THE LAW OF ANOMALOUS NUMBERS

FRANK BENFORD

Physicist, Research Laboratory, General Electric Company,
Schenectady, New York

(Introduced by Irving Langmuir)

(*Read April 22, 1937*)

ABSTRACT

It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9. A compilation of some 20,000 first digits taken from widely divergent sources shows that there is a logarithmic distribution of first digits when the numbers are composed of four or more digits. An analysis of the numbers from different sources shows that the numbers taken from unrelated subjects, such as a group of newspaper items, show a much better agreement with a logarithmic distribution than do numbers from mathematical tabulations or other formal data. There is here the peculiar fact that numbers that individually are without relationship are, when considered in large groups, in good agreement with a distribution law—hence the name " Anomalous Numbers."

A further analysis of the data shows a strong tendency for bodies of numerical data to fall into geometric series. If the series is made up of numbers containing three or more digits the first digits form a logarithmic series. If the numbers contain only single digits the geometric relation still holds but the simple logarithmic relation no longer applies.

An equation is given showing the frequencies of first digits in the different orders of numbers 1 to 10, 10 to 100, etc.

The equation also gives the frequency of digits in the second, third $\cdots$ place of a multi-digit number, and it is shown that the same law applies to reciprocals.

There are many instances showing that the geometric series, or the logarithmic law, has long been recognized as a common phenomenon in factual literature and in the ordinary affairs of life. The wire gauge and drill gauge of the mechanic, the magnitude scale of the astronomer and the sensory response curves of the psychologist are all particular examples of a relationship that seems to extend to all human affairs. The Law of Anomalous Numbers is thus a general probability law of widespread application.

PART I: STATISTICAL DERIVATION OF THE LAW

IT has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms of the low numbers 1 and 2 are apt to be more stained and frayed by use than those of the higher numbers 8 and 9. Of

PROCEEDINGS OF THE AMERICAN PHILOSOPHICAL SOCIETY,
VOL. 78, NO. 4, MARCH, 1938

course, no one could be expected to be greatly interested in the condition of a table of logarithms, but the matter may be considered more worthy of study when we recall that the table is used in the building up of our scientific, engineering, and general factual literature. There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.

Methods and Terms

Before presenting the data collected while investigating the possible existence of a distribution law that applies to numerical data in general, and to random data in particular, it may be well to define a few terms and outline the method of attack.

First, a distinction is made between a digit, which is one of the nine natural numbers 1, 2, 3, $\cdots$ 9, and a number, which is composed of one or more digits, and which may contain a 0 as a digit in any position after the first. The method of study consists of selecting any tabulation of data that is not too restricted in numerical range, or conditioned in some way too sharply, and making a count of the number of times the natural numbers 1, 2, 3, $\cdots$ 9 occur as first digits. If a decimal point or zero occurs before the first natural number it is ignored, for no attention is to be paid to magnitude other than that indicated by the first digit.

The Law of Large Numbers

An effort was made to collect data from as many fields as possible and to include a variety of widely different types. The types range from purely random numbers that have no relation other than appearing within the covers of the same magazine, to formal mathematical tabulations that admit of no variation from fixed laws. Between these limits one will recognize various degrees of randomness, and in general the title of each line of data in Table I will suggest the nature of the source. In every group the count was continuous from the beginning to the end, or in the case of long tabulations, to a sufficient number of observations to insure a fair average.

The numbers counted in each group is given in the last column of Table I.

TABLE I

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit 1	2	3	4	5	6	7	8	9	Count
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	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	±0.2	±0.2	±0.2	±0.2	±0.3	—

At the foot of each column of Table I the average percentage is given for each first digit, and also the probable error of the average. These averages can be better studied if the decimal point is moved two places to the left, making the sum of all the averages unity. The frequency of first 1's is then seen to be 0.306, which is about equal to the common logarithm of 2. The frequency of first 2's is 0.185, which is slightly greater than the logarithm of 3/2. The difference here, log 3 − log 2, is called the logarithmic integral. These resemblances persist throughout, and finally there is 0.047 to be compared with log 10/9, or 0.046.

The frequency of first digits thus follows closely the logarithmic relation

$$F_a = \log\left(\frac{a+1}{a}\right), \tag{1}$$

where F_a is the frequency of the digit a in the first place of used numbers.

TABLE II

OBSERVED AND COMPUTED FREQUENCIES

Natural Number	Number Interval	Observed Frequency	Logarithm Interval	Observed − Computed	Prob. Error of Mean
1	1 to 2	0.306	0.301	+0.005	±0.008
2	2 to 3	0.185	0.176	+0.009	±0.004
3	3 to 4	0.124	0.125	−0.001	±0.004
4	4 to 5	0.094	0.097	−0.003	±0.003
5	5 to 6	0.080	0.079	+0.001	±0.002
6	6 to 7	0.064	0.067	−0.003	±0.002
7	7 to 8	0.051	0.058	−0.007	±0.002
8	8 to 9	0.049	0.051	−0.002	±0.002
9	9 to 10	0.047	0.046	+0.001	±0.003

There is a qualification to be noted immediately, for Table I was compiled from numbers composed in general of four, five and six digits. It will be shown later that Eq. (1) is a distribution law for *large* numbers, and there is a more general equation that applies when considering numbers of one, two $\cdots$ significant digits.

