Applied Mathematical Sciences, Vol. 20, 2026, no. 4, 149 - 156
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2026.919328

On Benford Testing for Academic Fraud Detection

Marco Di Marzio

University of Chieti-Pescara, Italy

Stefania Fensore

University of Chieti-Pescara, Italy

Chiara Passamonti

University of Luxembourg, Luxembourg

Abstract

Academic fraud refers to various dishonest behaviors that violate standards of integrity and transparency. Unethical activities include research misconduct, plagiarism, peer review manipulation, and data fabrication or falsification. Because academic fraud undermines the credibility of scientific research, it is crucial to investigate it. In this paper, we analyze suspicious data contained in retracted papers about the use of chloroquine for COVID-19 treatment. To accomplish this, we use Benford's law, a well-known mathematical model employed to describe the first significant digit distribution in several collections of numbers. In fact, it is considered a useful tool in data analysis to detect frauds, identify financial statement manipulation and anomalies in income reporting, assess the integrity of experimental data, and verify the accuracy of reported health statistics. A typical way to check if a statistical population adheres to Benford's law is through hypothesis testing. We report some tests to check the Benfordness of data and apply them to a real dataset.

Mathematics Subject Classification: 62G10

Keywords: Academic Frauds, Benford's law, Statistical hypothesis testing.

1 Introduction

In 1938, physicist Benford, and 50 years before him, astronomer Newcomb noticed that the frequencies of the leading digit in several collections of data are not uniformly distributed but exhibit a heavy bias towards lower ones, such as 1, 2, and 3, over the higher digits, such as 7, 8, and 9. The *first significant decimal digit* of a positive real number x is defined as the first non-zero digit appearing in the decimal expansion of x. For example, the first significant digits ($f.s.d.$) of 1968 and 0.1968 are both 1. The so-called *Benford's law* well describes the frequency distribution of the first significant digit. This model is very common in many real-world datasets and, for this, it has attracted the interest of mathematicians, statisticians, economists, and engineers.

The most common application of Benford's law is the forensic auditing field to detect data manipulation. The key idea is that, given a collection of data which is expected to follow the Benford distribution, it is possible to identify fabricated or falsified data by comparing the frequency of their leading digits against the theoretical one predicted by Benford's law. This idea has been applied in financial contexts but also in economics (incomes, insurance claims, electricity/telephone bills), social (populations of cities, death rates) and environmental (lengths and flow rates of rivers) sciences and astrophysics (diameter of planets, stellar and galaxy distances). These numerous specialized applications have also significantly contributed to the advancement of Benford's law theory. An accurate way to investigate the conformity of data to Benford's law is hypothesis testing.

The paper is organized as follows. In Section 2 we recall the main features of Benford's law, while in Section 3 we present a couple of statistical tests to validate the Benfordness of a set of data. A real data case study concerning academic fraud is presented in Section 4.

2 On the Benford's law

Benford's law well describes the behavior of the first leading digit in various datasets and numerical sequences according to the following model, which is depicted in Fig. 1:

$$P(f.s.d. = d) = \log_{10}\left(\frac{d+1}{d}\right) \tag{1}$$

where $d = 1, \ldots, 9$. The barplot highlights that the digits do not follow a uniform pattern, but they are modeled by a logarithmic distribution. As an example, consider that the frequency of numbers with leading digit 1 is about 30%, while the frequency of the leading digit 9 is more or less 4.6%.

According to Theorem 4.2 of [3] a sequence of real numbers is Benford if and

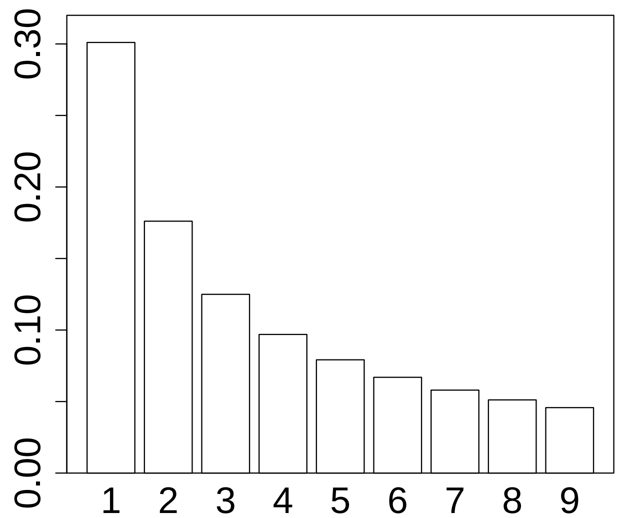

Figure 1: The frequency distribution of the first significant digit.

only if the decimal logarithm part of its absolute value, also called *mantissa* or *log-significand*, is uniformly distributed modulo one.

