Loi de Benford et généralisations

Franck Jeannot

Montréal, Canada, Avril 2023, AA801, v1.1

Abstract

Benford's law generalizations and implementations.

Keywords: Benford's law, generalization, python, fraud, forensics, complex numbers

1. Introduction

Selon une nouvelle étude (Eckhartt, 2023-04) [1] [2] de l'*Université de St Andrews (UK)*, les pratiques d'audit de l'industrie financière peuvent être adaptées pour identifier la fraude académique. L'article publié dans la revue *Research Integrity and Peer Review* montre comment des outils statistiques efficaces peuvent être utilisés par ceux qui examinent les études scientifiques pour aider à détecter et à enquêter sur les données suspectes.

L'article examine la loi de **Benford** comme moyen d'examiner la distribution de fréquence relative des chiffres principaux des nombres dans les ensembles de données, qui est utilisée dans la pratique de l'audit professionnel. Les chercheurs recommandent que les contrôles de fraude au sein des institutions scientifiques et des éditeurs soient améliorés pour éliminer plus efficacement les fraudeurs.

Les rétractations d'articles scientifiques ont atteint près de 5 000 dans le monde en 2022 selon *Retraction Watch* [3], soit près de 0,1% des articles publiés. Bien que peu courantes, les cas de fraude scientifique ont un impact disproportionné sur la confiance du public dans la science.

La loi de Frank Benford [4] [5], également connue sous le nom de *loi des nombres anormaux*, est une loi empirique qui décrit la distribution des chiffres significatifs dans de nombreux ensembles de données du monde réel. Cette loi a été observée pour la première fois en 1881 par l'astronome américain Simon Newcomb.

2. Loi de Benford historique sous forme mathématique

Soit un nombre, ou une variable, X, et $Y = Log(X)$ son logarithme en base 10. Y peut se décomposer, comme tout nombre, en sa partie entière, $E(Y)$, et sa partie fractionnaire, $F(Y)$. Par exemple, si $Y = 23,227$, $E(Y) = 23$ et $F(Y) = 0,227$. La version historique de la loi de Benford s'énonce ainsi : X suit une loi de Benford si la partie fractionnaire de Y suit une loi uniforme – autrement dit, si $F(Y)$ a autant de chance de valoir n'importe quelle valeur entre 0 et 1. Cette loi de Benford historique implique la version avec les **premiers chiffres significatifs** [6].

3. Limitations d'usage de la loi de Benford

1. La loi de Benford n'est pas une loi au sens scientifique mais une loi **empirique** : la plupart des données testées par les chercheurs ne suivent pas du tout la loi de Benford, et celles qui la suivent le font la plupart du temps de manière très approximative. Dans son article de 1938, Benford testait 20 séries de données, dont moins de la moitié se conformait à peu près à sa « loi ». Des auteurs tels que (Scott et Fasli, 2001) [7] ont testé 230 séries de données et ont constaté que moins de 13% vérifient la loi de Benford [6].
2. Certains ensembles de données peuvent suivre des **distributions différentes ou avoir des caractéristiques spécifique**s qui affectent la distribution de leurs chiffres.
3. Il est important de noter que les nombres peuvent être facilement manipulés et que les contrôles statistiques ne sont qu'une ligne de défense parmi d'autres pour détecter les fraudes.
4. **La taille de la plage de données ou taille d'échantillon**. Des travaux, dont ceux de Richard J. C. Brown (2005) [8], pointent que la qualité de l'ajustement des données à la loi de Benford semble [1] notamment dépendre du nombre d'ordres de grandeur de la plage de données, calculée comme $R = log_{10}(X_{max}/X_{min})$. Les ensembles de données avec des plages numériques plus importantes, en particulier quatre ordres de grandeur ou plus, sont plus susceptibles de suivre la Loi de Benford. la loi de Benford est plus susceptible de s'appliquer à des échantillons de grande taille. Les échantillons plus petits peuvent ne pas contenir suffisamment de données pour suivre la loi de Benford de manière précise.
5. **Type de données** : la loi de Benford est souvent utilisée pour analyser des données numériques qui sont distribuées de manière aléatoire, telles que des chiffres de factures, des populations ou des résultats financiers. Les données non numériques, ou les données qui ne suivent pas une distribution aléatoire, peuvent ne pas suivre la loi de Benford.
6. **Présence de valeurs aberrantes** : des valeurs aberrantes ou des erreurs de mesure peuvent affecter la distribution des données et donc leur conformité à la loi de Benford. Il est donc important de supprimer les valeurs aberrantes ou de les traiter correctement avant d'appliquer la loi de Benford.
7. **Échantillonnage biaisé** : si les données ont été sélectionnées de manière biaisée, cela peut affecter leur distribution et donc leur conformité à la loi de Benford.
8. **Interprétation incorrecte** : la loi de Benford ne doit pas être utilisée pour prouver une fraude ou une manipulation de données. Elle doit être utilisée comme un outil de diagnostic pour identifier des anomalies dans la distribution des données [9]..

