Proof that a Data Set that Conforms to Benford's law is not Always Sum Invariant with Respect to the First Digits

R C Hall, MSEE, BSEE

e-mail: rhall20448@aol.com

abstract

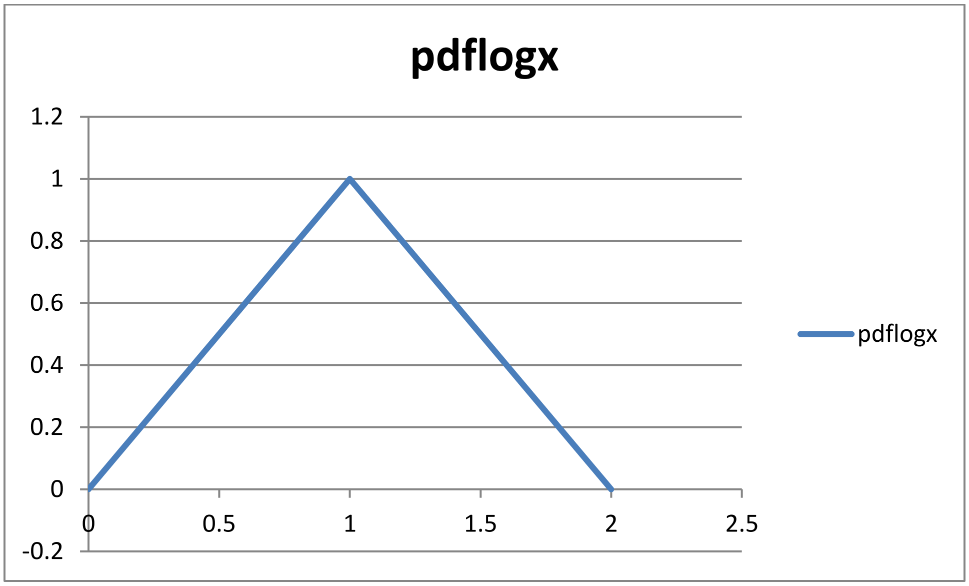
The summation test consists of adding all numbers that begin with a particular first digit or first two digits and determining its distribution with respect to these first or first two digits numbers. Most people familiar with this test believe that the distribution is a uniform distribution for any distribution that conforms to Benford's law i.e. the distribution of the mantissas of the logarithm of the data set is uniform U[0,1). The summation test that results in a uniform distribution is true for an exponential function (geometric progression) i.e. $y = a^{kt}$ but not necessarily true for other data sets that conform exactly to Benford's law.

Introduction

When the summation test is applied to real data such as population of cities, time intervals between earthquakes, and financial data, which all closely conforms to Benford's law, the summation test results in a Benford like distribution and not a uniform distribution. Citing *Benford's Law,* page 273, author Dr. Mark Nigrini, " The analysis included the summation test. For this test the sums are expected to be equal, but we have seen results where the summation test shows a Benford- like pattern for the sums." Citing *Benford's Law,* page 141, author Alex Kossovski, " Worse than the misapplication and confusion regarding the chi-sqr test, Summation Test stands out as one of the ***most misguided application in the whole field of Benford's Law***, attaining recently the infamous status of a fictitious dogma and leading many accounting departments and tax authorities astray." He also states on page 145, "Indeed **all** summation tests on actual statistical and random data relating to accounting data and financial data, census data, single-issue physical data, and so forth, show a strong and consistent bias towards higher sums for low digits, typically by a factor of 5 to 12 approximately in the competition between digit 1 and digit 9, there is not a single exception!"

The histograms of the logarithm of the aforementioned data tend to resemble a Normal distribution, which is the definition of a Log Normal distribution (the Central Limit theorem applied to random multiplications). It has been shown, both mathematically and empirically, that the summation test performed on data that conforms to a Log Normal distribution results in a Benford like distribution.

If the probability density function of the logarithm of a data set begins and ends on the x –axis and the distance between each integral power of ten (IPOT) consists of a straight line then the data set will conform exactly to Benford's law (see appendix A). Figure 1 illustrates such a data set.



Figure#1

The equation corresponding to figure#1 is expressed in line 1.

1) $\text{pdf}_{\log x} = \log(x)_{0\ to\ 1} + 2_{1\text{ to }2} - \log(x)_{1\text{ to }2}$

The probability density function (pdf) of x is expressed in line 12.

