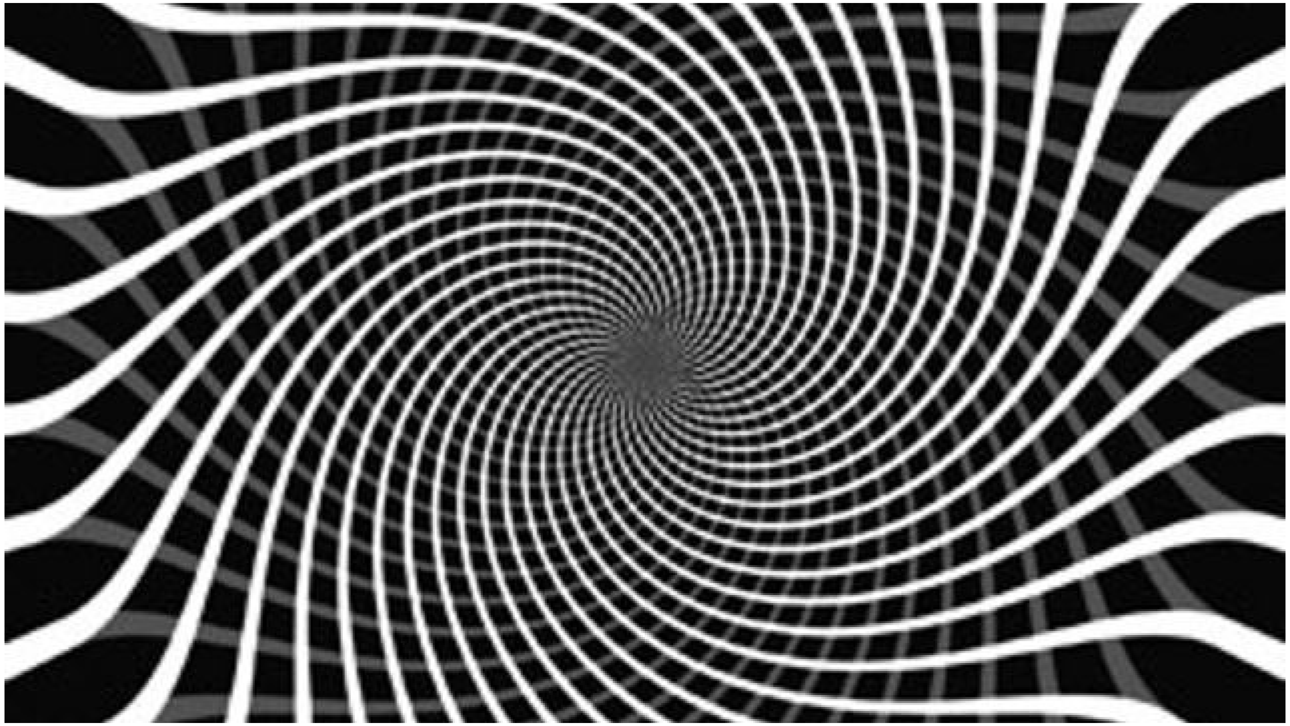
Is the whole Benford phenomenon merely an illusion?



This could be so because Benford’s Law is depended on our arbitrarily invented positional number system as it focuses on the symbolic digits of numbers.

Admittedly, a decisive digital pattern does exist for our positional number system.

Benford's Law for our number system

A number in a data set:

478,932

The first digit on the left:

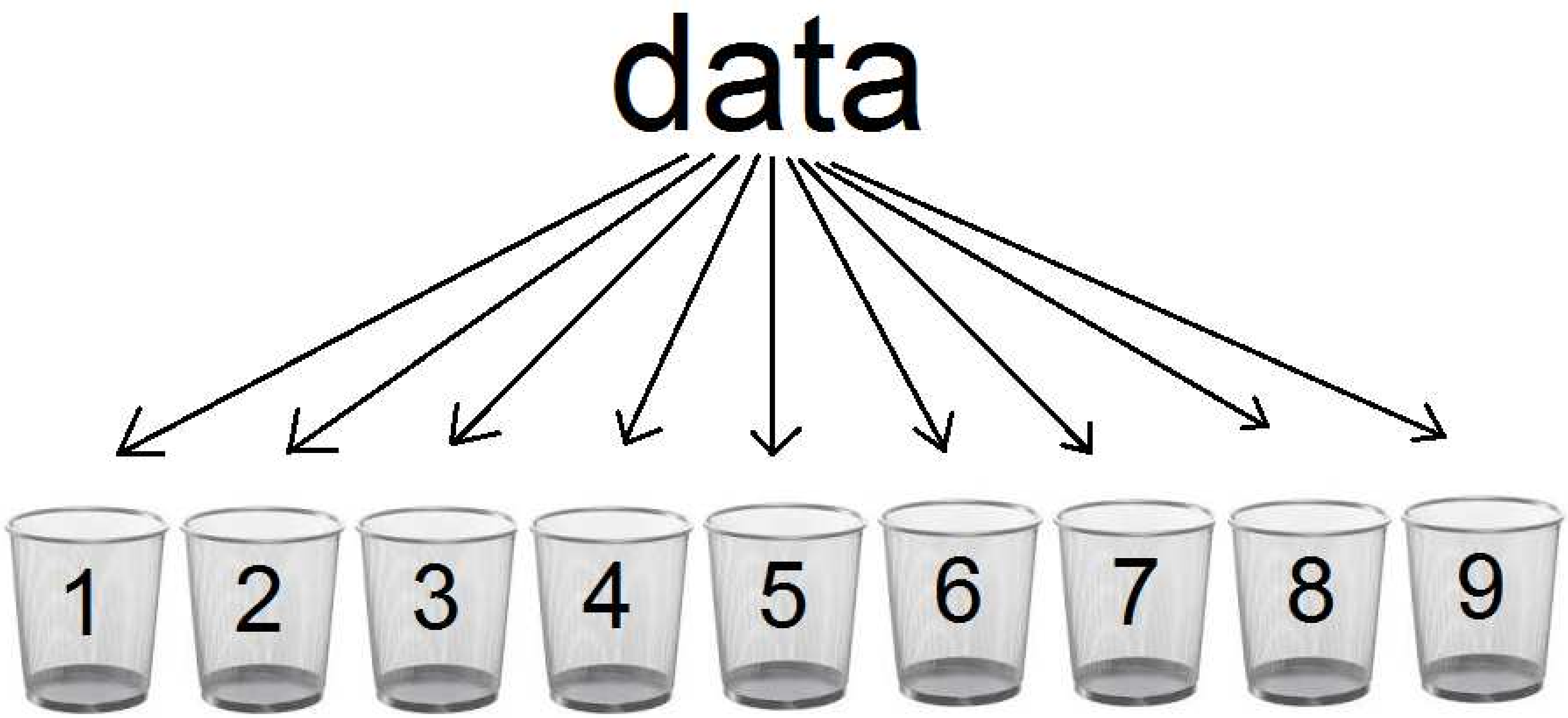
478,932

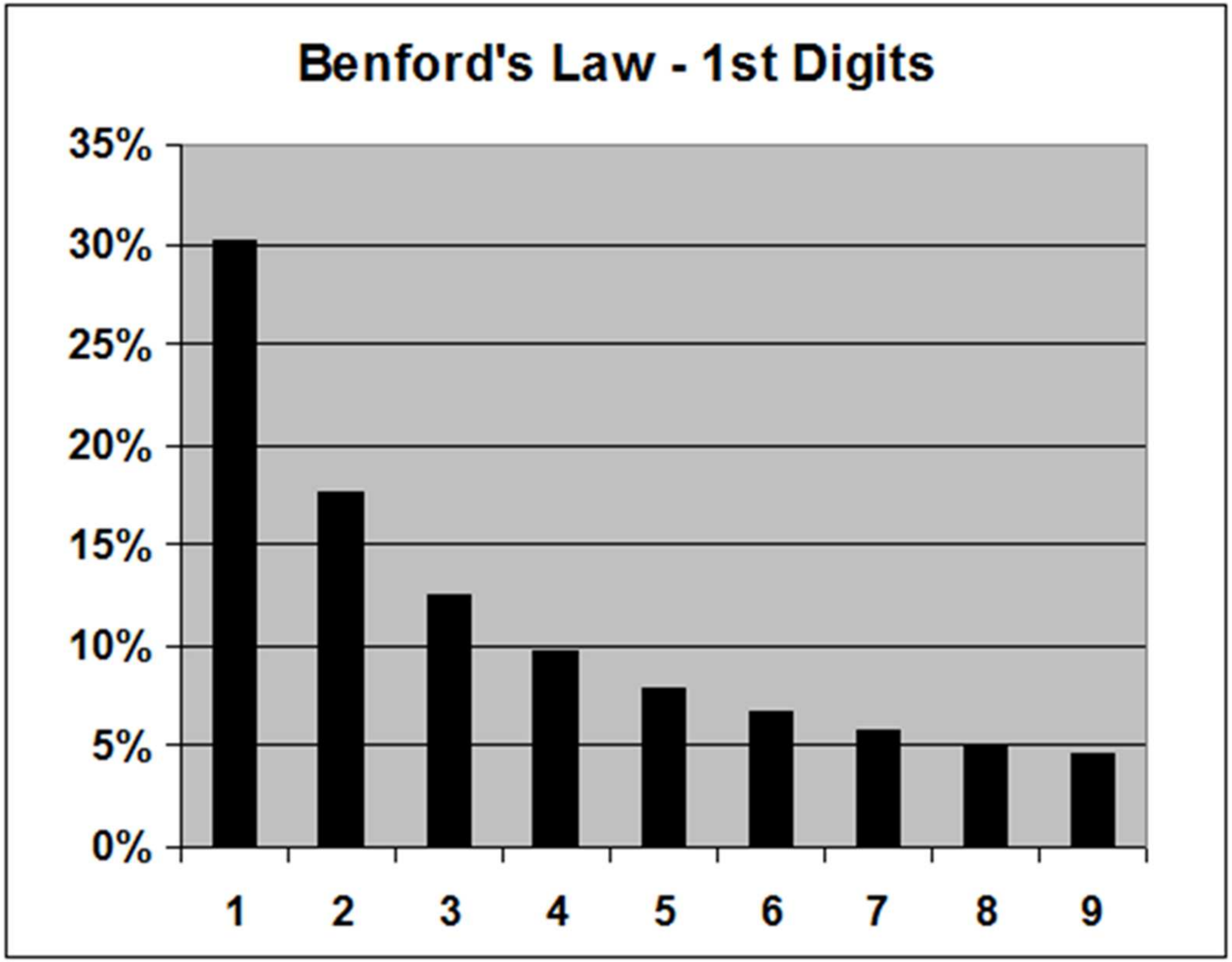
Data Set

285.29	185.35	2579.80	27.11
5330.22	1504.49	1764.41	574.46
1722.16	815.06	3686.84	1501.61
494.17	362.48	1388.13	1817.27
3516.80	5049.66	2414.06	387.78
4385.23	2443.98	2204.12	1224.42
1965.46	3.61	1347.30	271.23
3247.99	753.80	1781.45	593.59
1482.64	1165.04	4647.39	1219.19
251.12	7345.52	1368.79	4112.13

Focus on 1st digits

285.29	**1**85.35	**2**579.80	**2**7.11
5330.22	**1**504.49	**1**764.41	**5**74.46
1722.16	**8**15.06	**3**686.84	**1**501.61
494.17	**3**62.48	**1**388.13	**1**817.27
3516.80	**5**049.66	**2**414.06	**3**87.78
4385.23	**2**443.98	**2**204.12	**1**224.42
1965.46	**3**.61	**1**347.30	**2**71.23
3247.99	**7**53.80	**1**781.45	**5**93.59
1482.64	**1**165.04	**4**647.39	**1**219.19
251.12	**7**345.52	**1**368.79	**4**112.13


data
1
2
3
4
5
6
7
8
9


Benford's Law - 1st Digits
35%
30%
25%
20%
15%
10%
5%
0%
1
2
3
4
5
6
7
8
9

Data is random, but…

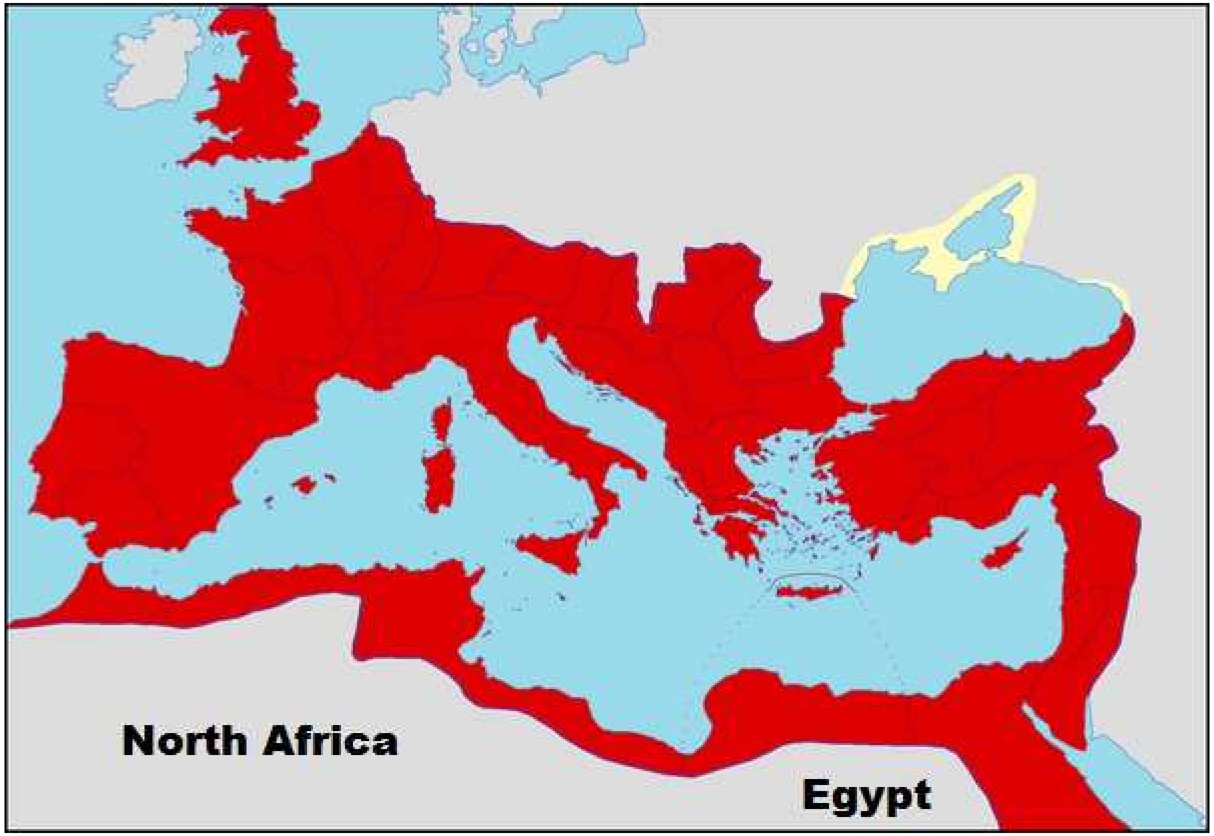
The 1st digit is not so random!

The 1st digit is almost predictable!

===============================

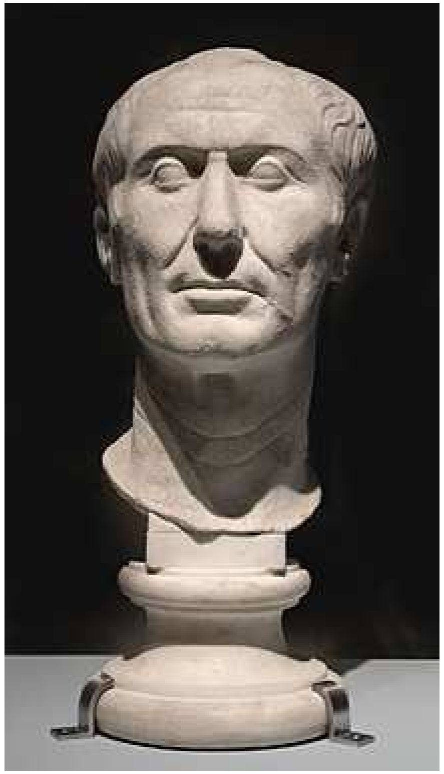
Let us consider other number systems:

Roman Empire Territory



27 BC – 395 AD

The Roman Empire



Emperor Julius Caesar, 100 BC – 44 BC

The Roman Empire

Pax Romana (Roman Peace), 27 BC - AD 180

The Roman Empire

Pax Romana (Roman Peace), 27 BC - AD 180

Roman Numerals

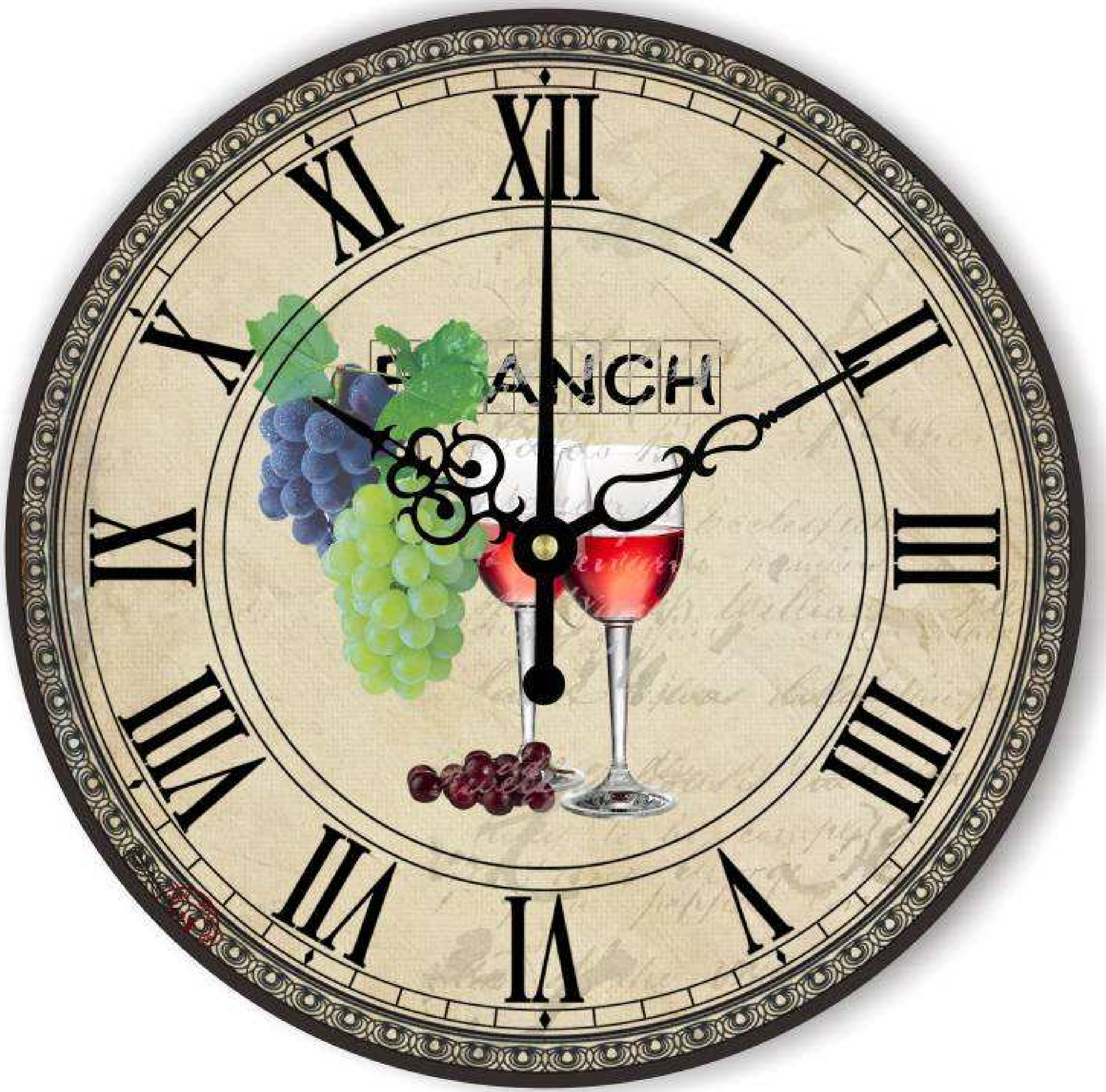
I 1	II 2	III 3	IV 4	V 5
VI 6	VII 7	VIII 8	IX 9	X 10
XI 11	XII 12	XIII 13	XIV 14	XV 15
XVI 16	XVII 17	XVIII 18	XIX 19	XX 20

Roman Numerals

90	= 50 + 10 + 10 + 10 + 10
	= L + X + X + X + X
	= LXXXX
90	= 100 – 10
	= C – X
	= XC

Terribly inefficient !

…yet elegant and beautiful…



Is there a ‘Benford-like-law’ for Roman Numerals?

Data Set

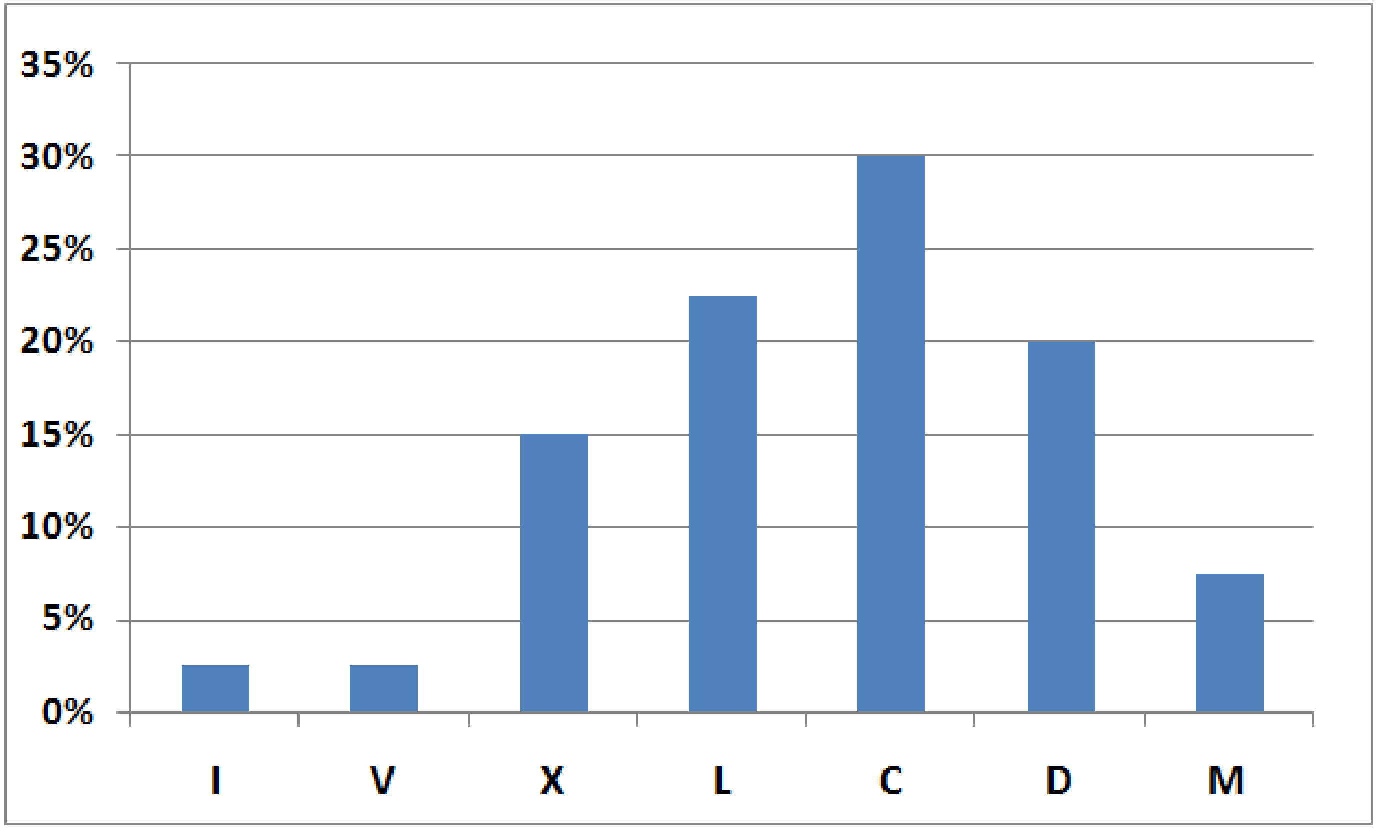
DLXXXV	CCLXXX	XVII	XCII
CDIV	MVIII	CDXLIII	VCIII
DCLXXVII	LDXV	LXXVII	CMVLIII
LMXLIII	DCCX	CCCVCIV	LXVII
DLXVI	LXXII	XLIV	LII
CCIX	MCMXI	CDV	CXXII
DCIX	LDXII	XXVIII	CDXLIII
DCVC	LXVII	CIX	DCLV
MMCLIII	IV	XXVI	CCCLXVI
DCII	LXXXIV	CCCXXI	XXXII

Focus on 1st numeral

DLXXXV	**C**CLXXX	**X**VII	**X**CII
CDIV	**M**VIII	**C**DXLIII	**V**CIII
DCLXXVII	**L**DXV	**L**XXVII	**C**MVLIII
LMXLIII	**D**CCX	**C**CCVCIV	**L**XVII
DLXVI	**L**XXII	**X**LIV	**L**II
CCIX	**M**CMXI	**C**DV	**C**XXII
DCIX	**L**DXII	**X**XVIII	**C**DXLIII
DCVC	**L**XVII	**C**IX	**D**CLV
MMCLIII	**I**V	**X**XVI	**C**CCLXVI
DCII	**L**XXXIV	**C**CCXXI	**X**XXII

I	V	X	L	C	D	M
1	5	10	50	100	500	1000

I	V	X	L	C	D	M
3%	3%	15%	23%	30%	20%	8%


35%
30%
25%
20%
15%
10%
5%
0%
I
V
X
L
C
D
M

NO!