1. http://dspace.library.uvic.ca:8080/bitstream/handle/1828/3031/Thesis.pdf?sequence=1

4. Généralisations de la loi de Benford

Il existe plusieurs auteurs qui ont proposé des généralisations de la loi de Benford. En voici quelques-uns :

- (1995) Hill donne dans son article *"A Statistical Derivation of the Significant-Digit Law"* [10] les fondations d'une généralisation (avec certaines limitations).
- (2006) Steven J. Miller et Daniel L. Nigrini ont proposé une généralisation de la loi de Benford pour les nombres complexes dans un article intitulé *"Order Statistics and Benford's Law" (Statistical Science, vol. 21, n° 4)* [11]. Ils ont montré que la loi de Benford peut être étendue aux nombres complexes, en utilisant les statistiques d'ordre pour définir une distribution de probabilité (Réf : 10).
- (2007) Diekmann (6), s'appuyant sur (Hill, 1995a) [10] utilise une loi généralisée qui permet la dérivation des distributions marginales des chiffres de deuxième ordre et ultérieurs.
- (2011) Arno Berger et Theodore P. Hill ont proposé une généralisation de la loi de Benford pour les distributions de probabilité continues dans un article intitulé *"A basic theory of Benford's Law" (Probability Surveys, vol. 8)* [12]. Ils ont montré que la loi de Benford peut être étendue aux distributions de probabilité continues en utilisant la théorie de l'information.
- (2012) Mark J. Nigrini a proposé une généralisation de la loi de Benford pour les nombres ayant une base différente de 10 dans son livre (340 p.) intitulé *"Benford's Law : Applications for Forensic Accounting, Auditing, and Fraud Detection" (John Wiley and Sons)* [13]. Il a montré que la loi de Benford peut être étendue aux nombres ayant une base différente de 10, en utilisant des coefficients de pondération appropriés pour les chiffres significatifs.
- (2018) Jeremy Quastel et Avi Wigderson ont proposé une généralisation de la loi de Benford pour les ensembles finis dans un article intitulé *"The Benford Distribution of Discrete Finite Sets" (Advances in Mathematics, vol. 329)*. Ils ont montré que la loi de Benford peut être étendue aux ensembles finis, en utilisant des techniques de théorie de l'information.

Ces généralisations montrent que la loi de Benford peut être appliquée à différents types de données, ce qui en fait un outil utile dans de nombreux domaines, y compris l'audit, la comptabilité, la finance et la théorie de l'information.

5. Loi de Benford des premiers et deux premiers digits

La loi de benford pour le premier chiffre significatif est :
Prob $(D_1 = d) = \log_{10}\left(1 + \frac{1}{d}\right)$ pour tous $d = 1, 2, \ldots, 9$ .Ici, D_1 indique le premier chiffre décimal significatif, soit :

$$\begin{aligned}
&D_1(\sqrt{2}) = D_1(\mathbf{1}.414) = 1, \\
&D_1\left(\pi^{-1}\right) = D_1(0.\mathbf{3}183) = 3, \\
&D_1\left(e^{\pi}\right) = D_1(\mathbf{2}3.14) = 2.
\end{aligned}$$

