Facoltà di Impresa e Management

Cattedra di Matematica

BENFORD'S LAW: MATHEMATICAL PROPERTIES AND FORENSIC ACCOUNTING APPLICATIONS

RELATORE:
Prof. Marco Dall'Aglio

CANDIDATO:
Federico Iudica
Matr. 159111

ANNO ACCADEMICO 2011/2012

ACKNOWLEDGEMENTS

I have been indebted in the preparation of this thesis to my supervisor, Professor Marco Dall'Aglio of the Department of Economics, whose patience and kindness, as well as his academic experience, have been extremely helpful at the time I needed it the most.

I am extremely grateful to the entire faculty here at LUISS Guido Carli, for the extraordinary and always professional support given to me both inside and outside classrooms. I constantly felt inspired by the intelligence and humanity surrounding me. It has been a place where I wanted to do well, not because of fleeting markers of success such as course grades, but because I saw so many were doing their best everyday.

The support and encouragement of many friends has been indispensable, and I would like particularly to acknowledge the contribution of Angelo, Edo, Leo, Ludo, Marco, Maria and Marta that help me so much to shape my personality toward being the person I want to be. And to my first friends that have always been there for me: Andrea, Daniele, Dario, Luca, Michele and Salvo. I'd also like to mention how the best and worst moments of my bachelor journey have been shared with two of the best persons I know: Alessandro Iodice and Stefania Tortora. I sincerely hope that the gifts of the friendships I made will continue to stand the test of time and distance.

Finally, I want to thank my parents. They have been a constant source of support – emotional, moral and of course financial – during my undergraduate years, and this thesis would certainly not have existed without them.

It is to them and to my sister Silvia that this thesis is dedicated.

CONTENTS

INTRODUCTION

It was in 1881 that the astronomer Simon Newcomb, after noticing how the logarithmic tables in the library were dirtier in the first pages than in the last ones, published a 2-page article in the American Journal of Mathematics stating that numbers in ‘nature’ have their leading digit distributed in a specific, non-uniform way. He came out with a law describing their expected distribution. His paper went unnoticed and was soon forgotten.

While working as a physicist at the General Electric Research Laboratories in Schenectady, New York, in the 1920s, Frank Benford, unaware of Newcomb’s paper, made the same simple observation. He saw that the first pages, of his log tables, which showed the logarithms of numbers with low first digits (1 and 2), were more worn than the last pages, showing logs of numbers with high first digits (8 and 9). He then speculated about the fact that numbers starting with low first digits appear more often. He tried to prove, succeeding, that significant digits, the number to the left of the decimal point, fitted to some sort of numerical pattern. This pattern would be remembered as Benford’s Law.

After collecting and assembling a total of 20,229 numbers from 20 lists of numbers extrapolated from extremely diverse sources, such as: population sizes, street addresses of “American Men of Science”, all the numbers in an issue of Reader’s Digest, geographical measures and scientific constants. He tabulated and analyzed the datasets. The results were published in a paper that appeared in 1938 on the pages of Proceedings of the American Philosophical Society called “*The law of Anomalous Numbers*”.

Benford's paper could have easily be ignored and lost just like Newcomb's, instead the next paper in that fortunate edition of the journal was one by a group of important physicists, Bethe, Rose and Smith. Gousmit (1977) assumes that many physicists might have been intrigued by the last phrase of Benford's article, stating: "*and the numbers but play a poor part of lifeless symbols for living things.*" Then some of them turned back to read the paper from the beginning.

The results of his effort "*to collect data from as many fields as possible*" showed that 30.6% of the numbers had a first digit 1. The first digit 2 occurred 18.5% of the time. In contrast only 4.7% of the numbers had first digit 9. The naive expectation is that the numbers are distributed approximately in a uniform way, so there should be as many threes as sevens and so on. Consequently, the expected probability for any first digit is on average 1/9. Instead, what the datasets show is a heavy skeweness towards the lower digits. A bias in favor of lower digits exists even for the digit in the second position, even if they are much less skewed.

The specific data of Benford's article are shown in Table 1:

Table 1

Title	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***Samp-les***
Rivers	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
Newspapers	30.0	18.0	12.0	10.0	8.0	8.0	6.0	5.0	5.0	100
Specific Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
Atom. Wgt	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
n^-1	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
Designs	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
Reader's Digest	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
n!	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
Deathrate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error	± 0.8	± 0.4	± 0.4	± 0.3	± 0.3	± 0.2	± 0.2	± 0.2	± 0.3	

This thesis is organized as follows: in part 1 the theoretical and mathematical aspects of the law describing the behavior of the first significant digits is taken into account, in part 2 some applications of the law, commonly denominated Benford's Law, to forensic accounting are introduced, with a focus on Digital Analysis and on a model, the Distortion Factor Model, both commonly used tools for the detection of frauds by the tax-collection agencies of many advanced countries. In part 3, a brief example of how Benford's Law has been applied on macroeconomic data of the European Union countries is presented. The paper concludes with a summary.

Part 1 - EXPECTED DIGIT FREQUENCIES

The next part of Benford's research was deriving the expected frequencies of the digits in the lists. He noticed that the first significant digits adhered to the following logarithm laws tremendously well.

$$Prob(D_1{=}d_1) = \log\left(1+\frac{1}{d_1}\right) \qquad d_1 \in \{1, 2, ..., 9\} \tag{1.1}$$

$$Prob(D_2{=}d_2) = \sum_{d_1=1}^{9} \log\left(1+\frac{1}{d_1 d_2}\right) \qquad d_2 \in \{0, 1, 2, ..., 9\} \tag{1.2}$$

$$Prob(D_1 D_2{=}d_1 d_2) = \log\left(1+\frac{1}{d_1 d_2}\right) \qquad d_1 d_2 \in \{10, 11, ..., 99\} \tag{1.3}$$

$$Prob(D_2{=}d_2 \mid D_1{=}d_1\ ) = \frac{\log\left(1+\frac{1}{d_1 d_2}\right)}{\log\left(1+\frac{1}{d_1}\right)} \tag{1.4}$$

Where D_1 represents the first digit, D_2 the second digit, and D_1D_2 the first two digits of a number; *Prob()* indicates the probability of obersving the event in the parentheses. (1.1) is the formula for first digit proportions, (1.2) the formula for second digit proportions; the formula for first two digits proportion is shown in (1.3) and the probability of the second digit being d_2 given that the first digit is d_1 in (1.4). Logarithms are assumed to be base 10 wherever something different is not specified.

For example, according to this formulae the probability of having 4 as the first digit is:

$$Prob(D_1{=}4) = \log\left(1+\tfrac{1}{4}\right) = 0{,}09691$$

And the expected frequency of a number starting with a 3 and a 4 is:

$$Prob(D_1 D_2{=}34) = \log\left(1+\tfrac{1}{34}\right) = 0{,}01258$$

The expected digit frequencies are shown in the table below:

	Position in number			
Digit	1st	2nd	3rd	4th
0		0,11968	0,10178	0,10018
1	0,30103	0,11389	0,10138	0,10014
2	0,17609	0,10882	0,10097	0,10010
3	0,12494	0,10433	0,10057	0,10006
4	0,09691	0,10031	0,10018	0,10002
5	0,07918	0,09668	0,09979	0,09998
6	0,06695	0,09337	0,09940	0,09994
7	0,05799	0,09035	0,09902	0,09990
8	0,05115	0,08757	0,09864	0,09986
9	0,04576	0,08500	0,09827	0,99820

The general form of the law is:

$$Prob\left(mantissa \leq \tfrac{t}{10}\right) = \log_{10} t \qquad \{1 \leq t < 10\} \tag{1.5}$$

Where the mantissa of a positive real number *x* is the unique number *r* in [1/10,1) with

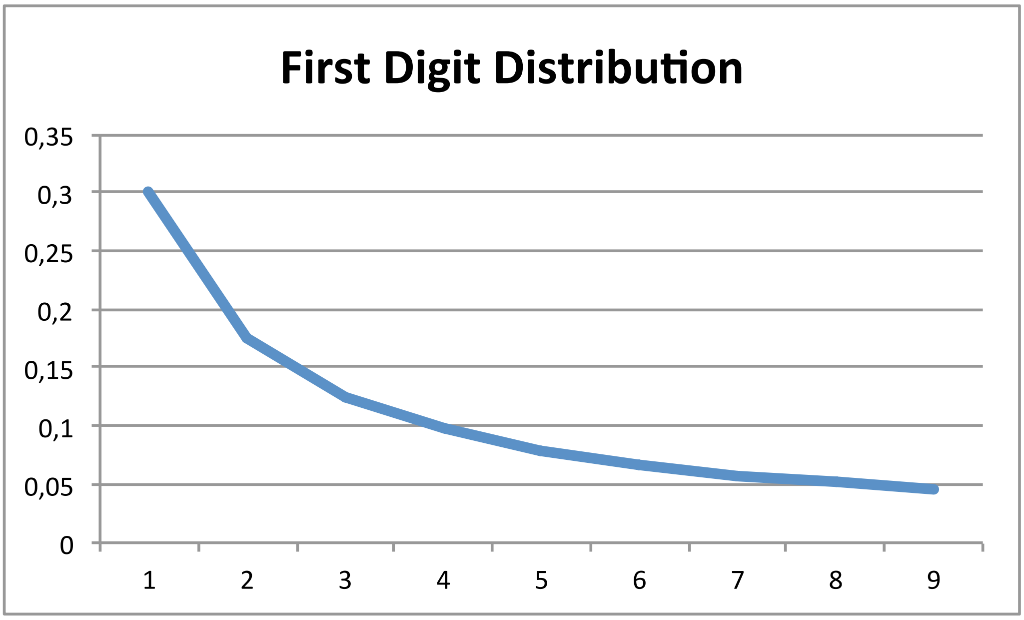
$x = r \times 10^n$ for some integer n; so the mantissae of 628 and 0,00000628 would be equal.