If we may assume the accuracy of Eq. (1), we then have a probability law of the most general nature, for it is a probability derived from "events" through the medium of their descriptive numbers; it is not a law of numbers in themselves. The range of subjects studied and tabulated was as wide as time and energy permitted; and as no definite exceptions have ever been observed among true variables, the logarithmic law for large numbers evidently goes deeper among the roots of primal causes than our number system unaided can explain.

Frequency of Digits in the qth Position

The second-place digits are ten in number, for here we must take 0 into account. Also, in considering the frequency

F_b of a second-place digit b we must take into account the digit a that preceded it. The logarithmic interval between two digits is now to be divided into ten parts corresponding to the ten digits 0, 1, 2, $\cdots$ 9. Let a be the first digit of a number and b be the second digit; then using the customary meaning of position and order in our decimal system a two-digit number is written ab, and the next greater number is written $ab + 1$.

The logarithmic interval between ab and $ab + 1$ is $\log (ab + 1) - \log \mathrm{ab}$, while the interval covered by the ten possible second-place digits is $\log (a + 1) - \log a$. Therefore the frequency F_b of a second-place digit b following a first-place digit a is

$$F_b = \log \left(\frac{ab + 1}{ab} \right) \Big/ \log \frac{a + 1}{a} \cdot \tag{2}$$

As an example, the probability F_b of a 0 following a first-place 5 in a random number is the quotient

$$F_b = \log \frac{51}{50} \Big/ \log \frac{6}{5} \cdot$$

It follows that the probability for a digit in the qth position is

$$F_b = \frac{\log \dfrac{abc \cdots p\,(q + 1)}{abc \cdots pq}}{\log \dfrac{abc \cdots o\,(p + 1)}{abc \cdots p}} \cdot \tag{3}$$

Here the frequency of q depends upon all the digits that precede it, but when all possible combinations of these digits are taken into account F_q approaches equality for all the digits 0, 1, 2, $\cdots$ 9, or

$$F_q \doteq 0.1. \tag{4}$$

As a result of this approach to uniformity in the qth place the distribution of digits in all places in an extensive tabulation of multi-digit numbers will be also nearly uniform.

TABLE III

FREQUENCY OF DIGITS IN FIRST AND SECOND PLACES

Digit	First Place	Second Place
0	0.000	0.120
1	0.301	0.114
2	0.176	0.108
3	0.125	0.104
4	0.097	0.100
5	0.079	0.097
6	0.067	0.093
7	0.058	0.090
8	0.051	0.088
9	0.046	0.085

Reciprocals

Some tabulations of engineering and scientific data are given in reciprocal form, such as candles per watt, and watts per candle. If one form of tabulation follows a logarithmic distribution, then the reciprocal tabulation will also have the same distribution. A little consideration will show that this must follow for dividing unity by a given set of numbers by means of logarithms leads to identical logarithms with merely a negative sign prefixed.

The Law of Anomalous Numbers

A study of the items of Table I shows a distinct tendency for those of a random nature to agree better with the logarithmic law than those of a formal or mathematical nature. The best agreement was found in the arabic numbers (not spelled out) of consecutive front page news items of a newspaper. Dates were barred as not being variable, and the omission of spelled-out numbers restricted the counted digits to numbers 10 and over. The first 342 street addresses given in the current *American Men of Science* (Item R, Table IV) gave excellent agreement, and a complete count (except for dates and page numbers) of an issue of the *Readers' Digest* was also in agreement.

On the other hand, the greatest variations from the logarithmic relation were found in the first digits of mathematical tables from engineering handbooks, and in tabulations of such closely knit data as Molecular Weights, Specific Heats, Physical Constants and Atomic Weights.

TABLE IV

SUMMATION OF DIFFERENCES BETWEEN OBSERVED AND THEORETICAL FREQUENCIES

Order	Item	Nature	Difference	Order	Item	Nature	Difference
1	D	Newspaper Items	2.8	11	N	Cost Data, Concrete	12.4
2	F	Pressure Lost, Air Flow	3.2	12	S	$n^1 \cdots n^8$, $n!$	13.8
3	G	H.P. Lost in Air Flow	4.8	13	L	Design Data Generators	16.6
4	R	Street Addresses, A.M.S.	5.4	14	B	Population, U. S. A.	16.6
5	P	Am. League, 1936	6.6	15	I	Drainage Rate of Rivers	21.6
6	Q	Black Body Radiation	7.2	16	K	n^{-1}, $\sqrt{n} \cdots$	22.8
7	O	X-Ray Voltage	7.4	17	H	Molecular Wgts.	23.2
8	M	*Readers' Digest*	8.4	18	E	Specific Heats	24.2
9	A	Area Rivers	9.8	19	C	Physical Constants	34.9
10	T	Death Rates	11.2	20	J	Atomic Weights	35.4

These facts lead to the conclusion that the logarithmic law applies particularly to those outlaw numbers that are without known relationship rather than to those that individually follow an orderly course; and therefore the logarithmic relation is essentially a Law of Anomalous Numbers.