No law is found here!

Distinct data sets yield distinct 1st-numeral proportions.

There exists no pattern!

WHY?

Just because Roman Numerals are inefficient?

NO!

That lack of a pattern has nothing to do with number-system-efficiency!

===============================

Ancient Egypt Territory

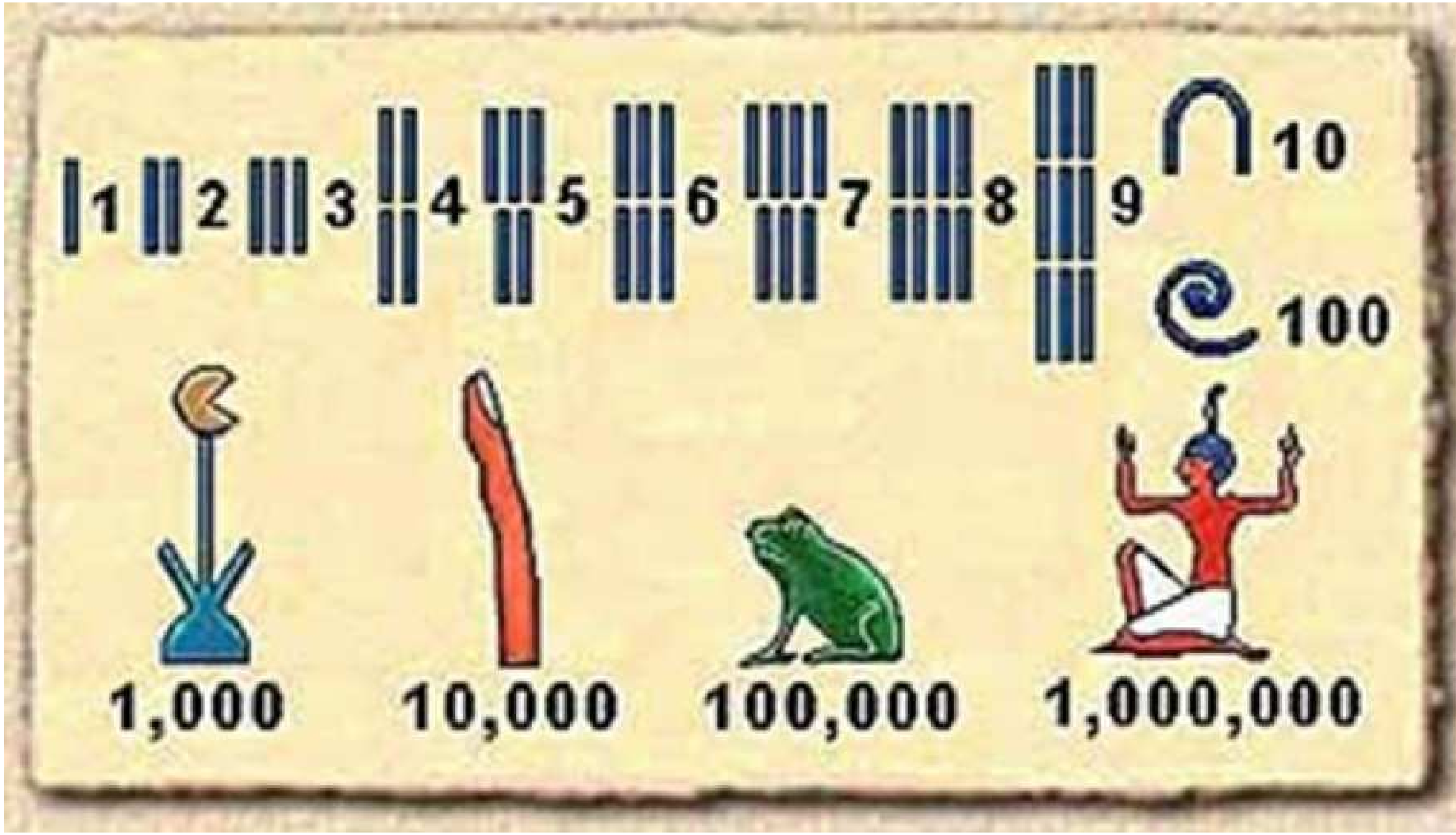
3000 BC – 30 BC

Ancient Egypt

Ancient Egypt

Ancient Egypt

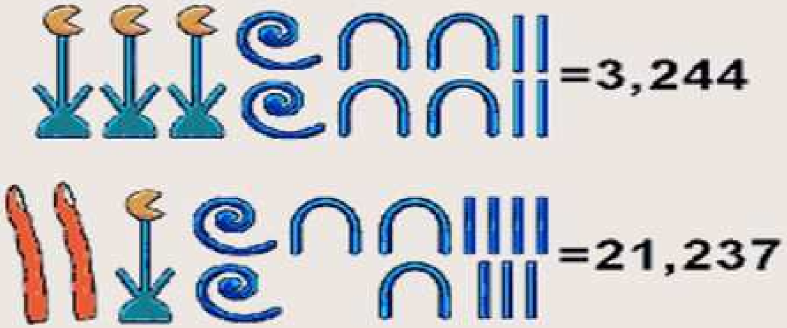
Egyptian Numerals



Egyptian Numerals

Egyptian number system

- Egyptian (as early as 3000 BCE)
  - How would you write 3,244?
  - How would you write 21,237?



Terribly inefficient !

Is there a‘Benford-like-law’ for Egyptian Numerals?

NO!

No law is found here!

There exists no pattern!

===============================

Positional Number System Base 10

An example from positional number system base 10:

7205.38 is **defined** as:

$7*10^{3} + 2*10^{2} + 0*10^{1} + 5*10^{0} + 3*10^{-1} + 8*10^{-2}$.

It **combines** multiplications (*) and additions (+) of powers of ten (10^N).

It's quite peculiar!

Positional number system base 10 is truly a sort of

a scheme

an algorithm

a process

a procedure

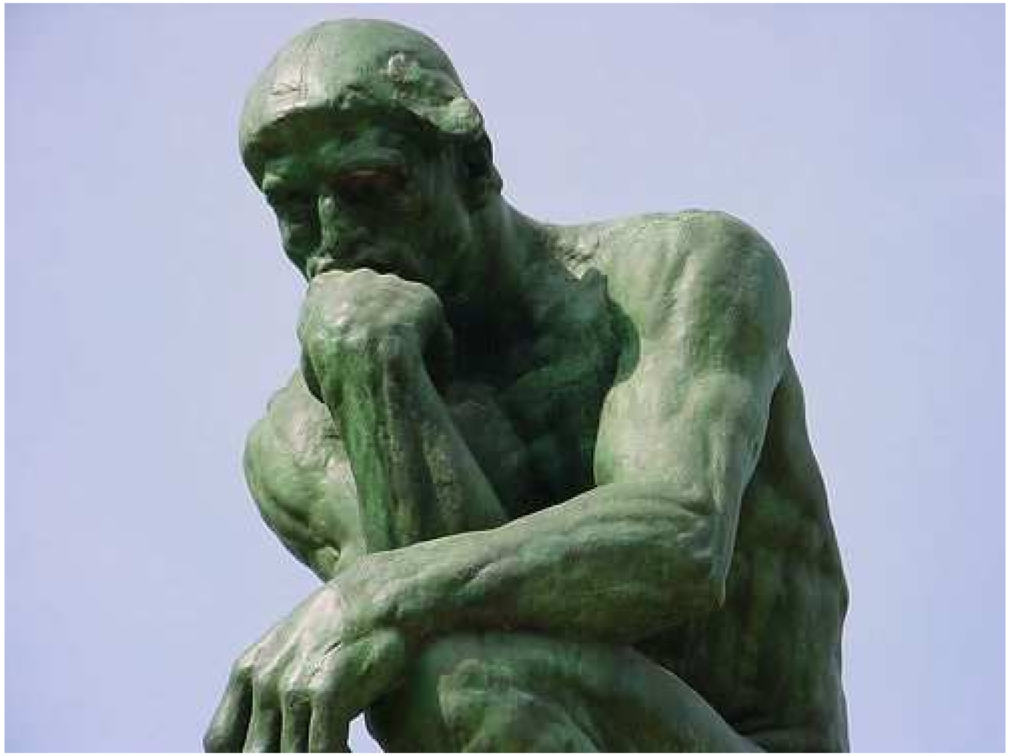
!!!

Should Benford's Law then be considered simply as **arbitrary**?!

Our positional number system, completed during the Renaissance Period is extremely efficient.

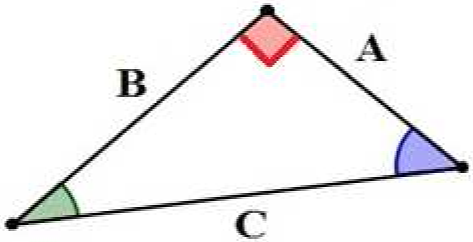
But it's still arbitrary!

Our positional number system was **invented** by us.

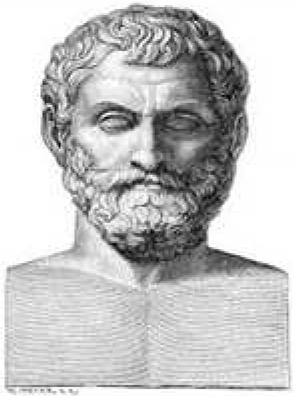


The meanings of the verbs to discover and to invent are distinct.

We discovered:



$$C^2 = B^2 + A^2$$


Thales


Pythagoras

We discovered:

$$F = MA$$
$$F = GM_1M_2/R^2$$


Isaac Newton

We invented:

Positional Number System

$$843.7 = 8*10^2 + 4*10^1 + 3*10^0 + 7*10^{-1}$$


Brahmagupta

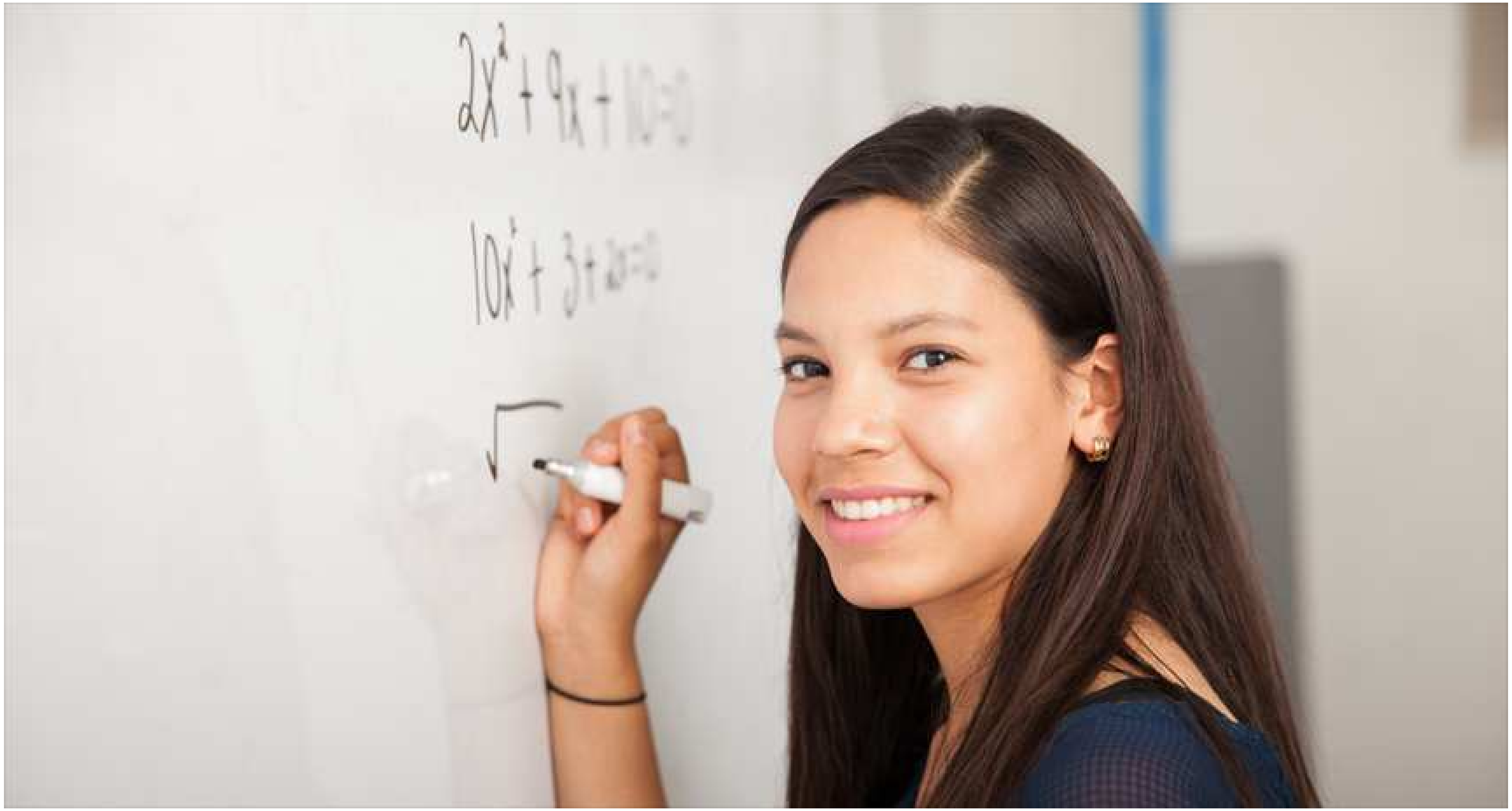

al-Khwarizmi

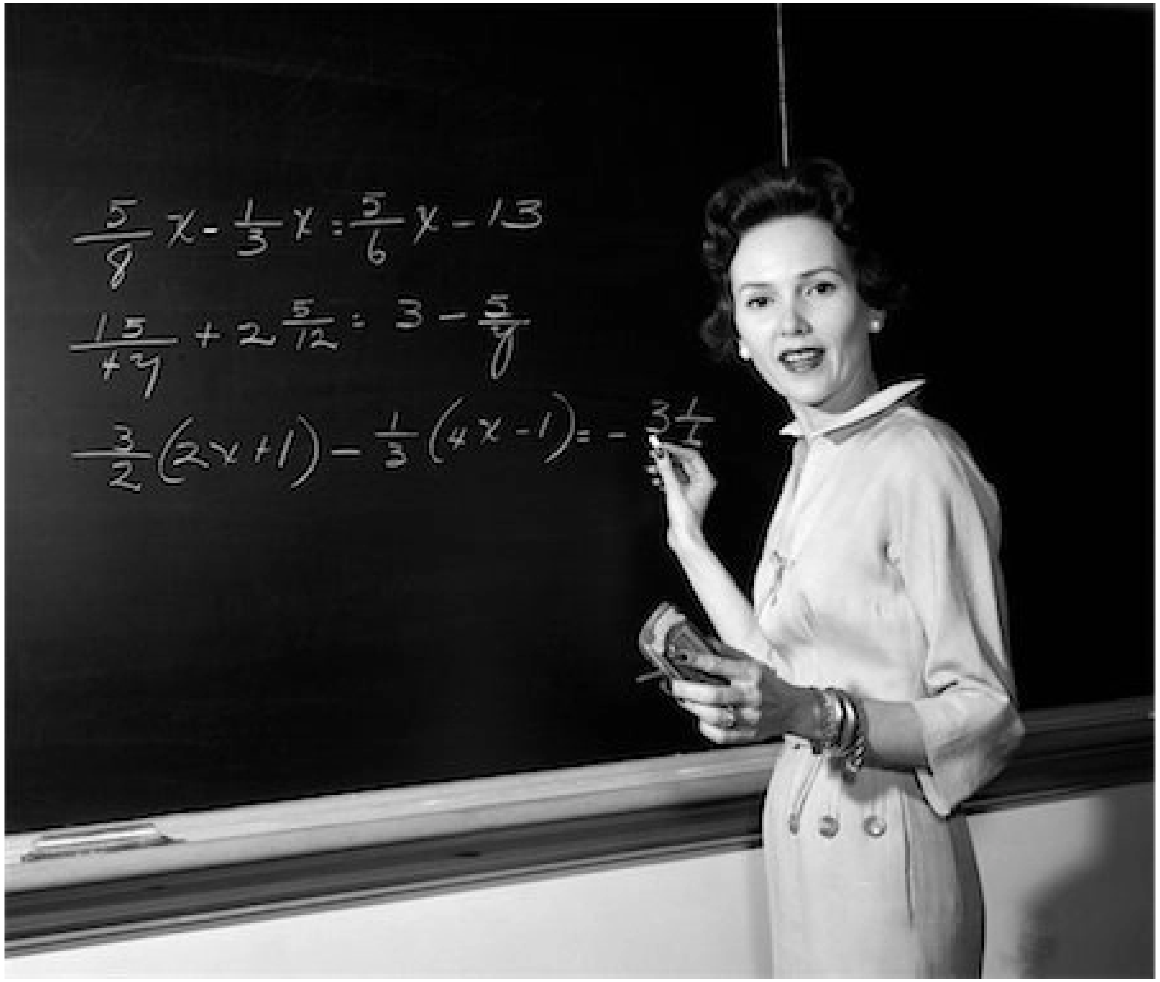
We are so used to reading, writing, calculating, and working with numbers, from very young age, that we tend to associate them with something ‘divine’ or ‘absolute’.


01234
56789


83+38=121
56+78
68+52=


1+1=2
3+4=7
6+6=12
3+7=10
4+8=12
6+7=13
3×2=6
6×7=42
4×4=16
8×3=24
3+x=
8+x=
x+1=6
4+x=7
8-x=5


$2x^2 + 9x + 10 = 0$


5/8 x - 1/3 x = 5/6 x - 13

This is why we tend to believe that our numbers are the ‘natural’ and the ‘only’ proper way to express quantities.

Other number systems seem ‘funny’ and ‘game-like’, or appear only as ‘intellectual exercise’.

We need to break out of this mathematical orthodoxy and dogma.

“STOP!”

“THIS IS HERESY!”

“Thou shall praise and respect our splendid and divine number system each and every day of your life! ”

But in reality our number system has **no** such divine mathematical aura!

Hence Benford's Law, being so intimately involved with our number system that it is actually stated in terms of its symbols (digits), is **arbitrary** just as well!

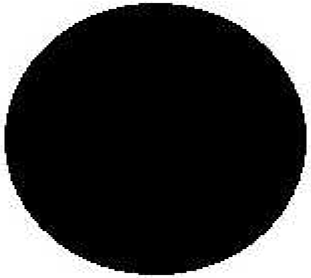
This realization leads one to suspect that $LOG_{10}(1+1/d)$ for the 1st digits does not account for the full story of the phenomenon, and that there exists possibly a more universal and non-arbitrary law.

Let us summarize:

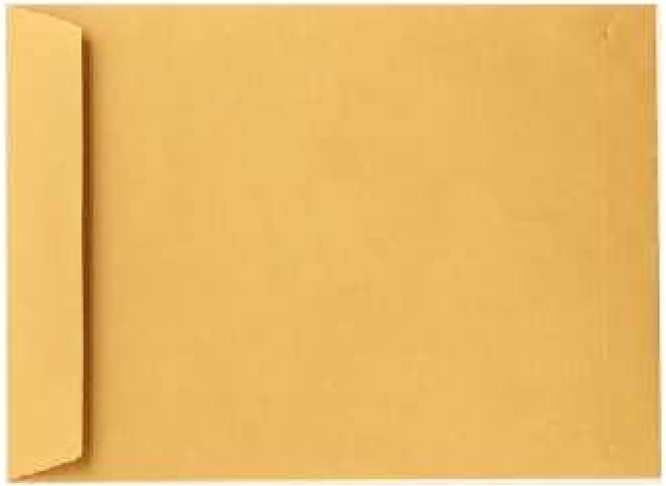
What's wrong with Benford's Law?