Letting D_1, D_2... Denote the base10 significant digit functions $D_1(0.628) = 6$, $D_2(0.628)=2$, the general significant digit law takes the following form:

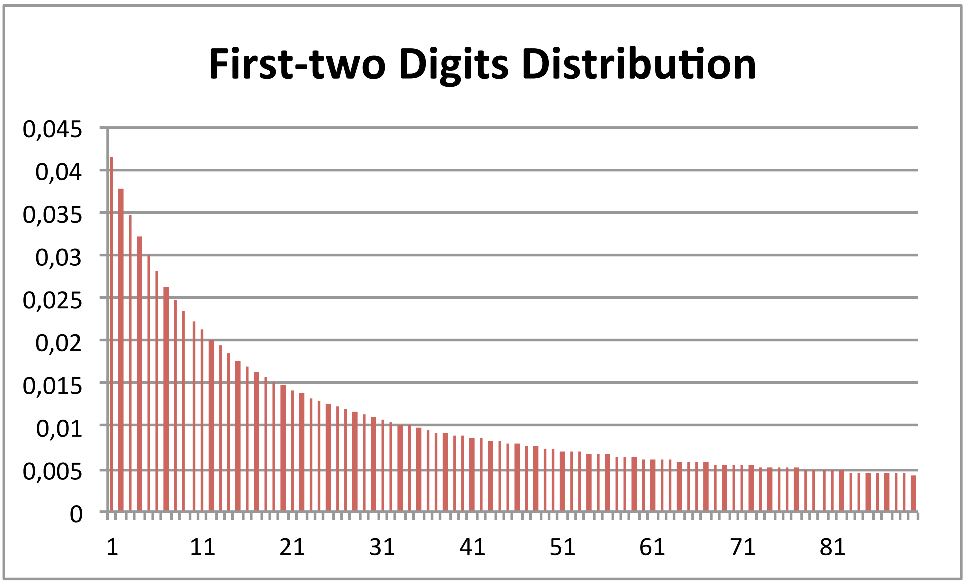
$$Prob(D_1 = d_1, \ldots, D_k = d_k) = \log\left[1 + \left(\sum_{i=1}^{k} d_i \times 10^{k-i}\right)^{-1}\right] \qquad (1.6)$$

With the unpredicted corollary that the significant digits are dependent from each other.

Graphically, the distribution of the First Digit is:



Whereas the First-two Digits distribution is:



Instead of considering the mantissae, in this paper, the analysis will use the scientific notation of numbers. Whereas mantissae are commonly used to investigate on more complex distributions and not just with the simple one made up by the first significant digit.

To get a formal definition of the first digit of a number, it is needed to resort to scienfic notation, very much in use in the fields of physics, chemistry, astronomy and wherever very big or very small numbers occur quite often. A number is in scientific notation when it is written as a number between 1 and 10 times a power of 10.

Scientific notation (11235) = 1.1235×10^4

The integer part of that number is called significand, in this case it's a 1. Every positive number can be converted without much difficulty to scientific notation, and most calculators and spreadsheets have functions that make this passage very easy.

First digit (x) = Abs (significand (a)) (1.7)

Where: $x = a \cdot 10^k$ and $0 \leq a < 10$, k integer

The definition can be adapted simply so that it defines the first two digits. The resulting statement will be that the first digits are going to be the 9 integers from 1 to 9, while the first-two digits are the 90 numbers 10, 11, 12,..., 99.

GEOMETRIC SEQUENCES AND A LITTLE FIBONACCI

In his article Benford noted how the best fits to the expected pattern were digits that could be associated with a geometric progression. Records in a dataset made up of natural numbers form a geometric sequence, especially when numbers derive from "*natural events or events of which man considers himself the originator*".

The usual mathematical representation for a geometric sequence is:

$$S_n = a \cdot r^{n-1}$$

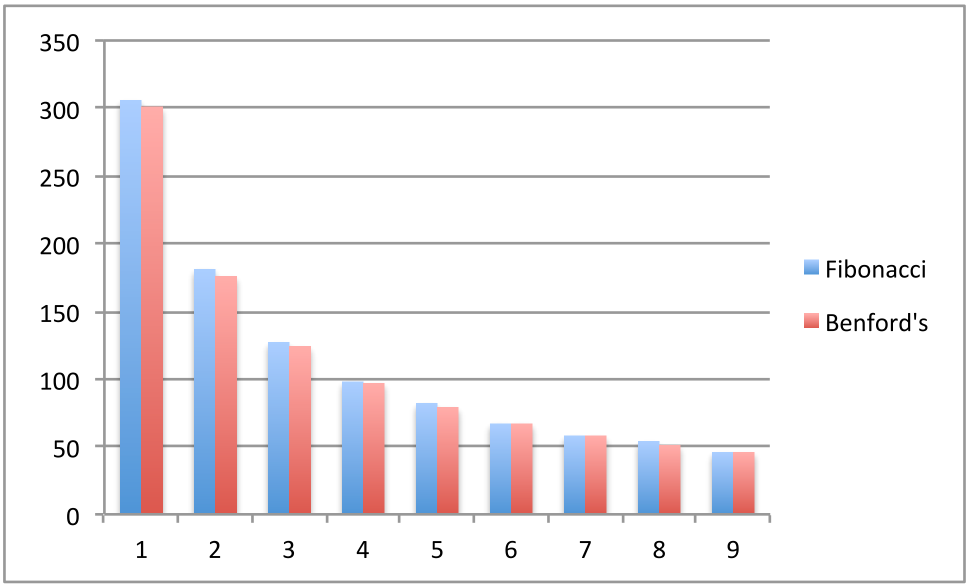
Where *a* is the first term, *r* is the common ratio and *n* represents the nth term.

It can be shown (and it will be in the next sections) how if the difference between the log of the upper bound and the log of the lower bound is an integer the set is conform to Benfors's Law. A set of records that follows Benford's Law is called Benford's Set. In the early 1970s Wlodarski (1971), soon followed by Sentence (1973) and Brady (1978), showed how the familiar Fibonacci sequence follows perfectly the first-digit law.

A Fibonacci series is, by definition, a sequence where, after the first two numbers 0 and 1, every following integer is the sum of the previous two. The sequence F_n of Fibonacci numbers is defined by the recurrence relation: $F_n = F_{n-1} + F_{n-2}$, with $F_0 = 0$ and $F_1 = 1$ as starting points.

Mathematically, the Fibonacci sequence is a geometrical progression where the common ratio tends quickly to the so called Golden Ratio, equivalent to 1.6180.

Using an Excel spreadsheet, with the first 1020 numbers of Fibonacci we obtain an extremely good fit.



The fit will tend to perfection as long as we add numbers to the sequence.

It could be considered interesting to note how a result obtainable after a few passages in Miscrosoft Excel today could be regarded as worth of publication just over thirty years ago.

AN INTUITIVE EXPLANATION

A first intuitive explation for the 'preference' the world seems to have for numbers with low first digits can be given simply considering the net assets of a mutual fund. Suppose the assets are 100 million and grow at a steady yearly rate of 10%.

The first digit one persists until the cumulative growth has reached one hundred percent, this would take 7.3 years compunding. Once at 200 millions reaching 300 millions would take much less, only 4.3 years. From 900 millions to go back to a number with first digit one would require the fund to grow only 11.1%, which would be achieve in a little more than one year. So the dataset composed by the net assets of the mutual funds over time would be mostly made of numbers with first digits recurring more often.