2) $\text{pdf}_{y\ 0\ to\ 1} = \log(x)_{0\text{ to }1} = \frac{\ln(x)}{\ln(10)}$

3) $y = \log(x) = \frac{\ln(x)}{\ln(10)}$

4) $\text{Pdf}_y\, dy = \text{pdf}_x\, dx$

5) $\text{Pdf}_x = \text{pdf}_y \frac{dy}{dx}$

6) $\frac{dy}{dx} = \frac{1}{x\ln(10)}$

7) $\text{pdf}_{x\ 1\ to\ 10} = \text{pdf}_{y}\frac{dy}{dx} = \frac{\ln(x)}{x\ln^2(10)}\ _{1\ to\ 10}$

8) $\text{pdf}_{y\ 1\ to\ 2} = 2\ _{1\ to\ 2}$

9) $\text{pdf}_{x} = \text{Pdf}_{y}\frac{dy}{dx} = \frac{2}{x\ln(10)}\ _{10\ to\ 100}$

10) $\text{pdf}_{y\ 1\ to\ 2} = -\log(x)_{1\ to\ 2} = \frac{-\ln(x)}{\ln(10)}\ _{1\ to\ 2}$

11) $\text{pdf}_{x} = \text{Pdf}_{y}\frac{dy}{dx} = -\frac{\ln(x)}{x\ln^2(10)}\ _{10\ to\ 100}$

12) $\text{pdf}_{x} = \frac{\ln(x)}{x\ln^2(10)}\ _{1\ to\ 10} + \frac{2}{x\ln(10)}\ _{10\ to\ 100} - \frac{\ln(x)}{x\ln^2(10)}\ _{10\ to\ 100}$

The probability distribution function is expressed in line 13 and beyond.

13) $\int_1^{100} \text{Pdf}\,x\ dx = \frac{1}{\ln^2(10)}\int_1^{10}\frac{\ln(x)}{x}dx + \frac{2}{\ln(10)}\int_{10}^{100}\frac{dx}{x}dx - \frac{1}{\ln^2(10)}\int_{10}^{100}\frac{\ln(x)}{x}dx$

14) $\frac{1}{\ln^2(10)}\int_1^{10}\frac{\ln(x)}{x}dx = \frac{\ln^2(10)}{2\ln^2(10)} = \frac{1}{2}$

15) $-\frac{1}{\ln^2(10)}\int_{10}^{100}\frac{\ln(x)}{x}dx = \frac{-1}{2\ln^2(10)}\ [\ln^2(100) - \ln^2(10)] =$

16) $\frac{-1}{2\ln^2(10)}[\ln(100) + \ln(10)][\ln(100) - \ln(10)] =$

17) $\frac{-1}{2\ln^2(10)}[3\ln^2(10)] = -\frac{3}{2}$

18) $\frac{2}{\ln(10)}\int_{10}^{100}\frac{dx}{x}dx =$

19) $\frac{2}{\ln(10)}\ln(10) = 2$

20) $\text{sum} = \frac{1}{2} - \frac{3}{2} + 2 = 1$

The probability distribution function for 1- 2 and 10 - 20 is expressed

In lines 21 through 29.

21) $\int_1^2 \mathrm{pdf}_{\,x}\,dx + \int_{10}^{20} \mathrm{pdf}_{\,x}\,dx = \frac{1}{\ln^2(10)} \int_1^2 \frac{\ln(x)}{x}\,dx + \frac{2}{\ln(10)} \int_{10}^{20} \frac{dx}{x}\,dx - \frac{1}{\ln^2(10)} \int_{10}^{20} \frac{\ln(x)}{x}\,dx =$

22) $\frac{\ln^2(2)}{2\ln^2(10)} + \frac{2}{\ln(10)}\ln(2) - \frac{1}{2\ln^2(10)} [\ln^2(20) - \ln^2(10)] =$

23) $\frac{1}{2\ln^2(10)} [\ln^2(2) – [\ln^2(20 - \ln^2(10)]] + \frac{2}{\ln(10)}\ln(2)=$

24) $\frac{1}{2\ln^2(10)} [\ln^2(2) – [(\ln(20) + \ln(10))(\ln(20) – \ln(10))] + \frac{2}{\ln(10)}\ln(2)=$

25) $\frac{1}{2\ln^2(10)} [\ln^2(2) – [(\ln(20) + \ln(10))\ln(2)] + \frac{2}{\ln(10)}\ln(2)=$

26) $\frac{1}{2\ln^2(10)} [\ln^2(2) – [(\ln(2) + 2\ln(10))\ln(2)] + \frac{2}{\ln(10)}\ln(2)=$

27) $\frac{1}{2\ln^2(10)} [\ln^2(2) – \ln^2(2) – 2\ln(10)\ln(2)] + \frac{2}{\ln(10)}\ln(2)=$

28) $\frac{1}{2\ln^2(10)} [– 2\ln(10)\ln(2)] + \frac{2}{\ln(10)}\ln(2)= \frac{\ln(2)}{\ln(10)}$

29) $\frac{\ln(2)}{\ln(10)} = \log_{10} 2$, which conforms to Benford's law

More generally (digits 1 … 9):

30) $\int_{d}^{d+1} \text{pdf } x \ dx + \int_{10d}^{10(d+1)} \text{pdf } x \ dx$ =

31) $\int_{d}^{d+1} \frac{\log(x)}{x\ln(10)} dx + \int_{10d}^{10(d+1))} \frac{2}{x\ln(10)} dx - \int_{10d}^{10(d+1)} \frac{\log(x)}{x\ln(10)} dx$

