**Modeling of disaster death tolls with Benford’s Law**

**Abstract**: In 1938, physicist Frank Benford showed that among many naturally occurring collections of numbers, the leading significant digit is expected to be small, a pattern which was first indicated by astronomer Simon Newcomb and is now known as Benford's law or Newcomb-Benford's law. According to the law, in a set where the law fits, most numbers begin with digit 1, accounting for over 30% of total numbers, and the frequency of numbers starting with later digits falls almost exponentially. This law has been shown to apply to multitude of natural data sets, including stock and house prices, population numbers, death rates, length of large objects, and mathematical constants, among others. In this paper, the law (for the first digit) has been applied to model the number of deaths due to four different disasters: earthquakes, wars, accidents, and floods, and significant match with the law has been found, which can be applied to analyze, model, and predict disaster death tolls. The compliance was also tested with Pearson's Chi-squared goodness-of-fit test as well with some other testing procedures specific to the law, including Leemis’s m.

**Key words:** *First digit law, Benford’s law, probability theory, disaster death tolls, disaster modeling.*

1. **Introduction**

The Benford’s law deals with the first significant digits of numbers. For example, the first significant digits of the numbers 976 and 167 are 9 and 1, respectively. Since a number can start with digits 1, 2, … up to 9, it is natural to assume that the probability of any of the nine numbers being in the first position of a randomly chosen number should be . However, in many real world data sets, it is seen that the frequencies of numbers starting with each digit (1-9) are in strong congruence with what is predicted by Benford’s law. [1] The anomalous expected frequencies according to the law are given in table 2.

The law, though popularized by Frank Benford, traces its back to an 1881 work by the mathematician and astronomer Simon Newcomb, who showed that ten digits in the logarithmic table do not occur with equal frequency. [3] He concluded that numbers in logarithmic tables obey the pattern, while anti-logarithms do not. He went on to propose a law that the probability of a single number N being the first digit of a number would be equal to .

The law has generated a great number of papers, dealing with matters including theoretical developments, verifying for numbers of bases other than decimal, explanation of the pattern, testing compliance with real-world data, scale invariance, [4] applications to various phenomena, including for legal and accounting purposes, price digit analysis, [19] detection of fraudulent scientific data, [20] as well as developing of statistical tests to verify agreement with the law. [21]

In his paper, Benford (1937) demonstrated several data sets to agree with the law, including surface areas of 335 rivers, 3259 US populations, 104 physical constants, 1389 specific heat values, 703 pressure data, 1165 Black Body data, and 418 death rates, among others. It was later shown [5] and [6] that data sets containing uniformly numbers of several orders of magnitude (e.g. populations) agree with the Benford’s law, while data sets with numbers with mainly within only one order of magnitude (e.g. IQ scores) mostly, if not always, defy the law. Many well-known integer sequences have been shown to comply with the law, among them being Fibonacci numbers (Washington, 1981), the powers of 2 (Raimi 1976), and continuous growth processes, especially exponential growth or decay mechanism.

Several benchmark testing procedures have been suggested to test for agreement with the Benford’s law. Pearson’s chi-squared test is the de facto test in this scenario, while Kolmogorovo-Smirnov and Kuiper tests are better for small sample sizes. [9] There are some tests which are specific to the law, like the test statistics proposed by Leemis et al (2000) and Cho et al (2007).

The goal of this paper is to model the number of deaths due to disaster with the Benford’s law along with carrying the relevant statistical tests, and to reexamine, to greater extent, the compliance of the law with populations of countries (testing for 263 countries and territories, both with small and large populations). The findings can be applied to detecting fraudulent data, disaster-related policy-making, as well to estimate the expected number of death tolls due to disasters.