Time	Assets
0	€ 100,00
1	€ 110,00
2	€ 121,00
3	€ 133,10
4	€ 146,41
5	€ 161,05
6	€ 177,16
7	€ 194,87
8	€ 214,36
9	€ 235,79
10	€ 259,37
11	€ 285,31
12	€ 313,84
13	€ 345,23
14	€ 379,75

WHEN DATA CONFORM TO BENFORD'S LAW

Not all data sets follow Benford's Law. And Benford himself did not give any suggestion on which lists of numbers follow the expected frequencies and which do not. Because of the link between geometric sequences and the law, the data should be approximately a geometric progression. Experience has not shown many near-perfect geometric sequences. So those data more likely to be compliant to Benford's distribution will be those:

I. Without built-in minimum or maximum values. Except if the lowest possible record is zero.

II. Where the records should describe the sizes of similar phenomena. Such as cities' population, market value of listed stocks or the flow rates of rivers.

III. Where the records need not to be assigned numbers. Like telephone numbers, car license plate numbers or flight numbers.

IV. That have more small numbers than larger ones, implying that the data should not be clustered around a mean value. This means the larger the (bounded) variance the better the fit.

SCALE INVARIANCE

A property usually associated to Benford's Law is scale invariance.

If the first digits obey to a fixed distribution law this fact should be independent of the system of unit chosen, since nature does not prefer either the metric system or the British system, the dollars or the euros. So if the areas of the world's nations or the lengths of rivers followed a law, nothing should change if those numbers are expressed in square miles or square kilometers. Then one would define the law as scale invariant under multiplication by a non-zero constant.

A definition of scale invariance is that: A probability measure P on R^+ is scale invariant if and only if

$$P\left(\bigcup_{n=-\infty}^{\infty}[1,t)\times 10^n\right) = \log_{10} t \quad \textit{for all } t\in[1,10). \tag{1.8}$$

Roger S. Pinkham, in 1961, in the second part of his paper 'On the distribution of First Significant Digits' suggested the principle that Benford's Law as a matter of fact implies scale invariance. If scale invariance means that if a vast amount of measurements *N* are made, and if *k* is any constant, except zero, then the first digits of the numbers *kN* will be distributed in the same fashion as those of *N*. Pinkham's idea was that this would have also hold true for logN and logkN. Since *logkN = logk + logN*. If one lets *b = logk*, this signifies that the first digits of *b + logN* are distributed in the same way as the first digit of *N*, and this requires a uniform distribution for *logN*. Specifically, a uniform distribution for *logN(mod1)*.

To define this more accurately, let be F(x) a cumulative distribution function, continuous and differentiable, such that *prob(x≤a)* = *F(a)*. F is an approximation which can describe any sample of observations at hand.

$$P\,(x \in D(p)) = \sum_{-\infty}^{+\infty} [F((p+1)10^k) - F(10^k)] \qquad (1.8.1)$$

Where

Let $$D_p = \bigcup_{n=-\infty}^{\infty} [10^n, (p+1)10^n).$$

denotes the set of all members of R+ whose standard decimal expansion begins with an integert ≤ p (p=1, ..., 9)

$$G(x) = F(10^k) \qquad (1.8.2)$$

So (8.1) becomes:

$$P\,(x \in D(p)) = \sum_{-\infty}^{+\infty} [G(k + \log(p+1)) - G(k)] \qquad (1.8.3)$$

If we define

$$H\,(x) = \sum_{-\infty}^{+\infty} [G(k + x) - G(k)] \qquad (1.8.4)$$

The statement in (1.8.3) can be rewritten:

$$P\,(x \in D(p)) = H(\log(p+1)) \qquad (1.8.5)$$

Benford's law is now simply:

H(x) = x so H(x) - x = 0 for x= log2, log3, ..., log9; (1.8.6)

Since we know that the cumulative distribution function for Benford's Law is

$\log_{10}(d_1+1)$

A sufficient condition for the exactness of Benford's Law is simply:

$$H(x) = x \text{ for all } x \in [0,1) \tag{1.8.7}$$

Denoting the derivatives of the functions F, G and H, by f, g, h (these are the densities of their respective distribution functions). Thus, for all $x \in [0,1)$,

From (8.4) we have:

$$h(x) = \sum_{-\infty}^{+\infty} g(x+k) \tag{1.8.8}$$

And the sufficient condition (8.7) becomes

$$h(x) = 1 \quad \text{for all } x \in [0,1) \tag{1.8.9}$$

In terms of random variables, if F is the cumulative distribution function for variable *x*, then G is the cumulative for variable log*x* and H for the variable logx(*mod1*).

Then it is necessary to find conditions on a variable *x* which assure that logx(*mod1*) is uniformly distributed on [0,1]. Because the uniform distribution is the only kind of distribution that remains unchanged when a constant is added.

Pinkham's contribution was to apply Fourier Transform theory and Abel's theorem to obtain an explicit bound on $|H(x) - x|$. So that it's easier to recognize distributions F whose corresponding H has property (8.6) and are consequently considerable as a Benford Set.

The explicit formulas describing this bound are:

$$H(x) - x = \sum_{k=-\infty, \neq 0}^{+\infty} R(k)[1\text{-}\exp(-i2\pi kx)] \qquad (1.8.10)$$

where $$R(k)=\left(4\pi^{2k2}\right)^{-1}\int_{-\infty}^{+\infty} \exp(i2\pi kt)dg(t)$$

The sufficient conditions for the validity of these formulae are that g be of bounded variation (var(g) <∞)

Since:

$$|\exp(i2\pi kx)| = 1$$

$$|1\text{-}\exp(-i2\pi kx)| \leq 2 \qquad (1.8.11)$$

$$|R(k)| \leq \left(4\pi^2 k^2\right)^{-1} var(g)$$

$$|H(x) - x| \leq \sum_{k=0} 2\left(4\pi^2 k^2\right)^{-1} var(g)$$

Summing the series in (8.10) gives:

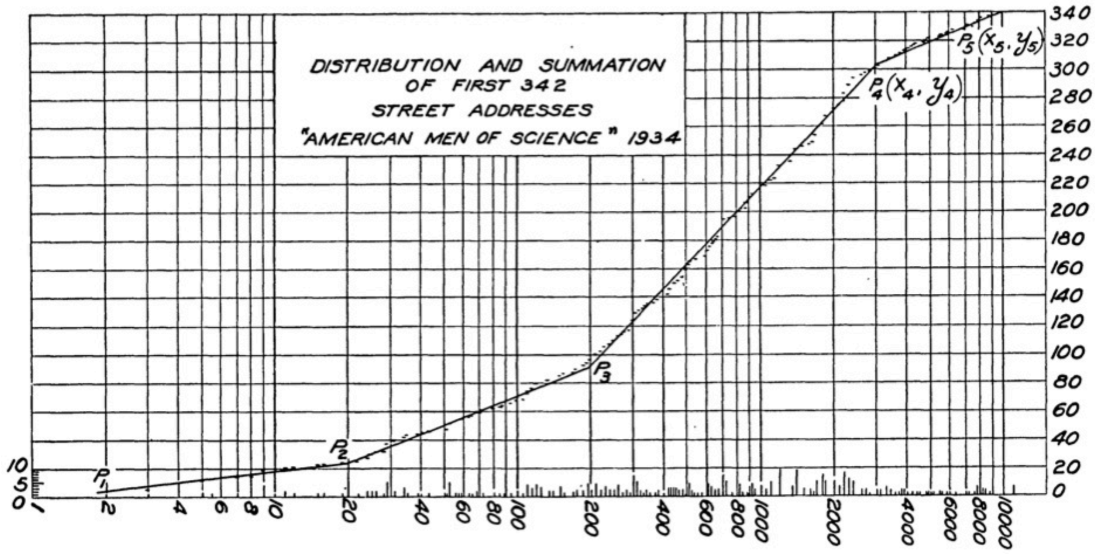
$$|H(x) - x| \leq \left(\tfrac{1}{6}\right) var(g) \qquad (1.8.12)$$

where g is the density of function G

It is clear that the quality of the approximation of the set is generally very high and does not depend on the fine structure of F or G. And it is also evident that as the variance of G increases the approximation improves markedly. The latter, as explained in the last paragraph, is one of the conditions that one looks for when is looking towards understanding if a given set is Benford or not.

When looking for conditions on a variable x which assures that logx(*mod1*) is uniformly distributed on [0,1] one can also include 'natural' distributions that fail to exist here and there, and only request that $h(x) = 1$ for a finite number of points.

Suppose F is such that for some fixed $j \geq 0$, G turns out to be piecewise linear with slope $s(n)$ on each interval $(j + n, j + n + 1)$ and such that $\sum s(n) = 1$. Then $G' = g$ is a step-function such that (1.8.8) implies (1.8.9). To illustrate it, in the figure below one can graphically see G, even if the horizontal axis is labelled as if it was F, it would actually be G just by relabelling the abscissae with the logarithms of the number displayed. Those numbers comes from one of the sets Benford's used in his original paper, specifically those taken from American Men of Science, 1934, a biographical reference on leading scientists in the United States and Canada published as a series of books.