We place the real quantities in the physical world into arbitrary and artificial envelops (digital symbols**), and then we insist on counting those envelops, looking for patterns in the envelops – namely: Benford's Law!?**

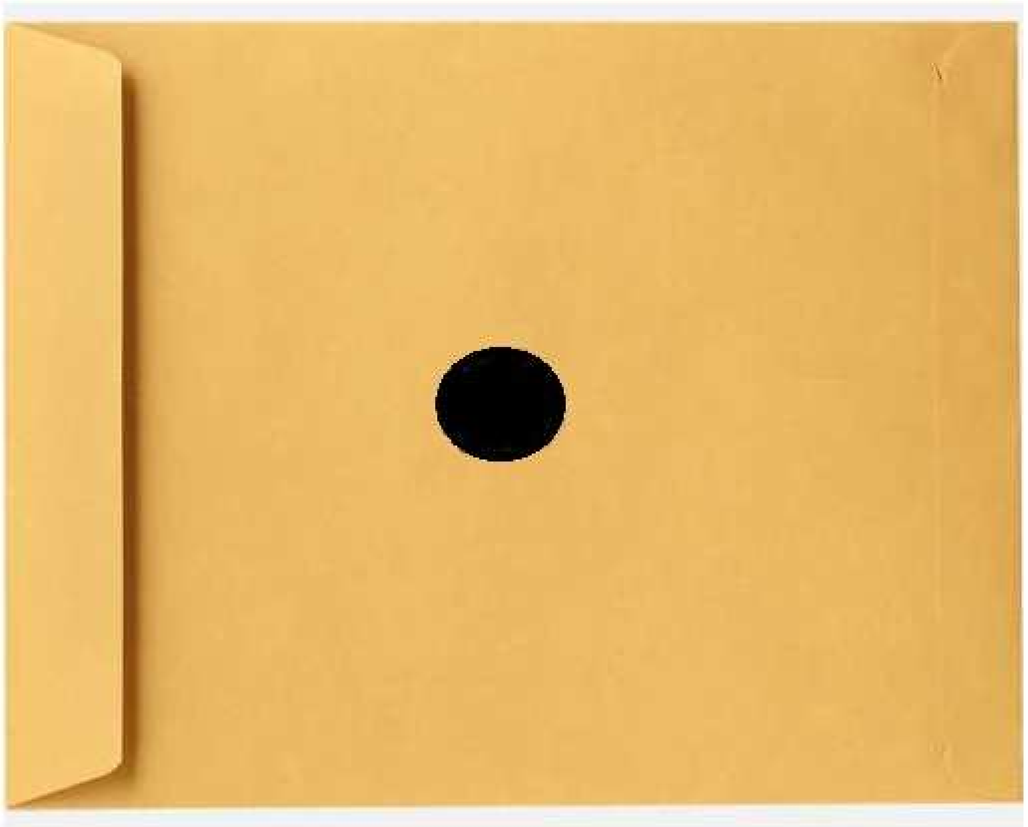
quantity



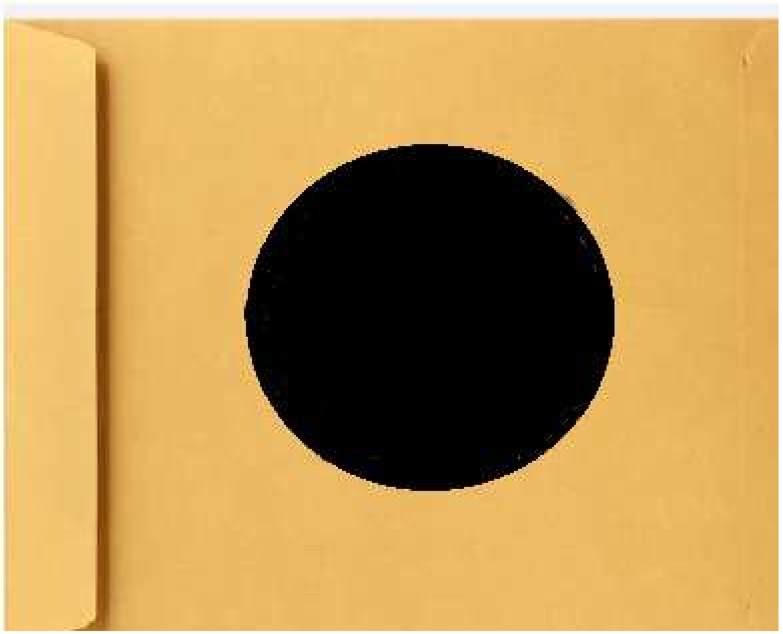
Digital representation



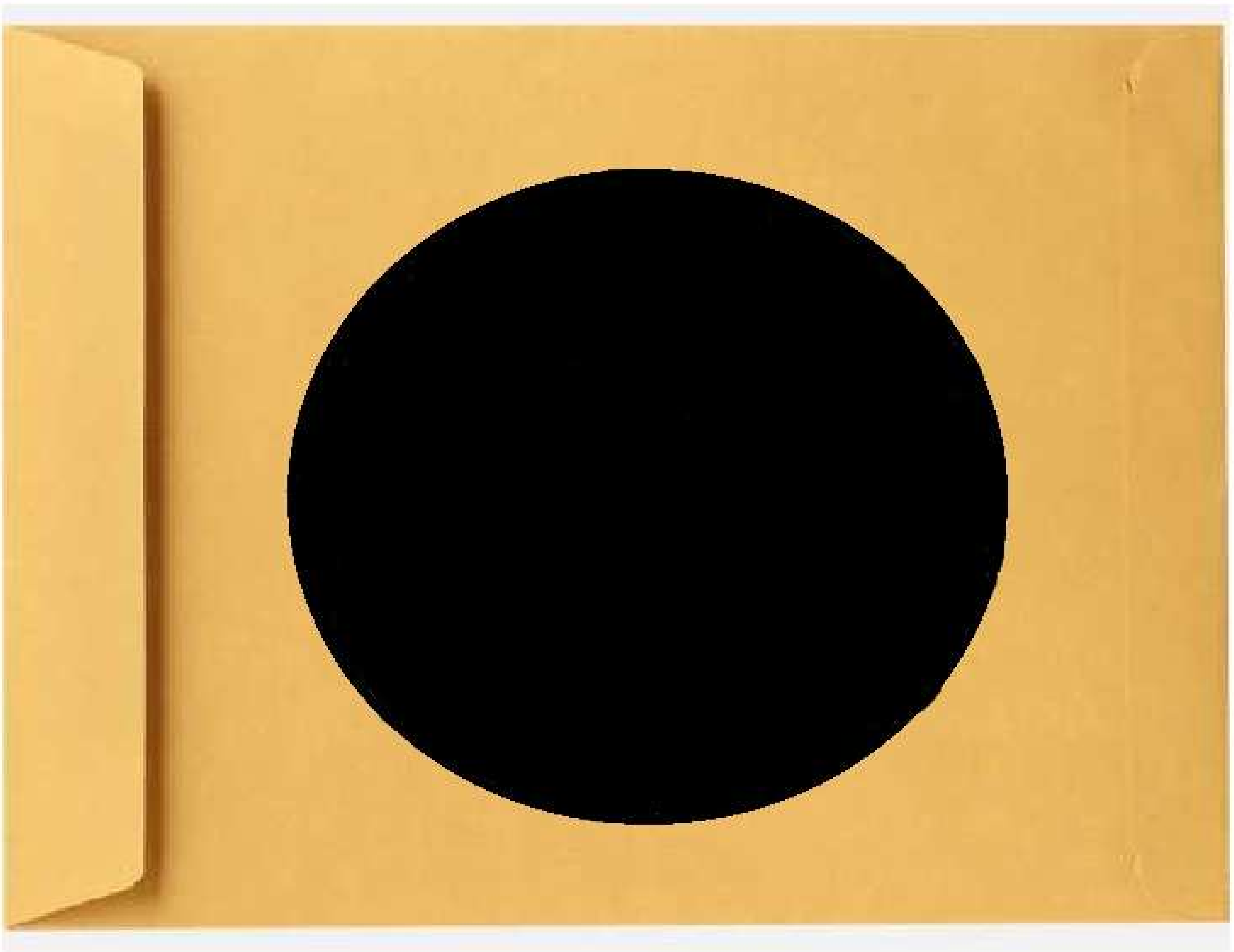
An envelop for a quantity:



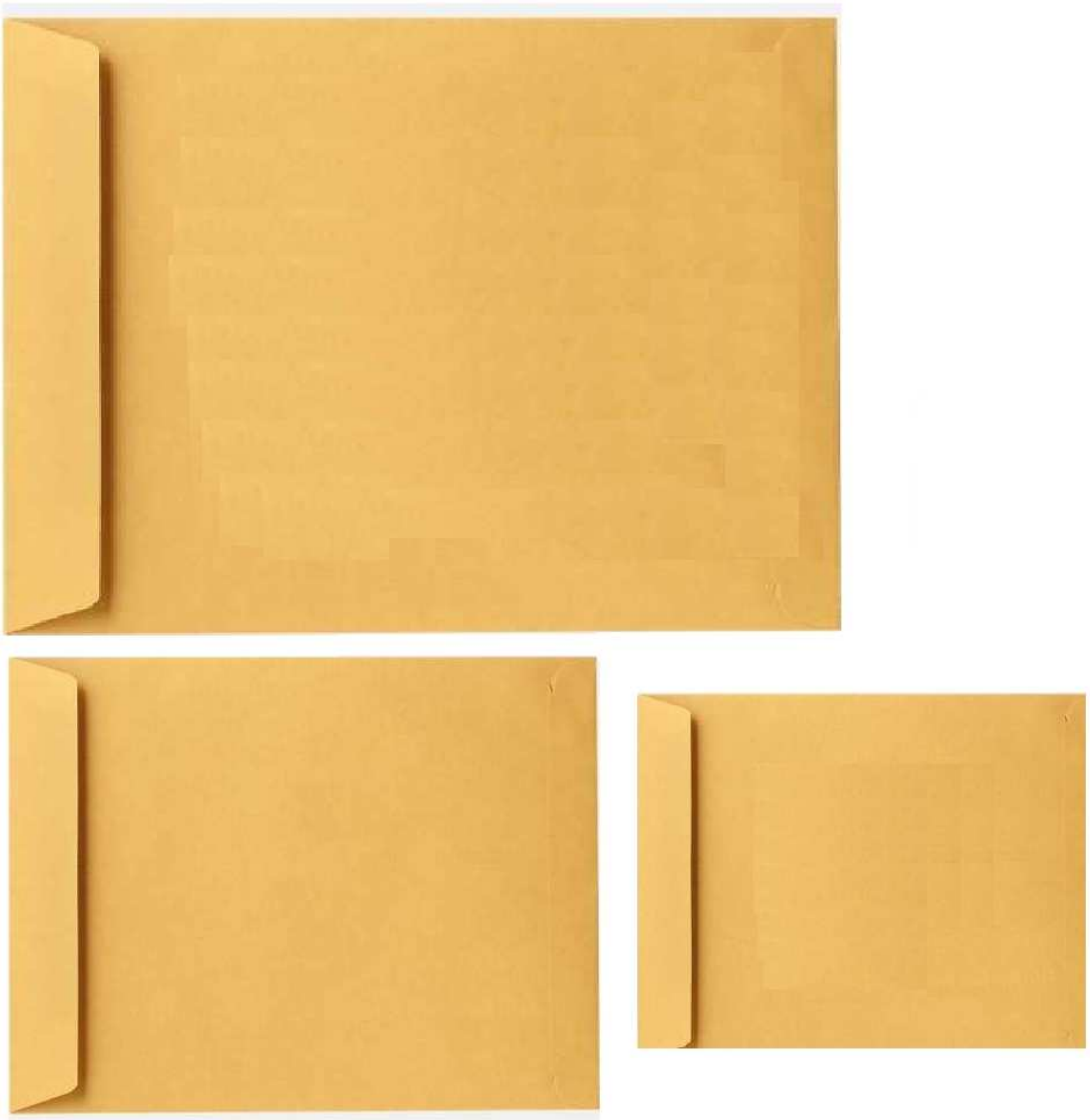
An envelop for a quantity:



An envelop for a quantity:



Benford's Law merely counts these envelops:

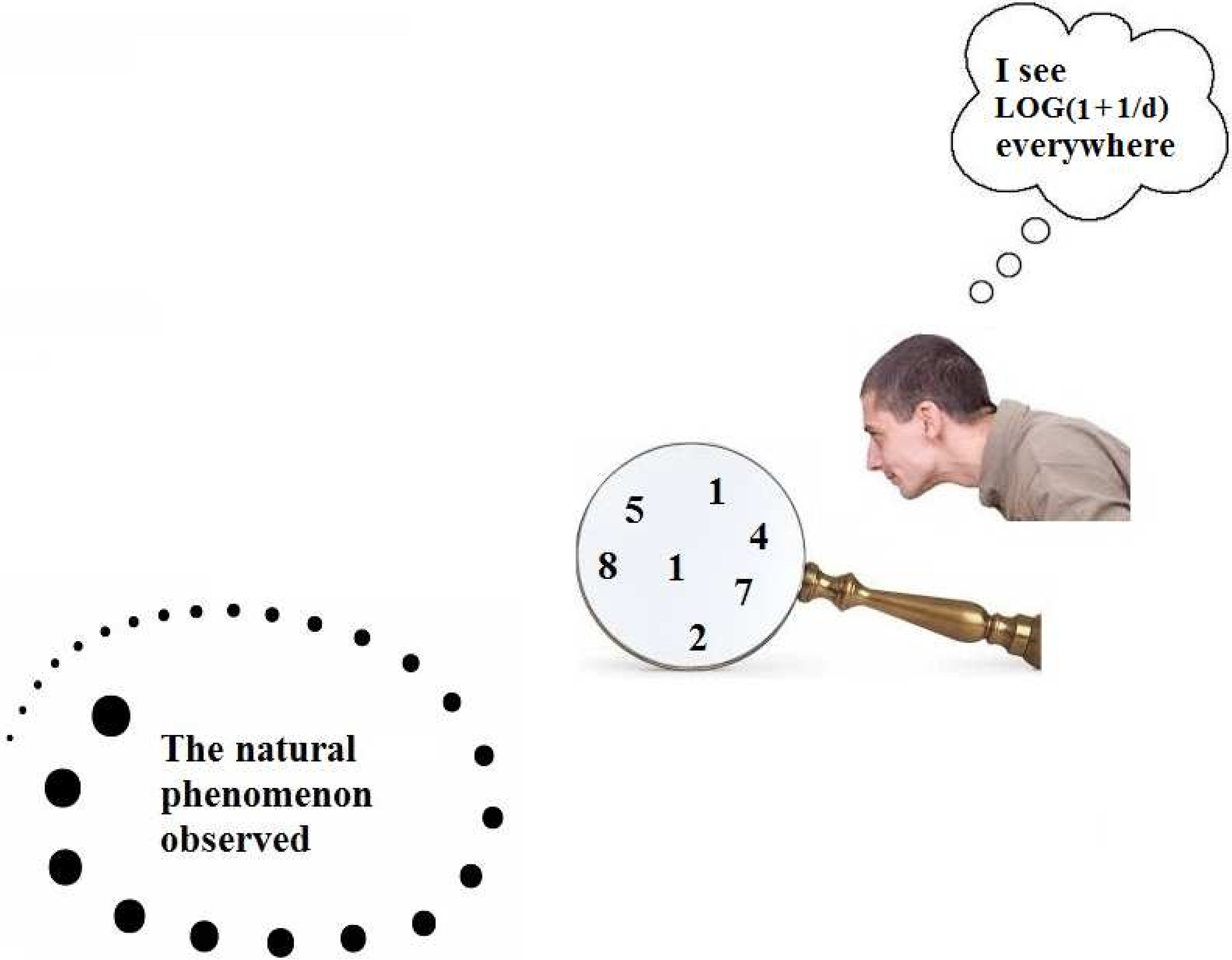


Physical Reality Versus Digital Perception

Benford's Law is highly prevalent in the physical world.

But first everything has to be recorded in our positional number system; then data is converted into 1st digits; and then $LOG_{10}(1+1/d)$ is found!

Our digits serve as a lens of sorts.


I see
LOG(1+1/d)
everywhere
5 1 4 8 1 7 2
The natural
phenomenon
observed

Two radically different interpretations of the Benford phenomenon are given:

<u>First:</u> **REAL & PHYSICAL**

<u>Second:</u> **ILLUSIVE & NUMERICAL**

Two radically different **interpretations** of the Benford phenomenon are given:

First:

This is truly a physical phenomenon existing independently of us and our way of recording data.

It is a physical law of nature.

Second:

This digital pattern found in physical data is simply due to our own peculiar way of counting values by way of their digital representations,

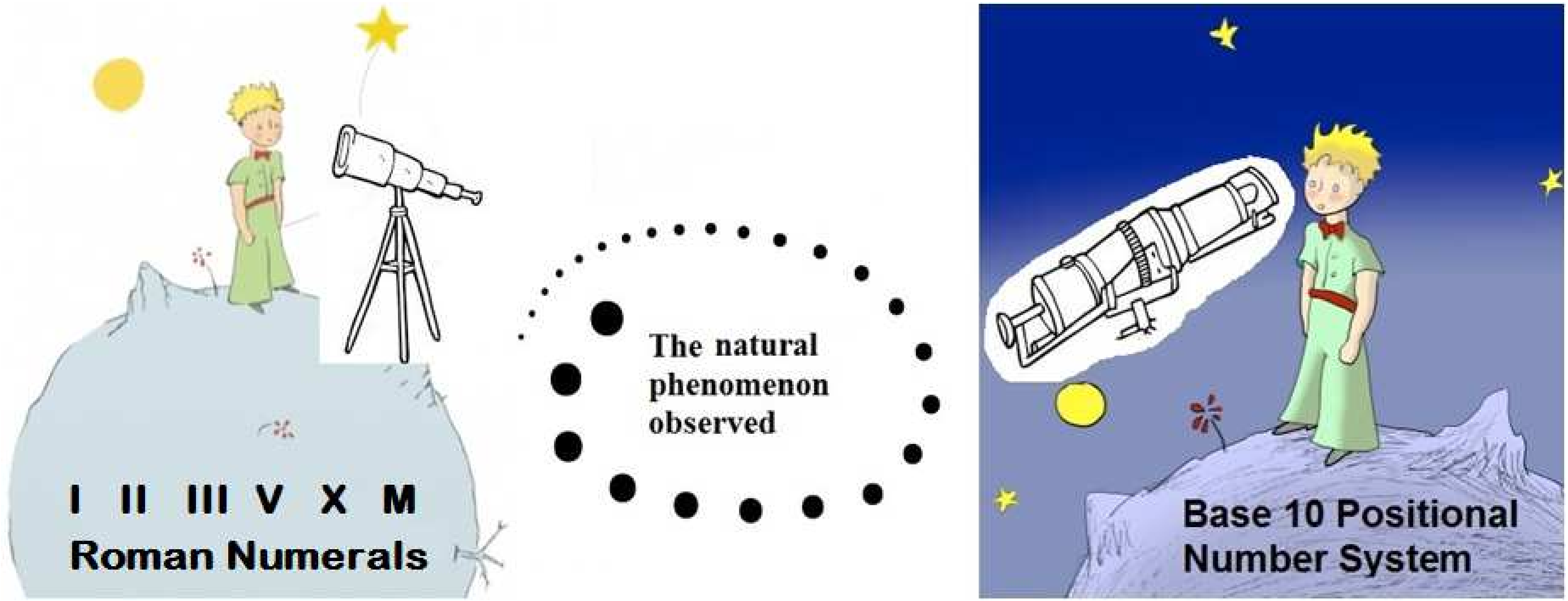
The phenomenon has NO independent physical existence outside our digital perception.

As an analogy for the *second interpretation*, a child wearing red eyeglasses may believe that every physical object in the world is red.

“Daddy, how come everything in the world is red?”

The red color on her eye glasses is arbitrary, and that's why the fact that everything appears red is arbitrary as well.

Had she been wearing green eye glasses, everything would then appear green.



"I can **not** observe any pattern in the data! Could anybody help me construct a measure such that all observers would agree upon? But I **refuse** to adapt another number system, I am emotionally attached to mine.

"I observe Benford's Law very clearly, 1^{st} digits are as in $LOG_{10}(1+1/d)$!"

It is necessary that they should all come up with a universal and primitive statistical measure agreed by all observers for this clearly and easily observable physical phenomenon.

In other words, that a singular quantitative statement should be formulated which would be identical for all planetary observers, being number system invariant.

And that singular quantitative statement is:

GLORQ !

The General Law of Relative Quantities

G.L.O.R.Q. (acronym)

THE IDEA: That universal and primitive measure to be agreed on by all planetary observers could be a mathematical expression relating to the commonly observed **histogram** of the data in question (as this shall be shown soon to be of such universal character).

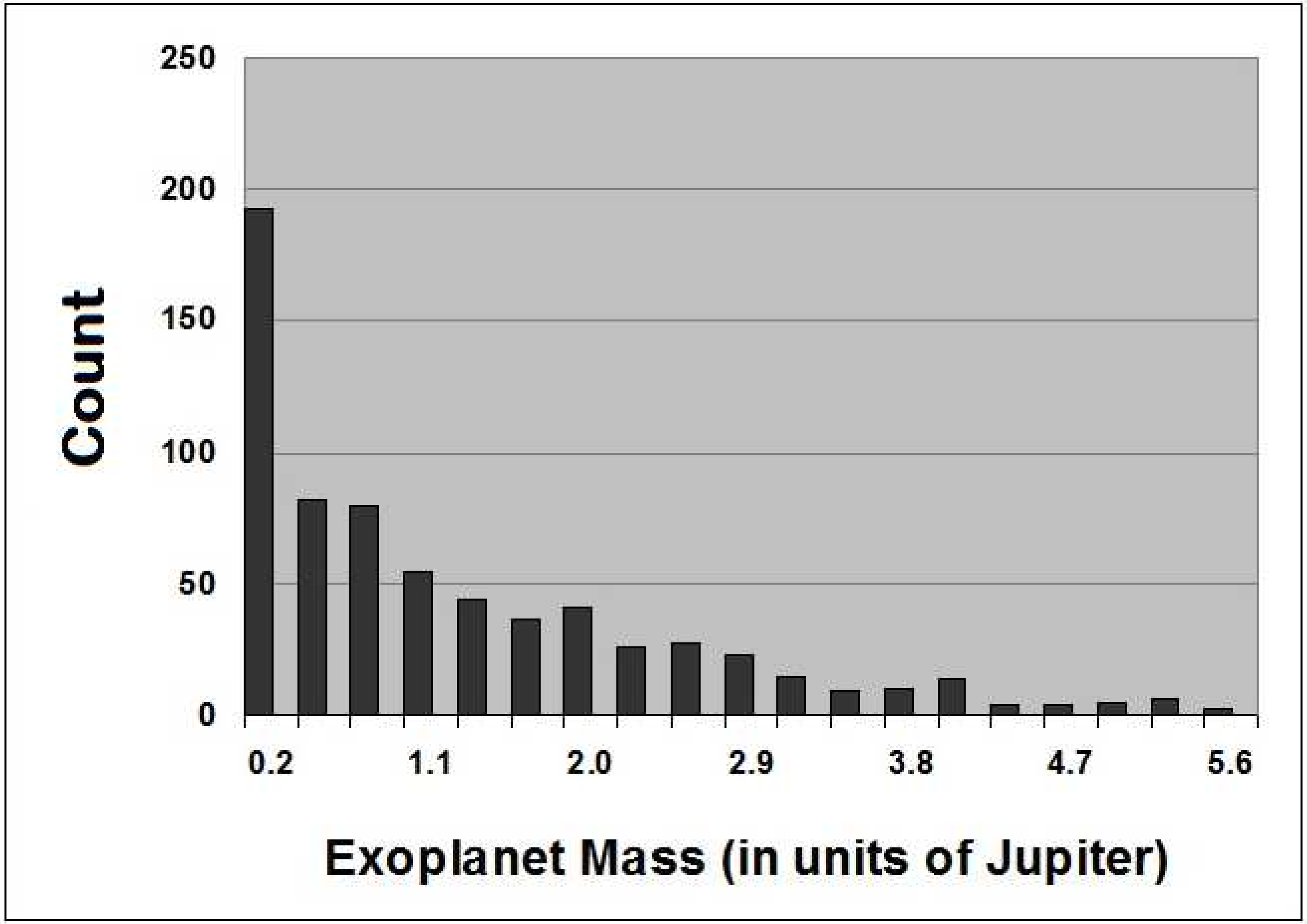
But **what aspect** of histograms could it be?

One characteristic common to all Benford obeying data sets is their overall skewed histogram falling on the right.

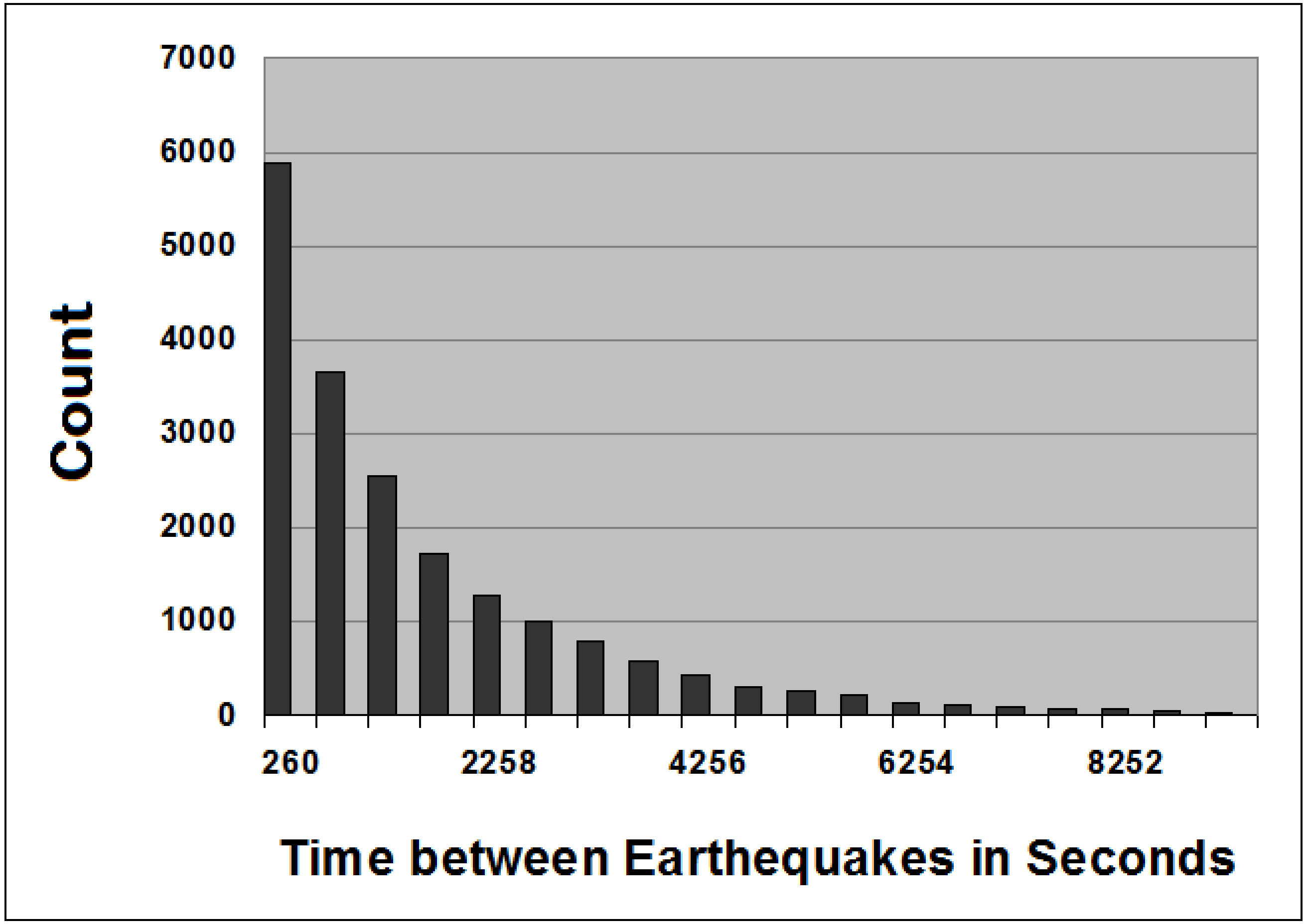
This implies having many small values, but only very few big ones.

