Benford’s Law and the Universe

Theodoros Alexopoulos
National Technical University of Athens

Stefanos Leontsinis
National Technical University of Athens
Brookhaven National Laboratory
(Dated: January 24, 2014)

Benford’s law predicts the occurrence of the $n^{th}$ digit of numbers in datasets originating from various sources all over the world, ranging from financial data to atomic spectra. It is intriguing that although many features of Benford’s law have been proven, it is still not fully mathematically understood. In this paper we investigate the distances of galaxies and stars by comparing the first, second and third significant digit probabilities with Benford’s predictions. It is found that the distances of galaxies follow the first digit law reasonable well, and that the star distances agree very well with the first, second and third significant digit.

I. INTRODUCTION

In 1881 the astronomer and mathematician S. Newcomb made a remarkable observation with respect to logarithmic books [1]. He noticed that the first pages were more worn out than the last. This led him to the conclusion that the significant digits of various physical datasets are not distributed with equal probability but the smaller significant digits are favored. In 1938 F. Benford continued this study and he derived the law of the anomalous numbers [2].

The general significant digit law [3] for all $k \in \mathbb{N}$, $d_1 \in \{1, 2, \ldots, 9\}$ and $d_k \in \{0, 1, \ldots, 9\}$, for $k \geq 2$ is

$$P(d_1, d_2, \ldots, d_k) = \log_{10} \left[ 1 + \left( \sum_{i=1}^{k} d_i \times 10^{k-i} \right)^{-1} \right]$$  \hspace{1cm} (1)

where $d_k$ is the $k^{th}$ leftmost digit. For example, the probability to find a number whose first leftmost digit is 2, second digit is 3 and third is 5 is $P(d_1 = 2, d_2 = 3, d_3 = 5) = \log_{10}(1 + 1/235) = 0.18\%$.

For the first significant digit can be written as

$$P(k) = \log_{10} \left( 1 + \frac{1}{k} \right), \quad k = 1, 2, \ldots, 9$$  \hspace{1cm} (2)

This law has been tested against various datasets ranging from statistics [4] to geophysical sciences [5] and from financial data [6] to multiple choice exams [7]. Studies were also performed in physical data like complex atomic spectra [8], full width of hadrons [9] and half life times for alpha and $\beta$ decays [10, 11].

An interesting property of this law is that it is invariant under the choice of units of the dataset (scale invariance) [12]. For example, if the dataset contains lengths, the probability of the first significant digits might be homologous in the case where the units are chosen to be meters, feet or miles.

Still, Benford’s law is not fully understood mathematically. A great step was done with the extension of scale to base invariance (the dependance of the base in which numbers are written) by Theodore Hill [13]. Combining these features and realising that all the datasets that follow Benford’s law are a mixture from different distributions, he made the most complete explanation of the law. Another approach in the explanation of the logarithmic law was examined by Jeff Boyle [14] using the Fourier series method.

A simple example of Benford’s law is performed on numerical sequences. It is already proven that the Fibonacci and Lucas numbers obey the Benford’s law [15]. The three additional numerical sequences considered in this paper are:

- Jacobsthal numbers ($J_n$), defined as
  \[- J_0 = 0 \]
  \[- J_1 = 1 \]
  \[- J_n = J_{n-1} + 2J_{n-2}, \quad \forall \ n > 1 \]

- Jacobsthal-Lucas numbers ($JL_n$), defined as
  \[- JL_0 = 2 \]
  \[- JL_1 = 1 \]
  \[- JL_n = JL_{n-1} + 2JL_{n-2}, \quad \forall \ n > 1 \]

- and Bernoulli numbers ($B_n$), defined by the contour interval
  \[- B_0 = 1 \]
  \[- B_n = \frac{n!}{2\pi i} \oint_{C=1} \frac{z^n}{e^z-1} dz \]

A sample of the first 1000 numbers of these sequences is used to extract the probabilities of the first significant digit to take the values 1, 2, ..., 9 and the second and third significant digits to be 0, 1, ..., 9. The results can be seen in figure [4]. Full circles represent the result from the
analysis of the Jacobsthal and Jacobsthal-Lucas numbers and the empty circles indicate the probabilities calculated from Benford’s formula (equation 1). It is clear that all three sequences follow Benford’s law for the first (black), second (red) and third (blue) significant digit.

