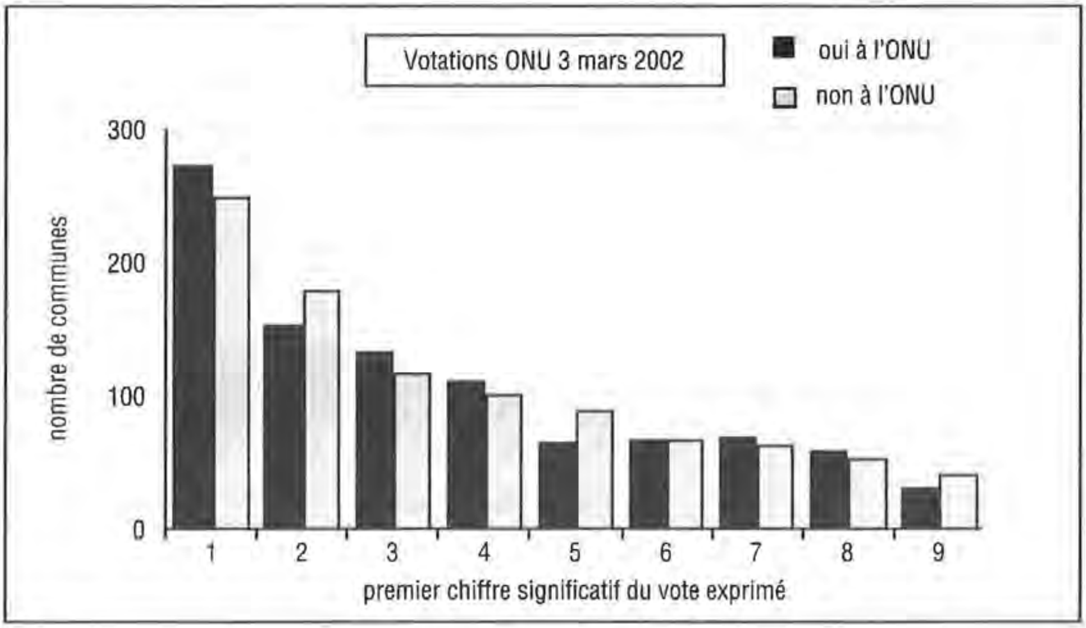
1. Par la suite, on parlera du premier chiffre significatif d'un nombre, c'est-à-dire de son premier chiffre différent de zéro.

Une première constatation se dessine: dans les magnifiques vitrines de Noël d'un magasin de votre ville, il y aura sûrement plus d'articles dont le premier chiffre significatif du prix est 1 que de ceux dont le premier chiffre significatif est 8. De même, il y en aura plus dont le premier chiffre significatif est 2 que de ceux dont le premier chiffre significatif est 3, etc.[2]

Un travail analogue a été proposé aux six élèves de neuvième année d'une classe de mathématiques pratiques dans les deux cas suivants:

1. La votation du 3 mars 2002 concernant l'adhésion de la Suisse à l'ONU.

Il s'agissait d'une part de se procurer les résultats de toutes les communes de Suisse romande et d'autre part de les dépouiller après s'être partagé le travail, l'objectif étant de compter les communes dont le nombre de «oui» exprimés commençait par 1,2, ...8 ou 9. Le même travail a été réalisé pour les «non» exprimés par les mêmes communes. Voici les résultats des 957 communes recensées:



2. Un domaine géographique.

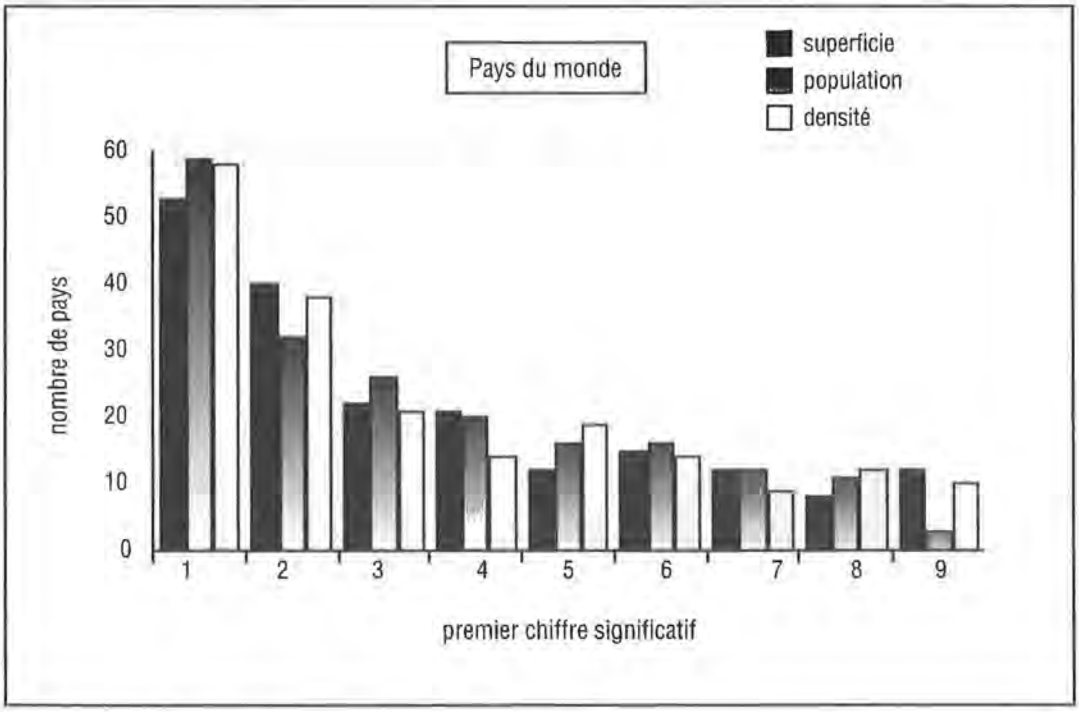
Les élèves étaient intrigués de savoir si la distribution précédente allait se répéter pour la superficie, la population et la densité des 195 pays du monde recensés à la mi-2001, en s'intéressant exclusivement au premier chiffre significatif des nombres de chacune des trois séries.