Selon la loi de Benford, la fréquence attendue d'une paire de chiffres significatifs spécifique $d1d2$ est en généralisant :

$$\text{Prob}\,(D_1 D_2 = d_1 d_2) = \log_{10}\left(1 + \frac{1}{d_1 d_2}\right); \quad d_1 d_2 \in \{10, 11, \ldots, 99\} \tag{1}$$

$$\begin{aligned}
P(\text{premiers deux digits sont } d_1 d_2) &= P(\text{premiers deux digits sont } 10d_1 + d_2) && (2)\\
&= P(\text{premiers deux digits sont } d) && (3)\\
&= \log_{10}\left(1 + \frac{1}{d}\right) && (4)
\end{aligned}$$

Cette généralisation, telle que dans l'article *An Introduction to Benford's Law, Chapter 1, Princeton University Press* [14] présente le même résultat sous un format soustractif :

$$\begin{aligned}
P(\text{premiers } n \text{ digits sont } d_1 d_2 d_3 \ldots d_n) &= P(\text{premiers digits sont } d) && (5)\\
&= \log_{10}(d+1) - \log_{10} d && (6)
\end{aligned}$$

Exemple :

$$\begin{aligned}
P(\text{premiers 2 digits sont 76}) &= \log_{10} 77 - \log_{10} 76 && (7)\\
&= \log_{10} \frac{77}{76} && (8)\\
&= \log_{10}\left(1 + \frac{1}{76}\right) && (9)
\end{aligned}$$

5.1. Application d'exemple python - Benford 2 digits

```
from math import log10

def benford_prob(d):
    return log10(1 + 1/d)


def benford_2digit_prob(d1, d2):
    if d1 == 0:
        return 0
    else:
        return log10(1 + 1/(10*d1 + d2))


d1 = int(input("Enter d1 (1-9): "))
d2 = int(input("Enter d2 (0-9): "))

# Calculate probabilities for single digit numbers
probs = [benford_prob(d) for d in range(1, 10)]
print(f"P({d1}) = {probs[d1-1]:.6f}")

# Calculate probabilities for two digit numbers
probs2 = [benford_2digit_prob(d1, d2) for d1 in range(0, 10)
          for d2 in range(0, 10)]
print(f"P({d1}{d2}) = {probs2[(d1*10+d2)]:.6f}")

# Print expected probabilities
expected = [benford_prob(d) for d in range(1, 10)]
expected2 = [benford_2digit_prob(d1, d2) for d1 in range(1, 10)
             for d2 in range(0, 10)]
expected2 += [benford_prob(d) for d in range(10, 100)]
print("Expected probabilities:")
print(" ".join([f"{p:.6f}" for p in expected]))
print(" ".join([f"{p:.6f}" for p in expected2]))
#
Enter d1 (1-9): 7
Enter d2 (0-9): 6
P(7) = 0.057992
P(76) = 0.005677
Expected probabilities:
0.301030 0.176091 0.124939 .....
```

6. Généralisations - Diekmann (2007)

Dans son article *"Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data"*, Diekmann (2007) [15] aborde l'application de la Loi de Benford pour détecter les fraudes ou les manipulations éventuelles dans les données scientifiques. L'article, publié dans le *Journal of Applied Statistics*, présente une méthode d'analyse des données basée sur la distribution des premiers chiffres selon la Loi de Benford, qui stipule que dans de nombreux ensembles de données d'origine naturelle, les premiers chiffres ne sont pas distribués de manière uniforme, mais suivent un modèle logarithmique.

L'auteur soutient que les données frauduleuses ou manipulées dévient souvent de la distribution attendue des premiers chiffres prédite par la Loi de Benford. L'article aborde la manière dont la Loi de Benford peut être utilisée comme un outil statistique pour identifier les motifs suspects dans les données, en comparant la distribution observée des premiers chiffres dans un ensemble de données avec la distribution attendue basée sur la Loi de Benford.