1. **Data and Methodology**
   1. **Data collection and processing**

The primary variable in the study is number of deaths from disasters, considering earthquakes, flood, accidents and wars. For all type of disasters, data were collected from all over world, focusing primarily on the deadliest cases on record with respect to number of deaths. Data on earthquakes were collected from various relevant sources, among them being the Unites States Geological Survey (USGS), International Association of Engineering Geology, National Geophysical Data Center etc., on accidents from various news sources including BBC, [13] the Age, [14] and Brisbane Times etc., on wars from [10], [11], and [12], among others, and on flood from Global Active Archive of Large Flood Events, [15], and the Dawn etc. Whenever an interval of estimate was found instead of a single number, the lowest estimate was considered. Among the types of accidents considered were structural fires and collapses, road, aviation and maritime accidents, explosions, industrial disasters, and sporting events. Additionally, gathered were data of populations of 263 countries and dependencies since 1960 until 2016, totaling 57 years of data for 263 regions. [17]

* 1. **Benford’s law**

A collection of numbers is said to obey the Benford’s law if the first digit d (where d ∈ {1, 2, ... , 9} has the probability of occurring

Where, P (d) is the percentage of numbers starting with the digit d.

The distribution of the leading digits, according to the law is shown in table 2.

The law can be stated for bases other than decimal as well for 2nd and later significant digits.

For any number system with base b, the law can be stated as follows:

Also, since the law can be generalized to digits other than the first, [18] the probability that d (d=0,1,…,9) is found as the nth digit is

For example, the probability that digit 2 is placed as the second digit is

* 1. **Test procedure**

Since the obtained frequencies in our analysis are not small (that is, they are all > 5), the chi-squared test of goodness-of-fit, a test procedure for checking observed frequencies with expected frequencies, can be assumed to be perfect to test for the compliance with the law. The relevant test statistic is as follows:

Where, = Pearson's cumulative test statistic, which asymptotically approaches a distribution.

= the number of observations of type i.

= the expected number of count of type i, as stated by the null hypothesis that the proportion of type i in the population is .

n = number of cells in the table.

For the above statistic, the null hypothesis is that there is no difference between observed and expected frequencies of numbers. If p-values obtained from the estimation are greater than 0.01 and 0.05, this will imply that the observed frequencies are in compliance with the expected frequencies (as predicted by Benford’s law) at 0.01 and 0.05 levels of significance, respectively.

Tests which are specific to this law are also available [22] [23] due to Leemis et al (2000) and Cho et al (2007).

The test statistic proposed by Leemis et al is known as max (m) or Lemis’s m and defined as below:

Where FSD is the first significant digit and N is the sample size.

**Table 1** Critical regions of Leemis’s m and Cho-Gaines’s test statistics at different levels of significance.

|  |  |  |  |
| --- | --- | --- | --- |
| α  Statistic | 0.10 | 0.05 | 0.01 |
| Leemis’ m | 0.851 | 0.967 | 1.212 |
| Cho-Gaines’ d | 1.212 | 1.330 | 1.569 |

And the test statistic, called the distance statistic (d) or Cho-Gaines’ d, proposed by Cho et al is as follows:

In table 1, α is the level of significance of the test, and if the observed value of the test statistics is greater than the values on the table, then it is decided that the data do not obey the Benford’s law in given level of significance.

1. **Results and discussion** 
   1. **. Testing of disaster data**

The studied dataset contains a total of 193 severe floods from all over the world, spanning more than a millennium, while a total of 197 significant wars have been considered. Included are both World War I and II, Gallic wars, Crimean War, Crusades, Vietnam War, Napoleonic Wars, and so on. Also, 190 cases of floods and 1538 cases of different accidents from all over the world were used in the analysis. The result is shown in table 2.

