Non-Newcomb-Benford Distributions



W.A. Kreiner Faculty of Natural Sciences University of Ulm

1. Introduction

In many tables of numbers that play a role in everyday life the first digits follow the Newcomb-Benford Law saying that, in the first position of a number, a 1 occurs 6.58 times more often than a 9 while the probability of occurrence of the other digits in the first position follows a continuous function [1, 2]. This holds especially for objects which grow or decrease exponentially in size or value [3, 4]. Examples are prices, wages, monetary assets, loss of value over time, but, to a good approximation, also, page numbers in literature citations, street addresses, population of municipalities, development aid, length of rivers, microorganisms in a culture, spread out of internet contents, atomic weights, the masses of exoplanets or meteorites, and many other distributions which span at least a few orders of magnitude [3, 5]. A narrow distribution does not meet the criteria for Benford's law [3]. A nearly Newcomb-Benford distribution has to be approximately invariant when scaled up by any factor up to 10.

However, there are distributions which deviate strongly from the NB Law, although examples of this kind seem to be rare [6, 7, 8].

In the following, fragments of minerals are investigated [6, 7] as well as masses of celestial objects [9]. The question is discussed how it comes to distributions of the first digits which are so different from the NB Law. In case of fractured minerals, a possible explanation is presented.

Another goal of the manuscript is to calculate the first digits´ Shannon entropy [10, 11]. It gives the average information content of one leading digit and, simultaneously, the minimum number of binary digits necessary to characterize their distribution [7,10, 11,12]. Analysing minor deviations from the NB Law can be used for portfolio formation and index tracking [13].

2. The Newcomb-Benford Distribution

If a quantity is exponentially increasing or decreasing in time, then the percentage of time that it has each first digit satisfies Newcomb-Benford Law [3]. An example is given in the appendix. Many distributions follow the NB Law more or less exactly.

For a more general approach to the problem of first digits´ probabilities one starts from the density function [7]

$$\mathbf{D(x) \sim \frac{1}{x^P} = x^{-P}} \qquad \text{[EQ 1]}$$

The probability W(a,b) of the object to exhibit a size between a and b then is

$$\mathbf{W\ (a,b) = A \int_a^b x^{-P} dx} \qquad \text{[EQ 2]}$$

with A being a constant. In case of exponential growth of a certain object, x means the time and P equals 1. Choosing a=1 and b=2 gives the proportion of 1s among the first digits. The bars in **Figure 1** refer to subsequent integers a, b, defining a “z-interval“.

$$\mathbf{W\ (a,b) = W(z) = A \int_a^b x^{-1} dx = A\ ln \frac{b}{a}} \qquad \text{[EQ 3]}$$

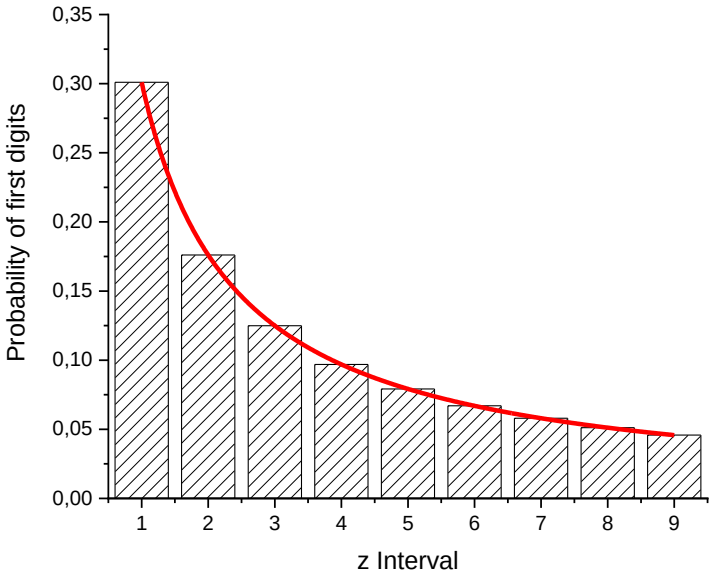


Figure 1 The Newcomb-Benford probability distribution of first digits [7]. The probability of a 1 to be the leading digit is 6.58 times as high as for a 9. The red curve shows W(z), [EQ 3].

[EQ 3] is scale invariant, i.e., one would obtain the same power law and the same ratio of 1s to the 9s in the first digital place if one would use other units to measure the objects´ properties, provided you choose comparable number intervals (e.g., one decade).

3. Non-NB Distributions

In case the first digits´ distribution would not follow the NB Law, it can be derived from the density function

$$\mathbf{D(x) = \frac{A}{x^P} = Ax^{-P}} \qquad \text{with } P \neq 1. \qquad \text{[EQ 1]}$$

The proportion of numbers falling into the range between x_1 and x_2 is calculated from

$$\mathbf{W(x_1, x_2)} = A \int_{x_1}^{x_2} D(x)dx = A \int_{x_1}^{x_2} x^{-P}\mathrm{dx} = \frac{\mathbf{A}}{\mathbf{1-P}} [\mathbf{x_2}^{\mathbf{1-P}} - \mathbf{x_1}^{\mathbf{1-P}}]$$

$$\mathbf{W(z)} = \frac{\mathbf{A}}{\mathbf{1-P}} [(\mathbf{z+1})^{\mathbf{1-P}} - \boldsymbol{z}^{\mathbf{1-P}}] \qquad \text{[EQ 4]}$$

In case the limits x_1 and x_2 are chosen as subsequent integers, $W(x_1, x_2) = W(z)$ is in proportion to the numbers starting with the digit $x_1 = z$. This law is scale invariant, too.