32) $\int_{d}^{d+1} \frac{\log(x)}{x\ln(10)} dx = \frac{1}{2\ln^2(10)} [(\ln^2(d+1) - \ln^2(d)] =$

33) $\frac{1}{2\ln^2(10)} [(\ln(d+1) + \ln(d))(\ln(d+1) - \ln(d))] =$

34) $\frac{1}{2\ln^2(10)} (\ln(d+1) + \ln(d))\ln(\frac{d+1}{d})$

35) $\int_{10d}^{10(d+1)} \frac{2}{x\ln(10)} dx = 2(\ln(10(d+1)) - \ln(10d))/\ln(10) = 2 \frac{\ln(\frac{d+1}{d})}{\ln(10)}$

36) $- \int_{10d}^{10(d+1)} \frac{\log(x)}{x\ln(10)} dx = \frac{-1}{2\ln^2(10)} [\ln^2(10(d+1)) - \ln^2(10d)] =$

37) $\frac{-1}{2\ln^2(10)} [(\ln(10(d+1)) + \ln(10d))(\ln(10(d+1)) - \ln(10d))] =$

38) $\frac{-1}{2\ln^2(10)} (\ln(10(d+1)) + \ln(10d))\ln(\frac{d+1}{d}) =$

39) $\frac{-1}{2\ln^2(10)} \ln(10 + \ln(d+1) + \ln(10) + \ln(d))\ln(\frac{d+1}{d})$

Combining 34, 35, 39 and simplifying:

40) $\frac{2}{\ln(10)}$ ln($\frac{d+1}{d}$) + ($\frac{1}{2\ln^2(10)}$) ln($\frac{d+1}{d}$) [ln(d + 1) + ln(d) –ln(10) - ln(d+1) – ln(10) - ln(d)] =

41) $\frac{2}{\ln(10)}$ ln($\frac{d+1}{d}$) + $\frac{1}{2\ln^2(10)}$ ln($\frac{d+1}{d}$)(-2ln(10)) = $\frac{\ln(\frac{d+1}{d})}{\ln(10)}$ = $\log_{10}(\frac{d+1}{d})$, which conforms to Benford's Law

The following arguments illustrate the probability density and distribution function of the expected value (which is the sum divided by the number of samples. i.e. pdf $_{\text{expected value of x}}$ = x pdf(x)

The expected value = $\frac{\int_{-\infty}^{\infty} x\mathrm{pdf}(x)dx}{\int_{-\infty}^{\infty} \mathrm{pdf}(x)dx}$, since $\int_{-\infty}^{\infty} \mathrm{pdf}(x)dx$ = 1

The expected value = $\int_{-\infty}^{\infty} x\mathrm{pdf}(x)dx$

The probability density function (pdf) of the expected value =

xpdf(x)

42) Pdf $_{\text{expected value}}$ = x$\frac{\log(x)}{x\ln(10)}$ $_{\text{1 to 10}}$ + x $\frac{2}{x\ln(10)}$ $_{\text{10 to 100}}$ - x $\frac{\log(x)}{x\ln(10)}$ $_{\text{10 to 100}}$

43) Pdf $_{\text{expected value}}$ =(log(x) $_{\text{1 to 10}}$ + 2 $_{\text{10 to 100}}$ – log(x) $_{\text{10 to 100}}$)/ln(10)

44) $\int_{1}^{100} \mathrm{pdf}(ev)dx = \frac{1}{\ln(10)}\int_{1}^{10} \log(x)\,dx + \frac{2}{\ln(10)}\int_{10}^{100} dx$ - $\frac{1}{\ln(10)}\int_{10}^{100} \log(x)\,dx$ =

45) $\frac{1}{\ln^2(10)}\int_{1}^{10} \ln(x)\,dx + \frac{2}{\ln(10)}\int_{10}^{100} dx - \frac{1}{\ln^2(10)} \int_{10}^{100} \ln(x)\,dx$

46) $\frac{1}{\ln^2(10)} \int_1^{10} \ln(x)\,dx = \frac{1}{\ln^2(10)}$ [10 ln(10) – 10 – (ln(1) – 1)] = 2.6454

47) $\frac{2}{\ln(10)} \int_{10}^{100} dx$ = 78.173

48) -$\frac{1}{\ln^2(10)} \int_{10}^{100} \ln(x)\,dx$ = -$\frac{1}{\ln^2(10)} 100$ [ln(100) – 100 – (10ln(10) – 10)] = -65.5409

49) combining 46, 47, 48 = 15.2775

50) $\frac{1}{\ln^2(10)} [\int_d^{d+1} \ln(x)\,dx - \int_{10d}^{10(d+1)} \ln(x)\,dx\,] + \frac{2}{\ln(10)} \int_{10d}^{10(d+1)} dx$ =