The remarkable feature of this figure, and of any other distribution that roughly complies with Benford's Law, inheres in its general shape which astoundingly resembles that of a probability density function of a uniform distribution.

The immediate application of scale invariance in economics is that Benford's Sets will keep following this distribution even if the currency of data change.

A question that may easily come up in the reader could be: why does the natural world presents so many data sets that more or less conform to Benford's Law and are therefore scale-invariant?

To answer this question, let's go a little bit back and restate that changing the measuring unit is equivalent to multiplying by a factor k. Hence, the original values N will then become N'=kN and the corresponding distribution should be identical to the starting one, expect from a constant scaling factor A(k) that may depend upon k but not on N. One can state invariance as:

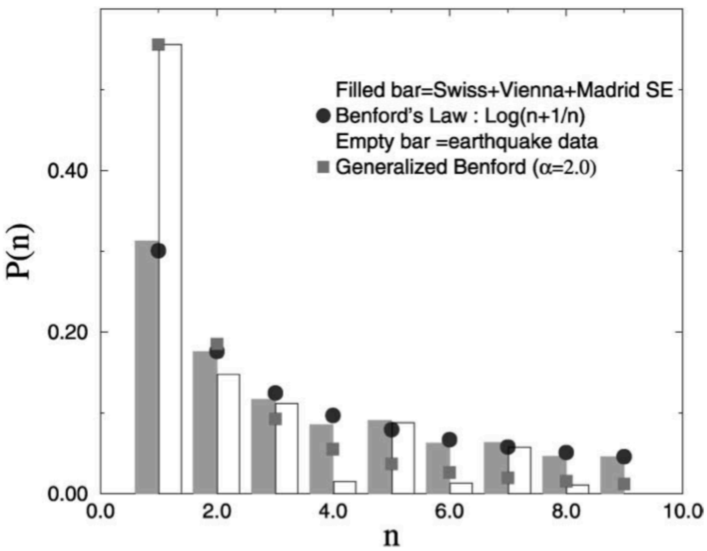
$$P(N') = P(kN) = A(k)P(N)$$

The general solution to this equation is this is the following power distribution law:

$$P(N') = N'^{-\alpha} = k^{-\alpha} \times N^{-\alpha}$$

In an example by Pietronero et al (2001) data from the magnitude of earthquakes in California are taken into account, the Richter scale has an alpha of circa 2. It's worth saying that the larger the alpha the more lower-digits one obtains. If alpha is one perfect scale invariance occurs.

Datas are taken from the combination of first digits from Zurich, Vienna and Madrid Stock Exchanges and



The case of alpha = 1 corresponds to a uniform distribution in logarithmic space. Benford Sets have an alpha of circa one. The precedent question then becomes why nature presents so many data sets with alpha ≃1?

One of the mostly used theorems in statistical probability is the Central Limit Theorem which states that let $x_1, ..., x_n$ be a set of random variables with mean u_i and bounded variance σ_i then if $S=S(x_1, ..., x_n)$ for $n \to \infty$, S is distributed Normally. If we consider a random variable

N which changes over time t according to a Brownian Dynamics so that:

$$N(t+1) = N(t) + e$$

where *e* is a random variable

The probability distribution P(N,t) to have a value N after t steps is a Gaussian with variance $\sigma \sim t^{(1/2)}$ which diverges in an infinite time limit. This is very far from scale invariance. Clearly, many systems do not follow such a dynamical description. Fluctuations are not linked to some external dynamic parameter, in most cases fluctuations are relative to the value of the random variable N, in the following way:

$$N(t+1) = e \cdot N(t)$$

where *e* is stochastic, positive and definite.

Undoubtedly this is very different from the Browninan Motion. But with a simple trasformation *e* = *loge*, which means taking the variable in logarithmic space. One can see how there's a relation betwen the two.

$$\log N(t+1) = \log e + \log N(t)$$

This means that for $t \to \infty$ the distribution P(logN) approaches a uniform distribution and by putting it back to linear space we have

$$\int P(\log N) d(\log N) = \mathrm{C} \int \frac{1}{N} dN$$

With C the normalization factor. This gives P(N) followsN^{-1} as the distribution of variables values N. Accordingly the distribution of first digits n will follow an ideal Benford law with alpha=1. What we recover is a Brownian dynamics in a logarithmic space; i.e. , a random multiplicative process corresponds to a random additive process in logarithmic space.

In conclusion, one can finally answer the question previously asked. Benford's Law shows how so many real-life phenomena are nothing more than random walks in log-space, or random multiplicative processes, whereas Benford's pushed himself to write to claim that "*claiming that mere Man counts arithmetically, 1,2,3,4…, while Nature counts* e^0, e^x, e^{2x}, e^{3x}, *and so on.*" I won't force myself to consider as true such a vigorous statement. Although, I will consider to make my own the assertion that the natural state of affairs in the world is non-linear, that contributes to explain why Benford's Law is present in such a wide range of fields and in such a large amount of observable data.

BASE INVARIANCE

Another interesting property of Benford's Law is base invariance. This means that if this significant law exists then it should be valid when written in bases different from 10, which was the base this paper has been using so far.

Let

$$Prob(D_1 = d_1, ..., D_k = d_k) = \log\left[1 + \left(\sum_{i=1}^{k} d_i \times 10^{k-i}\right)^{-1}\right]$$

Be the General significant digit law as in (1.6) for all k ∈ N, d1 ∈{1, 2, ...,9} of which the First-Digit Law is just a special case, just like the other marginal significant digit laws.

An example:

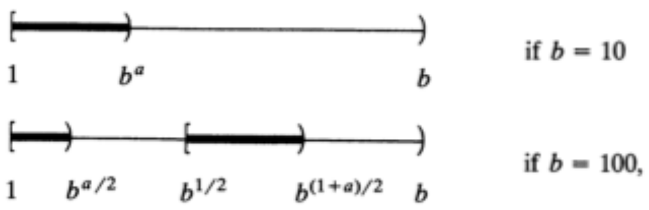
$$Prob((D_1, D_2, D_3) = (3,1,4)) = \log_{10}\left(1+\tfrac{1}{314}\right) = 0{,}0014$$

This is just a generalization of the statement that the significant digits are dependent and not independent as one may expect. Which is the intuition behind Benford's Law.

To demonstrate that (2) is the unique base invariant distribution we'll take into consideration the set of positive numbers S with, base 10, first significant digit less than 5. Using the decimal notation D_1 as above and letting D_1^{100} be the first significant digit base 100, one can see how:

$S = \{ 1 \leq D_1 < 5\} = \{1 \leq D_1^{100} < 5\} \cup \{10 \leq D_1^{100} < 50\}$

Which states graphically, as a subset of [1, b), the same set S is:



(where $a = \log_{10}5$. Thus if P is base invariant the two measures should be the same:

$$P([1, b^a)) = P([1, b^{a/2})) + P([b^{1/2}, b^{(1+a)/2}))$$

Which allows us to define P as base invariant if

$$P([1, 10^a]) = \sum_{k=0}^{n-1} P\left(10^{\frac{k}{n}}, 10^{\frac{(k+a)}{n}}\right) \qquad \textit{for all } n \in N \textit{ and all } a \in (0,1)$$

Letting P_L be the logarithmic probability defined in (1.6) and P_0 be the degenerate probability which assigns mass 1 to constant 1. It follows that:

$$P \textit{ is base-invariant} \Leftrightarrow P = qP_0 + (1 - q)P_L \qquad \textit{for some } q \in [0, 1]$$

Corollaries are that:

1. The logarithmic distriution (G) is the unique continuous base-invariant distribution;
2. Scale invariance implies base invariance, though the opposite is not true.

Thus, if there is a universal significant digit law and it is base invariant , then the constant 1 occurs with plausibly positive probability q, and otherwise the digits satisfy the logarithmic distribution law (1.6).

CONFORMITY TESTS

Among the several mothods used to verify conformity of an observations' set to Benford's Law the most used are those presented by Nigrini and Mittermaier (1997).

Let $f_i(T,N)$ be the observed relative frequency of a particular digit i in the context of conformity test T in a set of N records, and let $e_i(T)$ be the expected Benford probability as defined before.

First Digit Test (T=1): compares f_i with e_i for the first digit of numbers, i=1,..., 9.

Second Digit Test (T=2): compares f_i with e_i for the second digit of numbers, i=0, 1,...,9.

First-two Digits Test (T=3): compares f_i with e_i for the first digit of numbers, $i=i_1i_2$; $i_1i_2=10,..., 99$.

While the first two are mostly initial tests of reasonableness for a data set, the last one is applied intensively to select audit samples, as it will be described in the next sections.

Those tests can be utilized with individual statistics or with a collective statistic, one where all the relevant frequency deviations are joined in a single statistic.

A collective statistic to measure conformity to Benford's Law that ignores the number of records N is the Mean Absolute Deviation.

$$Stat\,(T, N) = \sum_i \frac{\left| f_i\,(T, N) - e_i\,(T) \right|}{n\,(T)}$$

Where n(T) is the number of feasible digits i in the context of conformity test T: n(1)=9, n(2)=10, n(3)=90.