4. The Shannon Entropy

Related to the occurrence of first digits, the entropy H is determined in the following way [7, 9, 10, 11]: If the first digits occur with different probabilities, $\mathbf{p_i}$, one forms the logarithmus dualis of the value corresponding to a particular digit, $\mathbf{ld\ p_i}$, and multiplies this with the probability of its occurrence, giving $\mathbf{p_i\ ld\ p_i}$. Taking the sum results in

$$\boldsymbol{H} = -\sum_{i=1}^{n}(\boldsymbol{p_i}\,\mathbf{ld}\,\boldsymbol{p_i}) \text{ [bits per digit, bpd].} \qquad \text{[EQ 5]}$$

with n being the total number of different digits (usually 9). H means the average information of a single digit and shows its highest value when all first digits are equally likely. If not, the value of H decreases. Therefore, H gives some information about the distribution of first digits and, simultaneously, the number of bits necessary for its transmission [7].

5. Mass of Meteorites.

The first digits´ distribution of the mass of 1077 meteorites [9] is given in **Table 1** and in **Figure 2**. Using equation [4], a value of P= 1.035(35) was obtained which corresponds to the NB Law within error limits. From this, for the 1s vs. 9s, a ratio of 7.03 has been obtained. From the observed values a ratio of 8.79 has been found.

Table 1 Mass of Meteorites. Occurrence of first digits.

1st digit	1	2	3	4	5	6	7	8	9	∑
Occurrence	334	187	146	116	77	74	58	47	38	**1077**

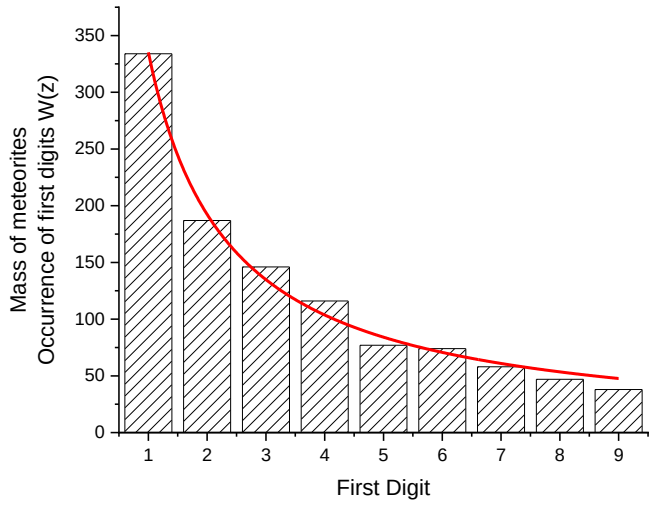


Figure 2 First digits´ distribution of the mass of 1077 meteorites. Data from [9].

6. Mass Distributions of Minerals

6.1 Fragments of Marble

The distribution of the first digits of fractions of minerals occasionally deviates strongly from the NB Law [6, 7]. Six samples of marble (irregularly shaped plates of 10 mm thickness, weighing between 4.70 and 51.69 grams) had been crushed in a hydraulic press and the weight of 1052 fragments between 10 and 99 mg has been taken. Fragments below 10 and above 99 mg have been omitted. From a fit of [EQ 4] (red curve in **Figure 3**) the value of

P =2.003(25) has been determined. **Table 2** gives the numbers of fragments sharing the same first digits.

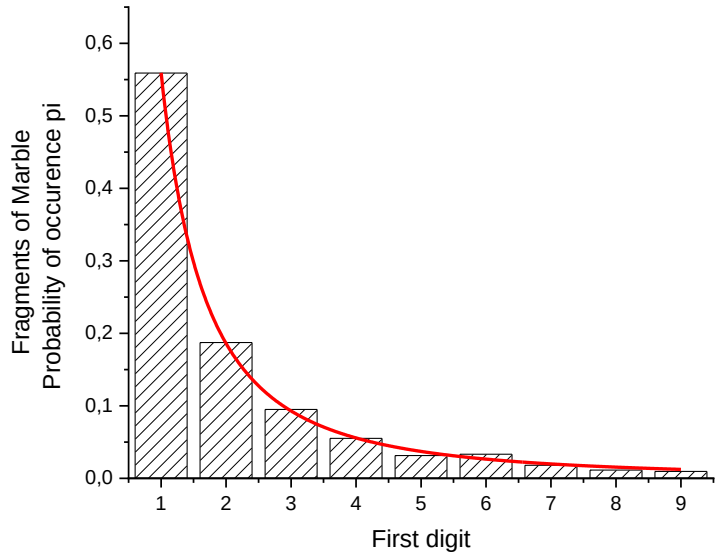


Figure 3 Fragments of six samples of marble [7]. Fit of the function W(z) to a bar graph giving the probabilities of first digits of the weight of 1052 fragments between 10 and 99 mg. The function fitted corresponds to a density function with P= 2.003(25).