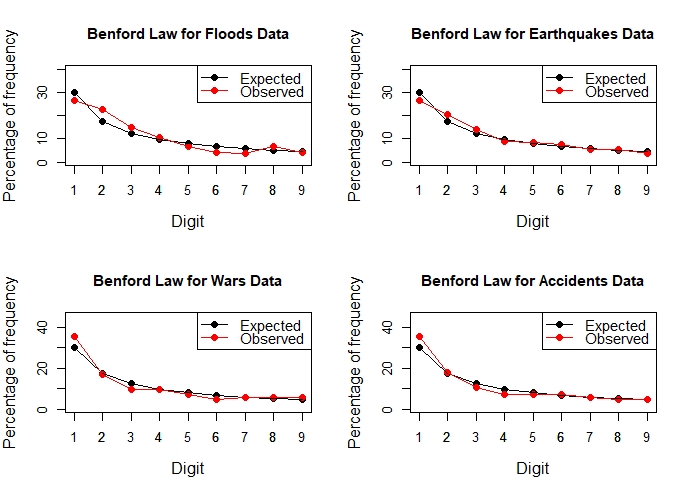
The table below and the succeeding plot summarize the result.

**Table 2** Summary of disaster death tolls with expected and observed frequencies of numbers starting with digits 1 through 9.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Digit | Expected percentage (by Benford’s Law, %) | Expected percentage  (Floods, %) | Expected percentage  (Earthquakes, %) | Expected percentage  (Wars, %) | Expected percentage  (Accidents, %) |
| 1 | 30.1 | 26.42 | 26.32 | 35.53 | 35.7 |
| 2 | 17.61 | 22.8 | 20.53 | 16.75 | 17.88 |
| 3 | 12.49 | 15.03 | 14.21 | 9.64 | 10.4 |
| 4 | 9.69 | 10.36 | 8.95 | 9.64 | 6.89 |
| 5 | 7.92 | 6.74 | 8.42 | 7.11 | 7.09 |
| 6 | 6.69 | 4.15 | 7.37 | 4.57 | 7.15 |
| 7 | 5.8 | 3.63 | 5.26 | 5.58 | 5.53 |
| 8 | 5.12 | 6.74 | 5.26 | 5.58 | 4.62 |
| 9 | 4.58 | 4.15 | 3.68 | 5.58 | 4.75 |
|  |  | Total observations: 193 | Total observations: 190 | Total observations: 197 | Total observations: 1538 |

**Figure 1** as well as table 2 and 3 suggest a strong agreement of the four studied data sets with Benford’s law. While the percentages of the digit 1 for floods and earthquakes are 26.42 and 26.32, both being less than the corresponding expected frequencies, the percentages of wars and accidents data sets for the same digit (1) exceed what is stated by Benford’s law, being 35.53 and 35.7, respectively i.e. the first digit percentages exceed expected percentage by almost 5%. All data sets, however, seem to, from the graphical illustrations, satisfy the law. The

assumption is confirmed by the p-values for all data sets except for accidents, the p-value being almost 0. For this data set, Pearson’s chi-squarest test as well as tests based on Leemis’ m and Cho-Gaines’ d differ from the graphical connotation. This contradiction might have resulted from the fact that although we have covered many different types of accidents, many other types remain to be explored. Still, the pattern of distribution of number to different digits conforms to Benford’s distribution. All other chi-squared p-values are well above 0.01, being no less than 0.28 and as high as 0.93, indicating clear and strong conformity with Benford’s law. Also, test results from Leemis’ m and Cho-Gaines’ d provide strong evidence in favor of conformity of floods, wars, and earthquakes death tolls with Benford’s law.



**Fig. 1 Expected and empirical percentages and frequencies of numbers starting with digits one through nine for disaster death estimates.** The red points and lines represent the observed percentages of numbers with each digits, while the black points and lines represent the expected and observed percentages of frequencies of numbers, respectively. X axis represents the digits and Y axis refers to the percentages of frequencies of numbers,

Table 3 reveals that according to Leemis’ m, the floods, earthquakes, and wars data sets are in strong conformity with the Benford’s law, while accidents data set is not, exactly as was suggested by the chi-squared tests. Also, by Cho-Gaines’ d measure, all data sets except the accidents satisfy the law. The agreement of three different tests lead us to conclude that death tolls due to floods, earthquakes, as well wars can be modeled by Benford’s law, while accidents death tolls may not agree with the law for small sample size.