51) $\frac{1}{\ln^2(10)} \int_d^{d+1} \ln(x)\,dx = \frac{1}{\ln^2(10)}$[(d+1)ln(d+1) – (d+1) – (dln(d) – d)] =

52) $\frac{1}{\ln^2(10)} [d\ln(d+1) +$ ln(d+1) – (d+1) - (dln(d) – d)] =

53) $\frac{1}{\ln^2(10)}$[dln($\frac{d+1}{d}$) + ln(d+1) -1]

54) - $\frac{1}{\ln^2(10)} \int_{10d}^{10(d+1)} \ln(x)\,dx\,] = \frac{-1}{\ln^2(10)}$[10(d+1)ln(10(d+1)) – 10(d+1) – (10dln(10d) – 10d)] =

55) $\frac{-1}{\ln^2(10)}$ [10dln(10(d+1)) + 10ln(10(d+1)) – 10d – 10 - 10dln(10d) + 10d] =

56) $\frac{-1}{\ln^2(10)} [10d\ln(\frac{d+1}{d})$ + 10ln(10(d+1)) – 10d – 10 + 10d] =

57) $\frac{-1}{\ln^2(10)}$ [10dln($\frac{d+1}{d}$) + 10ln(10(d+1)) -10]

58) $\frac{2}{\ln(10)} \int_{10d}^{10(d+1)} dx$ = 8.6859

59) combining 53, 57, 58 ; the result is 8.6859 + $\frac{1}{\ln^2(10)}$ [(-9d) ln($\frac{d+1}{d}$) - 9 ln(d+1) - 10ln(10) + 9]

60) The first digit distribution is:

61) (8.6859 + $\frac{1}{\ln^2(10)}$ [(-9d) ln($\frac{d+1}{d}$) -9 ln(d+1) – 10 ln(10) + 9])/15.2775

evaluated for d = 1, 2 ,3 ,4, 5, 6, 7, 8, 9

The following figure illustrates the theoretical (as opposed to empirical) result of first digit summation test

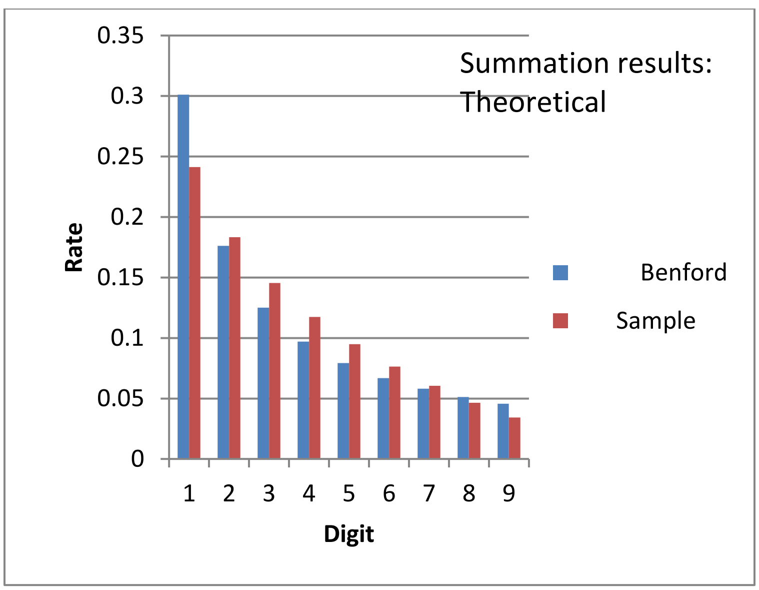


Figure #2

The results, while not completely conforming to Benford's law, are certainly neither uniform nor sum invariant.

In order to obtain empirical results one can generate random numbers that conform to the aforementioned probability density function and then perform the summation test i.e. add all the numbers generated that begin with a particular digit and compare the relative values.

Generate numbers conforming to the aforementioned pdf:

Pr refers to the probability distribution function as opposed to the probability density function

59) Pr 1 to 10 = $= \frac{1}{\ln^2(10)} \int_1^x \frac{\ln(x)}{x}$ dx = $\frac{\ln^2(x)}{2\ln^2(10)}$

60) Pr $_{10}$ = $\frac{\ln^2(10)}{2\ln^2(10)} = \frac{1}{2}$

61) Pr ≤ $\frac{1}{2}$; $\frac{\ln^2(x)}{2\ln^2(10)}$

62) ln(x) = $(2\text{Pr} * \ln(10))^{1/2}$

63) x = $e^{(2\text{Pr}*\ln^2(10))^{1/2}}$

If Pr > $\frac{1}{2}$

64) Pr – 0.5 = $\frac{1}{\ln(10)} \int_{10}^{x} \frac{2}{x}$ dx - $\frac{1}{\ln^2(10)} \int_{10}^{x} \frac{\ln(x)}{x}$ dx =

65) ln(10) (Pr – 0.5) = 2ln(x) – 2ln(10) - $\frac{1}{2\ln(10)}$[$\ln^2(x)$ - $\ln^2(10)$]