Table 2 Probabilities of Fragments of Marble

z interval	Number of fragments	p_i	$-p_i$ ld p_i
1	588	0.5589	0.4691
2	197	0.1873	0.4526
3	100	0.0951	0.3227
4	58	0.0551	0.2305
5	33	0.0314	0.1567
6	35	0.0333	0.1633
7	19	0.0181	0.1046
8	12	0.0114	0.0736
9	10	0.0095	0.0638
Σ	1052	1.0001	**H= 2.0370** bits per digit

Assuming the distribution would follow exactly this function, the statistical weights and the first digits´ probabilities can be obtained from

$$\mathbf{W\,(a,b) = W(z) = A}\int_{\mathbf{a}}^{\mathbf{b}} \mathbf{x^{-P}\,dx} = \frac{\mathbf{A}}{\mathbf{-1.003}}\left[\boldsymbol{b}^{\mathbf{-1.003}} - \boldsymbol{a}^{\mathbf{-1.003}}\right], \qquad \text{[EQ 6]}$$

where a,b are the integers defining the z intervals.

In this case, a 1 (a=1, b=2) is expected to occur as a leading digit 45.27 times more often than a 9 (a=9, b=10). For statistical reasons, the observed ratio is different (58.8). It is closer to the expected value for the ratio of the 1s to the 7s (calculated: 29.8, observed: 30.9).
From the probabilities $\mathbf{p_i}$ the value of the entropy H= 2.0370 has been found (**Table 2**).

6.2 Fragments of Weisser Jura

To give an impression of the average distribution (or number density) function of Weisser Jura, the average P values have been determined separately for each of the five decades [6].

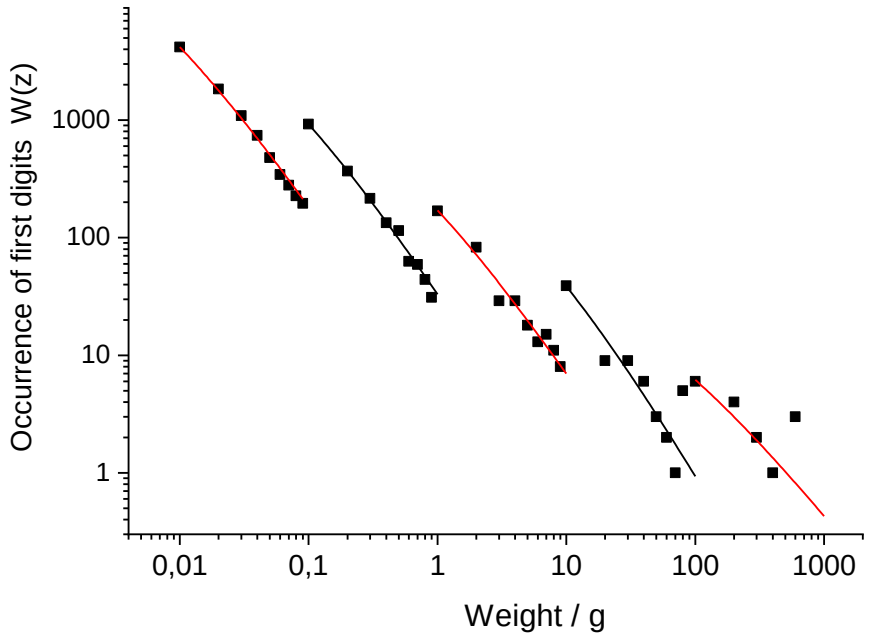


Figure 4 Plot of the average probability function W(z) of four samples of Weisser Jura [6]. The squares indicate the number of fragments per z interval.

Figure 4 shows the numbers of all fragments falling into particular z-intervals, starting at the left with z = [10mg, 20mg), up to to z = [600g, 700g) at the right. It shows kind of a saw tooth function**.** Then, using Eq. (4), P values have been fitted for each of the five decades [10 mg, 100 mg) till [1g, 9g), resulting in values of P= 1.571(15), P= 1.668(16), P= 1.602(59), P= 1.860(126), and P= 1.342(226). For the decade [10 mg, 100 mg), the ratio of 1s vs. 9s is 21.44.

6.3 Fragments of Granite

Samples of granite (13 square discs, 28g each, 3 more samples, slightly larger) have been crushed in a hydraulic press. The weight of 3849 fragments between 10 mg and 99 mg was taken [7]. Their distribution is given in **Figure 5**.