2. La relative fréquente apparition du premier chiffre significatif 9 chez Carrefour peut s'expliquer. Les dépliants publicitaires étaient constitués essentiellement d'offres promotionnelles, et l'on sait bien que psychologiquement, il vaut mieux proposer un article à FRS 9.95 qu'à FRS 10.25 par exemple.

Par exemple, pour la Jamaïque:

		premier chiffre significatif
Population:	2'5000'000 habitants	**2**
Superficie:	11'425 km2	**1**
Densité:	218 hab./ km2	**2**

Voici les résultats:



Et si la constatation effectuée précédemment devenait règle?

Nul doute que cette activité conduit les élèves à développer les mêmes attitudes que celles explicitées dans le texte introductif du domaine «Analyse de données» des futurs moyens d'enseignement de mathématiques de la Suisse romande pour les degrés 7 à 9:

«*...les élèves sont, plus particulièrement, amenés à*

- *prendre ou recevoir des informations;*
- *chercher et analyser des données;*
- *s'engager dans une recherche active individuelle ou collective;*
- *formuler, puis échanger, des constatations, c'est-à-dire...* »

Cette recherche offre aussi la possibilité de familiariser les élèves avec l'établissement de graphiques à l'aide de l'outil informatique:

Les résultats obtenus ont permis d'infirmer les croyances initiales des élèves: pour eux, dans une série de nombres se rattachant à un domaine précis, qu'un de ces nombres commence par 1, ou qu'il commence par 5, ou encore par 7, étaient des événements équiprobables.

De plus, cette activité procure l'occasion de jeter un pont vers l'histoire des mathématiques et des lois qui se sont mises en place au fil des siècles.

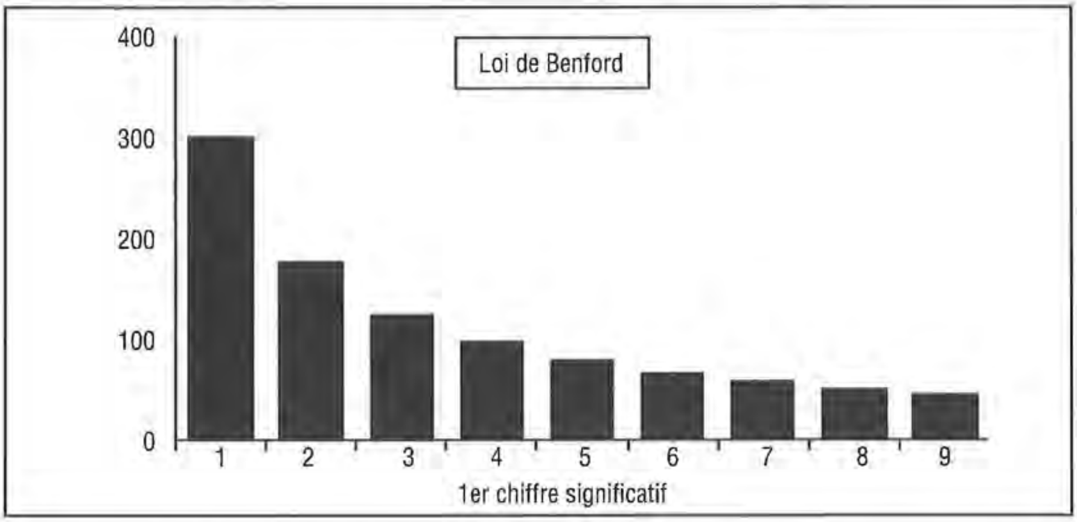
En effet, les exemples étudiés plus haut ne sont que des illustrations de la loi de Benford. En 1881, Simon Newcomb constate, en consultant différentes tables de logarithme dans des bibliothèques, que celles-ci sont plus sales dans les premières pages, correspondant aux petits chiffres. Il entrevoit l'explication suivante: les scientifiques travaillent plus souvent avec des nombres commençant par 1 que par 2, par 2 que par 3, ... mais, à l'époque, cette formule ne convaint personne.