PART II: GEOMETRIC BASIS OF THE LAW

The data so far considered have been composed entirely of *used* numbers; that is, numbers as they are used in everyday affairs. There must be some underlying causes that distort what we call the "natural" number system into a logarithmic distribution, and perhaps we can best get at these causes by first examining briefly the frequency of the natural numbers themselves when arranged in the infinite arithmetic series $1, 2, 3, \cdots n$, where n is as large as any number encountered in use.

Let us assume that each individual number in the natural number system up to n is used exactly as often as every other individual number. Starting with 1, and counting up to

10,000, for example, 1 would have been used 1,112 times, or 11.12 per cent of all uses. If the count is extended to 19,999 there are 9,999 1's added, and first 1's occur in 55.55 per cent of the 19,999 numbers. When number 20,000 is reached there is a temporary stopping of the addition of first 1's and 90,000 of the other digits are added to the series before

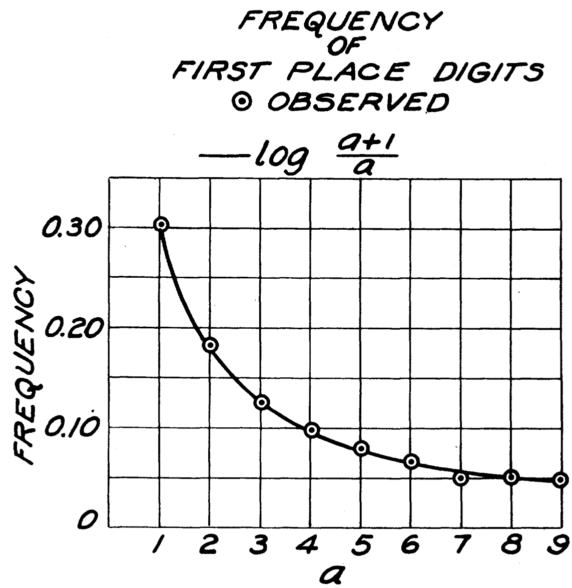

FIG. 1. Comparison of observed and computed frequencies for multi-digit numbers.

1's are again brought into the series at 100,000. At this point the percentage of 1's is again reduced to 11.112 per cent as illustrated in curve A of Fig. 2. This curve is F_n and log n plotted to a semi-logarithmic scale. If the equations for A are written for the three discontinuous but connected sections 10,000–20,000, 20,000–99,999 and 99,999–100,000 the area under the curve will be very closely 0.30103, where the entire area of the frame of coördinates has an area 1. But an integration by the methods of the calculus is merely a quick way of adding up an infinite number of *equally* spaced ordinates to the curve and from this addition finding the average height of the ordinates and hence the area under the curve. But if we are satisfied with a result somewhat short of the perfection of the integral calculus we may take a finite number of *equally* spaced ordinates and by plain arithmetic come to practically the same answer. By definition each point of A represents

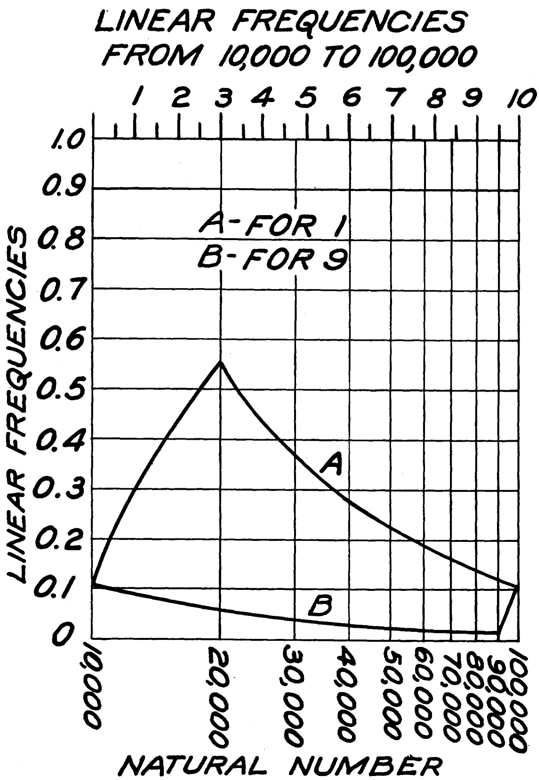

FIG. 2. Linear frequencies of the natural number system between 10,000 and 100,000.

the frequency of first 1's from 1 up to that point, and an integration (by calculus or arithmetic) under curve A gives the *average* frequency of first 1's up to 100,000. The finite number corresponding to equally spaced ordinates now represents a geometric series of numbers from 10,000 to 100,000, and it is substantially this series of numbers, in this and other orders of the natural number scale that lead to the numerical frequencies already presented.