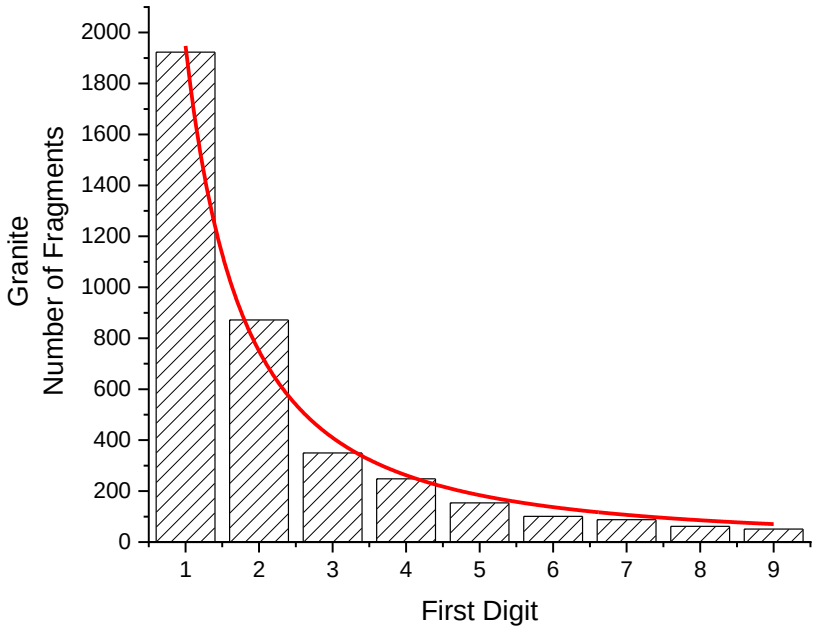


Figure 5. Fit of the function W(z) to a bar graph giving the probabilities of first digits of the weight of 3849 fragments of granite between 10 and 99 mg [7]. P= 1.748(81). From the function the probabilities for W(z= 1) and W(z= 9) have been calculated. Their ratio is 27.62. From the counted number of fragments the ratio is 37.71.

From a fit of ([EQ 4]; red curve in **Figure 5**) the value for the parameter P =1.748(81) has been determined. Assuming the distribution would follow exactly this function, the statistical weights and the first digits´ probabilities can be obtained from

$$\mathbf{W\,(a,b) = W(z) = A\int_a^b x^{-P}\,dx = \frac{A}{-0.748}\left[(z+1)^{-0.748} - z^{-0.748}\right]} \qquad \text{[EQ 7]}$$

From this, a 1 is expected to occur as a leading digit 27.62 times more often than a 9. From the counted number of fragments, the ratio is found to be larger (1923/51=37.71). From the probabilities of occurrence (**Table 3**) a value of H= 2.1817 bits per digit has been found.

Scale invariance: The distribution of the first digits would be close to the same power law when fragments were selected in the comparable 10-fold range 1.0 X 10^{-4} ounces to 9.9 X10^{-4} ounces.

Table 3 Probabilities of occurrence and first digits´ entropy of samples of Granite

z interval	Number of fragments	p_i	$-p_i \, ld \, p_i$
1	1923	0.4996	0.5002
2	872	0.2266	0.4853
3	350	0.0909	0.3145
4	248	0.0644	0.2549
5	154	0.0400	0.1858
6	102	0.0262	0.1378
7	88	0.0229	0.1246
8	62	0.0161	0.0959
9	51	0.0133	0.0827
∑	3849	1.0000	**H= 2.1817** bits per digit

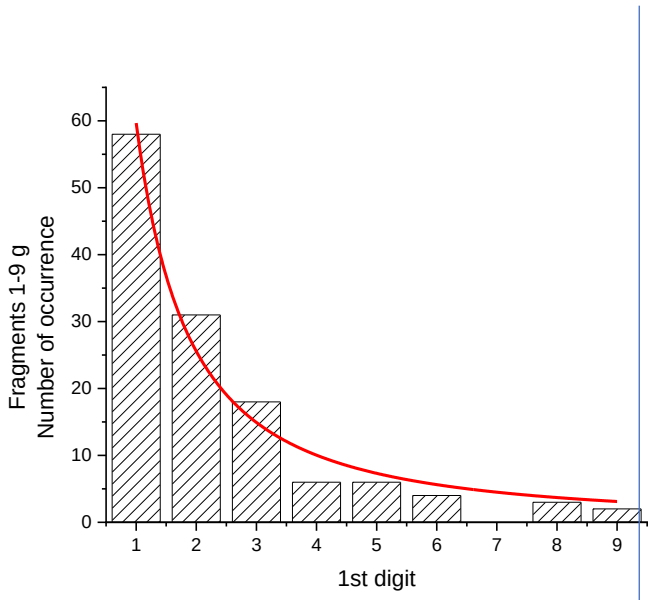
6.4 Fragments of Glass

647 fragments of two glass bottles, 1 liter each, were weighed. Fragments below 10 mg and above 9 grams have been omitted. The first digits´ distribution is shown in the diagrams of **Figures 6** and **7**. The P-values are given in **Table 4**.

Table 4 Fragments of glass. P-values and first digits´ratios.