Let X be an absolutely continuous random variable and F_X its distribution function. Now, s is the (base-10) log-significand function defined as

$$s(X) = \langle \log_{10} |X| \rangle,$$

where $\langle a \rangle = a - \lfloor a \rfloor$, and $\lfloor a \rfloor = \max\{n \in Z : n \leq a\}$, with the assumption that $s(0) := 0$. We can say that X is a Benford random variable if $s(X)$ has uniform distribution function

$$F_{s(X)}(u) = u, \qquad u \in [0, 1[\ . \tag{2}$$

An important property of Benford's law is the scale invariance one, which means that if X is a Benford random variable, also bX has the same distribution for all the scaling factors $b > 0$ (see [3]).

3 On Benfordness hypothesis testing

Considering the wide applicability of Benford's law in practice, it is crucial to investigate whether a collection of data adheres to the theoretical model. A typical way to check if a statistical population conforms to Benford's law is to implement hypothesis testing, where the "natural" null hypothesis to be tested is

$$H_0 : s(X) = U, \tag{3}$$

where U is a Uniform random variable on $[0,1[$.

Several formal tests on Benfordness exist. In this section we recall the most popular one and one of the last proposals which has been introduced by [2]. The simplest test, and also the most common in applications, is based on the first-digit Pearson statistic

$$\chi^2 = \sum_{d=1}^{9} \frac{(n\widehat{p}_d - np_d)^2}{np_d}, \tag{4}$$

for $d \in \{1, 2, \ldots 9\}$, where p_d is the theoretical first-digit probability for digit d defined in (1) when H_0 is true, while $\widehat{p}_d$ is its sample estimate. However, two main drawbacks of the χ^2, when used to validate Benfordness, have been arisen. First, it does not test the uniformity hypothesis, but the conformity of the first-digit distribution of X to (1). Moreover, the χ^2 is not a scale invariant statistic test.

To overcome these two disadvantages, starting from the series expansion of the distribution of the log-significand function $s(X)$, and truncating it at the Nth term, [2] recently derived a scale invariant likelihood ratio test on the *mantissa*. For each $u \in [0,1[$ there exists a vector of order $(N+1)$ of complex numbers, say $c = (c_0, c_1, \ldots, c_N)^T$, and the test statistic is given by

$$R_{N,n} = \prod_{i=1}^{n} (\widehat{c}^{\,*} T_{s(X_i)} \widehat{c}\,)^{-1}, \tag{5}$$

where c^* and $\widehat{c}$ are, respectively, the conjugate transpose and the maximum likelihood estimator of the vector c, and $T_u = (t_{k,j,u})$ is a Toeplitz matrix of order $(N+1)$ such that, for $k = 1, \ldots, N$ and $j = 0, \ldots, N-k$, $t_{k,j,u} = e^{2\pi \mathrm{i}(k-j)u}$.

4 A real case study: academic frauds

As shown by the Retraction Watch database (`https://retractionwatch.com`), the last two decades have been characterized by a sharp increase in retractions of scholarly papers due to some kind of authors' unethical behavior. A natural question arises on the reasons for which these misconducts are so common. The answer is rather trivial since there are many potential benefits for academics. First, the publication is the only tool for getting ahead and consequently for promotion and pay [11]; moreover, related to this, it's necessary to satisfy the need for academic prestige by individuals [1], not to mention that the increase in published work leads to higher performance expectations [5]. Importantly, there are no significant deterrents to avoid academic dishonesty [4].

Since the integrity of scientists, and consequently of their published works, is the cornerstone of science's credibility, it's a duty to investigate and detect academic fraud. Additionally, these misconducts led to another issue because governments spend millions to finance, unwittingly, fraudulent activities.