66) $2\ln^2(10)$(Pr- 0.5) = 4ln(10)ln(x) – $4\ln^2(10)$ + $\ln^2(10)$ - $\ln^2(x)$

67) $2\ln^2(10)$(Pr- 0.5) – 4ln(10)ln(x) + $3\ln^2(10)$ + $\ln^2(x)$ = 0

68) solve for ln(x)

69) ln(x) = (4ln(10) - $\sqrt{(4\ln(10))^2 - 4((\text{Pr} - 0.5)\, 2\ln^2(10) + 3\ln^2(10))}$)/2

70) x = $e^{((4\ln(10) - \sqrt{4\ln(10))^2 - 4((\text{Pr}-0.5)2\ln^2(10) + 3\ln^2(10)}}$)/2)

Utilizing an Excel spread sheet one can generate these numbers and compute the various Benford's tests.

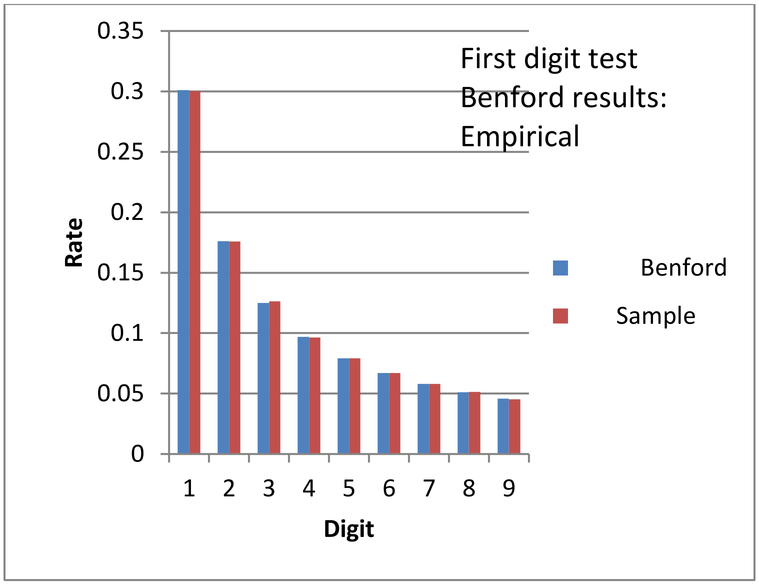


Figure #3

The first digit test evaluating conformance to Benford's law derived from the aforementioned randomly generated numbers indicates very close conformance.