In the following sections we examine the distances of stars and galaxies and compare the probabilities of occurrence of the first, second and third significant digit with Benford’s logarithmic law. If the location of the galaxies in our universe and the stars in our galaxy are caused by uncorrelated random processes, Benford’s law will not be followed because each digit would be equiprobable to appear. To our knowledge this is the first paper that attempts to correlate cosmological observables with Benford’s law.

II. APPLICATION ON THE UNIVERSE

Cosmological data with accurate measurements of celestial objects are available since the 1970s. We examine if the frequencies of occurrence of the first digits of the distances of galaxies and stars follow Benford’s law.

A. Galaxies

We use the measured distances of the galaxies from references 16, 87. The distances considered on this dataset are based on measurements from type II Supernova and all the units are chosen to be Mpc. The type-II supernova (SNII) radio standard candle is based on the maximum absolute radio magnitude reached by these explosions, which is $5.5 \times 10^{23}$ ergs/s/Hz. The total number of galaxies selected is 702. The results can be seen in figure 2(a) where with open circles we notate Benford’s law predictions and the measurements with the circle. Unfortunately due to lack of statistics the second and the third significant digit cannot be analyzed. The trend of the distribution tends to follow Benford’s prediction reasonably well.

B. Stars

The information for the distances of the stars is taken from the HYG database 88. In this list 115 256 stars are included. The result after analysing this dataset can be seen in figure 2(b). The first (black full circles) and especially the second (red full circles) and the third (blue full circles) significant digits follow well the probabilities predicted by Benford’s law (empty circles).

III. SUMMARY

Benford law of significant digits was applied for the first time on cosmological measurements. It is shown that the distance of the stars follow well Benford law for the first, second and third significant digit. The probabilities of the first significant digit of galaxy distances is in good agreement with Benford’s predictions. It would be interesting for this study to be repeated in the distant future with more statistics and compare the second
Significant digit of galaxy distances

| First Digit Measured | First Digit Benford |
|----------------------|---------------------|
| 0.0                 | 0.05                |
| 0.1                 | 0.1                |
| 0.15                | 0.15               |
| 0.2                 | 0.2                |
| 0.25                | 0.25               |
| 0.3                 | 0.3                |
| 0.35                | 0.35               |
| 0.4                 | 0.4                |

First Digit Measured  | First Digit Benford

(a) FIG. 2. Comparisons of Benford’s law (empty circles) and the distribution of the first (black), second (red) and third (blue) significant digit of the distances of the (a) galaxies and (b) stars (full circles).

Significant digit of star distances

| First Digit Measured | First Digit Benford |
|----------------------|---------------------|
| 0.0                 | 0.05                |
| 0.1                 | 0.1                |
| 0.15                | 0.15               |
| 0.2                 | 0.2                |
| 0.25                | 0.25               |
| 0.3                 | 0.3                |
| 0.35                | 0.35               |
| 0.4                 | 0.4                |

First Digit Measured  | First Digit Benford

(b) FIG. 2. Comparisons of Benford’s law (empty circles) and the distribution of the first (black), second (red) and third (blue) significant digit of the distances of the (a) galaxies and (b) stars (full circles).

and third significant digit law on the galaxies. Another interesting update would be to repeat this study in the future with the same sample of universes. The question to be answered is whether the redshift will alter the results presented in this paper.

ACKNOWLEDGMENTS

We would like to thank I.P. Karananas for the lengthy discussions on this subject. We would like also to thank Emeritus Professor Anastasios Filippas for the valuable comments.

The present work was co-funded by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2007-1013. ARISTEIA-1893-ATLAS MICROMEGAS.