Withdrawn literature due to academic dishonesty invades a variety of fields, ranging from medicine and biological sciences to economics and accountability research [13]. There are several reasons behind the retraction: laboratory or data errors, the inability to submit replication, plagiarism or conflict of interests. However, the main reason linked to retraction is unethical behaviour by academics, that is perceived as increasing within the scientific community. Very often fraud regards employing manipulated data [6]. In this sense, a glaring case is the instance of Professor James Hunton, who has been charged with fabricating/misrepresenting data in his articles. He retracted 37 works to his name [8]. However, observing data manipulation is usually very difficult because data are fabricated in a way to elude ordinary science's self-correcting procedures [12]. In addition, even if one may argue that scientific fraudsters are a minority, unfortunately the actions of a few can have huge ramifications for the academic community. Indeed, a retracted paper may have a large number of citations before its retraction, in such a way that other scientists are misled in their own research.

First, [15] proposed the use of Benford's law to detect fraudulent data and suggested that the law could be employed to validate the integrity and "naturalness" of seemingly random scientific data in a social science context. He asserted that conformity to the law of first significant digits does not automatically mean authenticity, but nonconformity should raise suspicion. Recently, [8] considered Benford's law as a tool to discriminate between academic articles that have been retracted due to falsified/manipulated data and those that have not been retracted.

Currently, the pandemic due to COVID-19 led researchers to publish more and faster, undermining the accuracy and quality of works [14]. Hence, the scientific community has been swamped by a wave of withdrawals (at the moment, the Retraction Watch database contains about 300 retracted papers on COVID-19). An impressive case is the highly-cited Lancet article "*Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis*" [9]. After its publication, various issues arose about data genuineness and the study performed by Surgisphere Corporation whose founder is a paper co-author Desai. The co-authors, except Desai, planned an independent audit to evaluate the origination of the database elements, "to confirm the completeness of the database and to replicate the analyses presented in the paper". Unfortunately, panel independent reviewers withdrew from the peer-review process. The reason was that Surgisphere did not transfer "the full dataset, client contracts, and the full ISO audit report to their servers for analysis as such transfer would have violated client agreements and confidentiality requirements". As a result, the co-authors of

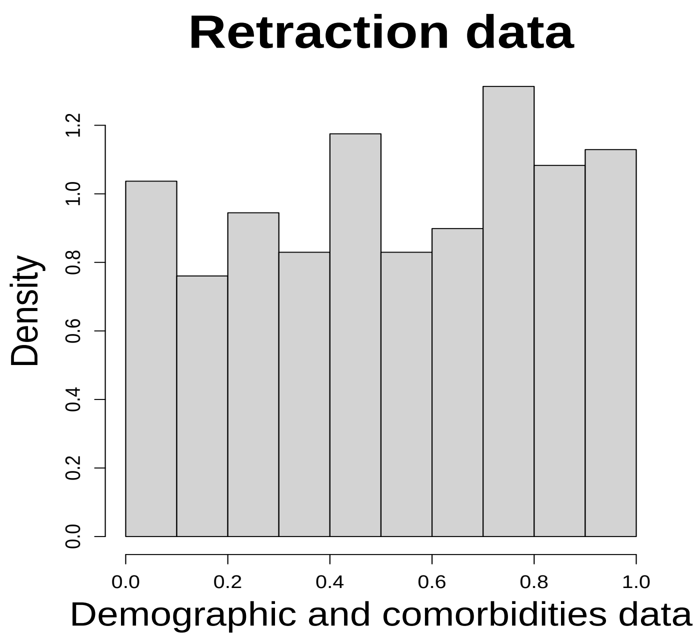

Figure 2: Density histogram of demographic and comorbidity data from the treatment-control study in [9].

Desai have concluded that they "can no longer vouch for the veracity of the primary data sources", and then retracted the paper. In addition, the same group of researchers retracted a further paper dealing with the cardiovascular diseases and drug therapies in COVID-19 [10].

Our experiment consists in checking if the results of Benfordness tests are able to somehow validate the retirement decision on the study performed by [9]. Obviously, it is not possible to assess the integrity of the original dataset, since it is not publicly available. However, it may be interesting testing whether the digits appearing in this paper conform Benford's law. More precisely, we consider the digits given in Tables 1 and 2 by [9]. These digits deal with the summary statistics of the treatment-control study. In particular, Table 1 contains the demographics and comorbidities of the patients in the considered population (survival or non-survival) during hospitalization. Moreover, Table 2 contains the patient demographics and characteristics by hydroxychloroquine (or chloroquine) treatment group. We extracted the digits which are present in the two tables. Since data arise from different sources and variables, the limit theorem by [7] should ensure that the digit distribution conforms Benford's law.