Curve B of Fig. 2 is for 9 as a first digit. The frequency of 9's decreases in the number range from 10,000 to 89,999 and then increases as 9's are added from 90,000 to 99,999, and an integration under curve B leads to a good numerical approximation to the logarithmic interval $\log 10 - \log 9$, as called for by the previous statistical study.

Geometric and Logarithmic Series

The close relationship of a geometric series and a logarithmic series is easily seen and hardly needs formal demonstration. The uniformly spaced ordinates of Fig. 2 form a geometric series of numbers for these numbers have a constant factor between adjacent terms, and this constant factor is determined in size by the constant logarithmic increment.

Semi-Log Curves

A geometric series of numbers plotted to a semi-logarithmic scale gives a straight line. In the original tabulation of observed numbers the line of data marked "R" is designated simply as "street addresses." These are the street addresses of the first 342 people mentioned in the current *American Men of Science*. The randomness of such a list is hardly to be disputed, and it should therefore be useful for illustrative purposes.

In Fig. 3 these addresses are first indicated by the height of the lines at the base of the diagram. The height of a line, measured on the scale at the left, indicates the number of addresses at, or near, that street number. Thus there were five addresses at No. 29 on various streets. In order to make the trend clearer, the heights of these lines were summed, beginning at the left and proceeding across to the right. It was found that four straight lines could be drawn among these summation points with fair fidelity of trend, and these four lines represent four geometric series, each with a different factor between terms. Each line will give the observed fre-

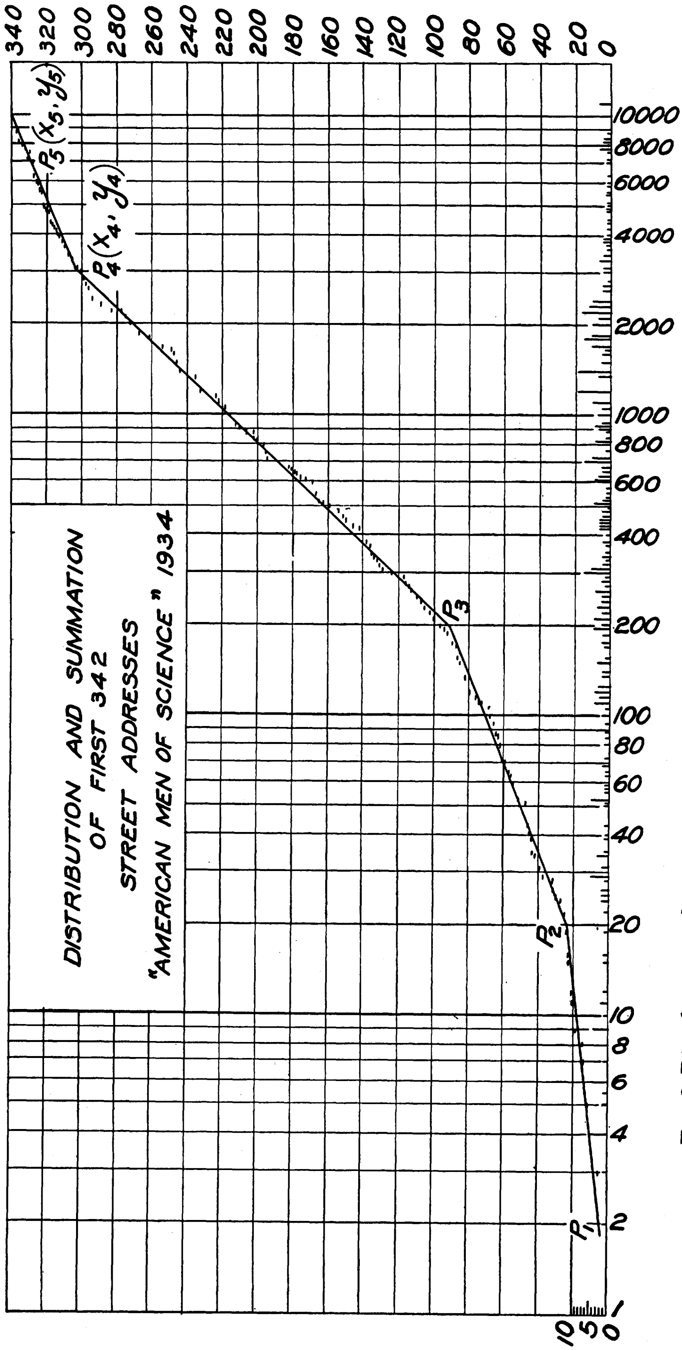

FIG. 3. Distribution and summation of first 342 street addresses, *American Men of Science*, 1934.

quency over the numerical range it covers, and hence satisfies the logarithmic relationship.