[1] S. Newcomb, Note on the frequency of use of the different digits in natural numbers, Am. J. Math., vol 4, No. 1/4, 39-40 (1881)
[2] F. Benford, The law of anomalous numbers, Proc. Am. Phil. Soc. 78, 551-572 (1938)
[3] Theodore P. Hill, The significant-digit phenomenon, The American Mathematical Monthly, vol 102, No. 4, 322-327 (1995)
[4] L. Shao and B. Q. Ma, First-digit law in nonextensive statistics, Phys. Rev. E, vol 82, 041110 (2010)
[5] A. S. De and U. Sen, Benford’s law detects quantum phase transitions similarly as earthquakes, Europhys. Lett., vol 95, 50008 (2011)
[6] P. Clippe and M. Ausloos, Benford’s law and Theil transform on financial data, [arXiv:1208.5896v1] (2012)
[7] F. M. Hopee, Benford’s law and distractors in multiple choice exams, [arXiv:1311.7609] (2013)
[8] J. C. Pain, Benford’s law and complex atomic spectra, Phys. Rev. E, vol 77, 012102 (2008)
[9] L. Shao and B. Q. Ma, First digit distribution of hadron full width, Mod. Phys. Lett. A 24, 3275-3282 (2009)
[10] B. Buck et al., An illustration of Benford’s first digit law using alpha decay half lives, Eur. J. Phys. 14, 59 (1993)
[11] NI Dong-Dong et al., Benford’s law and β-decay half-lives, Commun. Theor. Phys., vol 51, No. 4, 713-716 (2009)
[12] M. R. Wojcik, Notes on scale-invariance and base-invariance for Benford’s Law, [arXiv:1307.3620] (2013)
[13] T. P. Hill, Base-Invariance Implies Benford’s Law, Proceedings of the American Mathematical Society 123.3, 887-895 (1995)
[14] J. Boyle, An application of Fourier series to the most significant digit problem, The American Mathematical Monthly, vol 101, No. 9, pp. 879-886 (1994)
[15] J. Wlodarski, Fibonacci and Lucas numbers tend to obey Benford’s law, Westhoven, Federal Republic of Germany (1971)
[16] Gurugubelli et al., Photometric and spectroscopic evolution of type II-P supernova SN 2004A, Bulletin of the Astronomical Society of India, 36 (2-3), 79-97 (2008)
[17] Bartel N., Angular diameter determinations of radio supernovae and the distance scale, Supernovae as distance indicators; Proceedings of the Workshop, Cambridge, MA, September 27, 28, 1984 (A86-38101 17-90). Berlin
and New York, Springer-Verlag, 107-122 (1985)

[18] Bartel N., Hubble’s constant determined using very-long baseline interferometry of a supernova, Nature 318, 25-30 (1985)

[19] Chugai N. N., Supernova 1987A - Ha Profile and the Distance of the Large Magellanic Cloud, Astronomicheskii Tsirkulyar NO. 149/MAY (1987)

[20] Hoeflich, P., Model calculations for scattering dominated atmospheres and the use of supernovae as distance indicators, Nuclear astrophysics: Proceedings of the Workshop, Tegernsee, Federal Republic of Germany, Apr. 21-24, 1987 (A89-10202 01-90). Berlin and New York, Springer-Verlag, 307-315 (1987)

[21] R. V. Wagoner et al., Supernova 1987A: A Test of the Improved Baade Method of Distance Determination, BAAS 08/1988; 20:985 (1988)

[22] Bartel N., Determinations of distances to radio sources with VLBI, The impact of VLBI on astrophysics and geophysics; Proceedings of the 129th IAU Symposium, Cambridge, MA, May 10-15, 1987 (A89-13726 03-90). Dordrecht, Kluwer Academic Publishers, 175-184 (1988)

[23] Chilkurthy, M. et al., Type-II Supernova Photospheres and the Distance to Supernova 1987A, Atmospheric Diagnostics of Stellar Evolution, Chemical Peculiarity, Mass Loss, and Explosion. Proceedings of the 108th. Colloquium of the International Astronomical Union, held at the University of Tokyo, Japan, September 1-4, 1987. Lecture Notes in Physics, Volume 350, Editor, K. Nomoto; Publisher, Springer-Verlag, Berlin, New York, 1988. ISBN # 3-540-19478-9. LC #: QB806.118 1987, P. 295, (1987)

[24] Schmidt, B. P., Expanding Photospheres of Type II Supernovae and the Extragalactic Distance Scale, American Astronomical Society, 181st AAS Meeting, #107.04D; Bulletin of the American Astronomical Society, vol 24, p.1292 (1992)

[25] Schmidt-Kaler, T., The Distance to the Large Magellanic Cloud from Observations of SN1987A, Variable Stars and Galaxies, in honour of M. W. Feast on his retirement, ASP Conference Series, vol 30, B. Warner, Ed., p. 195 (1992)

[26] Brian P. Schmidt et al., The unusual supernova SN1993J in the galaxy M81, Nature 364, 600 - 602 (1993)

[27] Brian P. Schmidt et al., Type II Supernovae, Expanding Photospheres, and the Extragalactic Distance Scale, Thesis (PH.D.) - Harvard University, 1993. Source: Dissertation Abstracts International, Volume: 54-11, Section: B, page: 5717 (1993)

[28] Fernley J. et al., The absolute magnitudes of RR Lyrae from HIPPARCOS parallaxes and proper motions., Astron. Astrophys., vol 330, 515-520 (1998)