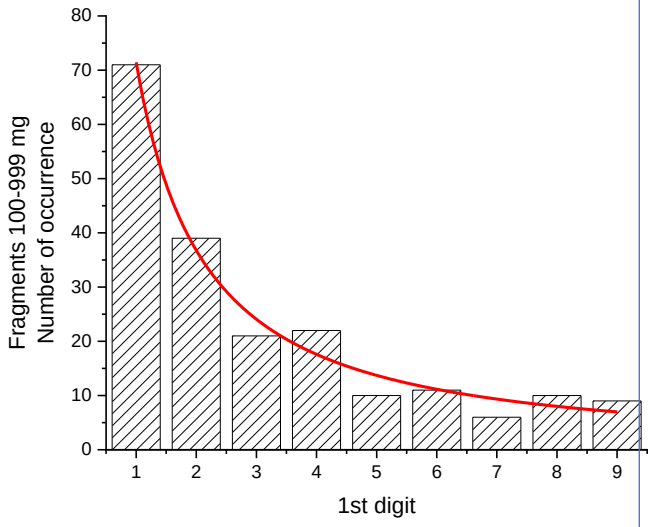
Range / Milligrams		1000 -9999	100 -999	10 - 99	10 – 9,999
Number of fragments		123	209	320	647
P-value		1.56(13)	1.232(73)	1.412(60)	1.391(27)
Number of 9s in relation to the number of 1s	From the function fitted	19.2	10.2	14.5	13.9
	Measured	29.0	7.9	8.9	11.3

From the distribution function fitted to the cumulated number of leading digits in the range between 10 and 9,999mg a negative power law exponent P= 1.391(27) as well as a ratio of the first digits (1s vs. 9s) of 13.9 has been found.



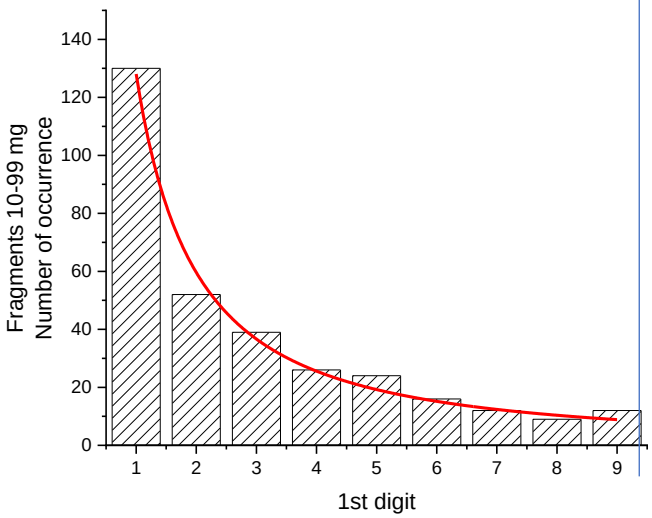
1 – 9 g

1 vs. 9 = 19.2 (calc.)



100 – 999 mg

1 vs. 9 = 10.2 (calc.)



10 – 99 mg

1 vs. 9 = 14.5 (calc.)

Figure 6 Fragments of glass. Number of occurrence of first digits for different weight ranges. Ratio 1s vs. 9s (calculated).

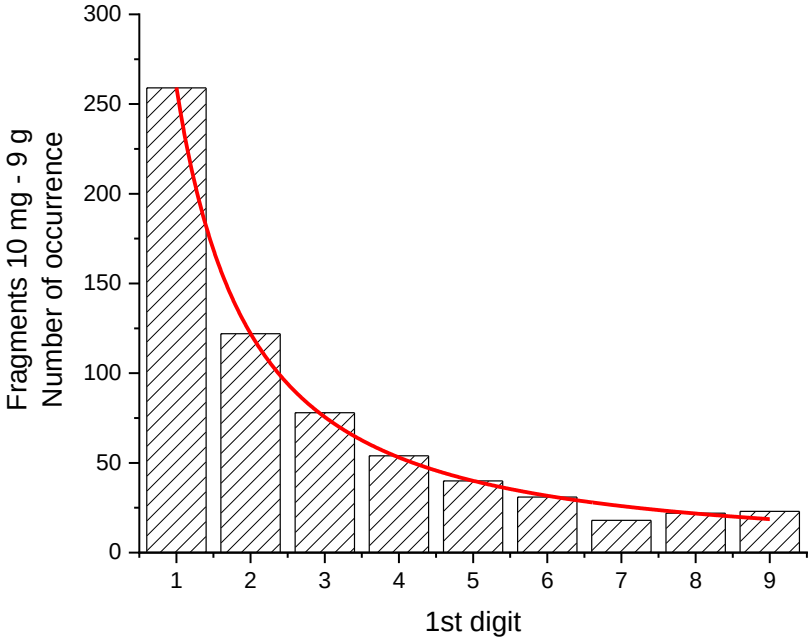


Figure 7 Accumulated number of fragments and 1st digits´ distribution between 10 mg and 9 grams. The calculated ratio of the 1st digits amounts to 13.9.

Table 5 Number of fragments and first digits´ entropy of samples of glass. The last two columns refer to the cumulated number of fragments.

z interval	Number of fragments				p_i	$-p_i$ ld p_i
	1000/ 9,999	100/ 999	10/ 199	10/ 9,999		
1	58	71	130	259	0.400	0.529
2	31	39	52	122	0.189	0.454
3	18	31	39	78	0.121	0.368
4	6	22	26	54	0.084	0.299
5	6	10	24	40	0.062	0.248
6	4	11	16	31	0.048	0.210
7	0	6	12	18	0.028	0.144
8	3	10	9	22	0.034	0.166
9	2	9	12	23	0.036	0.171
∑	128	209	320	647	1.002	**H= 2.** 589 bits per digit

7 Why is the 1 so strong in the majority?

It's kind of amazing that, when it comes to the weight of the fragments, the 1 occurs even more often as a leading digit than expected from the NB Law. To answer that you have to consider the probability that the mass breaks into a certain number of fragments with a certain weight, and thus with a certain first digit.