The Natural Numbers and Nature's Numbers

In natural events and in events of which man considers himself an originator there are plenty of examples of geometric or logarithmic progressions. We are so accustomed to labeling things 1, 2, 3, 4, $\cdots$ and then saying they are in natural order that the idea of 1, 2, 4, 8, $\cdots$ being a more natural arrangement is not easily accepted. Yet it is in this latter manner that a surprisingly large number of phenomena occur, and the evidence for this is available to everyone.

First, let us consider the physiological and psychological reaction to external stimuli.

The growth of the sensation of brightness with increasing illumination is a logarithmic function, as illustrated by Fechner's Law. The growth of sensation is slow at first while the rods of the retina are alone responsive, and a straight line on semi-logarithmic paper (the stimulus being on the logarithmic scale) can represent the intensity-brightness function in this region. When the cones come into action there is a sharp change in the rate of growth, and another straight line represents our working range of vision. When over-excitation and fatigue set in, a third line is needed; and thus three geometric series could be used to state the relation between illumination and the sensation of brightness. If the literature contained sufficient numerical references, the brightness function should give an extremely close approximation of the logarithmic law of distribution.

The sense of loudness follows the same rules, as does the sense of weight; and perhaps the same laws operate to make the sense of elapsed time seem so different at ages ten and fifty.

Our music scales are irregular geometric series that repeat rigidly every octave.

In the field of medicine, the response of the body to medicine or radiation is often logarithmic, as are the killing curves under toxins and radiation.

In the mechanical arts, where standard sizes have arisen from years of practical experience, the final results are often geometric series, as witness our standards of wire diameters and drill sizes, and the issued lists of "preferred numbers."

The astronomer lists stars on a geometric brightness scale that multiplies by 100 every five steps and the illuminating engineer adopts the same type of series in choosing the wattage of incandescent lamps.

In the field of experimental atomic physics, where the results represent what occurs among groups of the building units of nature, and where the unit itself is known only by mass action, the test data are statistical averages. The action of a single atom or electron is a random and unpredictable event; and a statistical average of a group of such events would show a statistical relationship to the results and laws here presented. That this is so is evidenced by the frequent use made of semi-log paper in plotting the test data, and the test points often fall on one or more straight lines. The analogy is complete, and one is tempted to think that the 1, 2, 3, $\cdots$ scale is not the natural scale; but that, invoking the base e of the natural logarithms, *Nature* counts

$$e^0, \quad e^x, \quad e^{2x}, \quad e^{3x}, \quad \cdots$$

and builds and functions accordingly.

Part III. Digital Orders of Numbers

The natural number system is an array of numbers in simple arithmetic series, but on top of this we have imposed an idea taken from a geometric series. Numbers composed of many digits are ordinarily separated into groups of three digits by interposing commas, and here we unknowingly give evidence of the use of these numbers on a geometric scale.

For convenience of description the natural numbers 1 to 10 are called the first digital order numbers, those from 10 to 100 the second digital order, etc. It will be noted that 10 is both the last number of the first order and the first number of the second order, and when an integration is carried out, as will be done later, 10 appears as both an upper and a lower limit, and it is thus used in this case as a boundary line rather than a unit zone in the natural number system.

In Fig. 4 the curves show the frequency with which the natural numbers occur in the Natural Number System, beginning at the left edge, where 1 is the only number, its frequency is 1; that is, until a second number is added 1 is the entire number system. When 2 is reached the frequency is 0.50 for 1

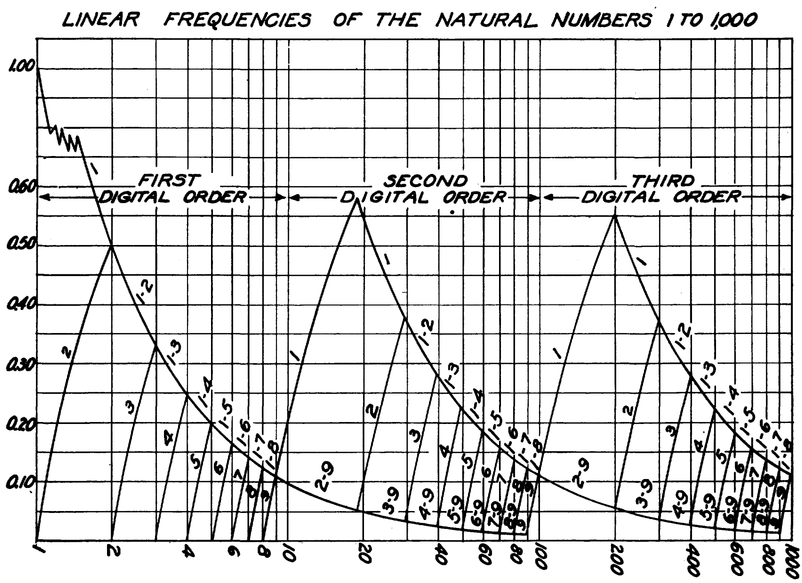

FIG. 4. Linear frequencies of the natural numbers in the first three orders.

and 0.50 for 2. At 5, for example, the frequency for each of the first 5 digits is 0.20, and the equal division continues until 9 is reached. At 10, the digit 1 has appeared twice and has a frequency of 0.20 against 0.10 for each of the other eight digits that have appeared but once.