The peculiarity of the MAD is that it doesn't follow any known distribution. Drake and Nigrini (2000) defined some critical values which can be used to conclude about the testing process.

MAD	Nonconformity
T = 1	> 0.012
T = 2	> 0.016
T = 3	> 0.0018

The MAD measures the accuracy in the same units of datas, so in our case those would be the proportions. This makes it pretty straight-forward to understand.

The MAD gives the average deviation between the Benford line and the heights of the bars. The higher it is, the larger the difference between actual and expected frequencies. Since it does not take into account the size of the data set can be used to compare the 'Benfordness' of two sets.

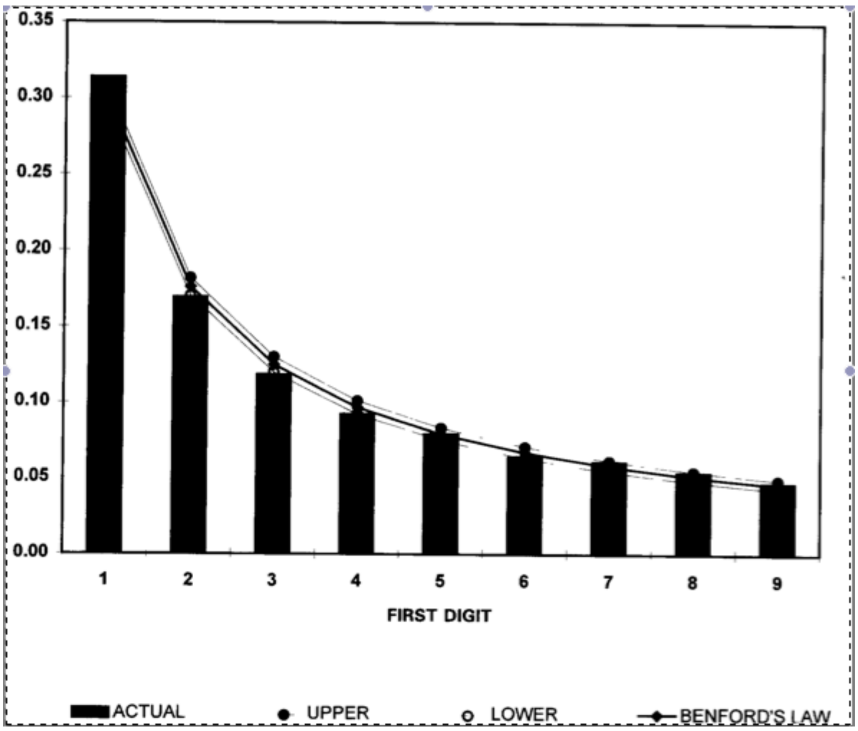
A more common statistical tool which can be applied is the Chi-Square statistic, given by

$$Stat_0\,(T,N) = N \sum_i \frac{\left[f_i\,(T, N) - e_i\,(T) \right]^2}{e_i\,(T)} \sim \chi^2_{[n(T) - 1]}$$

Where the null hypothesis corresponds to Benford's Law conformity.

Also individual statistics can sometimes prove useful. A Z-statistic consents to test whether or not the deviation of a particular feasible digit *i* from the expected Benford probability is significant. The Z-statistic is

$$Z_i(T,N) = \frac{\left[f_i(T,N) - e_i(T)\right]}{\sqrt{\frac{\left(e_i(T)\left[1-e_i(T)\right]\right)}{N}}} \sim N(0,1)$$



The null hypothesis is conformity of digit with the law. In operating this statistic one may compute n(T) statistics, one for each digit i. The frequencies shown by the thin lines above and below the Benford's Law line of the above graph are the upper and lower bounds for a significant ($p < 0.01$) difference as measured by the z-statistic.

ALL DISTRIBUTIONS LEAD TO BENFORD

Another intriguing quality of Benford's distribution is that if nonconforming data sets are multiplied between each other the resulting new data set will be more Benford than the previous two. If this process is done more than five times one obtains almost perfect conformity. In sum, numbers with any continuous distribution multiplied between each other will converge to the First-Digit Law.

The general rule seems to be that Benford's Law is similar to the Central Limit Theorem for products. This is particularly useful in accounting as many datas, take a company's payments or inventory data both obtained by multiplying the cost by the quantity. As a matter of fact, it follows that one should expect numbers resulting of successive multiplications, even if coming from uniform distributions, to adhere to Benford's Law.

Part 2 - ACCOUNTING APPLICATIONS

In 1972 Hal Varian, an economist, suggested that Benford's Law could be used as a test of honesty and 'naturalness' of purportedly random scientific data in a social science context. He affirmed that conformity to the law of the first significant digits does not automatically mean authenticity, but nonconformity should raise some skepticism. Unfortunately, his suggestion was not picked up before 1988 by Carslaw, who found that earnings numbers from New Zealand firms did not conform to the expected Benford distribution.

In his analysis Carslaw (1988) took into account psychological studies conducted by Gabor and Granger in 1966, that demonstrated the existence of key numbers that serves as reference points. The idea is that the human mind has a tendency to round up, or down, a number towards the closest reference point when assessing the magnitude of a number. For istance, when one observes a number like 7024 or 6986, a number, a factor of ten precisely, will be used as a yardstick in its perception and judgement, for this reason both numbers will presumably be reported as 7000. This is well-known in marketing and it's the explanation of the .99 € pricing phenomena.

If this is put into an accounting perspective, reference points will be used either by readers and analysts of financial statements or by the preparers of this information within a business organization. Since companies are (very often) results driven there will be pressures on management to reach an expected goal and to ensure that the first digit of a numeric goal is at least as large as the users' expectations. A smaller amount of attention will be put on meeting the second digit, even less on the third, and so on.

Because of this practice, figures reported in financial statements can be enhanced up to a significant degree especially in the mind of the user by just making sure that income numbers top those reference points, even artificially.

If this Phenomenon exists, there will be an atypical distribution of the second-from-the-left digits of income numbers by producing an abnormally low frequency of high digits and a compensating unusual high occurrence of lower digits. In this way the number will just exceed the reference point.

Carslaw (1988) reports that the frequency of occurrance of certain second digits, especially zero, contained in earnings numbers of New Zealand firms departs significantly from expectations (Benford's Law). Specifically, there's a higher than expected frequency of zeros and a less than expected frequency of nines. This provides evidence of goal oriented behavior when preparing end-of-the-year budgets.

The hypothesis of the existence of number targeting would signify that a somewhat wide deviation from the expected random distribution. Because in New Zealand the reported income may be either ordinary or net income, in contrast with North American procedures, an examination was performed on the one which was emphasized the most by the company in its report to shareholders.

In Table 2 a strong evidence emerges in favor of the numbers with lower second digits, while there's also a consistent lack of nines.

Thus, one cannot accept the alternative hypothesis of random distribution of the second digits.

By using a Chi-Squared test and individual Z-statistics he shows how most reference points are numbers just over multiples of 10^k (where $k > 0$) and no favor whatsoever was found for the number just below that threshold.

Table 2

Frequency of Second Digits for Emphasized Income Numbers

	Ordinary Income (n=319)			***Net Income (n=252)***	
Digit	Expected Distribution Percent	Observed Deviation Percent	Z-statistic	Observed Deviation Percent	Z-statistic
0	12.0	+8.7	+4.78***	+5.1	+2.49***
1	11.4	-1.7	-0.95	+1.3	+0.65
2	10.9	-3.1	-1.78*	+2.2	+1.12
3	10.4	+0.3	+0.18	+1.9	+0.99
4	10.0	-1.5	-0.89	-1.3	-0.69
5	9.7	-1.9	-1.15	-3.0	-1.61
6	9.3	+0.1	+0.06	-0.2	-0.11
7	9.0	+1.7	+1.06	-2.6	-1.44
8	8.8	+0.6	+0.37	-0.5	-0.28
9	8.5	-3.2	-2.05**	-2.9	-1.65*
χ^2	=	30.66***		15.02	

*	Significant at the .10 level
**	Significant at the .05 level
***	Significant at the .01 level

The conclusion that numbers in excess of factors of 10^k are especially common is then also tested by taking into account if the ownership of firms mattered in determining the amount of the deviation from the expected Benford's frequencies. In other words, corporations were also divided between those controlled by management and those controlled by owners. The owners were also divided between domestic and foreign. The findings are in Table 3.

The results are that domestic owner controlled companies are the ones were the abnormality of data is more pronounced and their income figures often are just in surplus over the key reference points. Manager controlled firms exhibit the same behavior, but at a lesser extent. Contrastingly, the distribution of second digits for foreign owned companies is a very good fit to the random distribution.