Table 6 a gives the number of first digits for the case, so one would divide the sample into the number of equal sized fragments given in the left column. For example, cutting a mass of 379426 grams (a weight chosen arbitrarily) into 380 pieces of equal size, one finds that the weight of each piece starts with a 9. This holds up to 421 fragments. Dividing the sample into smaller parts (between 422 and 474 fragments) results in an 8 in the first position. Between 1898 and 3794 fragments of equal size the leading digit will be a 1. Although the sample will probably never break into pieces of equal size, there will be at least a good chance for the majority of the fragments that their weight will start with a 1. A 9 will be rare. **Table 6 b** gives the probabilities of occurrence of the first digits. From **Table 6 c** one can see that, in this example, the ratio of 1s to 9s is 45265/ 1006 = 44.995. Starting from a mass with different weight produces quite similar results [7].

Figure 8 gives the distribution of probabilities among the z-intervals. From a fit (EQ 4) one obtains P= 2.0 for the exponent of the density function. From this value the ratio of the nines to the ones occurring as a first digit should be exactly 45. From the experiments ratios of 45.27 (marble), 19.63 (Weisser Jura), 27.62 (granite), and 13.9 (glass) have been calculated (Chapters 6.1 - 6.4).

How can one calculate the numbers given in **Table 6a**?

Suppose you divide the weight of a mass M by an integer, A, so that the weight of one fragment is 9.99... (≈ 10) grams. You choose another integer, B, so that a fraction weighs exactly 9.0 grams. How many integers are there in between that, when the mass is divided by them, give a leading 9? From

$\frac{M}{A} = 10$; $\frac{M}{B} = 9$ and $\frac{M}{10} = A$; $\frac{M}{9} = B$ it follows:

B – A = $M \cdot \left(\frac{1}{9} - \frac{1}{10}\right) = M \cdot 0.0111\ldots..$; with M= 379,426 g this gives 4216 fragments.

Correspondingly, the number of integers leading to fractions with a leading 1 is found from:

$\frac{M}{A} = 2$; $\frac{M}{B} = 1$ and $\frac{M}{2} = C$; $\frac{M}{1} = D$:

D – C = $M \cdot \left(1 - \frac{1}{2}\right) = M \cdot 0.5$ >> 189713 fragments.

As a result, we find that the number of leading 1s to the number of leading 9s is

$\frac{M \cdot 0.5}{M \cdot 0.0111\ldots.} = 45$ or the number of 2s to the 7s is 9.3. **Table 7** gives the probabilities.

Table 6a. Number of Fragments of a Mass of 379426 grams with identical first digits

Number of fragments	1	2	3	4	5	6	7	8	9
38 – 42									5
43 - 47								5	
48 – 54							7		
55 – 63						9			
64 – 75					12				
76 – 94				19					
95 – 126			32						
127 – 189		63							
190 – 379	190								
380 – 421									42
422 – 474								53	
475 – 542							68		
543 – 632						90			
633 – 759					127				
760 – 949				190					
950 – 1265			316						
1266 – 1897		632							
1898 – 3794	1897								
3795 – 4215									421
4216 – 4742								527	
4743 – 5420							678		
5421 – 6323						903			
6324 – 7588					1265				
7589 – 9485				1897					
9486 – 12647			3162						
12648 – 18971		6324							
18972 – 37942	18971								
37943 – 42158									4216
42159 - 47428								5270	
47429 – 54203							6775		
54204 – 63237						9034			
63238 – 75885					12648				
75886 – 94856				18971					
94857 – 126475			31619						
126476 – 189713		63238							
189714 – 379426	189713								
∑∑: 379389	**210771**	**70257**	**35129**	**21077**	**14052**	**10036**	**7528**	**5855**	**4684**

Table 6b. Probabilities of Occurrence of First Digits

	1	2	3	4	5	6	7	8	9
p_i	**0.55555**	**0.18518**	**0.09259**	**0.05556**	**0.03704**	**0.02645**	**0.01984**	**0.01543**	**0.01235**

Table 6c. Ratio 1vs. 9 and Entropy

Ratio 1s vs. 9s: 210771 /4684 = 44.998 ≈ 45	**Entropy H =** $-\sum_1^9 p_i\, ld(p_i)$ **= 2.0692**

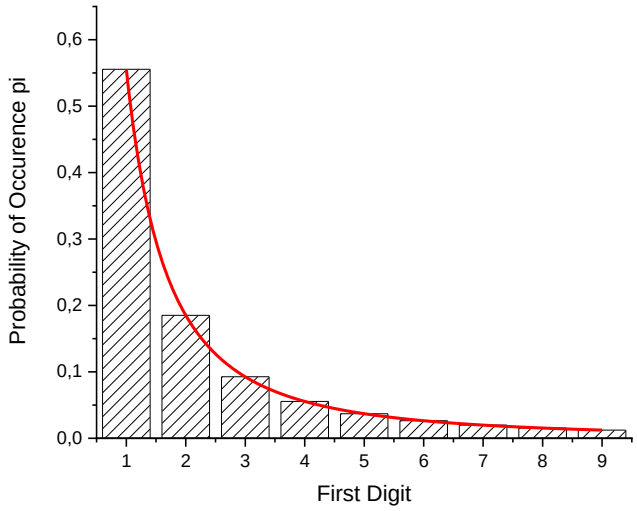


Figure 8. Probability of first digits when the sample is divided hypothetically into fragments of equal size as shown in **Table 6a**. From the fit of the probability function, W(z), for the density function exponent a value of P= 2.00003(4) is obtained.