Thus, the findings can be applied to detect artificial anomaly or fraud in data set related to disasters. Although, we have performed the analysis only for four types of disasters, it is expected that the law will hold for other disaster as well like landslide, volcanic eruptions, tornadoes, cyclones, tsunamis etc.

* 1. **Testing for populations data**

For this large-scale re-examination procedure, the world populations comprising of 263 countries and territories were considered.

**Table 3** Summary of statistical test of conformity

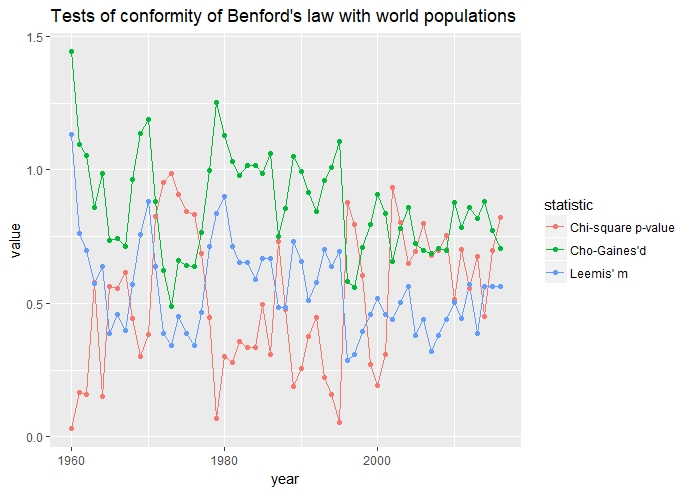
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data  Criteria | Floods | Earth-quakes | Wars | Acci-dents |
| Pearson’s chi-squared test p-value | 0.28 | 0.93 | 0.72 | <0.001 |
| Leemis’ m | 0.72 | 0.52 | 0.76 | 2.19 |
| Cho-Gaines’d | 1.10 | 0.73 | 0.94 | 2.62 |

Earlier in 2002, Sandron showed that populations of 198 countries or geopolitical entities as in the year 1997 followed Benford’s law. [16] Here, we have performed the analysis for 57 years: from 1960 to 2016, each year dealt separately, carrying out the above tests, and recording the associated p-values and other statistics. The resulting p-values are shown in figure 3.

Of the 57 years analyzed, no instance was found where the chi-squared test p-value was below 0.01, which implies world population data of different territories in any year conforming to Benford’s law, at 1% level of statistical significance, indicating a very strong compliance. The fact that the compliance is very strong is further justified from the finding that only one p-value is below 0.05 (in the case of the year 1960), which indicates that all other cases are very strongly congruent with Benford’s law.

Also examined were Leemis’ m and Cho-Gaines’ d measures. They have also been found to be uncompromising; at 1% level of significance, in stark dovetail with chi-squared test, no case could be concluded to defy the law—neither by Leemis’ m nor by Cho-Gaines’ d statistics. Thus, strong compliance with the law has been established.

Therefore, these findings can be applied to detect fraud or artificial anomaly in the figures related to populations of different territories at a given instant of time.



**Fig. 2 Chi-squared p-values and observed values of Leemis’ m and Cho-Gaines’ d test statistics in testing whether yearly world populations of 263 countries and territories conform to Benford’s law, with each year dealt separately. A total of 57 years were analyzed.**

**Conclusion**

We showed that the number of people dying in disasters around the world conforms well to the Benford’s law. The fact has been clearly illustrated by the plots as well as by Pearson’s chi-squared goodness-of-fit test and Leemis’s m and Cho-Gaines d measures. The tests have been performed for a total four types of disasters: floods, wars, earthquakes, and accidents. As far as the accidents are concerned, many different types were considered, including structural fires and collapses, road, aviation and maritime accidents, yet, although, the graph indicated conformity with the law, the p-values asnd

other tests indicated noncompliance, a fact possibly resulting from the inadequate types of accidents considered. In all other cases, p-values and m and d measures were in complete harmony with the graphs in establishing compliance of data with the law. The law has also been successfully shown to apply populations of territories in any year, concluded based on 57 years of data. The latter finding indirectly is associated with the former, since both are counts of human being. The finding can be employed for fraudulent or incorrect reporting of disaster death tolls, estimating probable number of deaths due to natural or artificial disasters etc.