Thomas (1989) is another researcher who did find a similar pattern in American firms: an excess of zeroes as second digits and a lack of nines in U.S. Net income data. He also found how the opposite effect is true for companies reporting losses on their balance sheets and that Earnings Per Share (EPS) were too often multiples of 5 cents and had ending digit 9 less often than expected. Nigrini (2010) demonstrated how this has revealed true even for companies subject to more or less recent financial scandals: AIG (excess of nines to cover losses) and Enron (excess of ones to exhibit better results).

Table 3

Frequency of Second Digits for Income Numbers for Owner and Manager Controlled Firms

	Domestic Owner (n=244)			***Foreign Owner (n=126)***		***Manager Controlled Firms (n=434)***	
Digit	Expected Distribution Percent	Observed Deviation Percent	Z-statistic	Observed Deviation Percent	Z-statistic	Observed Deviation Percent	Z-statistic
0	12.0	+9.3	+4.47***	-0.5	-0.18	+4.1	+2.63**
1	11.4	-1.1	-0.54	-0.6	-0.22	+0.6	+0.39
2	10.9	+0.6	+0.30	-4.7	-1.72*	-0.5	-0.33
3	10.4	-1.4	-0.71	+2.7	+1.01	+0.2	+0.14
4	10.0	-2.2	-1.15	-1.5	-0.57	-1.2	-0.83
5	9.7	-2.3	-1.22	+2.6	+1.00	-1.9	-1.34
6	9.3	+3.0	+1.61	-3.9	-1.53	+1.8	+1.30
7	9.0	+0.8	+0.44	+2.4	+0.96	-0.5	-0.37
8	8.8	-3.2	-1.93*	+1.2	+0.48	+0.6	+0.44
9	8.5	-3.5	-1.79*	+2.3	+0.94	-3.2	-2.39**
χ^2	=	29.77**		8.80		15.45	

*	Significant at the .10 level
**	Significant at the .05 level
***	Significant at the .01 level

U.S. STOCK INDEXES' DIGITS

Research has also focused on looking into the patterns in the digits of carefully followed stock market indexes. Eduardo Ley (1996), a former assistant to Varian, found how the series of one-day returns on the two most important U.S. Stock indexes: the Dow Jones Industrial Average (DJIA) and the Standard and Poor's 500 (S&P) reasonably agree with Benford's Law. He used datas going back from 1993 to January 1900, so the sample can be considered meaningful.

The test was conducted by using the daily returns. Hence, by letting p_t be the closing value of the stock index at time t, the day-to-day yield is defined as:

$$r_t = \frac{\ln p_{t+1} - \ln p_t}{d_t} \times 100$$

Where d_t is the number of trading days between t and $t+1$. For istance between friday ($=t$) and monday ($=t+1$), d_t will be equal to 3. Consequently, whenever there is a holiday r_t is the average rate of return. In 88.35% of the cases d_t was either 1 or 3.

If a basis of prior ignorance is assumed, one could think the probability to be equally likely. Instead, as it is shown in Table 4, the one-day returns of the two indexes agree approximately with Benford's Law.

Table 4

Digit	***Benford's Law***	***DJIA (1900-1993)***	***S&P 500 (1926-1993)***
1	30,10%	28,94%	29,17%
2	17,61%	16,78%	16,96%
3	12,49%	12,38%	13,42%
4	9,69%	9,99%	9,87%
5	7,92%	8,48%	7,76%
6	6,69%	7,23%	7,13%
7	5,80%	6,15%	5,60%
8	5,12%	5,32%	5,36%
9	4,58%	4,72%	4,73%
$\mathcal{X}^2$	=	71,98	43,46
n	=	33804	24126

In this context the law of the first significant digits could be interpreted as affirming that small movements in the DJIA and in the S&P are more likely than large ones.

Since the data information is very strong we can affirm that, in a long enough time frame, stock indexes' daily returns are part of the "*outlaw numbers without known relationship*" as Frank Benford called Benford's Sets.

DIGITAL ANALYSIS AND FRAUD DETECTION

But it's Mark J. Nigrini to be credited with the commencement of the era of the extensive use of Benford's Law in accounting, with the goal to detect fraud, just a few years after Carslaw and Thomas. Nigrini used digital analysis to help identify tax evaders.

Over the years, the use of analytical procedures and technology driven processes in the planning stage of an audit, commonly called Digital Analysis, has increased dramatically.
In the United States, the SAS (Statements on Auditing Standards), a list of guidelines for external auditors in regards to the audit of non-public companies, mentions Digital Analysis more than once.
SAS 56 (*Analytical Procedures)* requires auditors to use analytical procedures in planning the nature, timing and extent of other auditing operations. Digital Analysis is a reasonableness test of whether digit patterns of single numbers assembled to make a total conform to Benford's Law. One of the performed tests is the comparison of current year account balances with prior period account balances. This comparison usually does not make much sense, especially when a company is experiencing rapid growth or has made a massive acquisition of a business. A useful test is to compare the digit patterns of both years, indeed, those are expected to follow Benford's Law.
SAS 82 called "*Consideration of Fraud in a Financial Statement Audit*" requires to make evaluations about the level of risk of material misstatement due to fraud. Digital Analysis is a useful

tool for assessing the chance of fraud: by identifying the data that does not conform with expected frequencies the auditor could put more effort in the inspection of high risk areas.

At the same time though, SAS 99 cautions that relying solely on analytical procedures, especially those on highly aggregated data, can provide only limited indications of manipulation. If used correctly, Digital Analysis can lead to the identification of specific accounts where fraud might located.

In addition to the United States some European nations, notably Netherlands and Germany, have embraced Digital Analysis in the last 10 years. Since January 1, 2002, German fiscal authorities are able to demand company data in machine readable form, in order to analyze it by means of mathemathical procedures, even before the on-site tax audit. The important implication of putting in place a system like this one is that it cuts drastically the costs of the on-site investigation for companies as well as fiscal authorities.

This implication would make Digital Analysis extremely important in a country like Italy, where the tax gap (difference between the tax due and tax paid) is among the largest in the OECD group, a club of mostly rich countries, amounting to circa 20% of real GDP. Implementing analytical audit procedures to increase the effectiveness and efficiency of the tax audit should be a primary goal of tax-collecting agencies.

DIGITAL ANALYSIS TOOLS

Some Digital Analysis tools were presented in the first chapter, such as testing of: first digits, second digits and first-two digits. But in Nigrini and Mittermaier (2000) some more instruments are introduced: number duplication, multiples test, and last two digits test. Those screenings are also put to work in the paper and used on data provided by a NYSE listed oil company.

The first-two digits test, or FTD, assumes a predominant role in most cases, since often is the one test that releases more information when used on records extracted from inventories or invoices. The number duplication examination is an extension of the FTD test that focuses on the actual numbers that caused the positive spikes that may have occurred. To target audit attention on abnormal duplications, the frequencies of the actual numbers are tabulated and auditors should determine the reasons for the duplications. Another test is that of checking for multiples, which scrutinizes the phenomenon of rounding by tabulating the proportions of numbers that are multiples of 10, 25, 100 and 1000. Users would want to use such a test when rounding could signal estimation, in cases where estimation is not acceptable. The last-two digits test, LTD, is more targeted than the rounded number test and is relevant when auditors suspect that number invention might be occurring. Indeed, it reasonable to expect that in most cases that each of the LTD has an equal probability of occurring, hence their distribution should be very close to perfect uniformity. In a recent case, LTD testing was the tool used to discover a bar owner which was just making up the numbers on the receipt at the end of each working day. The excess of 40s and 60s generated suspicion that something was wrong at Internal Revenue Service, which eventually led to a fine of a considerable amount.

In Nigrini (1992) a scheme to show a mathematical link between fraud and Digital Analysis is suggested: the Distortion Factor Model. This is a model that uses digit patterns to signal if data appears to be overstated, or understated, and quantifies the magnitude of the distortion. Estimating the level of manipulation, or distortion, requires the confrontation between the mean of the actual numbers and the mean of the numbers in a Benford Set. Since relatively small, or relatively large numbers can make up a Benford Set there is no unique mean. A solution to this first problem may be to move the decimal point of each actual number so that every record is contained in the range [10, 100). For instance, number 12345 is collapsed to 12,345 whereas 5,4321 is expanded in 54,321.

Recalling from the previous chapter, one should remember how the key trait of Benford's Law is that when data are ordered from the smallest to the largest they resemble a geometric sequence. A geometric sequence written as $ar^{(n-1)}$ where the ratio r is the (n +1)th element divided by the nth element. When r is a function of the range [10, 100) and of N (number of observations), it is computed as follows:

$$r = 10^{\frac{(\log(ub)\text{-}\log(lb))}{N}}$$

Where *ub* and *lb* represent the upper and lower bound of the geometric sequence.