Table 7 Probabilities of occurrence of first digits in relation to z= 1 ($\mathbf{p_{rel}}$) and $\mathbf{p_i}$

z interval	**Number of Integers**		$\mathbf{p_{rel}}$	$\mathbf{p_i}$	$\mathbf{-p_i\ ld\ p_i}$
1	$M\left[1-\frac{1}{2}\right]$	0.5000	1.0000	0.5556	0.4711
2	$M\left[\frac{1}{2}-\frac{1}{3}\right]$	0.1666	0.3333	0.1852	0.4506
3	$M\left[\frac{1}{3}-\frac{1}{4}\right]$	0.0833	0.1667	0.0926	0.3179
4	$M\left[\frac{1}{4}-\frac{1}{5}\right]$	0.0500	0.1000	0.0556	0.2318
5	$M\left[\frac{1}{5}-\frac{1}{6}\right]$	0.0333	0.0667	0.0371	0.1763
6	$M\left[\frac{1}{6}-\frac{1}{7}\right]$	0.0238	0.0476	0.0264	0.1384
7	$M\left[\frac{1}{7}-\frac{1}{8}\right]$	0.0179	0.0357	0.0198	0.1120
8	$M\left[\frac{1}{8}-\frac{1}{9}\right]$	0.0139	0.0278	0.0154	0.0927
9	$M\left[\frac{1}{9}-\frac{1}{10}\right]$	0.0111	0,0222	0.0123	0.0780
Σ			1.8000	1.0000	**H= 2.0688** bits per dig

8. Results

Although most of the first digits´ distributions in listings encountered in daily life obey the Newcomb-Benford Law, there are rare examples where the first digits´ distribution deviates strongly from this (weights of mineral fragments, fragments of glass). Examples are given in **Table 8**, together with some values for the entropy of the first digits.

Table 8. Values for first digits´ ratio and entropy for different distributions

Data referring to	**Ratio of first digits (1 vs. 9)**	**Exponent P of the density power function**	**H/entropy/ bits per digit**
Even distribution of first digits	**1**	**0**	**3.170**
Newcomb-Benford distribution	**6.58**	**1**	**2.876**
Masses of meteorites (from the density function fitted)	**7.03**	**1.035(35)**	
Weights of pieces of marble (from the density function fitted)	**45.27**	**2.003(25)**	**2.037**
Weisser Jura (from the density function fitted)	**19.63**	**1.71(15)**	
Weights of pieces of granite (from the density function fitted)	**27.62**	**1.748(81)**	**2.182**
Weights of pieces of glass (from the density function fitted)	**13.89**	**1.391(27)**	**2.589**

Summary

There are distributions where the first digits´ probability of occurrence deviates significantly from the Newcomb-Benford Law. While, in the case of the NB Law, the 9 occurs 6.58 as frequently in the first place as a nine, this ratio is significantly above 10, in case of fragments of glass as well as for some minerals. In the case of marble, a ratio up to 45 has been found. Whether this is due to material constants or whether it can be attributed to the experimental conditions has to be left to further investigations Correspondingly, the numerical value of the first digits´ entropy H covers a fairly wide range, too.

Appendix. First digits´ distribution of exponentially growing objects

Suppose, several capitals start growing at different times t at a certain size (= 1) according to the function

$$\mathbf{S = exp\ [0.076753\ t].} \qquad \text{[EQ 8]}$$

The number 0.076753 has been chosen such that the first capital reaches the amount S= 10 after 30 years. This corresponds to an annual interest rate of 8%. **Figure 9** shows the size of the capitals as a function of time. The first one begins with the interest at t = 0, the others following in each case at intervals of two years. After 30 years, one will find more amounts whose size starts with a 1 than with a 9.

This can be shown in the following way: From the functions

n = 0.5t +1 (n= number of objects; t= time/years) and

S = exp [0.076753 t],

one obtains dn = 0.5 dt and dS = 0.076753 exp [0.076753 t] dt , and from

$$\frac{dn}{dS} = \frac{dt}{2{\cdot}0.076753\ dt\ S}\ ; \quad dn = \frac{dS}{2{\cdot}0.076753\ S} = 6.5144\,\frac{dS}{S},$$

one gets the number n of objects within a certain size range

$$n = 6.5154 \int_{S1}^{S2} \frac{dS}{S} = 6.5144\ [\ln S2 - \ln S1] \qquad \text{[EQ 9]}$$

After 30 years, there will be **n= 4.5154** assets exhibiting a size between 1 and 2 units, but only **n= 0.6864** between 9 and 10 units, their ratio being **6.58**, corresponding to the NB Law (**Figure A1**).