It will be observed that the curve rising from 9 on the scale of abscissæ is for only the digit 1, while the curve continuing downward from 9 is for the digits 2 to 9 inclusive. At 19 the frequency curve for 2 rises to join the curve for 1 at 29 and 1 and 2 have a common curve until 99 is reached and a third first 1 is about to be added to the series. At any ordinate the curves therefore tell the frequency of the total number of natural numbers up to that point.

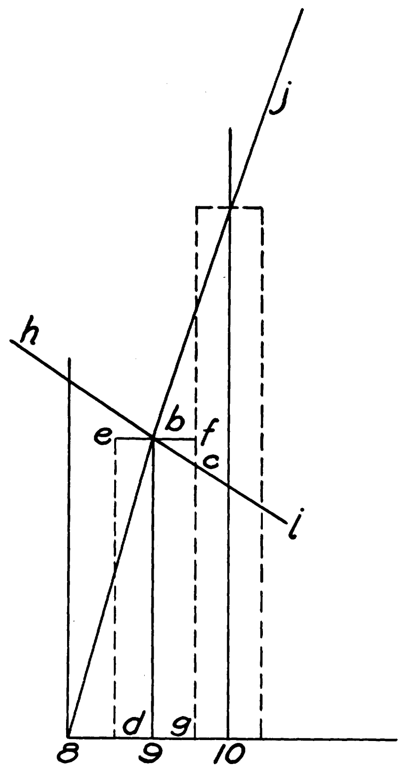

FIG. 5. Continuous and discontinuous functions in the neighborhood of the digit 9.

The curves are drawn as if we were dealing with continuous functions in place of a discontinuous number system. The justification for using a continuous form is that the things we use the number system to represent are nearly always perfectly continuous functions, and the number, say 9, given to any phenomenon will be used in some degree for all the infinite sizes of phenomena between 8 and 10 when we confine ourselves to single digit numbers.

An enlarged sketch of the linear frequency curves at the junction of the first and second orders is given in Fig. 5. The lines h–b and b–j are the computed ratios of 1 in this region, while the lines 8–b for the ratio of 9 begins at 8, for as soon as size 8 is passed there is a possibility of our using a 9, while for

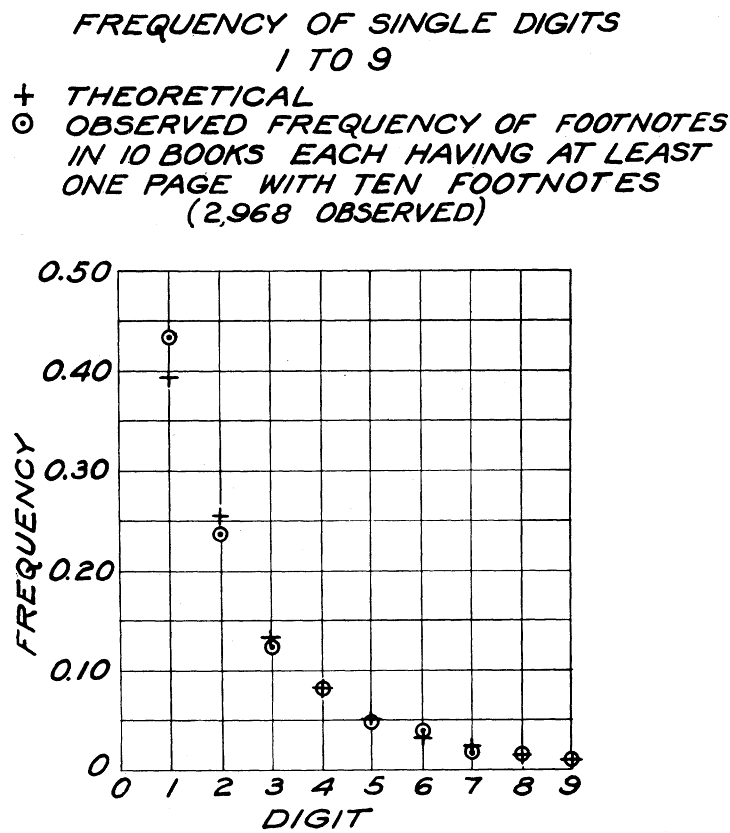

FIG. 6. Theoretical and observed frequencies of single digits.

size $8\frac{1}{2}$ the chances are about equal for calling it either 8 or 9. The summation of area under the curve 8–b–c is taken as the probability of using a 9 for phenomena in this region. This is about equivalent to knowing accurately the size of all phenomena in this region and deciding to call everything between 8.5

and 9.5 by the number 9. Once 9 is passed the curve for 1, *b–j*, begins to rise in anticipation of the phenomena between 9 and 10 that will be called 10.