For the sequence of collapsed values, a is fixed at 10 and log(ub) - log (lb) is fixed at 1. The average value of the series is dependent on N (r is a function of N) but it doesn't make much difference when N is greater than 500. Since the expected sum of the elements of any geometric series is ,

$$\mathrm{ES} = a \cdot \frac{(r^N - 1)}{(r - 1)}$$

The Expected Mean (EM) of elements spanning the [10, 100) interval is derived by substituting a=10, r = $10^{1/N}$ from equation (5) and then dividing by N,

$$\mathrm{E}M = \frac{90}{\left[N \cdot \left(10^{\left(\frac{1}{N}\right)} - 1\right)\right]}$$

The EM of any large (>500) Benford Set is approximately 39,08. If every digital combination had equal probability to be found the mean would be 55. The Distortion Factor Model is given by:

$$DF = \frac{(AM - \mathrm{E}M)}{\mathrm{E}M}$$

Multiplied by a hundred, it measures the percentage deviation of the AM from the EM. An overabundance of lower first digits signals how smaller numbers occurred more often compared to the Benford Set and the DF would be negative.

Hence, the steps to use the DIstortion Factor Model are:

1. Transform reported numbers to numbers in the range [10, 100); this would imply only to collapse larger number since it is often

better to eliminate the smaller figures which could interfere causing spikes due to immaterial amounts.

2. Compute the Actual Mean.
3. Compute the Expected Mean of a Benford Set.
4. Compute the DIstortion Factor.

The Distortion Factor Model is used on tax data to analyze fiscal evasion.

Cuccia (1994) notes how in the majority of research concerning compliance has found that the reporting decision, whether a taxpayer declares more or less than due to the authorities, is made at the time of filing. Nigrini (1992) suggests the following dichotomy of tax evasion:

I) Planned evasion (PE): the taxpayer prepares all year his steps and hides its traces to audit.

II) Unplanned Evasion (UPE): Evasion occures at the moment of preparing the return statement. Blatant adjustments (downward for income items, upward for deductible items) thought to be safe prior to an audit but detectable upon an audit. It is a behavioral act where the taxpayer fabricates a number in the tax return. The act is influenced by specific numbers acting as psychological barriers as specified by Carslaw.

The essential difference is in timing and skills the taxpayer puts in place. The core assumption at the base of the UPE is that taxpayers have a small amount of knowledge of the “average” return filed by people in homogeneous economic positions, and that the true number act as a reference point upon which to base the reported number. Consequently it is reasonable to think that the reported number is in the same $[10^k, 10^{k+1})$ range as the true value.

Thus, the DF could prove very useful in detecting unplanned fiscal evasion, shortly after the returns of the year are filed. Since evasion estimates are usually computable only after some years, the DF model formulates inexpensive extimates of the UPE. For obvious reasons, it wouldn't make much sense to use it to investigate on Planned Evasion, because even if data were checked through an on-site audit all the relevant papers would look clean.

Part 3 - BENFORD TOURS THE EU

When one begins to work with Benford's Law to detect accounting fraud he needs to look for support for the idea that accounting data should, habitually follow Benford's Law. One should expect conformity to Benford to be constant over time because of the scale invariance theorem. If the numbers increase by x% every year as a result of economic growth and by y% because of inflation, then a Benford Set will continue to be a Benford Set even with the numbers altering on a year-to-year basis.

A quite interesting paper published earlier this year by Gernot Brahler and a group of German researchers investigates through a Benford test on the quality of macroeconomic data relevant to the deficit criteria reported to Eurostat by the EU member states. Since the data are collected by national agencies and then reported to the European Statistic Agency (Eurostat), which until 2010 was not given the authority to inspect them more thoroughly and ensure their quality. Eurostat had no right to audit those informations the way regulators and private companies audit the financial statements of companies. This led to an incentive for an individual country to manipulate its economic statistics. Especially because the same data were the ones used to comply with the Stability and Growth Pact criteria of the European Union. On top of of that, countries could have been willing influence those numbers in order to obtain more favorable conditions on the capital markets.

The non-Benfordness of a set of observations does not imply manipulation, as said earlier, but significant non-conformity could mean poor data quality. A deviation from the law is not a conclusive proof as well as conformity doesn't prove cleanliness of the statistics. The concept is, once again, that non-conformity is a signal and could direct auditors to closer examination and additional testing. The use of Benford testing on macroeconomic indicators was put forward by Nye and Moul (2007), who were able to show how growth rates of GDP leads to sets agreeing with Benford's Law.

Brahler's approach was not of identifying datasets which disagree remarkably from Benford's Law, instead they ranked the EU member states as specified by the deviation of their data from the law. This ranking could be used to indicate the probability of manipulation in the countries' data. From the standing of each sovereign it is possible to determine the order and the extent of further auditing procedures.

The dataset used consisted of:

1. Government statistics: government deficits, surplus, debt and associated data.
2. Government finance statistics: government revenue, expenditure and main aggregates.
3. National accounts: GDP and main components, at current prices.
4. Financial accounts: balance sheets, financial transactions, asset and liabilities consolidated.

These categories are all related to public deficit, public debt and gross national product, thus, they are those use to compute the coefficients for the Stability and Growth Pact. The total is of 156 positions per country per year, from 1999 (Introduction of the Euro) to 2009. The total sample contains 39691 observations.

To measure the conformity to Benford's law a chi-squared test on the first digit was conducted. To make sure that the position in the ranking was not due to sample size, a Pearson correlation coefficient between sample size and chi-squared statistic was computed. The value of the coefficient (0,049) is insignificant. Benford distribution was the null hypothesis of the test, thereby a rejection of the test suggests manipulation. The null hypothesis is rejected at a 5% level of significance if the chi-squared statistic exceeds 15,5073.

Besides the standard chi-squared test another measure of distance of actual records from expected ones was calculated: dividing chi-squared by sample size n gives back a measure independent of sample size.

The results for each member state are shown in Table 5 below. Greece with a mean value of 17,74 leads every other nation for the magnitude of its deviation from the expected value. Belgium and Austria follow right after. The ranking determined by chi-squared and by chi-squared divided by n are significantly close.

Table 5

EU Sovereigns ranked by χ^2 statistics

	Mean size	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Mean χ2	χ2	χ2 /n
Greece	135.55	7.70	26.33**	16.55**	20.50**	12.37	9.45	16.36**	14.53	15.85**	27.88**	27.61**	17.74	1	1
Belgium	148.73	14.10	14.14	11.67	18.01**	39.27**	24.32**	23.87**	16.98**	6.48	8.09	12.40	17.21	2	2
Austria	139.36	14.73	32.00**	27.13**	7.84	17.40**	18.23**	13.06	14.96	4.69	7.76	9.89	15.25	3	3
Ireland	137.73	14.51	15.00	9.78	18.34**	14.75	10.17	9.23	9.14	18.74**	17.04**	23.90**	14.60	4	4
Finland	150.00	6.29	11.66	17.16**	19.01**	13.97	8.12	6.70	17.50**	13.48	19.38**	18.34**	13.78	5	9
Slovakia	132.91	4.37	21.01**	13.19	11.30	8.45	5.41	6.92	14.45	18.64**	23.82**	8.89	12.40	6	8
Germany	149.18	6.68	12.88	8.89	19.21**	12.68	7.20	5.95	16.35**	14.05	15.89**	16.33**	12.37	7	12
Italy	134.27	16.45**	13.55	5.47	13.13	13.30	10.46	12.77	14.57	14.76	9.12	12.45	12.37	8	7
Cyprus	117.55	7.80	12.73	17.07**	9.74	13.83	16.37**	11.85	9.65	8.46	7.08	16.96**	11.96	9	5
Slovenia	134.64	12.86	5.69	19.34**	10.46	23.91**	10.58	4.46	4.98	9.99	17.20**	6.35	11.44	10	10
Spain	137.09	11.50	9.92	13.63	10.88	6.92	23.84**	13.95	14.70	2.50	6.04	11.09	11.36	11	11
France	149.00	7.75	11.28	23.33**	8.39	10.10	7.24	6.25	11.30	8.73	6.62	21.24**	11.11	12	14
Malta	114.91	16.66**	9.94	2.80	4.96	10.05	7.67	12.39	15.57**	21.31**	7.57	9.65	10.78	13	6
Luxembourg	123.82	23.25**	16.99**	3.86	8.28	5.92	17.87**	9.04	9.36	3.34	6.89	8.87	10.33	14	13
Portugal	149.91	28.60**	12.43	22.64**	5.50	6.86	5.78	4.69	6.09	4.37	4.95	10.14	10.19	15	15
Netherlands	139.64	6.46	6.06	11.08	3.32	8.41	2.56	10.20	10.51	12.89	5.03	9.62	7.83	16	16

In Table 6 the distribution of the first digits for Greece is shown in detail. The higher proportions of digits one and two when juxtaposed to those predicted by Benford is noticeable. In massaged and adjusted data one could suppose a more uniform distribution, instead the pattern of the deviation is contingent on the specific way the data are misreported. It was Nigrini (1996) who pointed out that due to psychological barriers manipulated data will have lower digital combinations than the true values.
Problems with Greek data have been well-known for a long time (European Commision reports on the issue appeared in 2004 and 2010). It is of public knowledge now that the authorities repeatedly corrected debt figures at least for the three years going from 2005 to 2008. In 2008 the deficit was revised from 5% of GDP to 7,7%. In 2009 the planned deficit quotient of 3,7% was increased at first to 12,5% and then to 15,4%. Therefore it can be stated with a large degree of certainty that the fiscal statistics provided to Eurostat from the Hellenic Republic were not of the finest quality.