L'article aborde également les limites et les défis potentiels de l'utilisation de la Loi de Benford pour détecter les fraudes, tels que la nécessité de disposer de grands échantillons, les biais potentiels et autres considérations statistiques. Il met en avant l'importance d'utiliser la Loi de Benford comme un outil complémentaire en conjonction avec d'autres méthodes pour détecter les fraudes ou les manipulations potentielles dans les données scientifiques. On peut écrire :

Loi de Benford : généralisation (base Diekmann (2007) ; (Hill, 1995a [10]))

$$P\left(D_1 = d_1, \ldots, D_k = d_k\right) = \log_{10}\left[1 + \left(\Sigma d_i 10^{k-i}\right)^{-1}\right] \tag{10}$$

ou en clarifiant les limites de la somme :

Loi de Benford : généralisation (extrapolation base Diekmann (2007)

$$P\left(D_1 = d_1, \ldots, D_k = d_k\right) = \log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i 10^{k-i}\right)^{-1}\right] \tag{11}$$

Avec les variables aléatoires D_1, D_2, ...,D_k contenant les premiers digits (les plus à gauche, indépendants de la position de la virgule décimale), deuxièmes, ...,k_{iemes} chiffres significatifs, $d_1 = 1, 2, \ldots, 9$ et $d_j = 0, 1, \ldots, 9$ $(j = 2, \ldots, k)$. Cette « loi générale des chiffres significatifs » (Hill, 1995a) permet la dérivation des distributions marginales des chiffres de deuxième ordre et ultérieurs [2].

2. p 3/10 de [15]

6.1. Généralisations - Exemple

Si les chiffres suivent une distribution de Benford, la combinaison de chiffres significatifs 1028 (par exemple, 1.028 ou 0.001028) est attendue avec une probabilité de $\log_{10}[1 + 1/1028]$. Dans ce cas, nous avons k=4, car la combinaison de chiffres significatifs 1028 a 4 chiffres. Nous pouvons alors remplacer les valeurs de d_i par les chiffres correspondants dans 1028 soit : $d_1 = 1, d_2 = 0, d_3 = 2, d_4 = 8$.
On place les valeurs de cet exemple dans la formule de généralisation :

Loi de Benford : généralisation (base Diekmann : calcul d'exemple)

$$\mathrm{P}\left(D_{(1028)}\right) = \log_{10}\left[1 + \left(\sum_{i=1}^{k} d_i 10^{k-i}\right)^{-1}\right]$$

$$\mathrm{P} = \log_{10}\left[1 + \left(1 \cdot 10^3 + 0 \cdot 10^2 + 2 \cdot 10^1 + 8\right)^{-1}\right]$$

$$\mathrm{P}\left(D_{(1028)}\right) = \log_{10}\left[1 + \left(\frac{1028}{1000}\right)^{-1}\right]$$

$$\mathrm{P}\left(D_{(1028)}\right) = \log_{10}[1 + 1/1028]$$

Note : En utilisant la propriété de l'inverse d'un nombre élevé à la puissance -1, on peut dire que $\left(\frac{1028}{1000}\right)^{-1} = \frac{1}{1028}$. Donc, les deux expressions sont identiques et représentent la même valeur mathématique.

6.2. Exemples simplifiés en python

```
import math
def benford-diekmann(d, k):
  prob = math.log10(1 + (sum([d_i * 10 ** (k - i)
          for i, d_i in enumerate(range(1, k + 1))])) ** -1)
  return prob
```

Exemple inspiré de Turner (Matlab converti Python)[3]. Cet algorithme compte le nombre de chiffres de 0 à 9 dans toutes les positions de tous les nombres compris entre 1 et Max. Il le fait en séparant les nombres en sous-intervalles en fonction du nombre de chiffres dans leur représentation décimale et en comptant les chiffres pour chaque sous-intervalle. Ensuite, l'algorithme compte également le nombre d'occurrences du premier chiffre de chaque nombre dans les mêmes sous-intervalles. L'algorithme a une complexité de temps linéaire par rapport à Max et utilise des tableaux de compte pour stocker les résultats :