It has been noted that for high orders of numbers the areas under the curves of Fig. 2 are proportional to the frequency of use of the first digit. The same demonstration will now be made with the aid of the calculus in regions that are markedly discontinuous.

Selecting the third digital order, Fig. 4, the area under the 1-curve can be written

$$A_1''' = \int_{100}^{199} y_1\,dx + \int_{199}^{999} y_2\,dx + \int_{999}^{1000} y_3\,dx, \tag{5}$$

where the ordinates of the first rising section of the curve are

$$y_1 = \frac{a - 88}{a}. \tag{6}$$

The descending section of the curve has ordinates

$$y_2 = \frac{111}{a} \tag{7}$$

and the last rising section between 999 and 1,000 has ordinates

$$y_3 = \frac{a - 888}{a}. \tag{8}$$

The curves are plotted to semi-logarithmic coördinates and

$$x = \log a, \tag{9}$$

$$dx = da/a. \tag{10}$$

The integrals after making these substitutions give the value

$$A_1''' = \log_e \frac{1990}{999} + \frac{8}{1000}.$$

A similar operation yields for the 1-curve in the second digital order

$$A_1'' = \log_e \frac{190}{99} + \frac{8}{100}$$

and in the first order

$$A_1' = \log_e \frac{10}{9} + \frac{8}{10}.$$

From the symmetry running through these solutions and from the solutions for the eight other first digits, we can write the general equation for the Law of Anomalous Numbers

$$\left.\begin{aligned} F_1{}^r &= \left[\log_e \frac{10\ (2\cdot 10^{r-1} - 1)}{10^r - 1} + \frac{8}{10^r} \right] \frac{1}{N} \\ \underset{a\neq 1}{F_a{}^r} &= \left[\log_e \frac{(a+1)\ 10^{r-1} - 1}{a\ 10^{r-1} - 1} - \frac{1}{10^r} \right] \frac{1}{N} \end{aligned}\right\} \tag{11}$$

where $N = \log_e 10$ is the factor to convert the expressions from the natural logarithm system, base e, to the common logarithm system, base 10.

If high orders of r are considered, as was unwittingly done in the original statistical work, these expressions simplify by dropping the terms -1 in both numerator and denominator, and the numerical terms having 10^r in the denominator become negligible. Hence the general equations become

$$F_1{}^r = \log_{10} \frac{2}{1}, \tag{12}$$

$$\underset{a\neq 1}{F_a{}^r} = \log \frac{a+1}{a}, \tag{13}$$

but these two expressions no longer have a difference in form, and they may be merged into

$$F_a{}^r = \log_{10} \frac{a+1}{a}, \tag{14}$$

which was the relationship originally observed for multi-digit numbers.

In Table V numerical values are given for the theoretical frequencies of used numbers for the first, second, third and limiting digital orders.

TABLE V

THEORETICAL FREQUENCIES IN VARIOUS DIGITAL ORDERS

First Digit	First Order	Second Order	Third Order	Limiting Order
	1 to 10	10 to 100	100 to 1000	—
1	0.39319	0.31786	0.30276	0.30103
2	0.25760	0.17930	0.17638	0.17609
3	0.13266	0.12432	0.12487	0.12494
4	0.08152	0.09479	0.09669	0.09691
5	0.05348	0.07631	0.07889	0.07918
6	0.03575	0.06366	0.06662	0.06695
7	0.02352	0.05444	0.05764	0.05799
8	0.01456	0.04742	0.05078	0.05115
9	0.00772	0.04190	0.04537	0.04576

The frequencies of the single digits 1 to 9 vary enough from the frequencies of the limiting order to allow a statistical test if a source of digits used singly can be found. The footnotes so commonly used in technical literature are an excellent source, consisting of units that are indicated by numbers, letters or symbols.

The procedure of collecting data for the first-order numbers was to make a cursory examination of a volume to see if it contained as many as 10 footnotes to a page, for obviously no test of the range 1 to 9 could be made if the maximum number fell short of the full range. The numbers here recorded in Table VI are the number of footnotes observed on consecutive pages, beginning on page 1 and continuing to the end of the book, or until it seemed that a fair sample of the book had been obtained. The books used were the *Standard Handbook for Electrical Engineers*, *Smithsonian Physical Tables*, *Handbuch der Physik* and Glazebrook's *Dictionary of Applied Physics*.

In Table VI the observed percentages of single digits 1 to 9 are given along with the number of pages used in each volume and the number of footnotes observed. The frequency for 1 is seen to be 43.2 per cent as against the theoretical frequency of 39.3 per cent, and for the digit 9 the observations agree with theory with $F_9' = 0.8$ per cent.

