Note on the Frequency of Use of the Different Digits in Natural Numbers.

By Simon Newcomb.

That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9. The question naturally arises whether the reverse would be true of logarithms. That is, in a table of anti-logarithms, would the last part be more used than the first, or would every part be used equally? The law of frequency in the one case may be deduced from that in the other. The question we have to consider is, what is the probability that if a natural number be taken at random its first significant digit will be n, its second n', etc.

As natural numbers occur in nature, they are to be considered as the ratios of quantities. Therefore, instead of selecting a number at random, we must select two numbers, and inquire what is the probability that the first significant digit of their ratio is the digit n. To solve the problem we may form an indefinite number of such ratios, taken independently; and then must make the same inquiry respecting their quotients, and continue the process so as to find the limit towards which the probability approaches.

Let us suppose the numbers with which we commence to be arranged in periods according to the number of their digits, or, which is the same thing, according to the characteristics of their logarithms on the scale of which the basis is i, (i being 10 in the common system). Then, if two numbers are i^{c+s} and $i^{c'+s}$, c and c' being integers, the significant figures of the ratio will be independent of c and c', since changing these integers will only change the decimal point. We may, therefore, take both numerator and denominator of the ratio out of the same period.

Moreover, since both numerator and denominator are formed by the same process, we may suppose the law of distribution of the numbers from which they are selected to be the same. Our problem is thus reduced to the following:

We have a series of numbers between 1 and i, represented by fractional powers of i, say i^s, the distribution of the exponents s, and therefore of the numbers, being according to any arbitrary law. Since these exponents are formed by casting off all the integers from a series of numbers, we may suppose them arranged around a circle according to some law. Then, if we select 2^n exponents at random and call them s', s'', s''', etc., the final ratio, obtained in the manner we have described, will be

$$i^{s'-s''+s'''-s''''+\text{ etc.}}$$

The question is, what is the probability that the positive fractional portion of $s'-s''+s'''-s''''+$ etc., will be contained between the limits s and $s+ds$. It is evident that, whatever be the original law of arrangement, the fractions will approach to an equal distribution around the circle as n is increased, or the required probability will be equal to ds. But, the fractional part of $s'-s''+s'''-$ etc. is the mantissa of the logarithm of the limiting ratio. We thus reach the conclusion:

The law of probability of the occurrence of numbers is such that all mantissæ of their logarithms are equally probable.

In other words, every part of a table of anti-logarithms is entered with equal frequency. We thus find the required probabilities of occurrence in the case of the first two significant digits of a natural number to be:

Dig.	First Digit.	Second Digit.
0	. . .	0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

In the case of the third figure the probability will be nearly the same for each digit, and for the fourth and following ones the difference will be inappreciable.

It is curious to remark that this law would enable us to decide whether a large collection of independent numerical results were composed of natural numbers or logarithms.