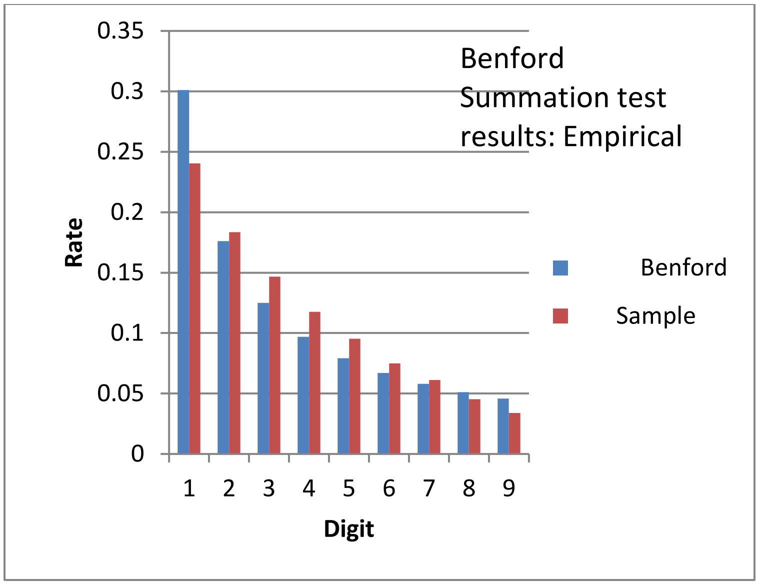


Figure #4

The results are almost identical to figure #2

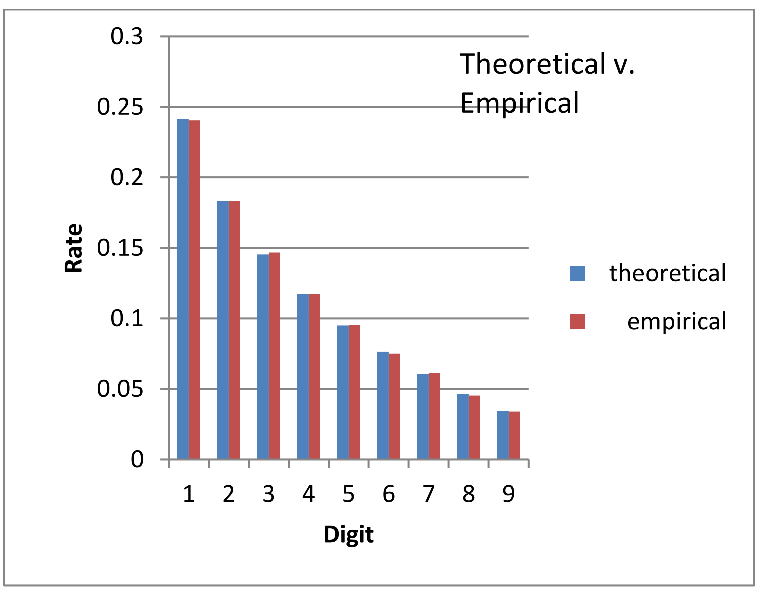


Figure #5

The results of the empirically derived summation test are very close to the theoretically derived results as figure #5 illustrates.

Conclusion

It has been clearly proven mathematically and demonstrated empirically that a data set that conforms exactly to Benford's law is not necessarily sum invariant i.e. the sum of all numbers that begin with any particular first digit (1,2,3,4,5,6,7,8,9) is the same for all digits. The mathematical and the empirical results are virtually identical.

Appendix A

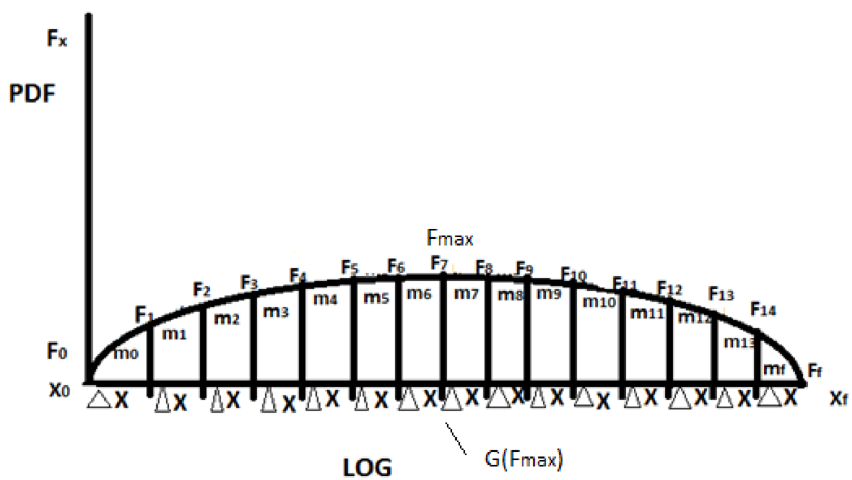
Proof that if the probability density function of the logarithm of a data set is continuous and begins and ends on the x-axis and the number of integral power of ten (IPOT) values approaches infinity then the probability density function of the resulting mantissas will be uniform and; therefore, the data set will conform to Benford's law

1) The probability density function of a data set that conforms to Benford's Law is k/x = $\frac{1}{\ln(10)x}$
2) The probability density function of the log of the same function is a uniform distribution,
   a. pdf(y)dy = pdf(x)dx
   b. Y = log(x) = $\frac{\ln(x)}{\ln(10)}$
   c. pdf(y) = pdf(x) $\frac{dx}{dy}$
   d. $\frac{dy}{dx} = \frac{1}{x\ln(10)}$

e. $\frac{dy}{dx}$ = xln(10)

f. pdf(y) = $\frac{x\ln(10)}{x\ln(10)}$ = 1 – Uniform Distribution

3) Therefore, if it can be shown that the pdf of the log of a function is uniform then the data set follows Benford’s Law.



4) Y = F(x)

5) Y’ = $\frac{d(F(x))}{dx}$

6) $\int_{Xo}^{Xf} Y'dx = \int_{Xo}^{Xo} F'(x)dx$ = F(Xf) – F(Xo) = 0

7) Avg Value of Y’ = $\frac{1}{Xf-Xo}\int_{Xo}^{Xf} Y'dx = \frac{0}{Xf-Xo}$

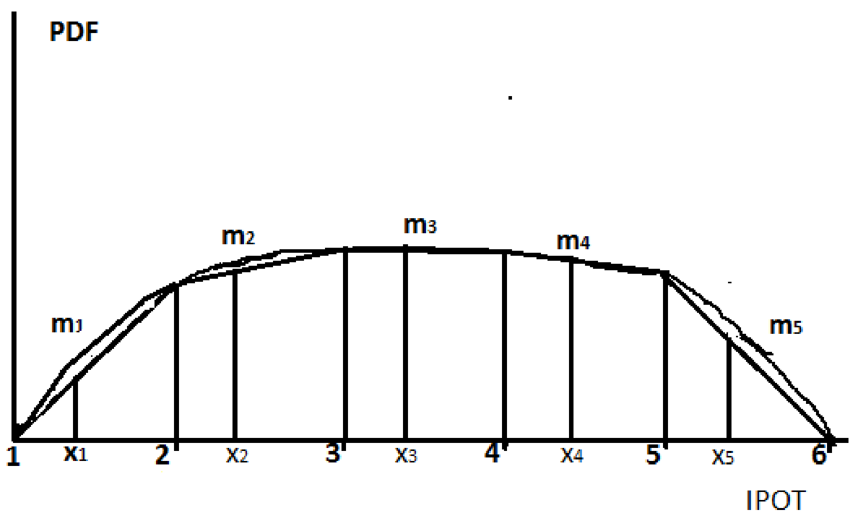
8) $F'_i(x) = \frac{F(i+1)-F(i)}{\Delta x}$; Δx → 0