This is a nearly universal feature in random data, being number system invariant. Therefore, a precise quantitative measure of such a fall in histograms may serve as a general law.

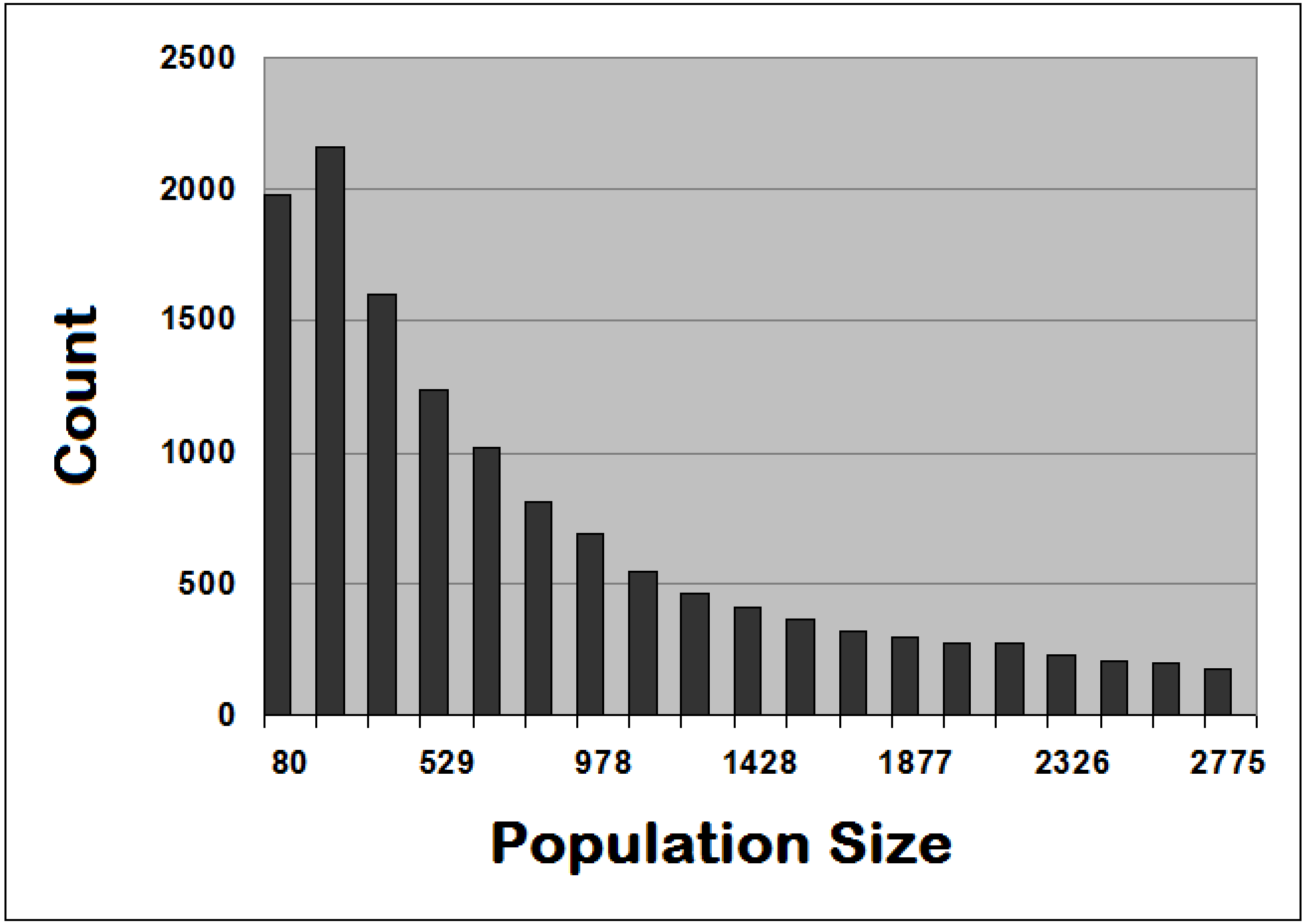
Histogram of the Mass of 800 Known Exoplanets



Histogram of Time between all 19,452 Earthquakes Occurring in 2012



Histogram of USA Population for all its 19,509 Cities and Towns in 2009



Indeed,

small is numerous,

big is rare.

Hence let’s change the agenda:

Instead of **digits**, let’s focus on **histograms** of data sets, and their quantitative structure preferring the small over the big.

===============================

Do histograms depend on the number system in use?

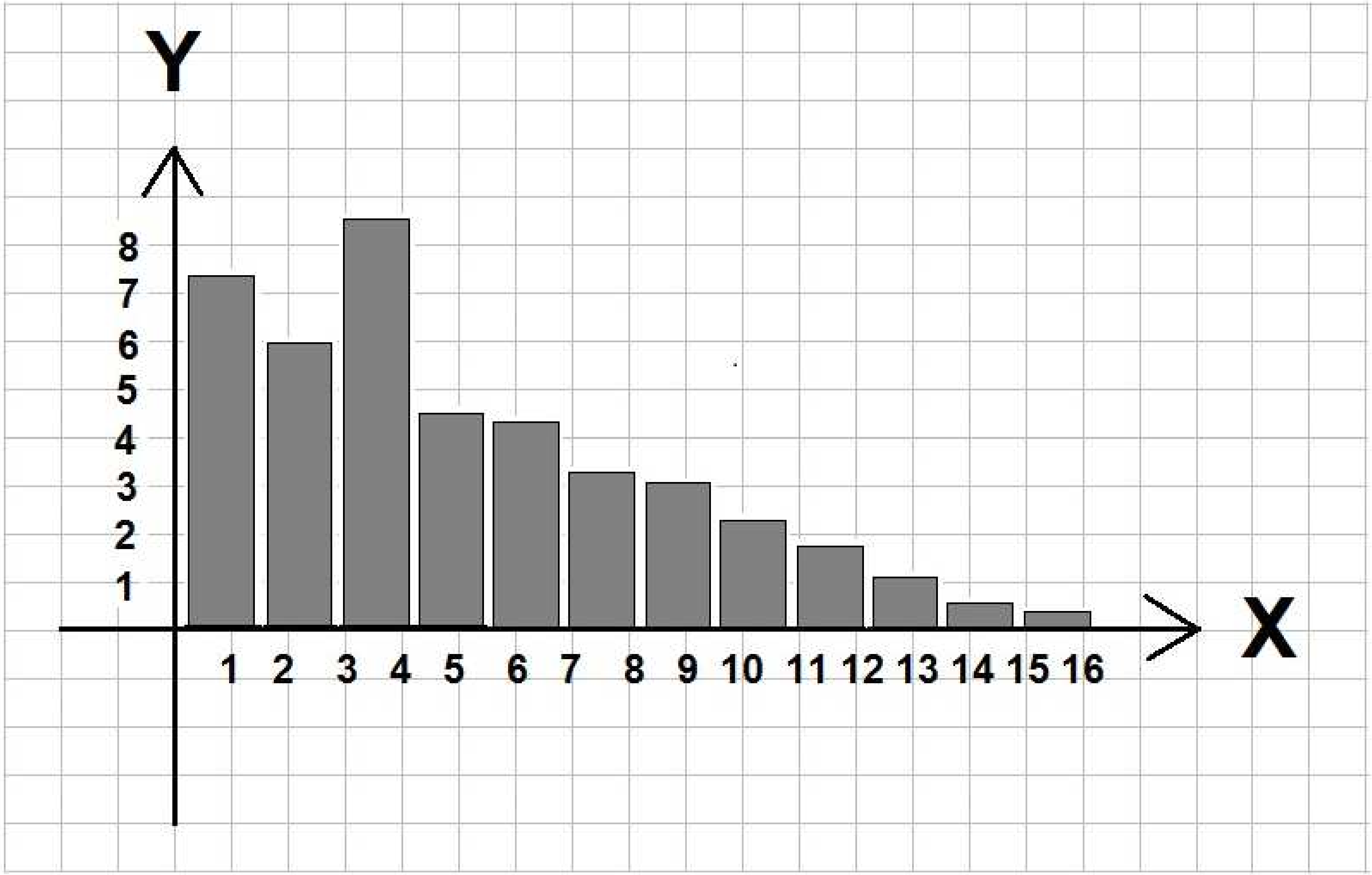
What happens to a histogram when we switch to another number system?

Does the visual picture of a histogram change?

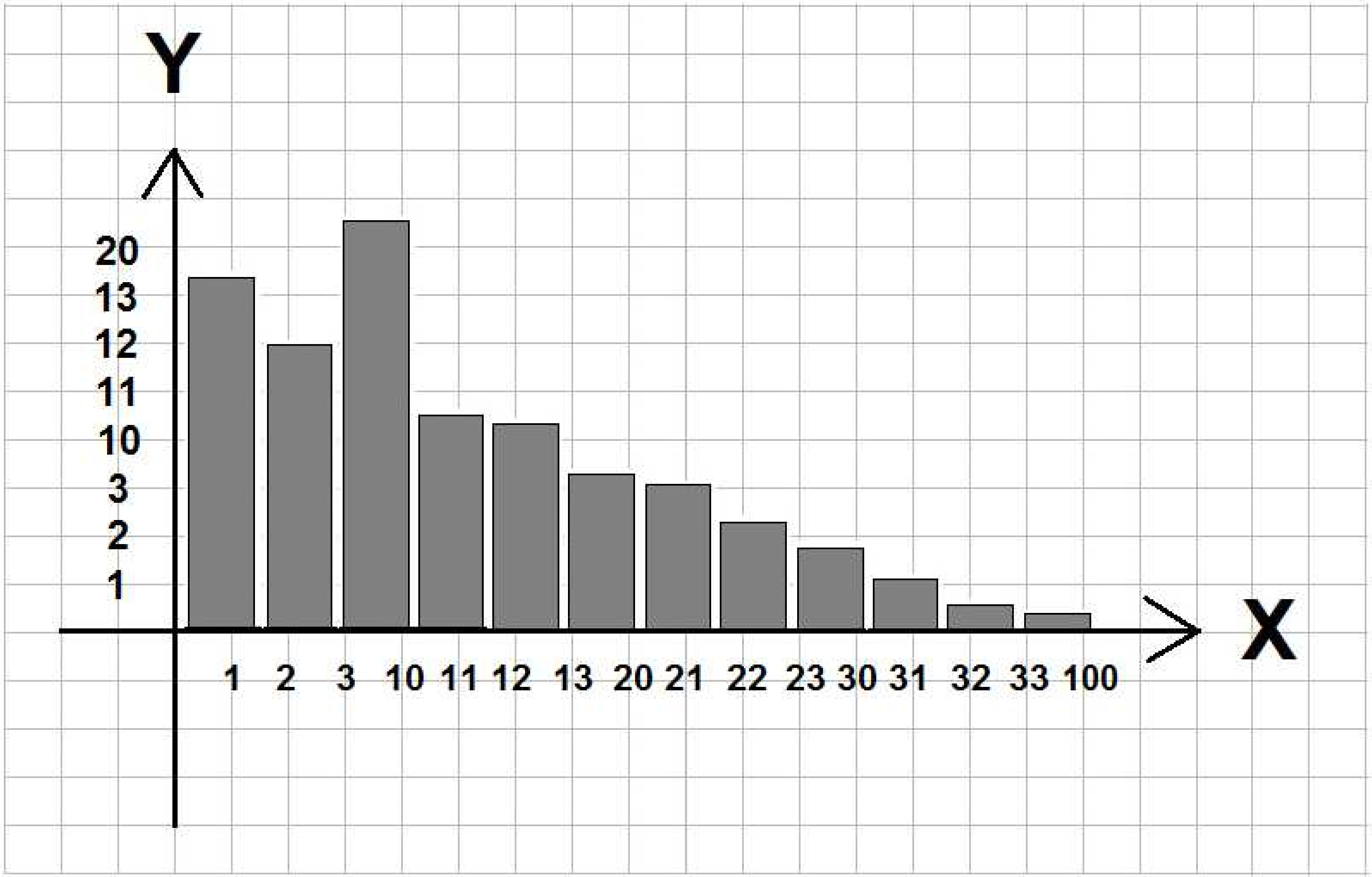
NO!

Here is the histogram of a singular real-life data set viewed through the prism of several number systems. Clearly, its visual aspect, the relative sizes of the bins, its shape, etc., are fixed (invariant).

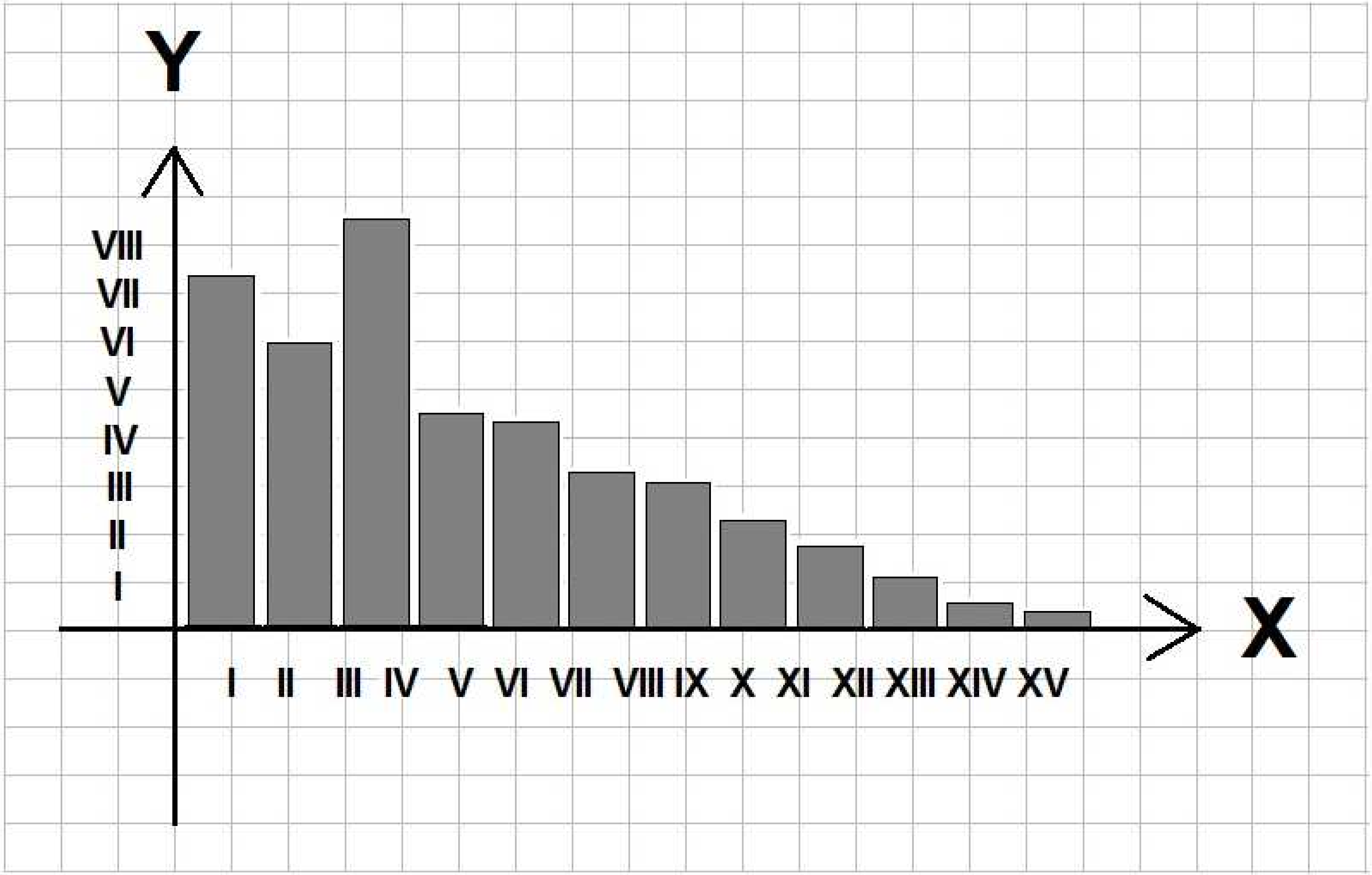
Positional Number System Base 10



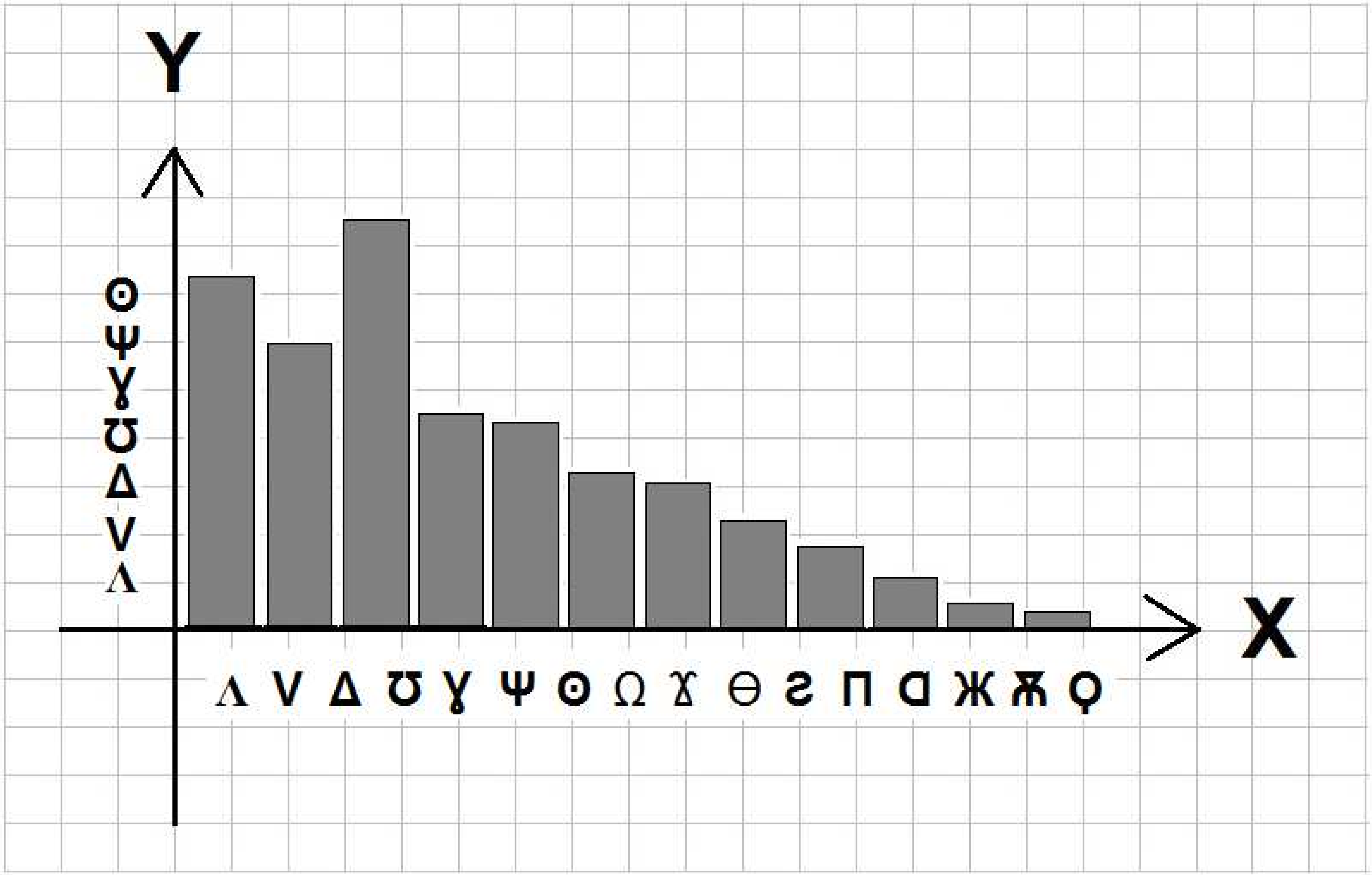
Positional Number System Base 4



Roman Numerals



Absent a Number System



Clearly, the message conveyed in a given histogram is universal, irrespective of the number system is use!

Histogram Invariance Principle

Since statistical distributions (PDF) are simply the continuous forms of discrete histograms (infinity refined), the principle is very general:

PDF - density distributions are number-system invariant!

===============================

The **‘histogram vista’** of Benford’s Law:

all numbers from **1 to 2** such as:
1.00, 1.15, 1.49, 1.76, 1.93, 1.97, 1.99
are with **first digit 1**.

all numbers from **10 to 20** such as
10.0, 13.8, 15.2, 16.8, 18.2, 18.8, 19.6,
are with **first digit 1**.

all numbers from **100 to 200** such as
100, 123, 141, 165, 176, 195, 197, 198
are with **first digit 1**.

Digit 1 leads on these sub-intervals:

etc. …

[1, 2),

[10, 20),

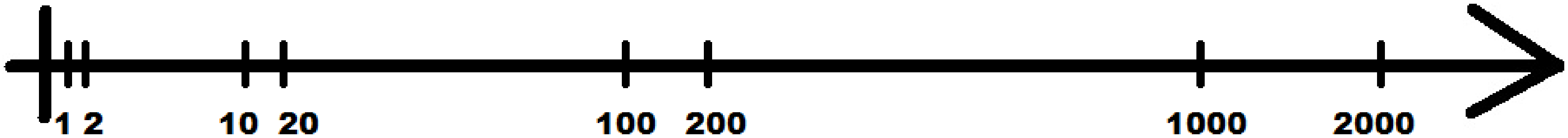
[100, 200) ,

[1000, 2000).

[10000, 20000).

…etc.

Digit 1 leads on these sub-intervals:



and these segments are **expanding** on the x-axis

The **'histogram vista'** of Benford's Law is:

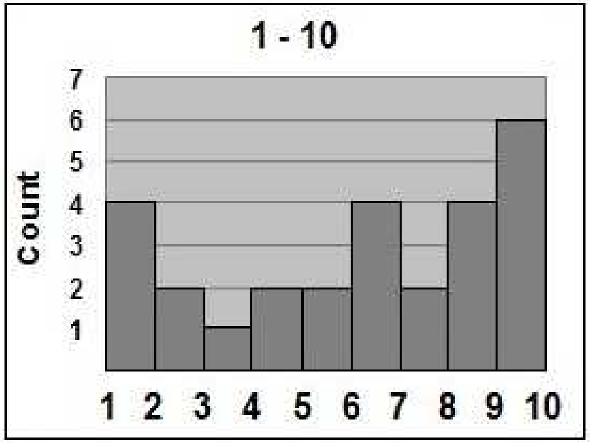
An infinite system of **9-bin** histograms, expanding by an **inflation factor of 10**, and all aggregated into a singular overall set of proportions.