**References**

1. Benford, F. 1938. The law of anomalous numbers. Proceedings of the American Philosophical Society. 78 (4): 551–572. JSTOR 984802.
2. Paul H. Kvam, Brani Vidakovic, Nonparametric Statistics with Applications to Science and Engineering, p. 158
3. Newcomb, S. 1881. Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics. American Journal of Mathematics, Vol. 4, No. 1. 4 (1/4): 39–40. doi:10.2307/2369148. JSTOR 2369148.
4. Roger S. Pinkham, On the Distribution of First Significant Digits, Ann. Math. Statist. Volume 32, Number 4 (1961), 1223-1230.
5. Steven W. Smith. The Scientist and Engineer's Guide to Digital Signal Processing, chapter 34, Explaining Benford's Law. (especially section 10).
6. Fewster, R. M. 2009. A simple explanation of Benford's Law (PDF). The American Statistician. 63 (1): 26–32. doi:10.1198/tast.2009.0005.
7. Washington, L. C. 1981. Benford's Law for Fibonacci and Lucas Numbers. The Fibonacci Quarterly. 19 (2): 175–177.
8. Raimi. Raimi, Ralph A. 1976. The First Digit Problem. American Mathematical Monthly. 83 (7): 521–538. doi:10.2307/2319349.
9. Stephens, M. A. 1970. Use of the Kolmogorov–Smirnov, Cramér–Von Mises and Related Statistics without Extensive Tables. Journal of the Royal Statistical Society, Series B. 32 (1): 115–122.
10. Nash. 1976. Darkest Hours. Rowman & Littlefield. p. 775. The Cambridge History of China: Alien regimes and border states, 907–1368, 1994, p.622, cited by White Barrett, David. World Christian Trends.
11. "Hope fades for water park victims". BBC News. 16 February 2004. Retrieved 13 September 2006.
12. Seven die at fun fair: Suddenly fire and smoke make the mock horror real. The Age. 11 June 1979. Retrieved 24 November 2016.
13. Brakenridge, G.R. Global Active Archive of Large Flood Events, Dartmouth Flood Observatory, University of Colorado, http://floodobservatory.colorado.edu/Archives/index.html.
14. Sandron, F. 2002. Les populations suivent-elles la loi des nombres anomaux?. Population, vol. 57,(4), 753-761. doi:10.3917/popu.204.0761.
15. Worldometers. Countries of the World. http://www.worldometers.info/geography/countries-of-the-world/
16. Theodore P. Hill, The Significant-Digit Phenomenon, The American Mathematical Monthly, Vol. 102, No. 4, (Apr., 1995), pp. 322–327.
17. Sehity, Tarek el; Hoelzl, Erik; Kirchler, Erich (2005-12-01). Price developments after a nominal shock: Benford's Law and psychological pricing after the euro introduction. International Journal of Research in Marketing. 22 (4): 471–480.
18. Diekmann, A. 2007. Not the First Digit! Using Benford's Law to detect fraudulent scientific data. J Appl Stat 34 (3) 321–329.
19. Morrow, J. 2010 Benford’s Law, Families of Distributions and a test basis, UW-Madison.
20. Leemis, L. M.; Schmeiser, B. W.; Evans, D. L. 2000. Survival distributions satisfying Benford's Law. The American Statistician. 54 (4): 236–241.
21. Cho, W. K. T.; Gaines, B. J. (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. The American Statistician. 61 (3): 218–223.
22. R Core Team 2018 . R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
23. Wickham, H. 2017. tidyverse: Easily Install and Load the 'Tidyverse'. R package version1.2.1.https://CRAN.R-project.org/package=tidyverse.