Une cinquantaine d'années plus tard, Franck Benford refait la même observation et retrouve la même équation que celle proposée par Newcomb. La loi de Benford est née:

La probabilité p qu'un nombre tiré d'un ensemble quelconque comporte un premier chiffre significatif égal à d est donnée par la formule suivante:

$$p_d = \log_{10}\left(1 + \frac{1}{d}\right), d = 1, 2, \dots, 9$$

La voici sous forme graphique, appliquée à une liste de mille nombres.

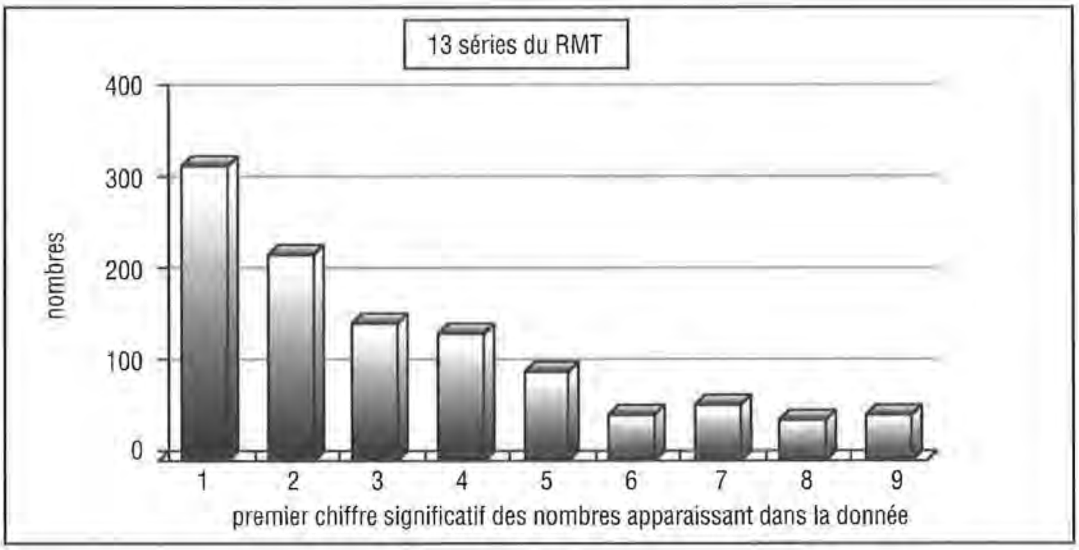


La ressemblance avec les exemples étudiés précédemment est vraiment frappante!

Pour revenir à l'Inn et au Tabeillon, sur un ensemble de mille cours d'eau, on peut donc s'attendre à ce que près de trois cents d'entre eux possèdent une longueur s'exprimant par un nombre dont le premier chiffre significatif est 1, c'est-à-dire qu'il y en aura environ quatre fois plus que ceux dont le premier chiffre significatif est 5.

Il n'est pas inutile de préciser que la loi de Benford est indépendante de l'unité choisie. Si, par exemple, on transformait en euros les prix des catalogues étudiés, on obtiendrait toujours la même distribution du premier chiffre significatif.

Pour les lecteurs encore perplexes, voici ce que fournit l'analyse de tous les nombres utilisés dans les énoncés de 13 séries du « Rallye mathématique transalpin » (1137 nombres recensés) et l'étude de leur premier chiffre significatif:



Etonnant, non?

Ajoutons quand même pour terminer que la loi ne s'applique malheureusement pas aux résultats de pur hasard, comme pour la loterie à numéros par exemple...

Pour en savoir plus:

«Le premier chiffre significatif fait sa loi», par Ted Hill, La Recherche N° 316, Janvier 1999
«Fraudeurs, méfiez-vous du 6!», par Véronique Parasote, Science et Vie junior, Août 1999

Réponses aux problèmes du numéro 203

p. 21, le problème de l'encadré «A propos de calcul de sommes»:

La suite cherchée est 8 ; 5 ; 16 ; 8 ; 5; 16 ; 8 ; 5 ; .... Son 2001e terme est 16, comme tous ceux dont le numéro d'ordre est un multiple de 3.

p. 45, les deux derniers cryptarithmes:

d) BON + PIED + BON + OEIL = PECHE
e) VENIRE + VIDERE + SCIRE + AUDIRE + DUCERE = VINCERE

d)

```
      6 7 3
    1 8 0 5
      6 7 3
  + 7 0 8 9
  ---------
  1 0 2 4 0
```

e)

```
    1 0 4 6 5 0
    1 6 3 0 5 0
      2 8 6 5 0
    9 7 3 6 5 0
  + 3 7 8 0 5 0
  -------------
  1 6 4 8 0 5 0
```

Pour d), il y a quatre groupes de deux solutions obtenues par permutations de D et L dans la colonne des unités:

2 x 673 + 1805 + 7089 = 2 x 673 + 1809 + 7085 = 10240
2 x 873 + 1605 + 7069 = 2 x 873 + 1609 + 7065 = 10420
2 x 687 + 1502 + 8054 = 2 x 687 + 1504 + 8052 = 10930
2 x 790 + 1524 + 9258 = 2 x 790 + 1528 + 9254 = 12362