The histogram for the log-significand values of the digits, depicted in Fig. 2, apparently shows a moderate departure from uniformity. Thus, we applied the testing procedure for assessing the Benford hypothesis suggested by [2]

Table 1: $p - values$ obtained performing the χ^2 and likelihood ratio tests.

Statistic test	$p - values$
χ^2	0.0491
$R_{1,n}$	0.0847
$R_{2,n}$	0.0429
$R_{3,n}$	0.1065

reported in Section 3, along with the usual χ^2 test on the first digit. Results are reported in Table 1.
Hence, even if the results are not conclusive, the obtained *p-values* give a hint for considering the dataset suspicious.

References

[1] Almer, E. D., Baldwin, A. A., Jones-Farmer, A., Lightbody, M., Single, L. E., Tenure-track opt-outs: Leakages from the academic pipeline. Advances in accounting education: Teaching and curriculum innovations, in: *Advances in Accounting Education: Teaching and Curriculum Innovations*, **19** (2016), 1-36. https://doi.org/10.1108/s1085-462220160000019001

[2] Barabesi, L., Cerioli, A., Di Marzio, M., Statistical models and the Benford hypothesis: a unified framework, *TEST*, **32** (2023), 1479-1507. https://doi.org/10.1007/s11749-023-00881-y

[3] Berger, A., Hill, T. P., *An introduction to Benford's law*, Princeton University Press, 2015. https://doi.org/10.23943/princeton/9780691163062.001.0001

[4] Cox, A., Craig, R., Tourish, D., Retraction statements and research malpractice in economics, *Research Policy*, **47** (2018), 924-935. https://doi.org/10.1016/j.respol.2018.02.016

[5] Glover, S. M., Prawitt, D. F., Summers, S. L., Wood, D. A., Publication benchmarking data based on faculty promoted at the top 75 US accounting research institutions, *Issues in Accounting Education*, **27** (2012), 647-670. https://doi.org/10.2308/iace-50140

[6] Hesselmann, F., Graf, V., Schmidt, M., Reinhart, M., The visibility of scientific misconduct: A review of the literature on retracted journal articles, *Current sociology*, **65** (2017), 814-845. https://doi.org/10.1177/0011392116663807

[7] Hill, T. P., A statistical derivation of the significant-digit law, *Statistical science*, (1995), 354-363. https://doi.org/10.1214/ss/1177009869

[8] Horton, J., Kumar, D. K., Wood, A., Detecting academic fraud using Benford law: The case of Professor James Hunton, *Research Policy*, **49** (2020), 1-19. https://doi.org/10.1016/j.respol.2020.104084

[9] Mehra, M. R., Ruschitzka, F., Patel, A. N., Retraction-Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis, *The Lancet*, (2020), 1-10. https://doi.org/10.1016/s0140-6736(20)31324-6

[10] Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D., Patel, A. N., Cardiovascular disease, drug therapy, and mortality in Covid-19, *New England Journal of Medicine*, **382** (2020), e102. https://doi.org/10.1056/nejmoa2007621

[11] Necker, S., Scientific misbehavior in economics, *Research Policy*, **43** (2014), 1747-1759. https://doi.org/10.1016/j.respol.2014.05.002

[12] Reich, E. S., The rise and fall of a physics fraudster, *Physics World*, **22** (2009), 24. https://doi.org/10.1088/2058-7058/22/05/37

[13] Sharma, K., Team size and retracted citations reveal the patterns of retractions from 1981 to 2020, *Scientometrics*, **126** (2021), 8363-8374. https://doi.org/10.1007/s11192-021-04125-4

[14] Soltani, P., Patini, R., Retracted COVID-19 articles: a side-effect of the hot race to publication, *Scientometrics*, **125** (2020), 819-822. https://doi.org/10.1007/s11192-020-03661-9

[15] Varian, H. R., Benfords law, *American Statistician*, **26** (1972), 65-66.

Received: March 23, 2026; Published: April 27, 2026