3. https://www.gatsby.ucl.ac.uk/~turner/TeaTalks/BenfordsLaw/benford.m

```
import numpy as np
import matplotlib.pyplot as plt
# Parameters
Max = 10000
# Initialize counts
count1 = np.zeros(Max)
count2 = np.zeros(Max)
count3 = np.zeros(Max)
# All digits count
for p in range(1, int(np.log10(Max))+1):
    nMax = int(Max / 10**p) - 1
    for n in range(nMax+1):
        st = 10**p * n + 10**(p-1)
        en = st + 10**(p-1) - 1
        count1[st:en+1] += 1
        if en + 10**(p-1) < Max:
            count2[st+10**(p-1):en+10**(p-1)+1] += 1
        if en + 2*10**(p-1) < Max:
            count3[st+2*10**(p-1):en+2*10**(p-1)+1] += 1
# Initialize first digits counts
count1st = np.zeros((Max, 1))
count2st = np.zeros((Max, 1))
# First digits count
for p in range(1, int(np.log10(Max))+1):
    lower = 10**(p-1)
    upper = lower + 10**(p-1) - 1
    i, = np.where((lower <= np.array(range(Max))) &
                  (np.array(range(Max)) <= upper))
    count1st[i] += 1
for p in range(1, int(np.log10(Max))+1):
    lower = 2*10**(p-1)
    upper = lower + 10**(p-1) - 1
    i, = np.where((lower <= np.array(range(Max))) &
                  (np.array(range(Max)) <= upper))
    count2st[i] += 1
```

Affichage des graphes de l'algorithme précédent :

```
# Plotting
plt.figure()
plt.plot(range(Max), count1, '-r', label='All Digits Count')
plt.plot(range(Max), count2, '-b', label='All Digits Count Shifted by 1')
plt.plot(range(Max), count3, '-k', label='All Digits Count Shifted by 2')
plt.legend()
plt.xlabel('Integer')
plt.ylabel('Count')
plt.title('Distribution of Digits Along Integer Lines')
plt.figure()
plt.plot(range(Max), count1st, '-r', label='First Digits Count')
plt.plot(range(Max), count2st, '-k', label='First Digits Count Shifted by 1')
plt.legend()
plt.xlabel('Integer')
plt.ylabel('Count')
plt.title('Distribution of First Digits Along Integer Lines')
plt.show()
```

7. Généralisations - Eckhartt, Ruxton (2023)

Dans leur article *"Investigating and preventing scientific misconduct using Benford's Law"*, les auteurs (Eckhartt, Gregory M. and Ruxton, Graeme D., 2023) [2] [1] mettent en avant l'importance de l'intégrité et de la confiance dans la recherche académique. Ils soulignent que les procédures de surveillance de la fiabilité de la recherche et d'enquête sur les cas de possible fraude de données ne sont pas bien établies. Les auteurs proposent une approche pratique pour enquêter sur les travaux suspectés de manipulation frauduleuse de données en utilisant la loi de Benford, en s'inspirant des pratiques d'audit financier. Ils fournissent une synthèse de la littérature sur les tests d'adhérence à la loi de Benford, aboutissant à la recommandation d'un seul test initial pour les chiffres dans chaque position des chaînes numériques d'un jeu de données. Ils recommandent également d'autres tests qui peuvent être utiles en cas d'hypothèses spécifiques sur la nature de la manipulation des données. Il est noté que leurs conseils diffèrent des implémentations courantes des tests de la loi de Benford. De plus, les auteurs appliquent leur approche à des données précédemment publiées, mettant en évidence l'efficacité de ces tests pour détecter des irrégularités connues. Enfin, ils discutent des résultats de ces tests en référence à leurs forces et limitations.