9) $\int_{Xo}^{Xf} F'(x)dx = 0$; $\sum_{i=0}^{N-1} \frac{F(i+1)-F(i)}{\Delta x} = 0$ as $\Delta X \rightarrow 0$

10) let $m(i) = = \frac{F(i+1)-F(i)}{\Delta x}$

11) $\sum_{i=0}^{N-1} m(i)\ \Delta X = 0$; $\Delta X \rightarrow 0$

Let's consider a simpler case.



12) Let $\Delta X = 1$

13) $m_1 + m_2 + m_3 + m_4 + m_5 = 0$

14) $\sum_{i=1}^{5} x_i = m_1x + m_1 + m_2x + m_1 + m_2 + m_3x + m_1 + m_2 + m_3 + m_4x +$

$m_1 + m_2 + m_3 + m_4 + m_5x = K$

15) $x(m_1 + m_2 + m_3 + m_4 + m_5) + m_1 + m_1 + m_1 + m_1 + m_2 + m_2 + m_2 + m_3 + m_3$

$+ m_4 = K$

16) $m_1 + m_2 + m_3 + m_4 + m_5 = 0$

17) $\sum_{i=1}^{5} x_i = 4m_1 + 3m_2 + 2m_3 + m_4 = K$ (constant)

18) AREA UNDER PDF = 1

18) $\int_1^6 f(x)\,dx = 1$

20) $\frac{m_1}{2} + m_1 + \frac{m_2}{2} + (m_1 + m_2) + \frac{m_3}{2} + (m_1 + m_2 + m_3) + \frac{m_4}{2} + (m_1 + m_2 + m_3 + m_4) + \frac{m_5}{2}$

$= 1$

21) $m_1 + m_2 + m_3 + m_4 + m_5 = 0$

22) $4m_1 + 3m_2 + 2m_3 + m_4 = 1$

Therefore K = 1

The sum of all functions at IPOT + x = 1 for any x.

The sum of all probability density functions of each mantissa value contained within all integral powers of ten respectively is equal to 1, which constitutes a uniform distribution

Which is the definition of a Benford distribution.

23) For the more general case:

24) $\sum_{i=1}^{r-1} m_i$ =

25) $m_1x + m_2 + m_2x + m_1 + m_2 + m_3x + \ldots\ldots m_1 + m_2 + m_3 + \ldots m_{r-1}x =$

K

26) $x(m_1+ m_2 + \ldots. + m_{r-1}) + (r\text{-}2)m_1 + (r\text{-}3)m_3 +..+ m_{r-2} = K$

27) $x(m_1 + m_2 + m_3 + m_{r-1}) = 0$

28) $(n\text{-}2)m_1 + (n\text{-}1)m_2 + \ldots.+ m_{r-2} = K$

29) $\frac{m_1}{2}+ m_1 + \frac{m_2}{2} + m_1 + m_2 + \frac{m_3}{2} + m_1 + m_2 + m_3 + .. + m_{r-2} + \frac{m_{r-1}}{2}$

$= K$

30) $\frac{1}{2}(m_1+ m_2+ m_3 + m_{r-1}) =0$

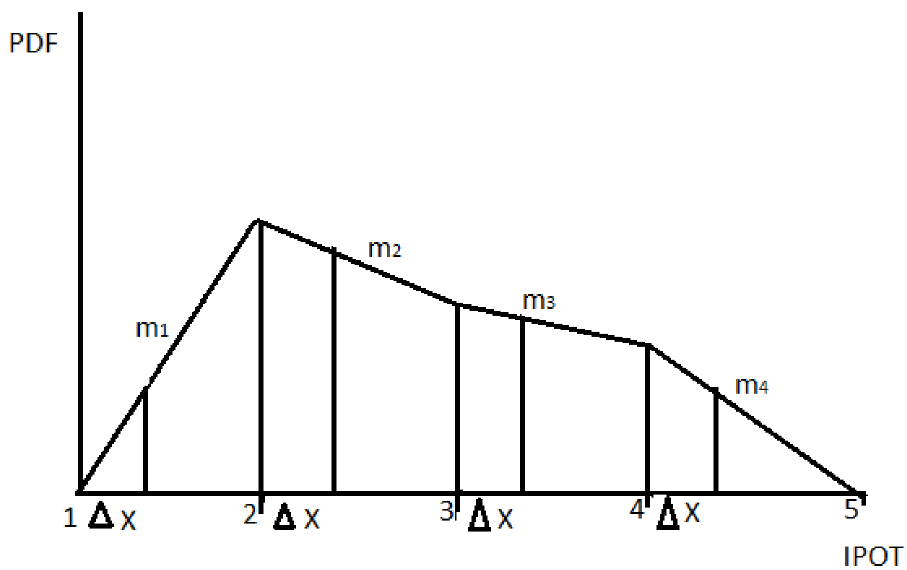
31) $(n\text{-}2)m_1 + (n\text{-}1)m_2 + \ldots.. + m_{r-2} = 1$