Although, contrary to one's expectations, the results do not seem to suggest that data reported by the so-called PIIGS, an acronym used to denominate the group of Eurozone countries sharing high debt and financial problems, are of lower than average quality.
Once again Benford's Law reveals to be a good indicator of manipulation that can direct audit on reported data to make the examination more efficient and effective.

Table 6

Comparison of Greek statistics (1999-2009) with Benford's Law

Digit	Benford	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	0.30	0.37	0.40**	0.42**	0.36	0.30	0.33	0.39**	0.31	0.34	0.26	0.29
2	0.18	0.22	0.27**	0.22	0.18	0.23	0.23	0.16	0.23	0.26**	0.34**	0.26**
3	0.12	0.06**	0.08	0.11	0.10	0.11	0.12	0.16	0.13	0.12	0.10	0.20**
4	0.10	0.09	0.03**	0.05**	0.07	0.08	0.10	0.05	0.08	0.08	0.11	0.10
5	0.08	0.09	0.10	0.04	0.06	0.04	0.04	0.06	0.04	0.04	0.05	0.02**
6	0.07	0.06	0.02**	0.05	0.04	0.06	0.05	0.05	0.07	0.08	0.05	0.04
7	0.06	0.04	0.02	0.04	0.06	0.04	0.02	0.01**	0.03	0.04	0.04	0.05
8	0.05	0.03	0.05	0.04	0.12**	0.06	0.06	0.04	0.03	0.03	0.04	0.03
9	0.05	0.05	0.03	0.05	0.03	0.08**	0.05	0.07	0.09**	0.01	0.01	0.00**
χ^2	7.70	26.33**	16.55**	20.50**	12.37	9.45	16.36**	14.53	15.85**	27.88	27.61**	
χ^2/n		0.08	0.20	0.13	0.14	0.09	0.07	0.12	0.10	0.11	0.20	0.20
Size (n)	101	132	132	142	142	141	140	141	140	140	140	

Part 4 - SUMMARY AND FUTURE RESEARCH

The first chapter reviewed the mathematics underlying Benford's Law. This included the logarithmic basis of the law, the scale invariance theorem dealing with multiplication by a constant, base invariance dealing with the validity of the law in counting systems different other than the decimal one, and mathematical manipulation that gave back a data set conforming to Benford's Law. The link with the Fibonacci sequence was also reviewed. Some of the primary tests associated with Benford's Law were introduced such as the first digit test and the first-two digits test, which are used for every analysis. The following chapter reviewed some of the applications of Benford's Law to economic data. Those ranged from stock market rate of return, to income numbers paid by public stocklisted companies in New Zealand. Next, Digital Analysis was introduced and various methods showing the relation of this technique with Benford's Law were reviewed, notably the Distortion Factor Model created by Mark Nigrini. Because of previous research it is implied that since many accounting data follow Benford's Law, manipulated data deviate from it. Examining those deviations with professional judgement can lead to suprising improvements in the quality of audits. The paper ends with a review of a study by German researchers who scritinized throughly macroeconomic data of European Union's nations, this pointed out how statistics provided by Greek authorities should have been audited long before 2009, when they were reported publicly to have been manipulated for years.

In sum, applications of Benford's Law to accounting are far more than the one proposed and new ones are emerging at a remarkably fast pace. Benford's analysis, when used correctly, is a useful tool for identifying suspect accounts for further analysis. This could improve substantially the efficiency and effectiveness of the planning stage of an audit, when targets for deep controls are chosen, reducing the loss of time and budget for tax agencies. Research in the field of accounting applications of Benford's Law is ongoing at remarkably fast pace. In the last ten years more than 500 papers were published, compared with only 200 since its discovery in 1938. My suggestion for future studies in the area is comparing the results of a set of tax audits when the planning is executed through a procedure that takes Benford's Law into account with the results of a number of tax audits executed on random targets. An empirical work of this kind would strongly favor the expansion of Digital Analysis techniques. Finally, a new paper that reviews all the applications of the law, from criminology to hydrology to finance, is much needed since the last one now belongs to more than fifteen years ago.

REFERENCES

American Institute of Certified Public Accountants (2007), *AICPA Professional Standards: U.S. Auditing Standards as of June 1*, 2007.

Benford F. (1938), *The law of anomalous numbers*, Proceedings of the American Philosophical Society, 78(4):551-572.

Brady, W. (1978), *More on Benford's Law*, Fibonacci Quarterly, 16, 51-52.

Carslaw C. (1988), *Anomalies in income numbers: Evidence of goal oriented behavior*, The Accounting Review, LXIII (2), pp 321-327.

Cuccia, A. (1994), *The economics of tax compliance: What do we know and where to we go?*, Journal of Accounting Literature, 13: 81-116.

Drake, P. And M. Nigrini (2000), *Computer assisted analytical procedures using Benford's Law*, Journal of Accounting Education, N° 18, pp 127-146.

Durtschi, C., W. Hillison and C. Pacini (2004), *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*, Journal of Forensic Accounting, Vol. V, pp 17-34.

European Commission (2010a), “*Report on Greek Government Deficit and Debt Statistics*”, available at http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=COM_2010_report_greek [state of 13/07/2010].

European Commission (2004), “*Report by Eurostat on the Revision of the Greek Government Deficit and Debt Figures*”, available at http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/GREECE/EN/GREECE-EN.PDF [state of 2010/07/13]. Gabor, A. , and C. W. J. Granger (1966), Price as an indicator of quality: report on an enquiry, Economica, pp 43-70.

Goudsmit, S. (1977), *Pitfalls in elementary probability*, Proceedings of the American Philophical Society, 121, 188-189.

Hill, T. (1995), *A statistical derivation of the significant-digit law*, Statistical Science, Vol. 10, N° 4, pp 335-356.

Hill, T. (1995), *Base-invariance implies Benford’s Law*, Proceedings of the American Mathematica Society, N° 123, pp 887-895.

Hill, T. (1995), *The significant-digit phenomenon*, American Mathematical Monthly, N° 102, pp 322-327.

Newcomb, S. (1881), *Note of the frequency of use of the different digits in natural numbers*, American Journal of Mathematics, 4:39-40.

Ley, E. (1996), *On the peculiar distribution of the U.S. Stock Indexes' Digits*, The American Statistician, Vol. 50, N° 4, pp 311-313.

Nigrini, M. (2000), *Continuous Auditing*, Ernst & Young Center for Auditing Research and Advanced Technology, University of Kansas.

Nigrini, M. And L. Mittermaier (1997), *The Use of Benford's Law as an Aid in Anlaytical Procedures*, Auditing: A Journal of Practice & Theory, Vol. 16, N° 2, pp 52-67.

Nigrini, M. (1994), *Using digital frequencies to detect fraud*, The White Paper, April, pp 3-6.

Nigrini, M. (1992), *A taxpayer compliance application of Benford's Law*, Journal of the American Taxation Association, pp 72-91.

Nigrini, M. (2011), *Benford's Law: Applications for Forensic Accounting, Auditng, and Fraud Detection*, Wiley & Sons.

Nye, J. and C. Moul (2007), "*The Political Economy of Numbers: On the Application of Benford's Law to International Macroeconomic Statistics"*, B.E. Journal of Macroeconomics 7.

Pietronero, L., Tosatti, E., Tosatti, V., and Vespignani, A. (2001), *Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf*, Physica A: Statistical Mechanics amd Its Applications, 293, 297-304.

Pinkham, R. (1961), *On the distribution of the first significant digits*, Annals of Mathematical Statistics, N° 32, pp 1223-1230.

Raimi, R. (1976), *The First Digit Problem*, The American Mathematical Monthly, Vol. 83, N° 7, pp 521-538.

Sentence, W. (1973), *A further analysis of Benford's Law*, Fibonacci Quarterly, 11, 490-494.

Thomas, J. (1989), *Unsual patterns in reported earnings*, The Accounting Review, N° 10, pp 773-787.

Varian, H. R. (1972), *Benford's Law*, The American Statistician, Vol. 26, pp 65-66.

Wlodarski, J. (1971), *Fibonacci and Lucas numbers tend to obey Benford's Law*, Fibonacci Quarterly, 9, 87-88.