7.1. Généralisations - équations équivalentes

On définit $P(d,i)$ qui représente la Loi de Benford pour les chiffres au-delà du premier, en tenant compte de la position i (où $i > 1$) et en utilisant la lettre k pour représenter les limites de la somme :

Loi de Benford : généralisation (base Echhartt, Ruxton 2023)

$$P(\mathrm{d} \mid i) = \sum_{k=10^{i-2}}^{10^{i-1}-1} \log_{10}\left(1 + \frac{1}{10k + d}\right)$$

8. Applications à l'analyse d'image forensique

L'article *"Analyzing Benford's Law's Powerful Applications in Image Forensics"* publié [16] dans *Applied Sciences* en 2021 par Diana Crișan et al. étudie la compression JPEG du point de vue de la loi de Newcomb-Benford. Cette loi stipule que dans un ensemble de nombres naturels, le premier chiffre a une distribution de probabilité qui décroît logarithmiquement. L'un de ses principaux domaines d'application est la compression JPEG des images, un domaine d'un grand intérêt pour des domaines tels que la criminalistique des images. L'article se concentre sur les moyens de détecter les images frauduleuses et les facteurs de qualité JPEG. En utilisant le canal de luminance de l'image et les coefficients JPEG, les auteurs décrivent une technique pour déterminer le facteur de qualité avec lequel une image JPEG est compressée.

Fu et al. (2007) ont proposé une généralisation de la loi de Benford dans *"A generalized Benford's law for JPEG coefficients and its applications in image forensics."* pour leur cas particulier d'analyse forensique d'images :

$$p(x) = N \log_{10}\left(1 + \frac{1}{s + x^q}\right), x = 1, 2, \ldots, 9$$

où N est un facteur de normalisation qui rend $p(x)$ une distribution de probabilité, s et q sont des paramètres de modèle pour décrire précisément les distributions pour différentes images et différents facteurs de compression Q, , lorsque $s = 0$ et $q = 1$, la formule se réduit à version historique de Benford.

9. Généralisation aux nombres complexes

Steven J. Miller et Daniel L. Nigrini ont proposé [11] une généralisation de la loi de Benford pour les nombres complexes dans un article publié en 2006 intitulé *"Order Statistics and Benford's Law" (Statistical Science, vol. 21, n° 4)*. Ils ont montré que la loi de Benford peut être étendue aux nombres complexes, en utilisant les statistiques d'ordre pour définir une distribution de probabilité.

Voici l'équation généralisée proposée par Steven J. Miller et Daniel L. Nigrini pour la loi de Benford étendue aux nombres complexes :

$$P(d) = \log_{10}(1 + |\omega|^2) - \log_{10}(1 + |\omega/d|^2), \tag{12}$$

où d est le premier chiffre significatif d'un nombre complexe, et ω est une variable complexe aléatoire avec une distribution uniforme sur l'ensemble du cercle unité dans le plan complexe. On note que cette formule est similaire à celle de la loi de Benford pour les nombres réels, mais avec l'ajout de la variable complexe ω.

10. Tests de vraisemblance - LR - Likelihood Ratio

Dans sa thèse *"Testing Benford's Law with the first two significant digits"*(Wong, 2010) [17] introduit l'usage de tests ou "Likelihood Ratio"(LR), qui sont des tests statistiques basés sur le rapport de vraisemblance. Le test de rapport de vraisemblance est une technique statistique couramment utilisée pour comparer deux modèles, l'un étant un sous-modèle de l'autre, afin de déterminer si le modèle plus complexe est justifié par les données ou s'il y a suffisamment de preuves pour le rejeter en faveur du modèle plus simple. Dans les tests de loi de Benford, le test de rapport de vraisemblance est utilisé pour comparer la distribution observée des chiffres significatifs à la distribution théorique de Benford, afin de déterminer si la distribution observée est cohérente avec la loi de Benford ou si elle diffère significativement de celle-ci. Ce type de tests peut être approfondi avec la revue de l'article *"Assessing Conformance with Benford's Law : Goodness-Of-Fit Tests and Simultaneous Confidence Intervals"* (Lesperance, 2016) [18].