For example:

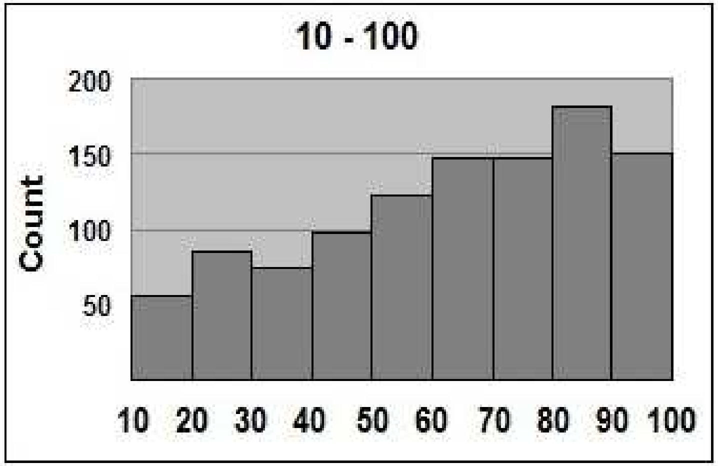
Data on USA population regarding all its 19,509 cities and towns in the 2009 census survey:

From :	1	10	100	1000	10000	100000	All Data	All Data
Up to:	10	100	1000	10000	100000	1000000	Count	Proportion
Bin 1	4	56	1565	2718	1222	168	5733	29.4%
Bin 2	2	86	1429	1437	537	47	3538	18.1%
Bin 3	1	75	1116	843	290	16	2341	12.0%
Bin 4	2	98	941	624	171	11	1847	9.5%
Bin 5	2	123	813	460	153	8	1559	8.0%
Bin 6	4	148	721	388	101	8	1370	7.0%
Bin 7	2	148	626	311	75	4	1166	6.0%
Bin 8	4	181	502	292	60	3	1042	5.3%
Bin 9	6	150	489	212	45	2	904	4.6%
Number of Cities	27	1065	8202	7285	2654	267	19500	19500
Data Proportion	0.1%	5%	42%	37%	14%	1.4%	100%	100%

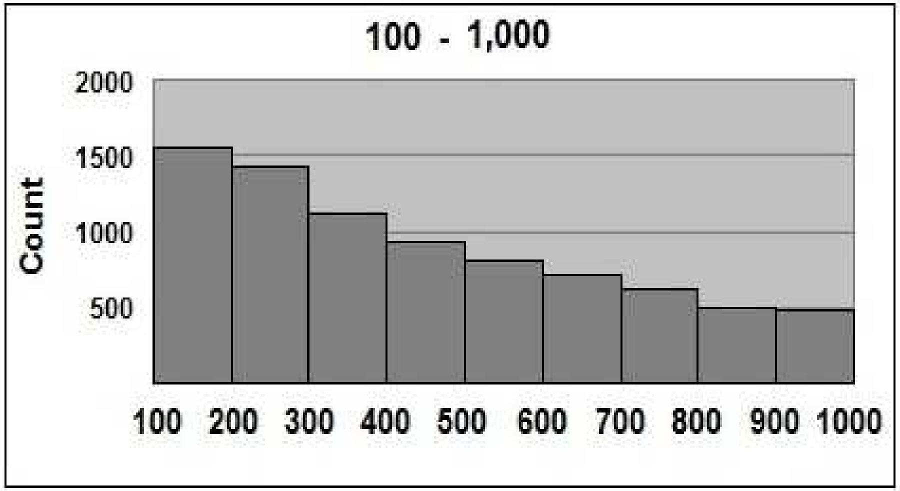
US Population on (1, 10)



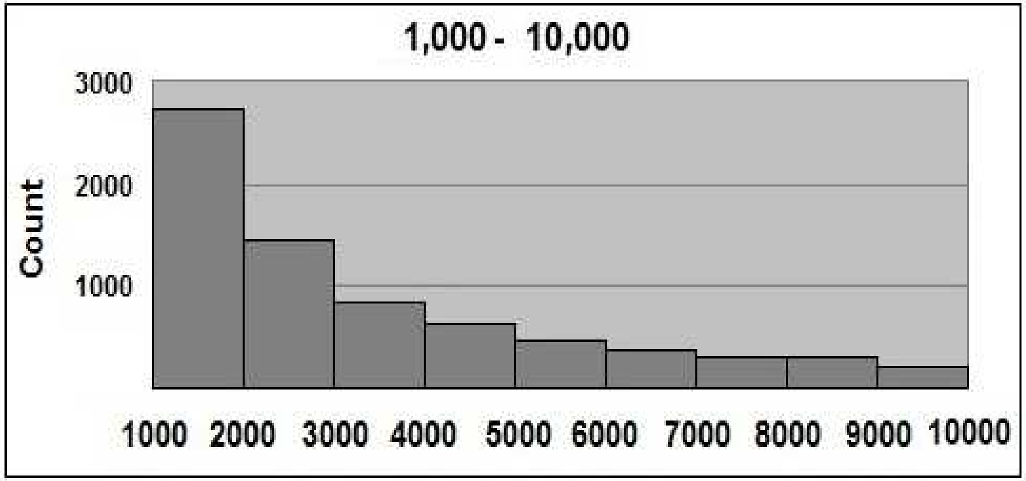
US Population on (10, 100)



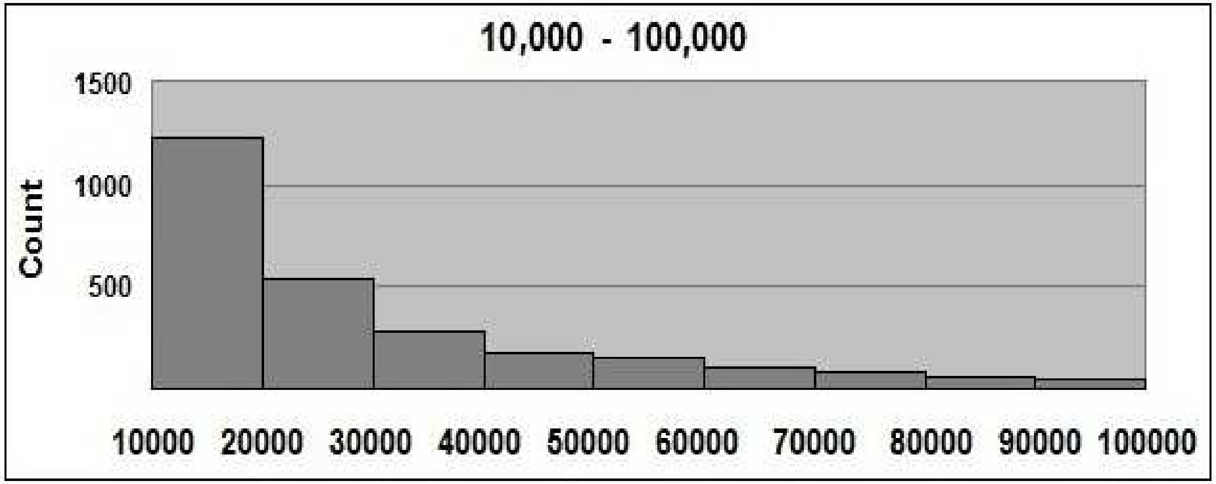
US Population on (100, 1000)



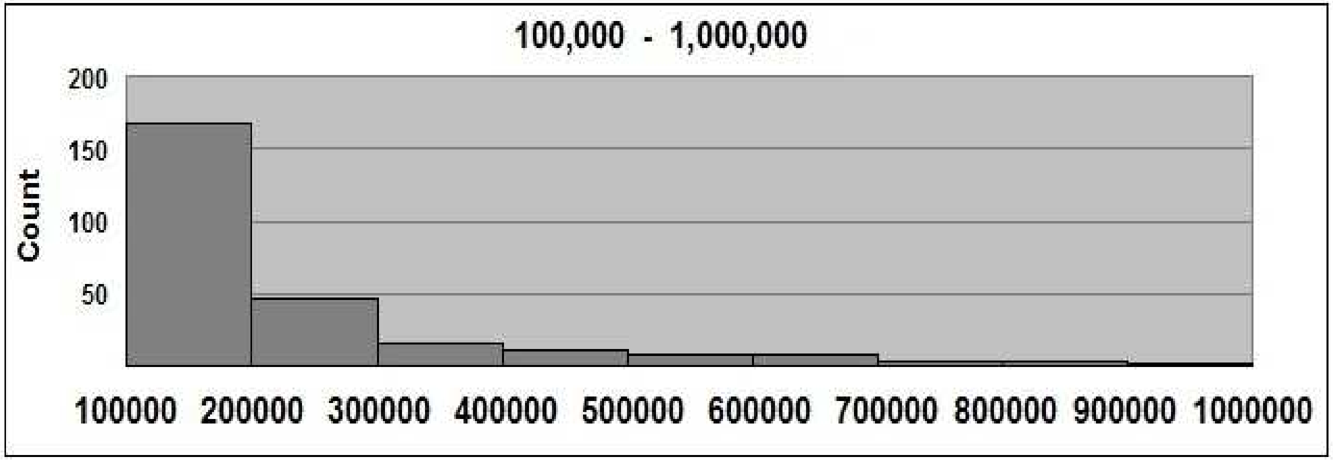
US Population on (1000, 10000)



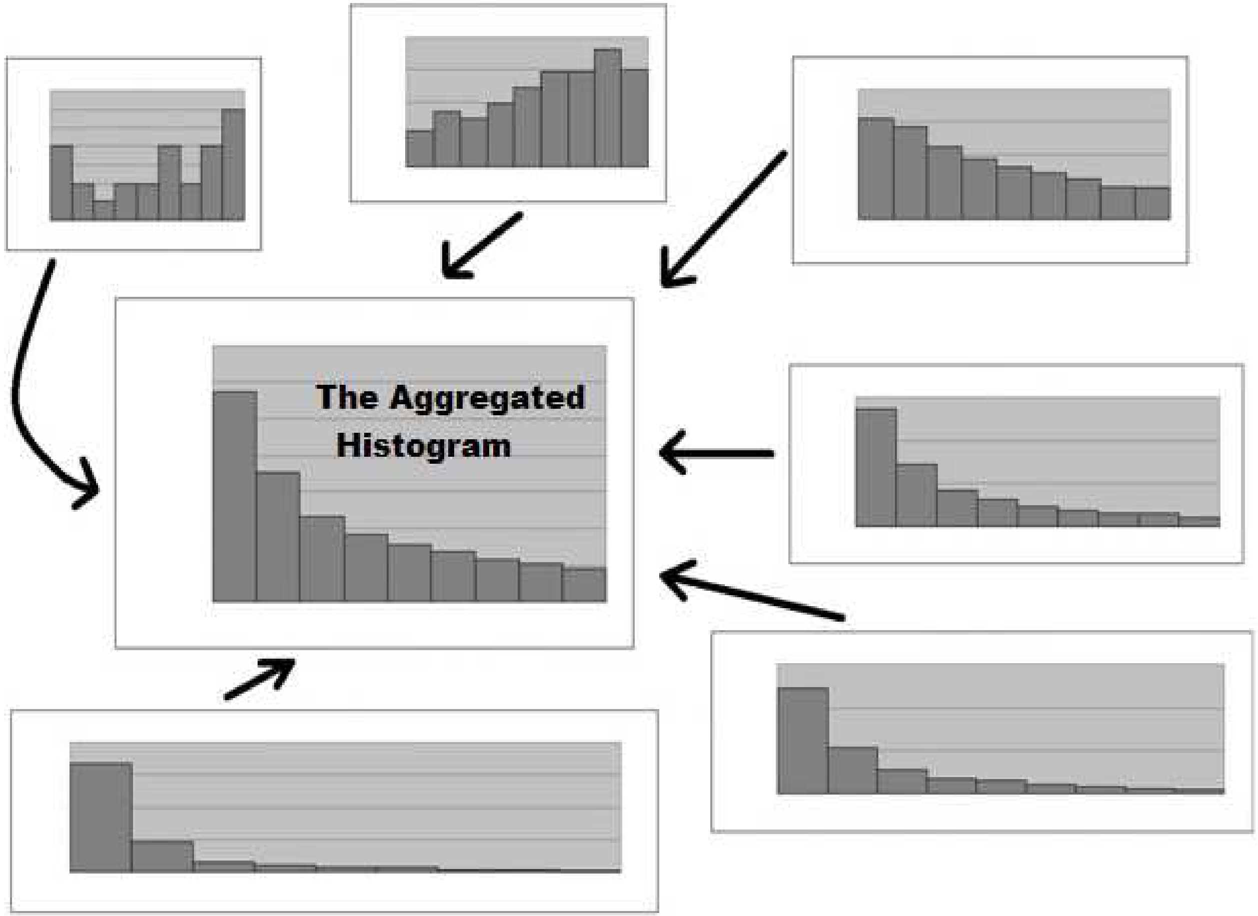
US Population on (10000, 100000)



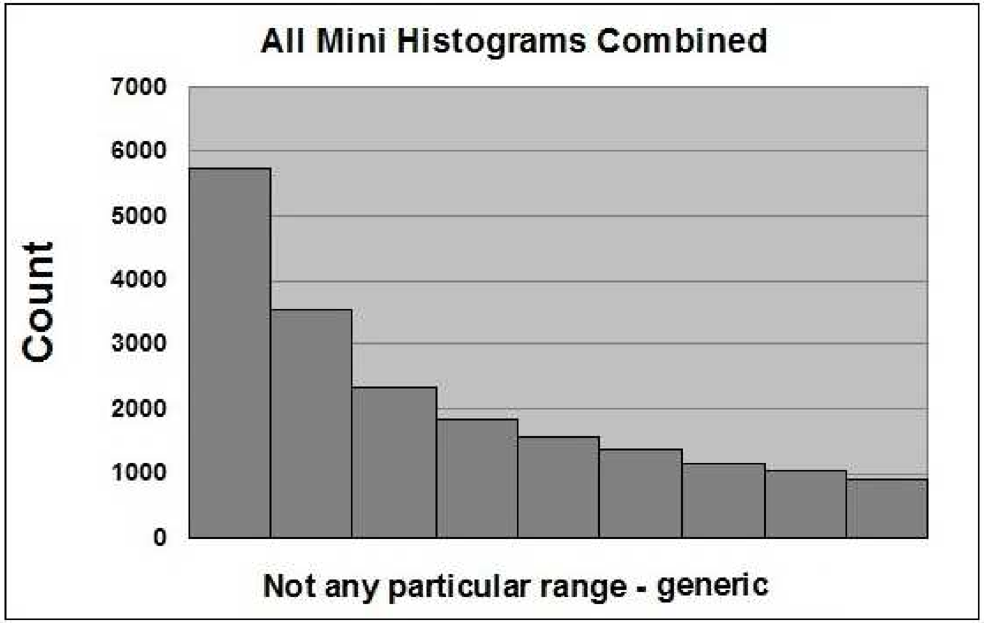
US Population on (100000, 1000000)



Fusing all the histograms into a singular aggregated "histogram" (bar chart):



Fusing these histograms together:



Digital distribution of any data set is nothing but the aggregated 'histogram' of the various 9-bin histograms, constantly expanding and inflating by a factor of 10, standing between 0.01, 0.1, 1, 10, 100, 1000, 10000, and so forth.

In the continuous case Benford's Law is nothing but the aggregated areas under the curve of the 9-sub-intervals standing between integral powers of ten such as 0.01, 0.1, 1, 10, 100, 1000, 10000, and so forth.

$$Prob(1st\ digit\ is\ d) = \sum_{int=-\infty}^{+\infty} \int_{(d)*10^{int}}^{(d+1)*10^{int}} f(x)dx$$

Int = the set of all the integers $\mathbb{Z}$

This vista begins to exonerate Benford's Law from the arbitrariness of our unique number system!

Benford's Law now begins to stand on a solid foundation!

Benford's Law now begins its journey of becoming independent of any number system!

Surely these histograms are deliberately constructed over a very particular partition of the entire x-axis range according to the cyclical way first digits occur in our number system, yet:

this is an exogenous issue!

We can now easily imitate the histogram structure within Benford's Law and generalize it and free it from our number system!

===============================

GLORQ

I. An infinite set of histograms

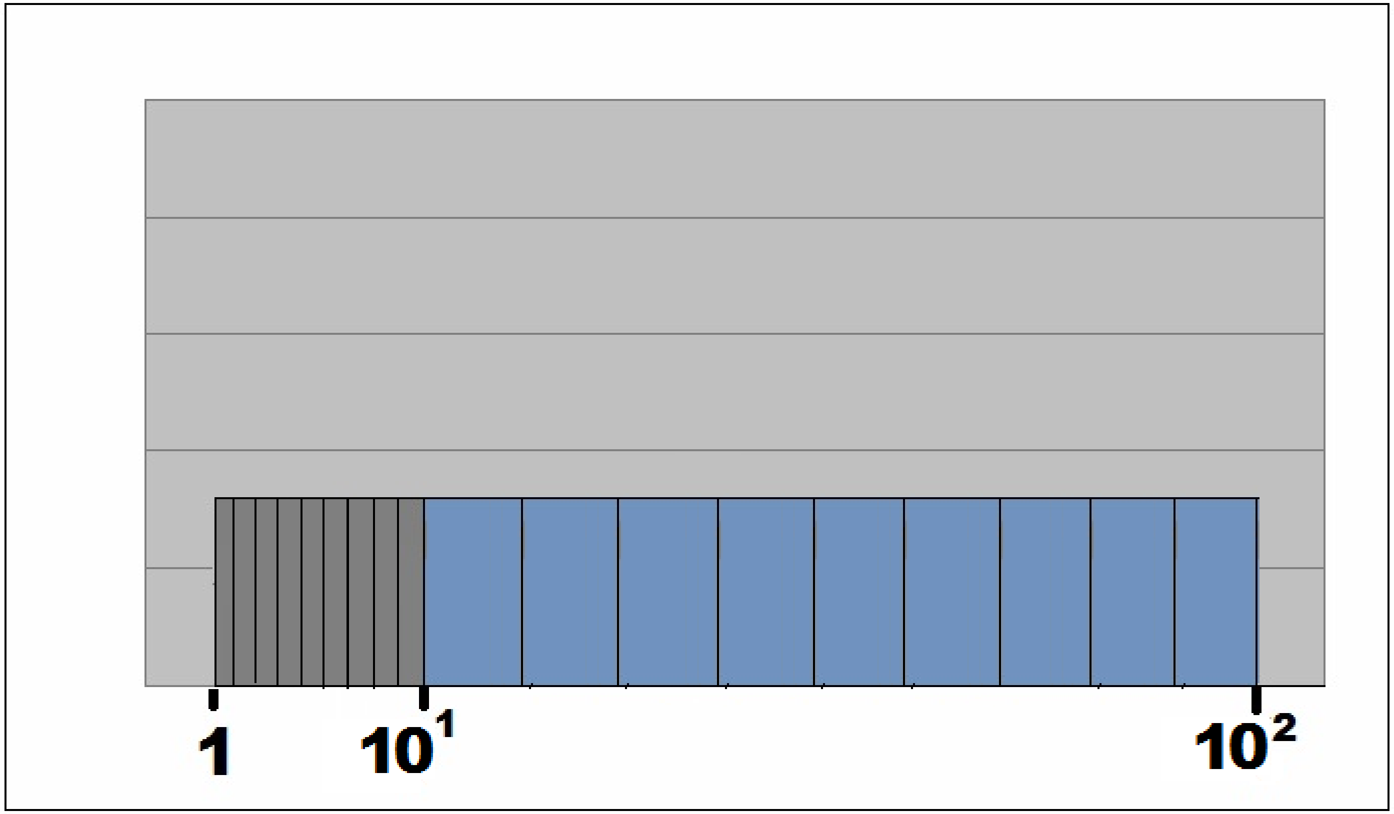
II. Each histogram is with **D** bins

III. All constantly expanding by inflation **F**

IV. No relationship exists between D and F and they are **free** to assume any value

Number System Base 10

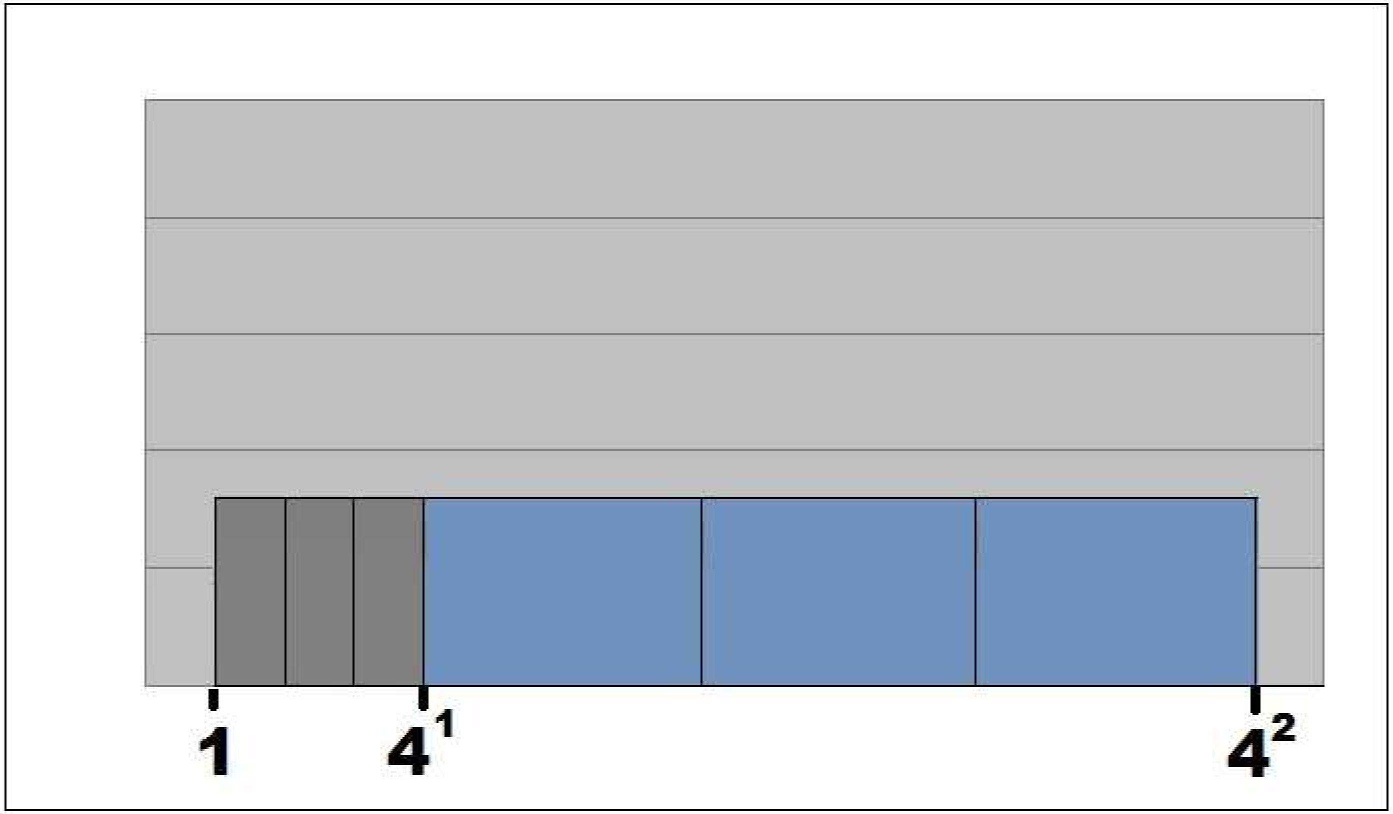
There are **D = 9** bins, expanding by a factor of **F = 10**



10 = 9 + 1

Number System Base 4

There are **D = 3** bins, expanding by a factor of **F = 4**.



4 = 3 + 1

Base **10** number system is a scheme of **10 = 9 + 1**

Base **4** number system is a scheme of **4 = 3 + 1**

Our number system is **restricted** to:

F = D + 1

Our number system is **restricted** to:

(Base) = (# of 1st Digits) + (1)

Number System

F = the base

D = the # of 1st digits

GLORQ

F = the inflation factor

D = number of bins in each histogram

For GLORQ, there is **no reason** whatsoever we should restrict **D** and **F** as such, hence:

(Inflation F) ≠ (D # of Bins) + (1)

(Inflation F)   and   (D # of Bins)

are two independent values

without any strict relationship

let us free ourselves of our number system!

Let us be totally flexible in how we choose D and F!

Let's try any D and F combination!