[29] Panagia, N., New Distance Determination to the LMC, Memorie della Societa Astronomia Italiana, vol 69, p.225 (1998)

[30] Garnavich, P. et al., Supernova 1987A in the Large Magellanic Cloud, IAU Circ., 7102, 1 (1999)

[31] Panagia N., Distance to SN 1987 A and the LMC, New Views of the Magellanic Clouds, IAU Symposium #190, Edited by Y.-H. Chu, N. Suntzeff, J. Hesser, & D. Bohlender. ISBN: 1-58381-021-8, p.549 (1999)

[32] Leonard, Douglas C. et al., A Study of the Type II Plateau Supernova 1999ig and the Distance to its Host Galaxy, NGC 3184, The Astronomical Journal, vol 124, Issue 5, pp. 2490-2505 (2002)

[33] Dessart, L. et al., Quantitative spectroscopic analysis of and distance to SN1999em Astronomy and Astrophysics, vol 447, Issue 2, pp.691-707 (2006)

[34] Takats V. J., Improved distance determination to M 51 from supernovae 2011dh and 2005cs, Astronomy & Astrophysics, vol 540, id.A93, 7 pp. (2012)

[35] Sonnenorn G. et al., The Evolution of Ultraviolet Emission Lines from Circumstellar Material Surrounding SN 1987A, The Astrophysical Journal, vol 477, 848-864 (1997)

[36] Gould A. et al., Upper Limit to the Distance to the Large Magellanic Cloud, The Astrophysical Journal, vol 494, 118-124 (1998)

[37] Romaniello M. et al., Hubble Space Telescope Observations of the Large Magellanic Cloud Field Around SN 1987A: Distance Determination with Red Clump and Tung of the Red Giant Branch Stars The Astrophysical Journal, vol 530, 738-743 (2000)

[38] Hamuy M. et al., The Distance to SN 1999em from the Expanding Photosphere Method, The Astrophysical Journal, vol 558, 615-642 (2001)

[39] Mitchell R. C. et al., Detailed Spectroscopic Analysis of SN 1987A: The Distance to the Large Magellanic Cloud Using the Spectral-Fitting Expanding Atmosphere Method, The Astrophysical Journal, vol 574, 293-305 (2002)

[40] Bartel N. et al., SN 1997C VLBI: 22 Years of Almost Free Expansion, The Astrophysical Journal, vol 591, 301-315 (2003)

[41] Leonard D. C. et al., The Cepheid Distance to NGC 1637: A Direct Test of the Expanding Photosphere Distance to SN 1999em, The Astrophysical Journal, vol 594, 247-278 (2003)

[42] Nugent P. et al., Toward a Cosmological Hubble Diagram for Type II-P Supernovae, The Astrophysical Journal, vol 645, 841-850 (2006)

[43] Baron E. et al., Reddening, Abundances and Line Formation in SNe II, The Astrophysical Journal, vol 662, 1148-1155 (2007)

[44] Bartel N. et al., SN 1993J VLBI. IV. A Geometric Distance to M81 with the Expanding Shock Front Method, The Astrophysical Journal, vol 668, 924-940 (2007)

[45] Dessart L. et al., Using Quantitative Spectroscopic Analysis to Determine the Properties and Distances of Type II Plateau Supernovae: SN 2005cs and SN 2006bp, The Astrophysical Journal, vol 675, 644-669 (2008)

[46] Pozzański D. et al., Improved Standardization of Type II-P Supernovae: Application to an Expanded Sample, The Astrophysical Journal, vol 694, 1067-1079 (2009)

[47] Jones M. I. et al., Distance Determination to 12 Type II Supernovae Using the Expanding Photosphere Method, The Astrophysical Journal, vol 696, 1176-1194 (2009)

[48] Kasen D. et al., Type II Supernovae: Model Light Curves and Standard Candle Relationships, The Astrophysical Journal, vol 703, 2205-2216 (2009)

[49] D’Andrea C. B. et al., Type II-P Supernovae from the SDSS-II Supernova Survey and the Standardized Candle Method, The Astrophysical Journal, vol 708, 661-674 (2010)

[50] Olivares F. E. et al., The Standardized Candle Method for type II Plateau Supernovae, The Astrophysical Journal, vol 715, 833-853 (2010)

[51] Roy R. et al., SN 2008in - Bridging the Gap Between Normal and Faint Supernovae of Type IIP, The Astrophysical Journal, vol 736, 76 (2011)