- LR-multinomial est un test de rapport de vraisemblance pour comparer la distribution observée à une distribution multinomiale de Benford.
- LR-decreasing est un test qui mesure la décroissance de la fréquence des chiffres de la loi de Benford.
- LR-generalized Benford est un test généralisé qui permet de tester les distributions autres que les lois de Benford.
- LR-Rodriguez est un test qui prend en compte les deux premiers chiffres significatifs des nombres, plutôt que seulement le premier chiffre

Références

[1] University of St Andrews, Accountants' tricks can help identify cheating scientists, says new study (Apr 2023).
URL https://phys.org/news/2023-04-accountants-scientists.html

[2] Eckhartt, Gregory M. and Ruxton, Graeme D., Investigating and preventing scientific misconduct using Benford's Law, Research Integrity and Peer Review 8 (1) (2023) 1. doi:10.1186/s41073-022-00126-w.
URL https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-022-00126-w

[3] retractionwatch.
URL https://retractionwatch.com/

[4] Franck Jeannot, Expérimentations sur la loi de Benford (08 2017).
URL https://franckybox.com/wp-content/uploads/Benfords_law.pdf

[5] Franck Jeannot, Loi de Benford et implémentation python (04 2023).
URL https://franckybox.com/wp-content/uploads/Benford_law_python.pdf

[6] afis.org, La loi de Benford : raccourcis médiatiques (2012).
URL https://www.afis.org/La-loi-de-Benford-raccourcis-mediatiques

[7] Pd Scott and Maria Fasli, Csm-349 - benford's law: An empirical investigation and a novel explanation, 2001.
URL https://repository.essex.ac.uk/8664/

[8] R. J. C. Brown, Benford's Law and the screening of analytical data: the case of pollutant concentrations in ambient air., Analyst 130 (9) (2005) 1280–1285. doi:10.1039/b504462f.
URL https://pubmed.ncbi.nlm.nih.gov/16096674/

[9] William Goodman, The promises and pitfalls of Benford's law (2016).
URL https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1740-9713.2016.00919.x

[10] T. P. Hill, A Statistical Derivation of the Significant-Digit Law, Statistical Science 10 (4) (1995) 354 – 363. doi:10.1214/ss/1177009869.
URL https://doi.org/10.1214/ss/1177009869

[11] Steven J. Miller and Mark J. Nigrini, Order statistics and benford's law, International Journal of Mathematics and Mathematical Sciences 2008 (2008) 1–19. doi:10.1155/2008/382948.
URL https://doi.org/10.1155%2F2008%2F382948

[12] Berger, Arno and Hill, Theodore, A basic theory of benford's law, Probability Surveys 8 (01 2011). doi:10.1214/11-PS175.
URL https://www.researchgate.net/publication/254211701_A_Basic_Theory_of_Benford's_Law

[13] M. J. Nigrini, Benford's Law : Applications for forensic accounting, auditing, and fraud detection, Vol. 586, John Wiley and Sons, 2012.

[14] Turner, An Introduction to Benford's Law, Chapter 1, Princeton University Press (2005).
URL https://assets.press.princeton.edu/chapters/s10526.pdf

[15] Diekmann, Andreas, Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data, Journal of Applied Statistics 34 (2007) 321–329. doi:10.1080/02664760601004940.
URL https://www.researchgate.net/publication/24083679_Not_the_First_Digit_Using_Benford's_Law_to_Detect_Fraudulent_Scientif_ic_Data

[16] D. Crișan, A. Irimia, D. Gota, L. Miclea, A. Puscasiu, O. Stan, H. Valean, Analyzing benford's law's powerful applications in image forensics, Applied Sciences 11 (23) (2021). doi:10.3390/app112311482.
URL https://www.mdpi.com/2076-3417/11/23/11482

[17] Stanley Wong, Testing Benford's Law with the first two significant digits, 2010.
URL http://dspace.library.uvic.ca:8080/bitstream/handle/1828/3031/Thesis.pdf?sequence=1

[18] M. Lesperance, W. J. Reed, M. A. Stephens, C. Tsao, B. Wilton, Assessing conformance with benford's law: Goodness-of-fit tests and simultaneous confidence intervals, PLOS ONE 11 (3) (2016) 1–20. doi:10.1371/journal.pone.0151235.
URL https://journals.plos.org/plosone/article/citation?id=10.1371/journal.pone.0151235