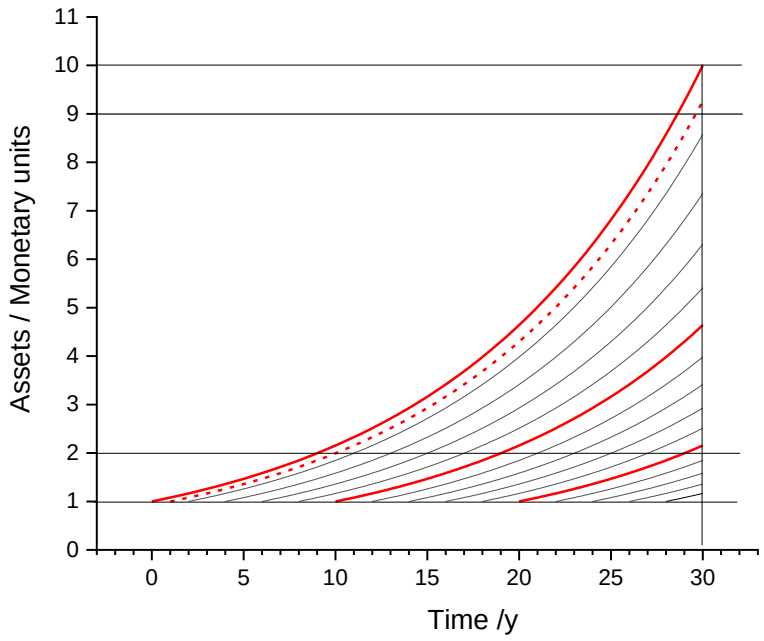


Figure A1 Capital growing with an annual interest rate of 8%.

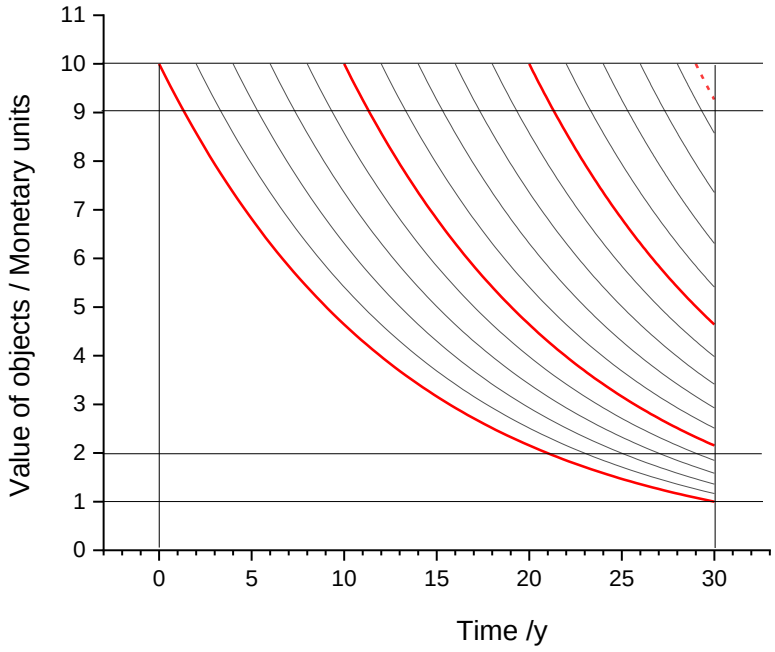


Figure A 2 Value of objects decreasing due to annual depreciation of 8%.

References

1. Newcomb, S. Note on the Frequency of Use of the Different Digits in Natural Numbers. Am. J. Math. **1881**, 4, 39–40.

2. Benford, F. The Law of Anomalous Numbers. Proceed. Am. Phil. Soc. **1938**, 78, 551–572.

3. Available online: https://en.wikipedia.org/wiki/Benford%27s_law (accessed on 28 October 2022).

4. Hill, T.P. A Statistical Derivation of the Significant-Digit Law. Statist. Sci. **1995**, 10, 354–363.

5. Biau, D. The First-Digit Frequencies in Data of Turbulent Flows. Phys. A Stat. Mech. Its Appl. **2015**, 440, 147–154.

6. Kreiner, W.A. On the Newcomb-Benford Law. *Z. Nat.* **2003**, *58a*, 618–622.

7. Kreiner, W. A. First Digits´ Shannon Entropy. *Entropy* **2022**, *24*, 1413; Available online: https://doi.org/10.3390/e24101413

8. Stylometry and Numerals Usage: Benford's Law and Beyond. Available online: https://encyclopedia.pub/entry/17716 (accessed on 28 October 2022).

9. Available online: https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh (accessed on 28 October 2022).

10. Shannon, C.E. A Mathematical Theory of Communication. Bell Syst. Tech. J. **1948**, 27, 379–423.

11. Shannon, C.E. Prediction and Entropy of Printed English. Bell Syst. Tech. J. **1951**, 30, 50– 64.

12. Available online: https://www.tf.uni kiel.de/matwis/amat/mw1_ge/kap_5/advanced/t5_3_2.html (accessed on 28 October 2022).

13. Posch, P.N.; Kreiner, W.A. Analysing digits for portfolio formation and index tracking. J. Asset Manag. **2006**, 7, 69–80.