In general the agreement with theory is as good as the computed probable errors of the observation.

TABLE VI

COUNT OF FOOTNOTES

Volume	Pages Used	1	2	3	4	5	6	7	8	9	Total Count
		Frequencies, in Per Cent									
1. *S. H. E. E.*	All	55.1	22.7	12.3	5.0	2.4	1.7	0.3	0.3	0.3	586
2. *Sm. Phy. Ta.*	All	56.3	22.1	6.6	6.1	5.0	2.2	1.1	0.6	0.0	181
3. *H. der Phy.*	360	52.8	23.6	8.5	5.5	4.0	3.2	0.8	0.0	1.6	127
4. *H. der Phy.*	360	37.2	25.7	12.1	9.5	4.8	5.2	2.6	2.2	0.9	230
5. *H. der Phy.*	365	29.7	26.6	14.6	11.0	8.0	5.9	1.8	1.0	1.4	287
6. *H. der Phy.*	361	19.5	17.4	17.7	11.9	11.3	9.2	6.1	5.8	1.1	293
7. *H. der Phy.*	360	33.0	27.5	11.8	10.7	4.3	5.9	2.8	2.4	1.6	254
8. *H. der Phy.*	360	56.8	23.2	6.7	7.6	2.4	1.4	0.5	1.4	0.0	211
9. Glazebrook I	All	49.5	22.3	13.7	6.9	2.3	1.5	1.5	1.5	0.8	394
10. Glazebrook V	All	41.7	25.2	13.4	9.1	4.7	3.2	1.7	0.5	0.5	405
Observed Ave.		43.2	23.6	11.8	8.3	4.9	3.9	1.9	1.6	0.8	2968
Predicted Ave.		39.3	25.7	13.3	8.1	5.3	3.6	2.4	1.5	0.8	
Difference		+3.9	−2.1	−1.5	+0.2	−0.4	+0.3	−0.5	+0.1	0.0	
Probable Error		±3.0	±0.6	±0.7	±0.5	±0.6	±0.5	±0.4	±0.4	±0.4	

Summation of Frequencies

One of the conditions that must be met by these expressions for the frequencies of the integers is that, in any one order, the sum of the frequencies must equal unity; that is, the sum of their probabilities must equal certainty.

Selecting the first-order digits, Eq. 11, and remembering the logarithmic rule that the sum of the logarithms of a group of numbers is equal to the logarithm of their combined products, we have the probability P'

$$P' = \log_{10} \frac{10 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9}{9 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8} + \left[\frac{8}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} - \frac{1}{10} \right] \frac{1}{N},$$

which reduces to

$$\begin{aligned} P' &= \log_{10} 10 + 0 \\ &= 1. \end{aligned}$$

In a similar manner from the complete set of equations indicated by Eq. 11 we have

$$\begin{aligned} P'' &= \log_{10} \frac{190 \cdot 29 \cdot 39 \cdot 49 \cdot 59 \cdot 69 \cdot 79 \cdot 89 \cdot 99}{99 \cdot 19 \cdot 29 \cdot 39 \cdot 49 \cdot 59 \cdot 69 \cdot 79 \cdot 89} \\ &\quad + \left[\frac{8}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} - \frac{1}{100} \right] \frac{1}{N} \\ &= \log_{10} 10 + 0 \\ &= 1 \end{aligned}$$

and similar proof can be worked out for the other orders.

Summary of Part III

Single digits, regardless of their relation to the decimal point and also regardless of preceding or following zeros, have a specific natural frequency that varies sharply from the logarithmic ratios. The second digital order, which is composed of two adjacent significant digits, has a specific frequency approximating the logarithmic frequency; and for three or more associated digits the variation from the latter frequency would be extremely difficult to find statistically.

The basic operation

$$F = \int \frac{da}{a}$$

or

$$F = \sum \frac{\Delta a}{a}$$

in converting from the linear frequency of the natural numbers to the logarithmic frequency of natural phenomena and human events can be interpreted as meaning that, on the average, these things proceed on a logarithmic or geometric scale. Another way of interpreting this relation is to say that small things are more numerous than large things, and there is a tendency for the step between sizes to be equal to a fixed fraction of the last preceding phenomenon or event. There is no necessity or implication of limits at either the upper or the lower regions of the series.

If the view is accepted that phenomena fall into geometric series, then it follows that the observed logarithmic relationship is not a result of the particular numerical system, with its base, 10, that we have elected to use. Any other base, such as 8, or 12, or 20, to select some of the numbers that have been suggested at various times, would lead to similar relationships; for the logarithmic scales of the new numerical system would be covered by equally spaced steps by the march of natural events. As has been pointed out before, the theory of anomalous numbers is really the theory of phenomena and events, and the *numbers* but play the poor part of lifeless symbols for living things.