For example, let us consider:

F = D + 8

F = D + 2

F = D – 5

F = 7*D

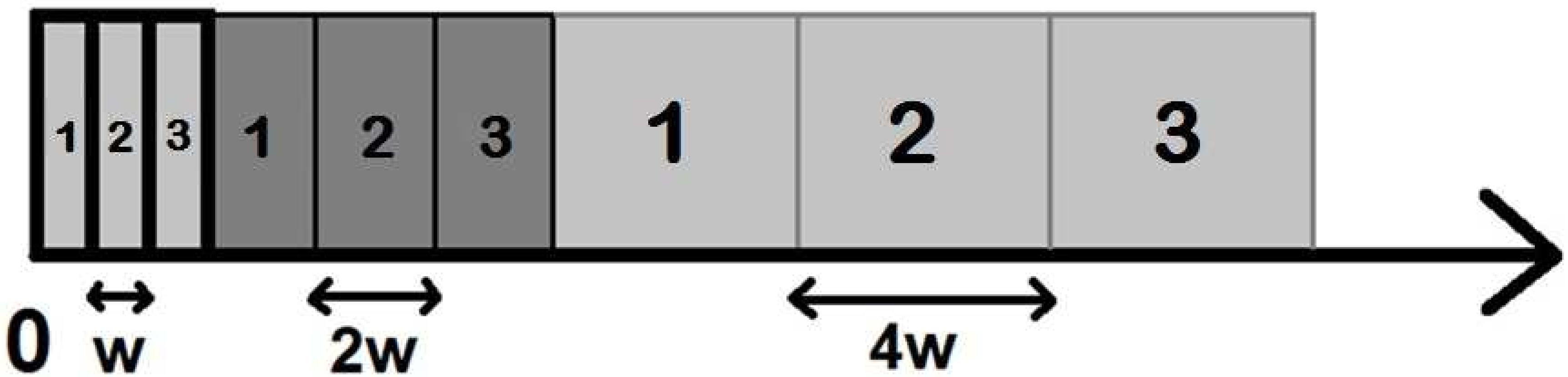
F = Any Arbitrary Number

We begin on the left from the 0 origin with an infinitesimally small **bin width** called **w** (approaching zero width in a limiting sense).

w —> 0

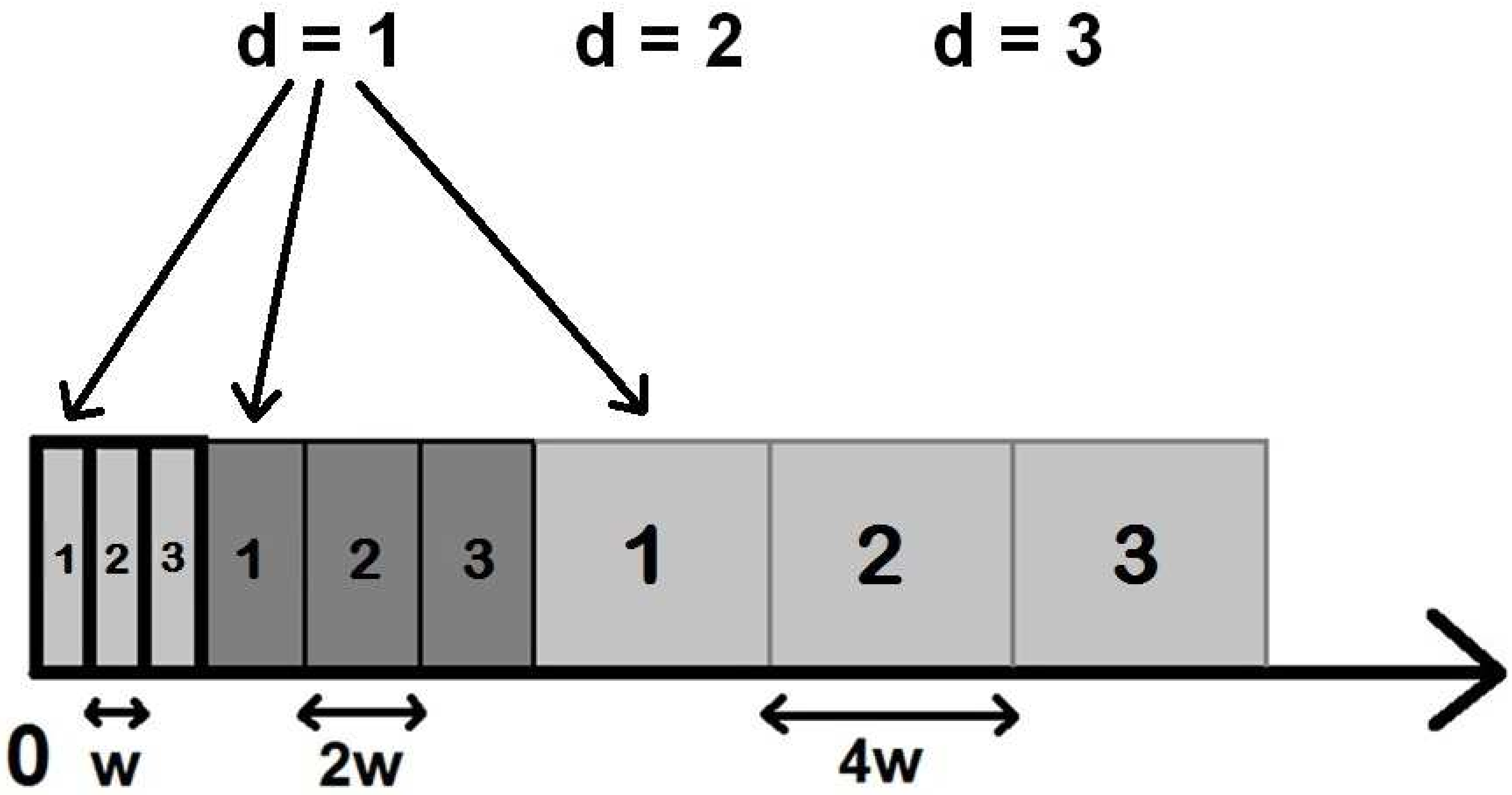
For example:

D = 3 F = 2

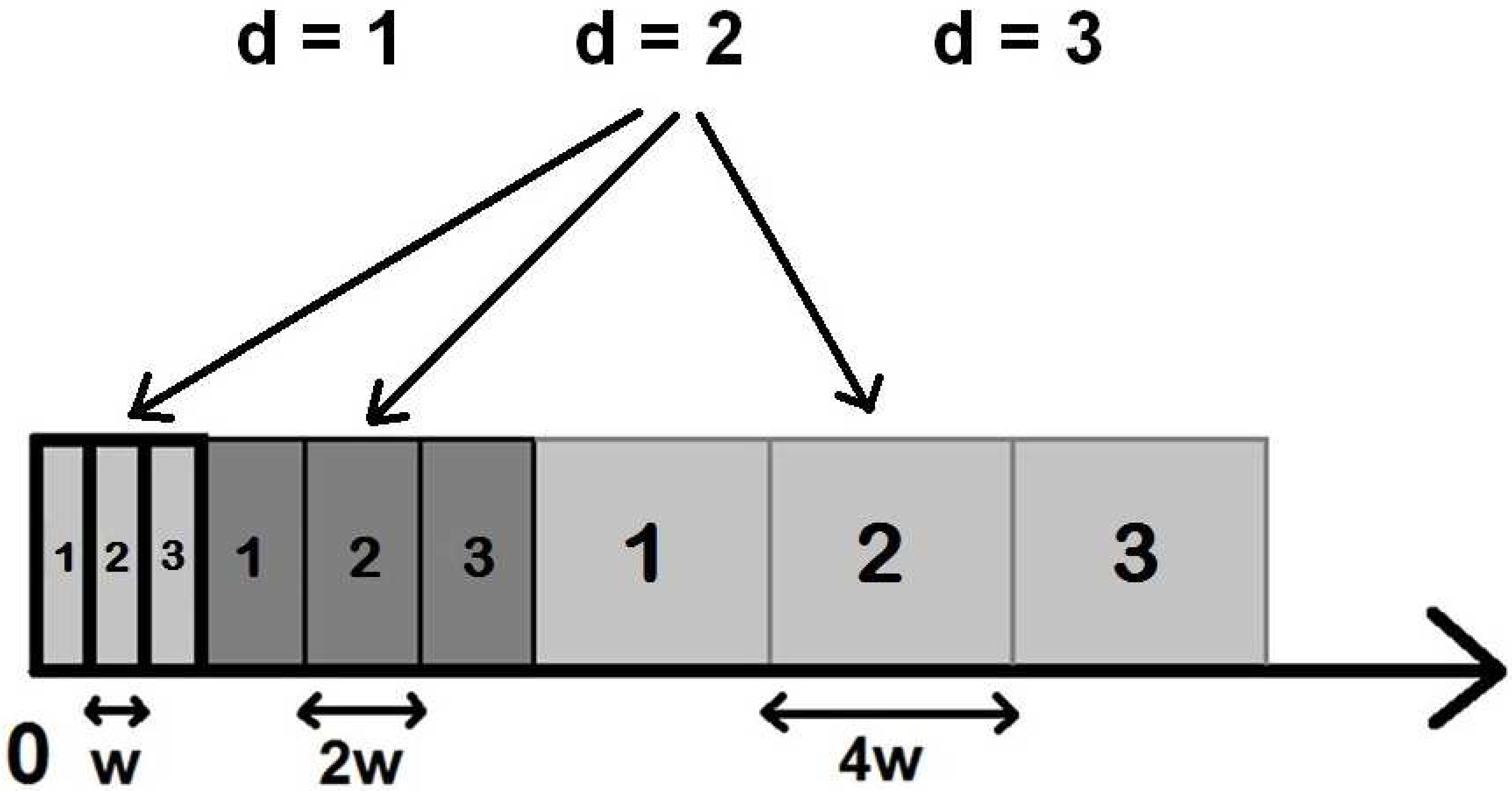


An infinite set of 3-bin histograms expanding by an inflation factor of 2.

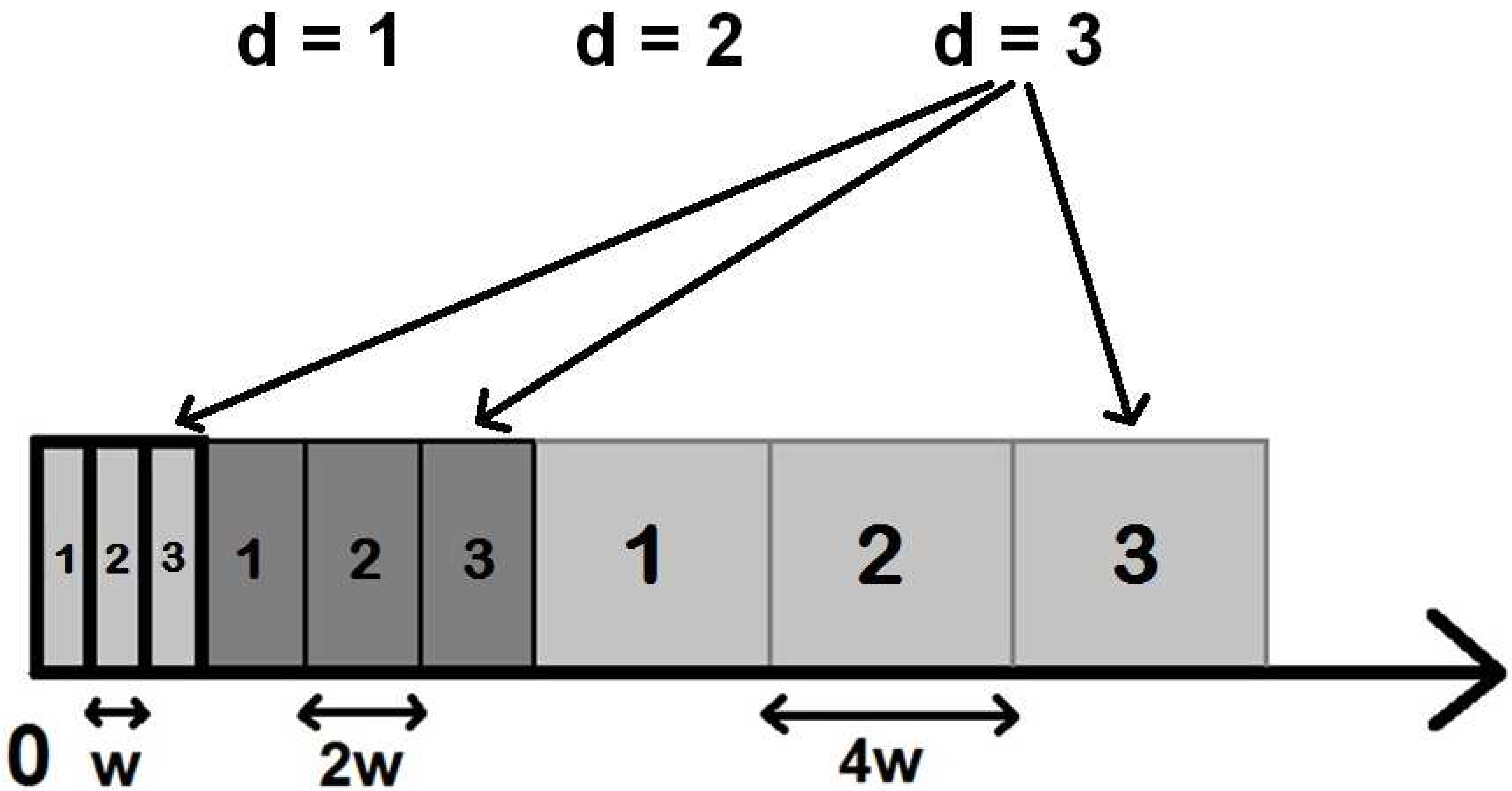
D = 3   F = 2



D = 3 F = 2



D = 3   F = 2



Lower case **d** signifies bin-rank.

d = 1

d = 2

d = 3

...

d = D

D = 3 F = 2

Are we imitating our own positional number system?

NO!

This can NOT be interpreted as a number system!

here $F < D$

but number systems are always with

$F > D + 1$

Let us **empirically** examine **real-life data** for any consistent pattern in bin scheme results:

A 3-bin scheme, with an expansion factor 11,
namely: D = 3 F = 11,
starting at the origin,
with an initial small width W = 0.002,
yields:

Time between all 19,452 earthquakes in 2012	{0.636, 0.221, 0.143}
USA population of all 19,509 cities in 2009	{0.603, 0.242, 0.154}
Price List of 8079 items www.mdhelicopters.com	{0.606, 0.248, 0.145}
Exponential 0.5% Growth, 3233 Periods from 600	{0.618, 0.226, 0.157}
USA Market Capitalization on Jan 1, 2013	{0.610, 0.238, 0.152}

A 7-bin scheme, with an expansion factor 4,
namely: D = 7 F = 4,
starting at the origin,
with an initial small width W = 0.007,
yields:

Time between earthquakes in 2012	{0.262, 0.184, 0.144, 0.122, 0.112, 0.092, 0.084}
US population, 19,509 cities in 2009	{0.257, 0.188, 0.152, 0.123, 0.108, 0.091, 0.082}
Catalog 8079 items mdhelicopters.com	{0.255, 0.190, 0.141, 0.121, 0.118, 0.091, 0.084}
Exp 0.5% Growth, 3233 Periods from 600	{0.263, 0.178, 0.143, 0.128, 0.109, 0.095, 0.084}
US Market Capitalization, Jan 1, 2013	{0.240, 0.192, 0.145, 0.132, 0.110, 0.098, 0.084}

A 4-bin scheme, with an expansion factor 8,
namely: D = 4   F = 8,
starting at the origin,
with an initial small width W = 0.0008,
yields:

Data Set	Bin A	Bin B	Bin C	Bin D
Time Between Earthquakes	48.3%	25.0%	15.3%	11.5%
USA Population Centers	48.9%	23.1%	16.0%	12.0%
LOG Symmetrical Triangular (1, 3, 5)	48.8%	24.1%	15.4%	11.7%
k/x over (1, 1000000)	49.3%	21.7%	16.3%	12.7%
Exponential Growth, B=1.5, F=1.01	47.9%	23.7%	15.6%	12.8%
Lognormal, Location=5, Shape=1	49.1%	23.3%	15.6%	12.1%
Lognormal, Location=9.3, Shape=1.7	48.6%	23.7%	15.8%	11.9%
Varied Data - Hill's Model	46.3%	25.3%	16.2%	12.2%
Chain U(U(U(U(U(0, 5666)))))	47.8%	24.1%	16.1%	12.0%

A 7-bin scheme, with an expansion factor 3, namely: D = 7 F = 3, starting at the origin, with an initial small width W = 0.0008, yields:

Data Set	Bin A	Bin B	Bin C	Bin D	Bin E	Bin F	Bin G
Time Between Earthquakes	22.4%	18.1%	15.5%	13.1%	11.7%	10.1%	9.0%
USA Population Centers	22.6%	18.9%	15.4%	13.1%	10.9%	9.9%	9.1%
LOG Symmetrical Triangular (1, 3, 5)	23.0%	17.8%	15.0%	13.0%	11.5%	10.2%	9.4%
k/x over (1, 1000000)	21.5%	17.9%	15.2%	13.4%	12.5%	10.3%	9.2%
Exponential Growth, B=1.5, F=1.01	22.6%	18.0%	15.0%	12.9%	11.7%	10.4%	9.4%
Lognormal, Location=5, Shape=1	23.1%	18.1%	14.9%	13.2%	11.4%	10.2%	9.1%
Lognormal, Location=9.3, Shape=1.7	22.8%	18.3%	15.2%	12.9%	11.5%	10.2%	9.2%
Varied Data - Hill's Model	22.0%	20.1%	15.6%	13.1%	10.3%	9.5%	9.4%
Chain U(U(U(U(U(0, 5666)))))	23.2%	17.8%	15.6%	13.3%	10.8%	10.2%	9.1%

It works!

Proportions are consistent across data sets.

Other D and F combinations, and using several other real-life physical data sets, also gave remarkably stable proportions.

We have found a genuine pattern in data independently of any number system!

The goal of the **scientist** in this case here is to explore and come up with a **generic mathematical expression** that would encompass all possible D and F cases.

Some empirical results of a variety of D and F combinations:

D	F	Bin A	Bin B	Bin C	Bin D	Bin E	Bin F	Bin G
3	2	41.3%	32.1%	26.7%				
3	3	46.1%	31.0%	22.9%				
3	4	50.0%	29.9%	20.1%				
3	5	53.1%	27.8%	19.1%				
3	6	55.9%	26.4%	17.7%				
3	11	63.6%	22.1%	14.3%				
4	2	32.1%	26.7%	22.2%	19.0%			
4	3	37.2%	26.5%	19.9%	16.5%			
4	4	41.4%	25.9%	18.4%	14.3%			
4	5	43.3%	24.6%	18.0%	14.2%			
4	12	53.5%	21.0%	14.9%	10.7%			
5	2	26.7%	22.2%	19.0%	16.9%	15.2%		
5	5	36.5%	22.5%	17.0%	13.3%	10.6%		
5	9	42.0%	22.5%	15.6%	11.1%	8.8%		
7	2	19.0%	16.8%	15.2%	13.6%	13.1%	11.5%	10.7%
7	4	25.9%	18.4%	14.3%	13.0%	10.6%	9.8%	8.0%
7	8	32.6%	19.0%	13.5%	11.1%	9.3%	8.0%	6.4%

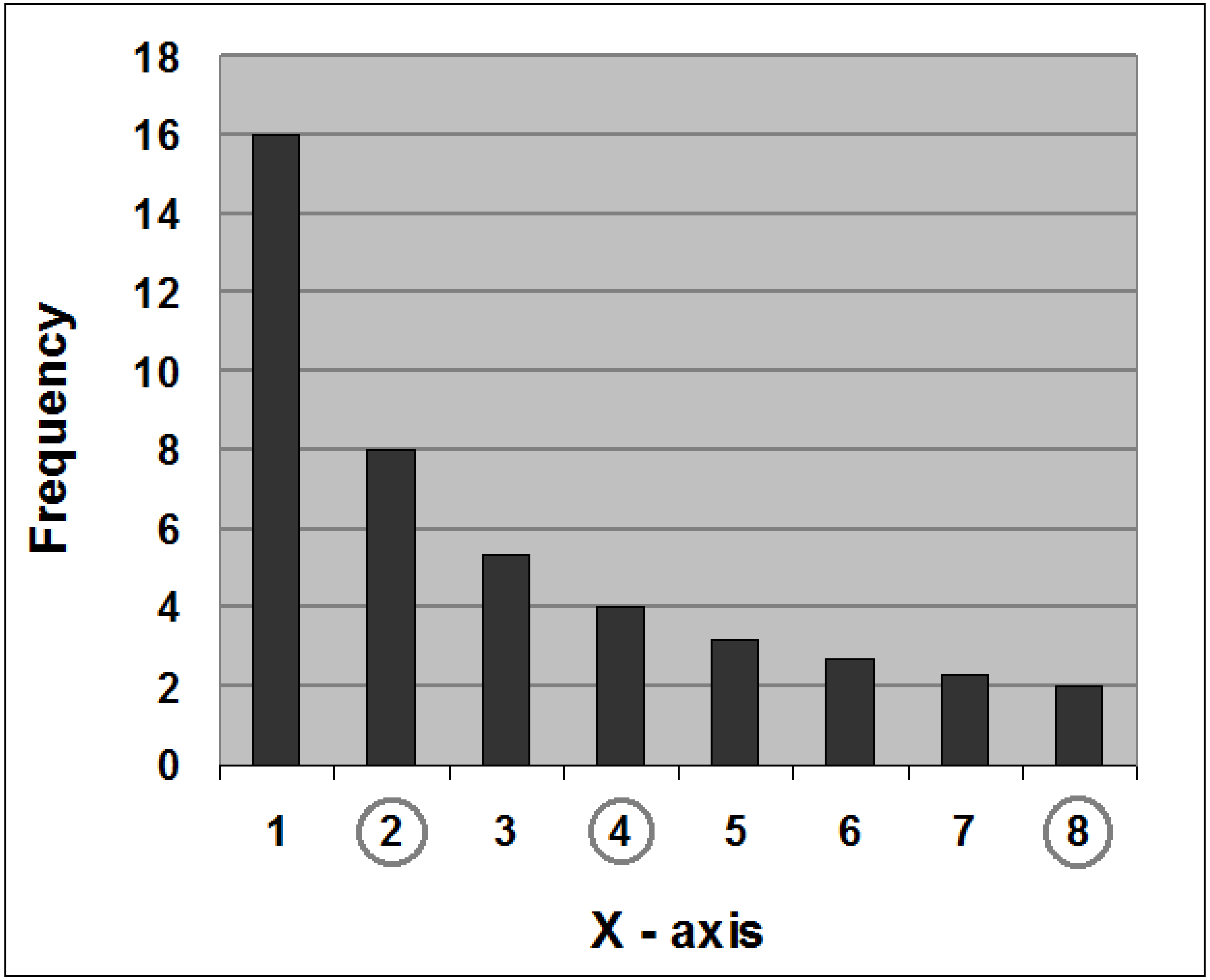
We seek a mathematical law that would encompass all these D & F cases of the previous table, and including any other possible combinations of D & F.

Philosophically, a sound approach would not merely attempt to find out **inductively** what is the best or most fitting expression in the approximate, but rather argue this by way of a conceptual postulate which would lead to an exact mathematical expression **deductively** – all the while closely agreeing with empirical results from real-life physical data sets.

The Postulate:

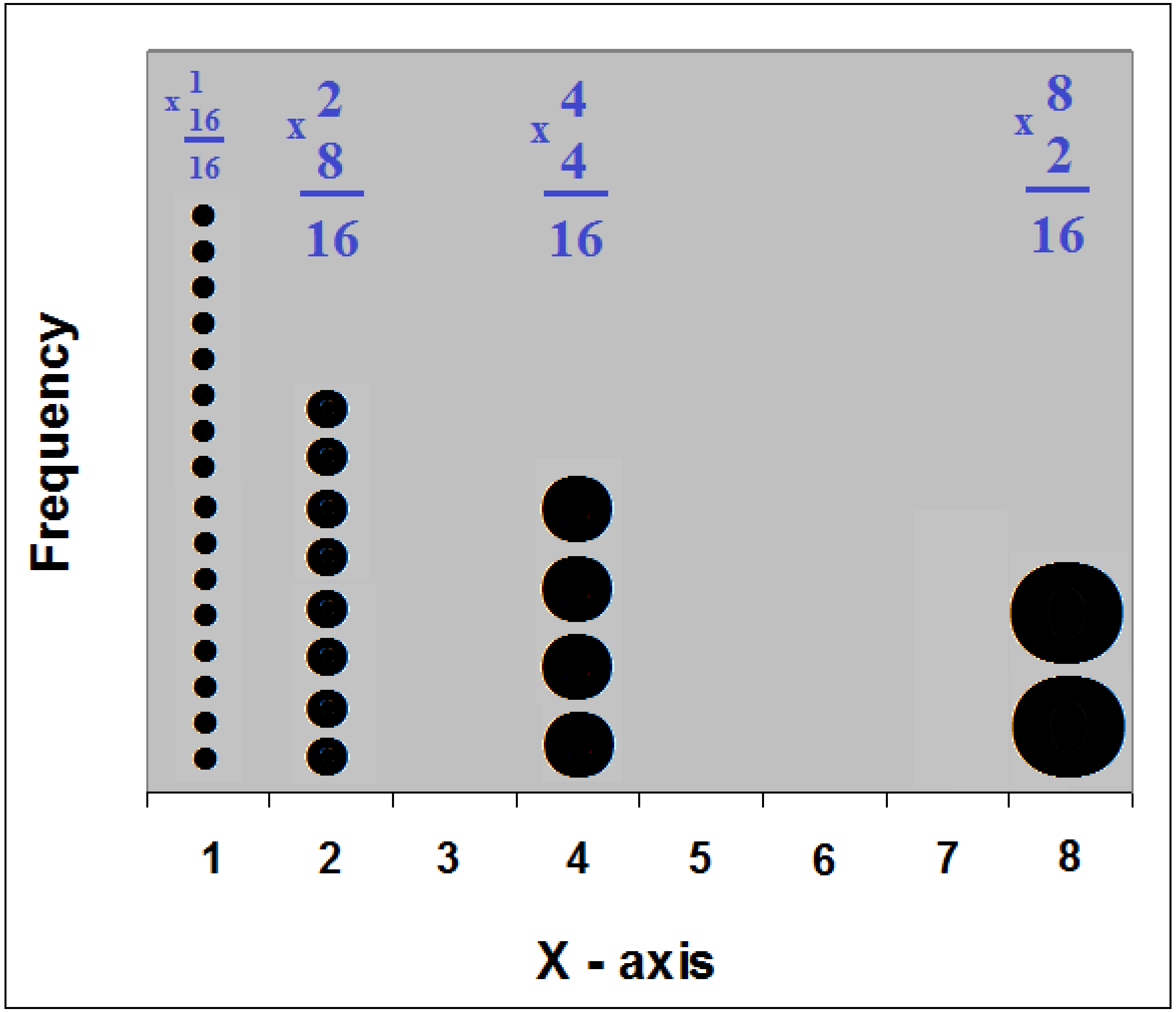
The generic pattern in how relative quantities are found in nature is such that the frequency of quantitative occurrences is inversely proportional to quantity.

