

Détection de fraudes et loi de probabilité de Newcomb-Benford

FERNANDEZ Christelle PONCHEELE Clément EL KAÏM Laura
Encadré par M.DUCHARME

5 *mars* 2021

RESUME DU PROJET EN QLQ LIGNES

REMERCIEMENTS

Table des matières

Introduction.	2
#Introduction #Les origines/Description #Experimentations ##Benford ##Pas Benford	
#Tests	

Introduction.

La fraude est une pratique répandue dans de nombreux domaines. Il peut être tentant pour un être humain de tricher, si cela peut impliquer pour lui une position plus confortable dans la société, telle qu'une réduction de charges, ou même un avantage sur un de ses concurrents. Il semblerait donc logique que des personnes cherchent à déceler ces fraudes.

Les données transmises par un individu ou un organisme peuvent faire l'objet de modifications, c'est de ce type de fraudes auquel nous nous intéresserons ici, et plus particulièrement la modification du premier chiffre significatif de nombres pris dans un certain ensemble de données. De telles modifications entraînent un changement de la répartition des chiffres présents naturellement. Il est évident que pour qu'une fraude soit plus crédible, le fraudeur ne modifiera qu'une partie des données, afin d'apporter plus de vérité à son mensonge. Si nous connaissons la répartition des chiffres présentés dans un ensemble de données arbitraires, il est donc techniquement possible de savoir si un nombre a été modifié ou non. De manière intuitive nous pourrions penser que les nombres sont répartis de manière uniforme. Qu'en est-il vraiment ? C'est la question que se posera en premier lieu l'astronome, mathématicien, économiste et statisticien canadien Simon Newcomb. En 1881, celui-ci fournira une première approche au principe statistique qui se fera appeler *Loi de Benford*.

Cette découverte mise de côté pendant plusieurs années, ce n'est qu'en 1938 que l'ingénieur et physicien américain Frank Benford pensera être le premier à l'initiative de cette loi. Les articles de Newcomb et Benford proposeront finalement les mêmes résultats. D'après leurs travaux, les nombres ne sont pas répartis de manière uniforme.

Dans ce projet, nous nous intéressons au premier chiffre significatif, ce dernier étant le premier chiffre d'un nombre qui n'est pas un zéro (par exemple : 1 pour 1234 et 2 pour 0,0234). D'après Newcomb-Benford, pour toute liste de nombres prise dans un ensemble de données arbitraires, la répartition de ces chiffres ne se suivrait pas une loi uniforme. Celle-ci suivrait la loi de N-B, que nous expliquerons plus en détail par la suite.

Cette loi nous dit que, dans une liste de données dites naturelles, la probabilité d'avoir le chiffre i comme premier chiffre significatif est de $\log_{10}(1 + \frac{1}{i})$. Par exemple, le chiffre 1 serait présent à 30 alors que le 9 à seulement 4,6. Nous verrons également que la loi de Benford ne s'applique évidemment pas à toute série statistique.

La détection de fraudes a pour objectif de suspecter des potentielles modifications sur des données naturelles.

La loi de Newcomb-Benford ou