32) $K=1$

33) The sum of mantissa values at IPOT + x = 1 for any x

34) The resultant probability density function of the mantissas is a uniform distribution whose amplitude is equal to 1 and, therefore a Benford distribution

Proof that if the probability density function of the Logarithm a data set is continuous and begins and ends on the x-axis and the number of integral power of ten values approaches infinity then the sum of probability distributions of all fixed intervals from all IPOT (ΔX) equals the interval Itself (ΔX).



-

1) $\sum_1^4 \int_i^{i+\Delta} pdf\ dx = \frac{1}{2}\mathrm{m_1}(\Delta\mathrm{x})^2 + \mathrm{m_1}\Delta\mathrm{x}\ + \frac{1}{2}\mathrm{m_2}(\Delta\mathrm{x})^2 + (\mathrm{m_1} + \mathrm{m_2})\Delta\mathrm{x} + \frac{1}{2}\mathrm{m_3}(\Delta\mathrm{x})^2 + (\ \mathrm{m_1} + \mathrm{m_2} + \mathrm{m_3})\Delta\mathrm{x}\ + \frac{1}{2}\mathrm{m_4}(\Delta\mathrm{x})^2 = \mathrm{K}$

2) $\frac{1}{2}(\Delta\mathrm{x})^{\ 2}\ (\mathrm{m_1} + \mathrm{m_2} + \mathrm{m_3} + \mathrm{m_4}\ ) + (3m_1\ + 2\mathrm{m_2} + \mathrm{m_3})\Delta\mathrm{x} = \mathrm{K}$

3) $\mathrm{m_1} + \mathrm{m_2} + \mathrm{m_3} + \mathrm{m_4} = 0$

4) $3\mathrm{m_1} + 2\mathrm{m_2} + \mathrm{m_3} = 1$

5) $(3m_1\ + 2\mathrm{m_2} + \mathrm{m_3})\Delta\mathrm{x} = \Delta\mathrm{x}$

6) $\sum_1^4 \int_i^{i+\Delta x} \mathrm{pdf\ dx} = \Delta\mathrm{x}$

In General:

7) $\sum_{i=1}^{r-1} \int_i^{i+\Delta x} \mathrm{pdf\ dx} = \frac{1}{2}(\Delta\mathrm{x})^2(\ \mathrm{m_1} + \mathrm{m_2} + \mathrm{m_3} + \ldots + \mathrm{m_{r-1}}\ ) +$

8) $[(\mathrm{n} - 2)\mathrm{m_1} + (n - 1)\mathrm{m_2} + \ldots + \mathrm{m_{r-2}}]\Delta\mathrm{x}\ = \Delta\mathrm{x}$

It can be easily shown that the fixed intervals don't have to start and end on an interval power of ten such as 10,100,1000 or 1,2,3 on a LOG plot as long as the fixed intervals are all offset by a power of ten.

For instance, the left most interval starting point, where the curve intersects the x-axis, could be 2 with each succeeding interval 10 times the previous interval i.e. 20,200,2000 etc. The data would still conform to Benford's Law with digit 1 contained in intervals 10-20, 100-200, 1000-2000; digit 2: 2-3,20-30,200-300;digit 3: 3-4,30-40,300-400;digit 4: 4-5,40-50,400-500;digit 5:5-6,50-60,500-600;digit 6:6-7,60-70,600-700;digit 7:7-8,70-80,700-800;digit 8:8-9,80-90,800-900;digit 9:9-10,90-100,900-1000. The first digit starts in the tens and ends in the 1000s; all of the others start in the single digits and end in the 100s. It's still the same result obtained by having the IPOT at each interval such as 1,10,100,1,000 etc.

This would explain why data sets that span many orders of magnitude conform very closely to Benford's law and data sets that span fewer orders of magnitude do not. This also explains why several other distributions such as gamma, beta, Weibull and exponential probability density functions conform fairly closely to Benford's law and why Gaussian or Normal distributions do not (the pdf of the logarithm of a Gaussian data span a very limited number of IPOTs. i.e.

X* $\frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-(\mathrm{x}-\mathrm{u})^2/2\sigma^2}$, the $e^{-(\mathrm{x}-\mathrm{u})^2/2\sigma^2}$ term falls too rapidly.