The Postulate



Doubling of X $\Longrightarrow$ Frequency is reduced by half

The Postulate evens the totals



Total = X*Frequency = Constant

This leads to the explorations of results from bin systems fitting **k/X** distribution on (W, ∞).

k/X is defined from **W** up to **infinity**.

W does **not** have to be small.

$$pdf(x) = k(1/X)$$

We shall impose discrete histograms onto the k/X continuous curve.

The subsequent tedious mathematical work then involves calculating definite integrals of **k/X** cycle by cycle, and having sufficiently large number of such results in order to enable us to decipher the eventual limit as the number of cycle goes to infinity.

Five features are involved in this construction:

(I) Avoidance of an upward explosion start of the k/X density at the origin 0 which would have been undefined due to a division by 0.

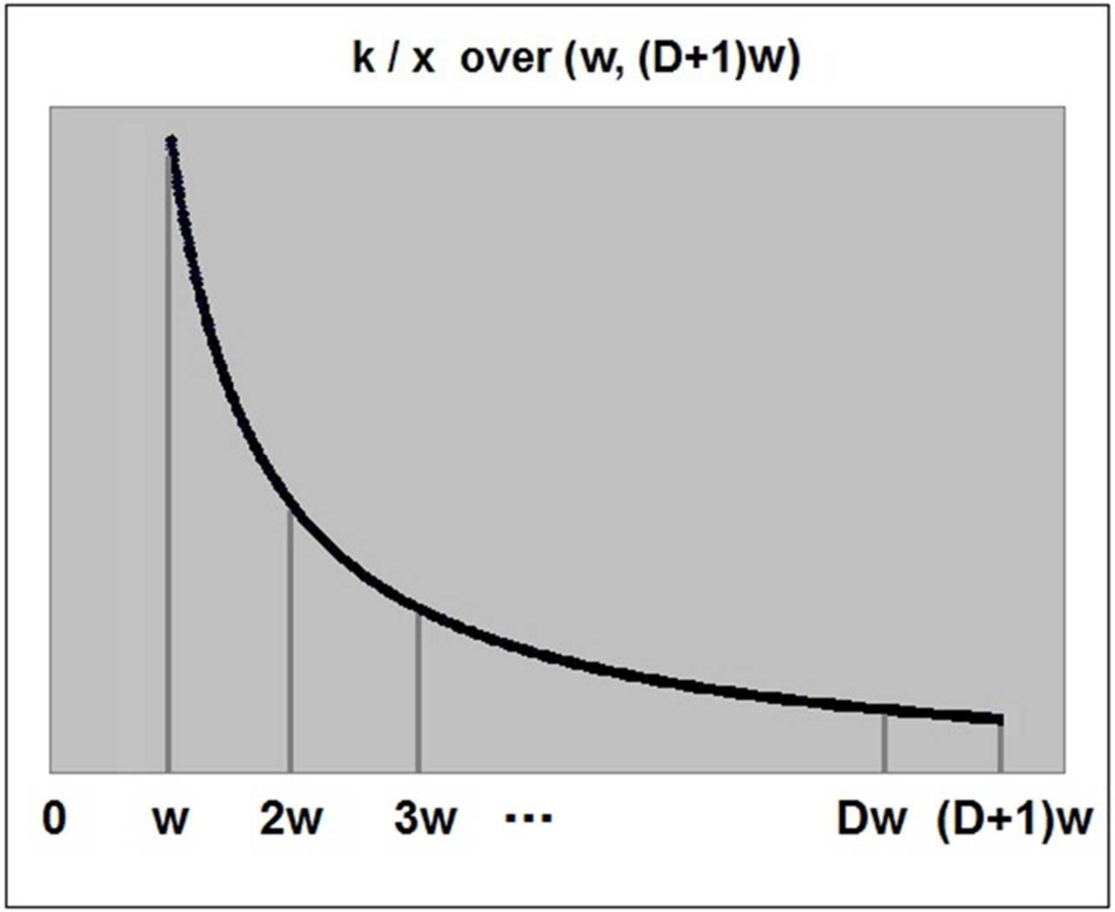
(II) Equal spacing (width) of all bins.

(III) Equality between the 1st bin width and the separation of the defined range from the 0 origin. Namely, that the length of the step from the origin to the launch of K/X is also the width of each bin in the 1st cycle. Algebraically it is expressed as **(2w - w) = (w - 0)**.

(IV) No coordination is employed or attempted whatsoever with any number system or digits on the x-axis below.

(V) Only positive numbers are involved.

One cycle



Equating the entire area to one, we obtained:

$$\int_{w}^{(D+1)w} \frac{k}{x} dx = 1$$

k[ln((D+1)w) - ln(w)] = 1

k[ln(D+1) + ln(w) - ln(w)] = 1

k[ln(D+1)] = 1

k = 1/ln(D + 1)

Evaluating the portion of area hanging over bin #d (with d running from 1 to D, as in digits), we obtain:

$$P(d) = \int_{dw}^{(d+1)w} \frac{k}{x} dx$$

P(d) = [1/ln(D+1)]*[ln(d+1) + ln(w) – ln(d) – ln(w)]

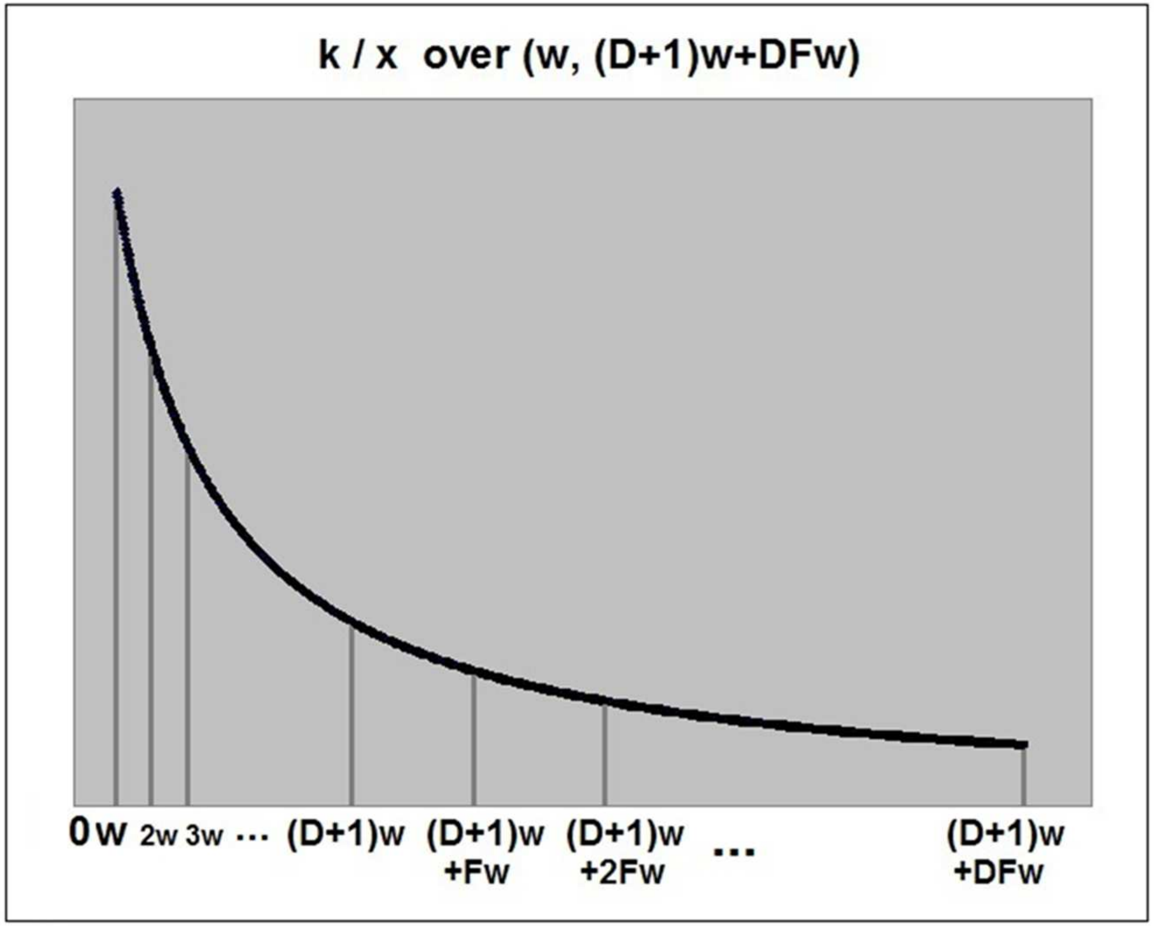
P(d) = [1/ln(D+1)]*[ln(d+1) – ln(d)]

P(d) = [1/ln(D+1)]*ln[(d+1)/(d)]

and finally:

P(d) = ln(1+1/d) / ln(D+1).

Two cycles



Equating the entire area to one, we obtained:

$$\int_{w}^{(D+1)w+DFw} \frac{k}{x}\,dx = 1$$

k[ln(w*[(D+1) + (DF)]) - ln(w)] = 1

k[ln(w) + ln[(D+1) + (DF)] - ln(w)] = 1

k[ln[(D+1) + (DF)]] = 1

k = 1/ ln(1 + D + DF)

Evaluating the **first** portion of area (1st cycle) hanging over bin #d (d running from 1 to D as in digits), we obtain:

$$P1(d) = \int_{dw}^{(d+1)w} \frac{k}{x} dx$$

P1(d) = [1/ln(1 + D + DF)]*[ln(d+1) + ln(w) – ln(d) – ln(w)]

P1(d) = [1/ln(1 + D + DF)]*[ln(d+1) – ln(d)]

Evaluating the **<u>second</u>** portion of area (2nd cycle) hanging over bin #d (with d running from 1 to D, as in digits), we obtain:

$$P2(d) = \int_{(D+1)w+(d-1)Fw}^{(D+1)w+(d)Fw} \frac{k}{x} dx$$

P2(d) = [1/ln(1 + D + DF)]*[ln((D+1) + dF) + ln(w) – ln((D+1) + (d-1)F) -ln(w)]
P2(d) = [1/ln(1 + D + DF)]*[ln((D+1) + dF) – ln((D+1) + (d-1)F)]

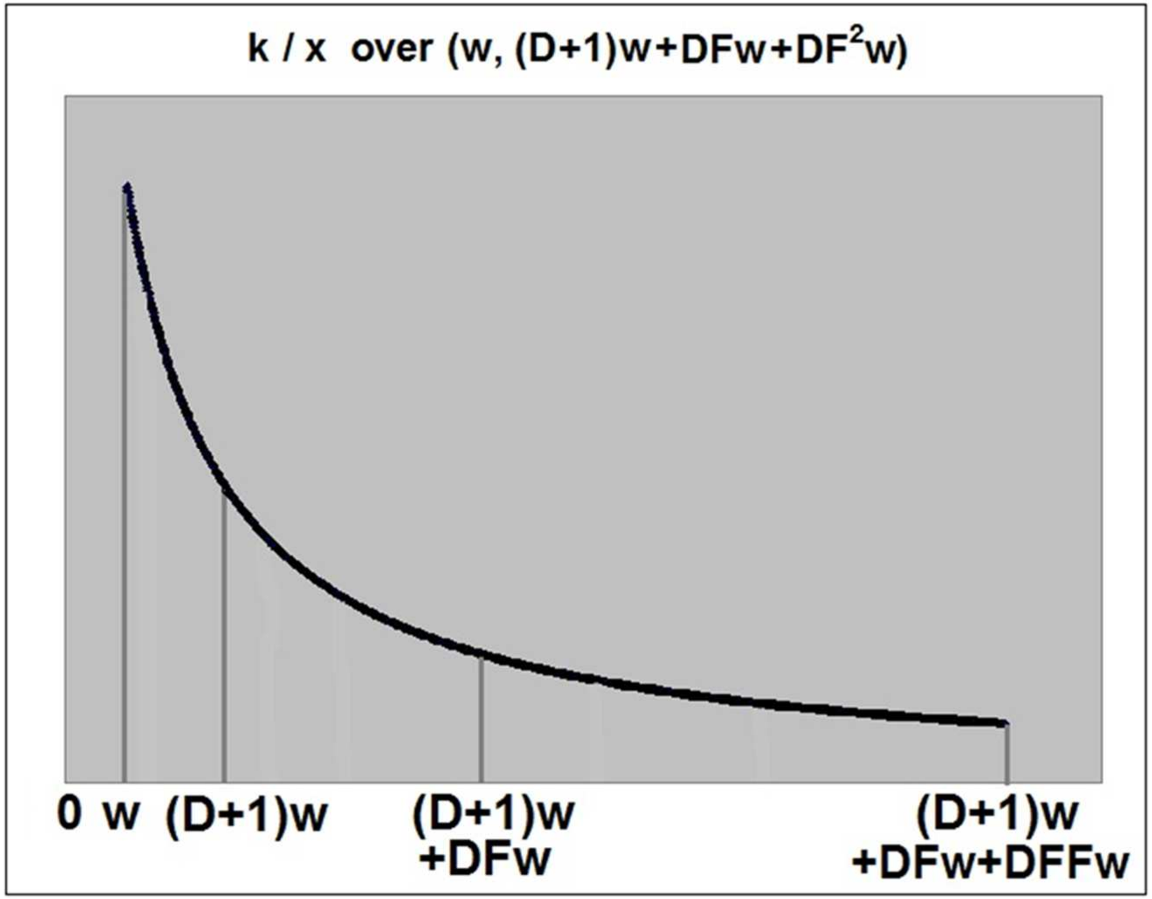
Combining both areas, namely P(d) = P1(d) + P2(d), we get:

P(d) = [1/ln(1 + D + DF)]*[ln(d+1) – ln(d) + ln((D+1) + dF) – ln((D+1) + (d-1)F)]

Applying the identity LOG(A) - LOG(B) = LOG(A/B) we finally get:

P(d) = [ln(1 +1/d) + ln((1+D+dF)/(1+D+(d-1)F))] / [ln(1 + D + DF)]

Three cycles



We need to evaluate the following definite integrals:

$$\int_{dw}^{(d+1)w} \frac{k}{x}\,dx + \int_{(D+1)w + (d-1)Fw}^{(D+1)w + (d)Fw} \frac{k}{x}\,dx + \int_{(D+1)w + DFw + (d-1)FFw}^{(D+1)w + DFw + (d)FFw} \frac{k}{x}\,dx$$

Equating the entire area to one, we obtained:

$$\int_{w}^{(D+1)w+DFw+DF^2w} \frac{k}{x} dx = 1$$

$$k[\ \ln(w) + \ln[(D+1) + DF + DF^2] - \ln(w)] = 1$$

$$k[\ \ln[(D+1) + DF + DF^2]\ ] = 1$$

$$\mathbf{k = 1/\ ln(1 + D + DF + DF^2)}$$

Evaluating the **first and second** portion of areas yields the same results as in the once-expanding bin system except for the different k constant expression here.

Evaluating the **third** portion of area hanging over bin #d (with d running from 1 to D, as in digits), we obtain:

$$P3(d) = \int_{(D+1)w+DFw+(d-1)F^2w}^{(D+1)w+DFw+(d)F^2w} \frac{k}{x} dx$$

$P3(d) = k \,(\ln[w] + \ln[(D+1) + DF + (d)*F^2] - \ln[w] – \ln[(D+1) + DF + (d-1)*F^2])$
$P3(d) = k \,(\ln[(D+1) + DF + (d)*F^2] – \ln[(D+1) + DF + (d-1)*F^2])$
$P3(d) = k* \ln([(D+1) + DF + (d)*F^2] / [(D+1) + DF + (d-1)*F^2])$

Combining all 3 areas, namely P(d) = P1(d) + P2(d) + P3(d), we finally get:

$$P(d) = k*\mathbf{ln}(1 + 1/d) + k*\mathbf{ln}([1+D+dF] / [1+D+(d-1)F]) + k*\mathbf{ln}([(D+1) + DF + (d)*F^2] / [(D+1) + DF + (d-1)*F^2])$$

$$P(d) = \frac{\ln(1 + 1/d) + \ln\left(\frac{[1 + D + (d)F]}{[1 + D + (d-1)F]}\right) + \ln\left(\frac{[1 + D + DF + (d)*F^2]}{[1 + D + DF + (d-1)*F^2]}\right)}{\ln(1 + D + DF + DF^2)}$$

Infinite cycles

Algebraic expressions for bin proportions of K/X distribution for higher expansion orders perfectly follow the above (clear) pattern as a sequence of ever increasing terms in the numerator and in the denominator.

The first 4 elements of this infinite sequence, beginning with a non-expanding bin system, and ending with a bin system having four cycles, are as follow:

$$\frac{\ln\left(\frac{[1 + (d)]}{[1 + (d-1)]}\right)}{\ln(1 + D)}$$

$$\frac{\ln\left(\frac{[1 + (d)]}{[1 + (d-1)]}\right) + \ln\left(\frac{[1 + D + (d)F]}{[1 + D + (d-1)F]}\right)}{\ln(1 + D + DF)}$$

$$\frac{\ln\left(\frac{[1 + (d)]}{[1 + (d-1)]}\right) + \ln\left(\frac{[1 + D + (d)F]}{[1 + D + (d-1)F]}\right) + \ln\left(\frac{[1 + D + DF + (d)F^2]}{[1 + D + DF + (d-1)F^2]}\right)}{\ln(1 + D + DF + DF^2)}$$

$$\frac{\ln\left(\frac{[1+(d)]}{[1+(d-1)]}\right) + \ln\left(\frac{[1+D+(d)F]}{[1+D+(d-1)F]}\right) + \ln\left(\frac{[1+D+DF+(d)F^2]}{[1+D+DF+(d-1)F^2]}\right) + \ln\left(\frac{[1+D+DF+DF^2 + (d)F^3]}{[1+D+DF+DF^2+(d-1)F^3]}\right)}{\ln(1 + D + DF + DF^2 + DF^3)}$$

What is the limit?

Does it exist?

Can we find a close form expression?

With assistance from the distinguished mathematician **George Andrews**, a closed form expression for the limit of the infinite sequence is obtained in the F > 1 case, enabling us to succinctly express the general law of relative quantities.

George Andrews from Pennsylvania State University is well-known for his extensive work on Ramanujan's Lost Notebook. He is considered to be the world's leading expert in the theory of integer partitions.

The pages of the mathematical derivation scripted by **George Andrews** in Sep 2013 follows:

Applying the finite geometric series formula

$$1+X+X^2+\cdots+X^{n-1} = \frac{X^n-1}{X-1},$$

the n^{th} term in your sequence can be written as

$$S_n = \frac{\displaystyle\sum_{j=0}^{n-1} \ln\left(\frac{1+\frac{D(F^j-1)}{F-1}+dF^j}{1+\frac{D(F^j-1)}{F-1}+(d-1)F^j}\right)}{\ln\left(1+\frac{D(F^n-1)}{F-1}\right)}$$

$$= \frac{\displaystyle\sum_{j=0}^{n-1} \ln\left(\frac{\left(\frac{D+d(F-1)}{F-1}\right)F^j+1-\frac{D}{F-1}}{\left(\frac{D+(d-1)(F-1)}{F-1}\right)F^j+1-\frac{D}{F-1}}\right)}{\ln\left(\frac{D}{F-1}F^n+1-\frac{D}{F-1}\right)}$$

Let $A = \frac{D + d(F-1)}{F-1}$, $B = \frac{D + (d-1)(F-1)}{F-1}$,

$C = \frac{D}{F-1}$ and $E = 1 - \frac{D}{F-1}$.

Thus we may write your sequence as

$$S_n = \frac{\displaystyle\sum_{j=0}^{n-1} \ln\left(\frac{AF^j + E}{BF^j + E}\right)}{\ln\left(CF^n + E\right)}$$

Let $f = \frac{1}{F}$, then

$$S_n = \frac{\displaystyle\sum_{j=0}^{n-1} \ln\left(\frac{A}{B}\,\frac{(1+\frac{E}{A}f^j)}{(1+\frac{E}{B}f^j)}\right)}{\ln\left(CF^n\left(1+\frac{E}{C}f^n\right)\right)}$$

$$= \frac{n\ln\frac{A}{B} + \ln\displaystyle\prod_{j=0}^{n}\frac{(1+\frac{E}{A}f^j)}{(1+\frac{E}{B}f^j)}}{n\ln F + \ln C + \ln\left(1+\frac{E}{C}f^n\right)}.$$

If $F > 1$ then $0 < f < 1$ and

$$\prod_{j=0}^{\infty} \frac{\left(1 + \frac{E}{A} f^j\right)}{1 + \frac{E}{B} f^j}$$

is a convergent infinite product

Hence $\lim_{n \to \infty} S_n = \dfrac{\ln \frac{A}{B}}{\ln F}$.

Note that we have assumed $F > 1$.

If $F = 1$, then

$$S_n = \frac{\sum_{j=0}^{n-1} \ln\left(\frac{1 + jD + d}{1 + jD + (d-1)}\right)}{\ln(1 + nF)}$$

$$= \frac{\sum_{j=0}^{n-1} \ln\left(1 + \frac{1}{1 + jD + (d-1)}\right)}{\ln(1 + nF)}$$

$$= \frac{\displaystyle\sum_{j=0}^{n-1} \left(\frac{1}{jD+d} + O\left(\frac{1}{(jD+d)^2} \right) \right)}{\ln n + \ln\left(F + \frac{1}{n}\right)}$$

The integral test tells us that

$$\sum_{j=0}^{n-1} \frac{1}{jD+d} \sim \frac{1}{D} \ln n$$

So

$$\lim_{n\to\infty} S_n = \frac{1}{D} \qquad \text{when } F = 1.$$

Let us narrate clearly Andrews‘ derivation:

The 4th term of the sequence expressed earlier, denoted as S_4 is:

$$\frac{\ln\left(\frac{[1+(d)]}{[1+(d-1)]}\right) + \ln\left(\frac{[1+D+(d)F]}{[1+D+(d-1)F]}\right) + \ln\left(\frac{[1+D+DF+(d)F^2]}{[1+D+DF+(d-1)F^2]}\right) + \ln\left(\frac{[1+D+DF+DF2 + (d)F^3]}{[1+D+DF+DF2+(d-1)F^3]}\right)}{\ln(1 + D + DF + DF^2 + DF^3)}$$

Employing the **finite** geometric formula for the terms involving F, namely:

$$1 + X + X^2 + X^3 + \ldots + X^N = (X^{N+1} - 1) / (X - 1) ,$$

the nth term in the sequence is then:

$$S_N = \frac{\displaystyle\sum_{j=0}^{N-1} \ln\left(\frac{1 + \dfrac{D\left(F^j - 1\right)}{(F-1)} + (d)F^j}{1 + \dfrac{D(F^j - 1)}{(F-1)} + (d-1)F^j}\right)}{\ln\left(1 + \dfrac{D(F^N - 1)}{(F-1)}\right)}$$

Pulling together all the coefficients of F^{POWER}, we get:

$$S_N = \frac{\displaystyle\sum_{j=0}^{N-1} \ln\left(\frac{1+\left(\frac{D+(d)(F-1)}{(F-1)}\right)F^j - \frac{D}{F-1}}{1+\left(\frac{D+(d-1)(F-1)}{(F-1)}\right)F^j - \frac{D}{F-1}}\right)}{ln\left(1+\left(\frac{D}{(F-1)}\right)F^N - \frac{D}{F-1}\right)}$$

In order to obtain a more compact expression, let us define:

$$A = \frac{D + d(F-1)}{F-1}$$

$$B = \frac{D + (d-1)(F-1)}{F-1}$$

$$C = \frac{D}{F-1}$$

$$E = 1 - \frac{D}{F-1}$$

$$S_N = \frac{\sum_{j=0}^{N-1} \ln\left(\frac{AF^j + E}{BF^j + E}\right)}{ln(\ CF^N + E\ )}$$

Since in our context $F \geq 1$, there is no hope of obtaining any obvious convergence in terms such as F^j or F^N, hence we define **$f = 1 / F$** , creating a quantity f such that $0 < f \leq 1$ holds, and which may hopefully let terms such as f^{j} or f^{N} converge.

$$S_N = \frac{\displaystyle\sum_{j=0}^{N-1} \ln\left(\frac{A\frac{1}{f^j} + E}{B\frac{1}{f^j} + E}\right)}{ln\left(C\frac{1}{f^N} + E \right)}$$

$$S_N = \frac{\displaystyle\sum_{j=0}^{N-1} \ln\left(\frac{A\frac{1}{f^j} + E}{B\frac{1}{f^j} + E} * \frac{f^j}{f^j} \right)}{ln\left(C * \frac{1}{f^N} + E * \frac{C}{C} * \frac{f^N}{f^N} \right)}$$

$$S_N = \frac{\displaystyle\sum_{j=0}^{N-1} \ln\left(\frac{A + E * f^j}{B + E * f^j} \right)}{ln\left(C * \frac{1}{f^N} * \left(1 + E * \frac{f^N}{C} \right) \right)}$$

General Logarithmic identity:

$$\ln(X_1) + \ln(X_2) + \ln(X_3) + \ldots + \ln(X_N) = \ln(X_1 X_2 X_3 \ldots X_N)$$

$$\sum_{j=1}^{N} \ln(x_j) = \ln\left(\prod_{j=1}^{N} x_j\right)$$

$$S_N = \frac{\sum_{j=0}^{N-1} \ln\left(\frac{\left(1 + \frac{E}{A} * f^j\right) A}{\left(1 + \frac{E}{B} * f^j\right) B} \right)}{ln\left(C * F^N * \left(1 + \frac{E}{C} f^N\right)\right)}$$

$$S_N = \frac{N * ln\left(\frac{A}{B}\right) + ln\left(\prod_{j=0}^{N-1} \frac{\left(1 + \frac{E}{A} * f^j\right)}{\left(1 + \frac{E}{B} * f^j\right)} \right)}{N * ln(F) + ln(C) + ln\left(1 + \frac{E}{C} f^N\right)}$$

It is only at this late stage that we let N go to infinity!

For F > 1 as in the normal case of expanding bin scheme, 0 < f < 1, therefore:

$$\prod_{j=0}^{\infty} \frac{\left(1 + \frac{E}{A} * f^{j}\right)}{\left(1 + \frac{E}{B} * f^{j}\right)}$$

is a convergent infinite product since $\sum_{j=0}^{\infty} f^{j}$ is converging.

as N $\rightarrow \infty$ $\quad \frac{\left(1 + \frac{E}{A} * f^{j}\right)}{\left(1 + \frac{E}{B} * f^{j}\right)} \rightarrow \frac{\left(1 + \frac{E}{A} * 0\right)}{\left(1 + \frac{E}{B} * 0\right)} \rightarrow \frac{(1+0)}{(1+0)} \rightarrow 1$

The term ln(**C**) is ln($\frac{D}{F-1}$), and it is finite and insignificant.

The term $ln\left(1+\frac{E}{C}f^N\right)$ is zero as N → ∞.

$$\lim_{N\to\infty} ln\left(1+\frac{E}{C}f^N\right) = ln\left(1+\frac{E}{C}0\right) = ln(1+0) = ln(1) = 0$$

Finally: $\lim_{N\to\infty} S_N = \dfrac{\ln(\frac{A}{B})}{\ln(F)}$

Using the definition of A and B above, we get:

$$\frac{\ln\left(\frac{\left(\frac{D+d(F-1)}{F-1}\right)}{\left(\frac{D+(d-1)(F-1)}{F-1}\right)}\right)}{\ln(F)}$$

which is further reduced by canceling out the two (F – 1) terms in the numerator to arrive at:

The General Law of Relative Quantities:

$$\frac{\ln\left(\frac{D + d(F-1)}{D + (d-1)(F-1)}\right)}{\ln(F)}$$

The 1st Miracle:

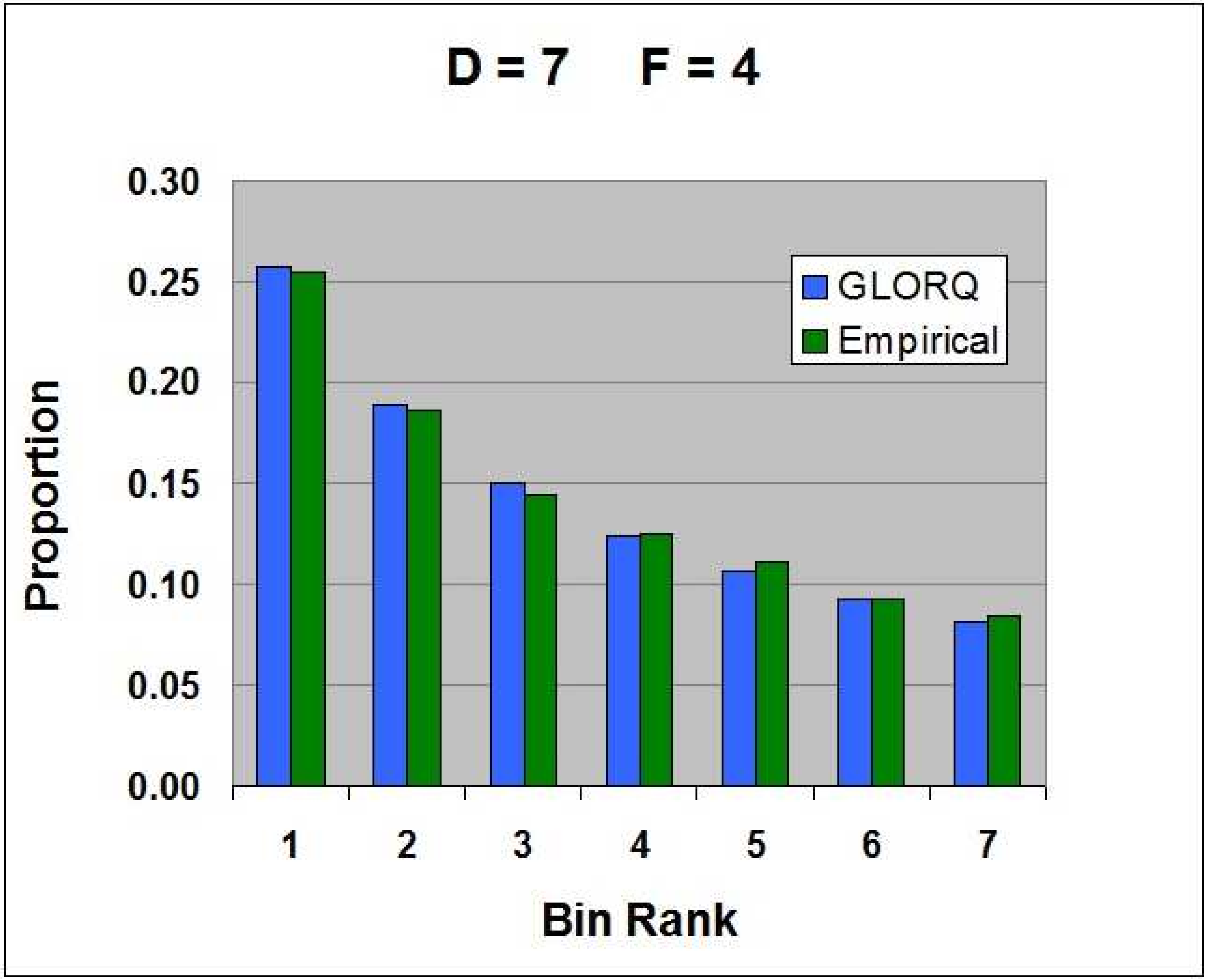
Empirical results from real-life physical data sets strongly confirm the general law:

D = 3 and F = 11

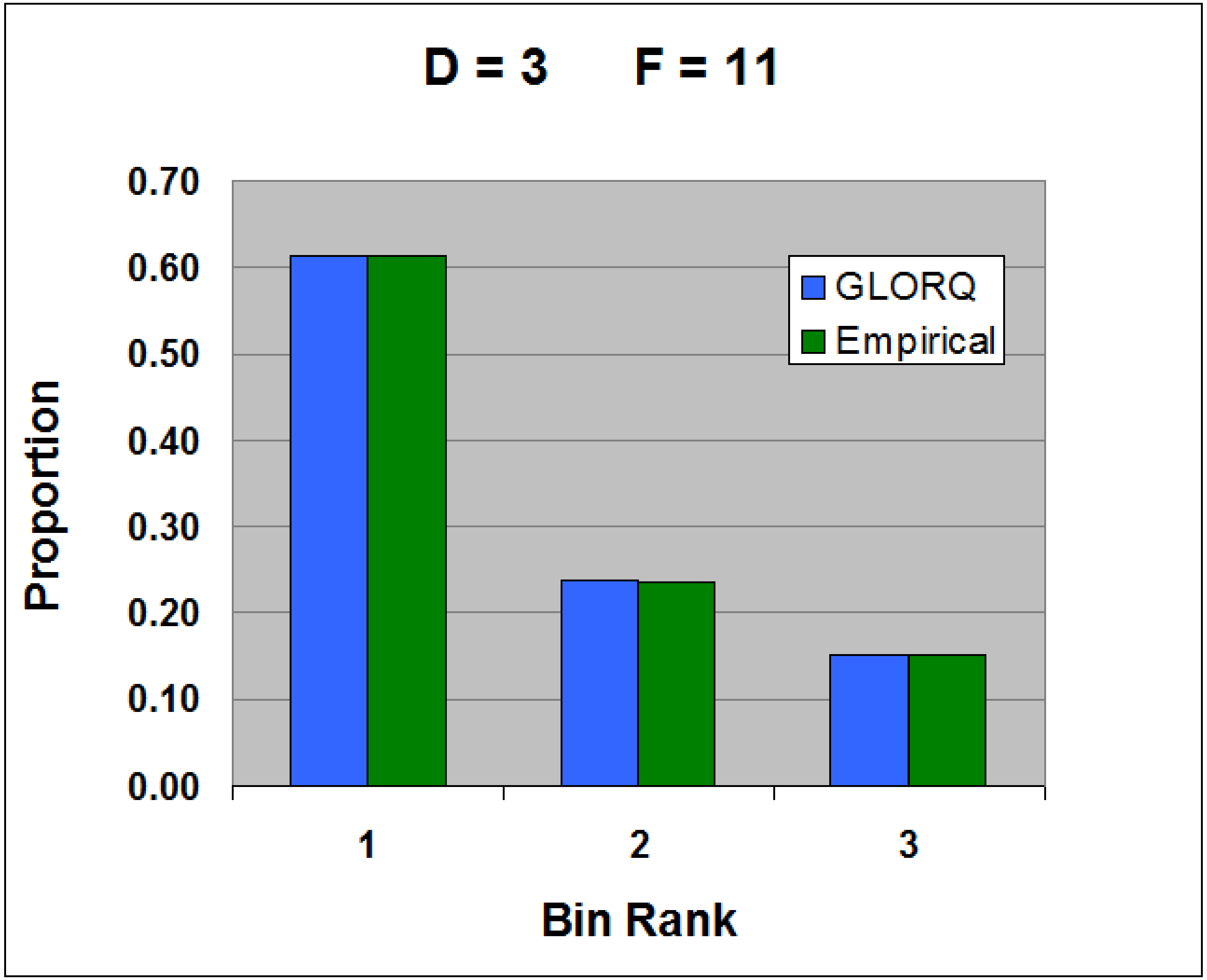
Time between all 19,452 earthquakes in 2012	{0.636, 0.221, 0.143}
USA population of all 19,509 cities in 2009	{0.603, 0.242, 0.154}
Price List of 8079 items www.mdhelicopters.com	{0.606, 0.248, 0.145}
Exponential 0.5% Growth, 3233 Periods from 600	{0.618, 0.226, 0.157}
USA Market Capitalization on Jan 1, 2013	{0.610, 0.238, 0.152}
ln((**3** + d(**11** – 1))/(**3** + (d – 1)(**11** – 1))) / ln(**11**)	**{0.612, 0.238, 0.151}**

D = 7 and F = 4

Time between earthquakes in 2012	{0.262, 0.184, 0.144, 0.122, 0.112, 0.092, 0.084}
US population, 19,509 cities in 2009	{0.257, 0.188, 0.152, 0.123, 0.108, 0.091, 0.082}
Catalog 8079 items mdhelicopters.com	{0.255, 0.190, 0.141, 0.121, 0.118, 0.091, 0.084}
Exp 0.5% Growth, 3233 Periods from 600	{0.263, 0.178, 0.143, 0.128, 0.109, 0.095, 0.084}
US Market Capitalization, Jan 1, 2013	{0.240, 0.192, 0.145, 0.132, 0.110, 0.098, 0.084}
ln((**7** + d(**4**–1))/(**7** + (d–1)(**4**–1))) / ln(**4**)	**{0.257, 0.189, 0.150, 0.124, 0.106, 0.092, 0.082**}



Using the **Average** of the 5 empirical data sets.



Using the **Average** of the 5 empirical data sets.

The 2nd Miracle:

Digital Benford's Law is simply a special case of the general law when bin schemes are constructed under the constraint F = D + 1.

The term F is then substituted by (D + 1) everywhere in expression of the general law:

$$\textbf{GLORQ} = \frac{\ln\left(\frac{D + d(F-1)}{D + (d-1)(F-1)}\right)}{\ln(F)} = \frac{\ln\left(\frac{D+d(D+1-1)}{D + (d-1)(D+1-1)}\right)}{\ln(D+1)} =$$

$$\frac{\ln\left(\frac{D+d(D)}{D + (d-1)(D)}\right)}{\ln(D+1)} = \frac{\ln\left(\frac{D(1+d)}{D(1+(d-1))}\right)}{\ln(D+1)} = \frac{\ln\left(\frac{1+d}{1+(d-1)}\right)}{\ln(D+1)} =$$

$$\frac{\ln\left(\frac{1+d}{d}\right)}{\ln(D+1)} = \frac{\ln\left(1+\frac{1}{d}\right)}{\ln(D+1)} = \frac{\ln\left(1+\frac{1}{d}\right)}{\ln(BASE)} = \frac{\mathrm{LOG}\left(1+\frac{1}{d}\right)}{\mathrm{LOG}(10)} =$$

$$\frac{\mathrm{LOG}\left(1+\frac{1}{d}\right)}{1} = \textbf{Benford's Law}$$

Benford's Law is merely a sideshow to this physical law of nature which can be measured and detected by ways other than our own digital perceptions.

We are no longer seduced and blinded by the incredible efficiency of our number system, and we are able to acknowledge its arbitrariness.

==

The General Law of Relative Quantities does **not** seek or need any *statistical theory* to establish its empirically validated discoveries, rather the approach is purely **scientific** aided by mechanizations and tools from **pure mathematics**.

Statistical theory can always be added as extra machinery after the establishments of GLORQ, but since real-life data sets strongly and nearly universally confirm GLORQ with very small deviations of empirical from theoretical, it follows that *statistical considerations could only marginally contribute some minor additions* to the whole edifice of GLORQ.

==

GLORQ implies that

Proportion (**d**) > Proportion (**d + 1**)

Corresponding to the fact that big sizes are rare and small sizes are numerous.

Corresponding to the fact that the histogram is falling to the right.

The GLORQ expression could be re-written via simple algebraic manipulations in order to emphasis its skewed quantitative configuration:

GLORQ original expression:

$$\frac{\ln\left(\dfrac{D + d(F-1)}{D + (d-1)(F-1)}\right)}{\ln(F)}$$

Expanding a bit the denominator of the numerator:

$$\frac{\ln\left(\dfrac{D + d(F-1)}{D + (d)(F-1) - (F-1)}\right)}{\ln(F)}$$

Subtracting (F – 1) and adding (F – 1) on top:

$$\frac{\ln\left(\dfrac{D + d(F-1) - (F-1) + (F-1)}{D + (d)(F-1) - (F-1)}\right)}{\ln(F)}$$

Further reducing the numerator:

$$\frac{\ln\left(1+\frac{(F-1)}{D+(d)(F-1)-(F-1)}\right)}{\ln(F)}$$

Simplifying the denominator of the numerator:

$$\frac{\ln\left(1+\frac{(F-1)}{D+(d-1)(F-1)}\right)}{\ln(F)}$$

Hence the GLORQ expression is **inversely proportional to d**.

“The General Law of Relative Quantities (GLORQ) hinges on a very subtle mathematical limit, and Alex E. Kossovsky enlisted my assistance in its mathematical derivation. I am not an expert on Benford's Law; I am a pure mathematician, however, my experience over the years is that when intricate mathematics is required in a theory, then it often follows that the theory will stand on its own merits. I can assure the readers that the mathematics behind GLORQ is valid and sufficiently surprising that it merits serious consideration”.

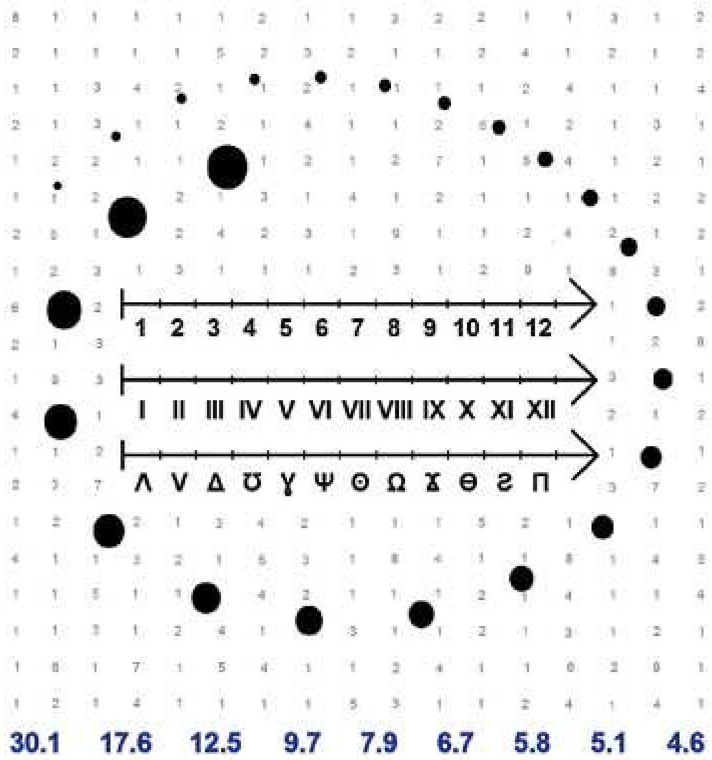
George Andrews

END

Benford's Law

Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications

Alex Ely Kossovsky



World Scientific

Benford's Law

Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications

Contrary to common intuition that all digits should occur randomly with equal chances in real data, empirical examinations consistently show that not all digits are created equal, but rather that low digits such as {1, 2, 3} occur much more frequently than high digits such as {7, 8, 9} in almost all data types, such as those relating to geology, chemistry, astronomy, physics, and engineering, as well as in accounting, financial, econometrics, and demographics data sets. This intriguing digital phenomenon is known as Benford's Law.

This book represents an attempt to give a comprehensive and in-depth account of all the theoretical aspects, results, causes and explanations of Benford's Law, with a strong emphasis on the connection to real-life data and the physical manifestation of the law. In addition to such a bird's eye view of the digital phenomenon, the conceptual distinctions between digits, numbers, and quantities are explored; leading to the key finding that the phenomenon is actually quantitative in nature; originating from the fact that in extreme generality, nature creates many small quantities but very few big quantities, corroborating the motto "small is beautiful", and that therefore all this is applicable just as well to data written in the ancient Roman, Mayan, Egyptian, and other digit-less civilizations.

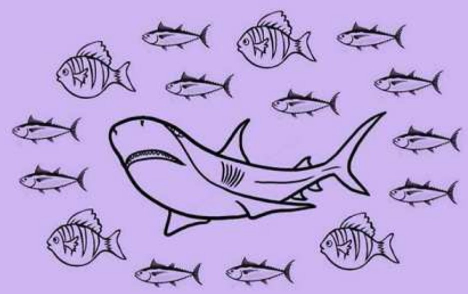
Fraudsters are typically not aware of this digital pattern and tend to invent numbers with approximately equal digital frequencies. The digital analyst can easily check reported data for compliance with this digital law, enabling the detection of tax evasion, Ponzi schemes, and other financial scams. The forensic fraud detection section is written in a very concise and reader-friendly style; gathering all known methods and standards in the accounting and auditing industry; summarizing and fusing them into a singular coherent whole; and can be understood without deep knowledge in statistical theory or advanced mathematics. In addition, a digital algorithm is presented, enabling the auditor to detect fraud even when the sophisticated cheater is aware of the law and invents numbers accordingly. The algorithm employs a subtle inner digital pattern within the Benford's pattern itself. This newly discovered pattern is deemed to be nearly universal, being even more prevalent than the Benford phenomenon itself, as it is found in all random data sets, Benford as well as non-Benford types.

World Scientific
www.worldscientific.com
9089 hc

ISBN 978-981-4583-68-8
9 789814 583688

SMALL IS BEAUTIFUL

Why the Small is Numerous but the Big is Rare in the World



Alex Ely Kossovsky

Why are there more poor people with small bank accounts than rich people with big bank accounts? Why is the small almost always more numerous than the big in the world? Empirical examinations of real-life data overwhelmingly confirm the existence of such uneven size proportions in favor of the small. There are more small planets and stars than big ones in the cosmos. There are more small molecules than big molecules in the chemical world. There are more small families with few children than big families with many children. In geological data, there are more small rivers than big rivers, and there are more harmless small earthquakes than devastating big ones.

There are by far many more small creatures than big creatures in the biological world. There are only about two million big whales swimming the oceans, yet there are over 300 billion small birds flying the sky. Tiny little ants are even more abundant, with estimates of over 100 trillions of them walking the earth! In number theory as well, there are more small prime numbers than big prime numbers for integers. In census data, there are more villages than towns, more towns than cities, and more cities than metropolises. In history, there were more small wars with low death toll than horrific big wars with high death toll such as WWII.

The vast list of topics & disciplines obeying this quantitative law of nature confirms the fact that the phenomenon is nearly universal. This book discusses in detail several real-life case studies; presents three distinct explanations for the phenomenon; and numerically quantifies the small is beautiful phenomenon in order to obtain an exact measure indicating by how much the relatively small is more numerous than the relatively big.

ISBN 9780692912416
90000 >
9 780692 912416

STUDIES IN BENFORD'S LAW

Arithmetical Tugs of War, Quantitative Partition Models, Primes Numbers, Exponential Growth Series, and Data Forensics

Alex Ely Kossovsky

This book explores the most recent research results and discoveries in the field of Benford's Law with a strong emphasis on relevant real-life physical, financial, accounting, scientific, and demographic data, while tying in diverse and seemingly unrelated areas of mathematics such as prime numbers, quantitative partition models, and exponential growth series. In addition, the book explores the resultant quantitative and digital configurations of a mixture of arithmetical operations on random variables, such as when additions and multiplications are involved within a single expression of some real-life physical or financial stochastic process. The book also includes innovative techniques in forensic digital analysis in the context of data fraud detection.

ISBN 9781723283257

9 781723 283257 90000

https://www.amazon.com/Benfords-Law-Quantities-Detection-Applications/dp/9814583685

Aug 2014

https://www.amazon.com/Small-Beautiful-Numerous-Rare-World/dp/069291241X

June 2017

https://www.amazon.com/dp/172928325X

April 2019

I take this opportunity to thank the mathematician Steven Miller for his input, insight, and comments, for past 12 years, and for always being ready and available out there virtually via emails.

I also take this opportunity here to thank the physicist Don Lemons for his input, comments, and support. His tiny yet seminal 2-page article from 1986 titled *"On the Number of Things and the Distribution of First Digits"* was an inspiration.

We all owe debt of gratitude to the distinguished mathematician Ted Hill for dedicating himself to the field of Benford's Law longer than anyone else, and for his innovative and thought-provoking approach in the field.