ON APPLICATION OF INFORMATION-THEORETIC NOTIONS IN THE THEORY OF NUMBERS

by M. Dutta, *A/31 C.I.T. Buildings, Singheebagan, Calcutta 7,*
and M. Sen, *Department of Pure Mathematics, University of Calcutta,*
35 Ballygunge Circular Road, Calcutta 19

(Communicated by S. N. Bose, F.N.I.)

(*Received June 30, 1966*)

In the decimal representation of numbers the occurrences of particular digits may be looked upon as independent random events; and hence entropy can be defined for real numbers after Shannon. Such an introduction of the concept of entropy for real numbers makes some notions like the definitions associated with normalcy very simple and the proofs of some well-known propositions easy, straightforward and neat. Moreover, it provides some measure of normalcy which may yield some new interesting problems in the theory of numbers.

1. Introduction

The notion of entropy was first introduced by Clausius in thermodynamics in 1865 (Zeise 1944). Boltzmann's work (Boltzmann 1923) particularly his discussions of the H-theorem and of the principle named after him demonstrated the statistical significance of entropy. The fundamental role of entropy in classical thermodynamics and also its statistical significance became evident in subsequent developments due to Gibbs (1902), Planck (1887, 1927) and others.

In a mathematical development of communication theory, Shannon (1948) defined entropy in terms of the probability of events without restricting the notion of events in any way. This generalization of the notion of entropy stimulated a good deal of research among a large number of workers in various fields of knowledge. The major part of this development is generally referred to as Information Theory (Khinchin (1957)). In this paper, first we shall look upon the decimal representation of a number between 0 and 1 in the scale of r as a representation of a suitable random event, so that we can define entropy for a number, after Shannon. After defining entropy, the usual methods of Information Theory and also of statistical thermodynamics will be used conveniently. Most of the known theorems connected with the decimal representations of numbers will be deduced in this way. Also we shall try to formulate in terms of this entropy, some problems which have not been formulated or discussed earlier in the literature (Hardy and Wright 1960; Niven 1956).

2. Interpretation of Decimal Representations of Real Numbers as Sample Points of a Random Event

Let us consider a random event with r possible outcomes which are indexed by numbers 0, 1, 2, . . . , $r-1$. Let us consider repeated trials and let us assume that the trials are independent with respect to the event (Dutta and Pal 1963). Then the outcomes of the trials, repeated for an arbitrary number of times, will be represented by the digits of a decimal representation of real numbers between 0 and 1 in the scale r. In the case of Bernoulli's trials evidently the outcomes will be represented by the decimals in binary scale where 1 stands for success and 0 for failure. If the trials are repeated finitely we shall get finite decimals; otherwise we shall get infinite decimals. For convenience, a finite decimal representation of a number will be looked upon as a representation in infinite decimals in which zero is repeated as an infinite number of times. On the assumption of the principle of statistical regularity (Cramer 1949; Dutta and Pal 1963) we can speak of the probability of occurrence of each of the digits in the decimal representation.

3. Entropy of a Number

Now as the occurrence of a particular digit may be looked upon as a random event, we can speak of the probability of occurrence of ith digit as ${}_rp_i$. Hence, after Shannon the entropy of a number x represented in the r-ary decimal is

$$ {}_rH(x) = -\Sigma_r p_i \log {}_rp_i \quad \ldots \quad (1) $$

where ${}_rp_i$ is the probability of occurrence of the ith digit in the decimal representation of x in the scale of r and i ranges from 0 to $r-1$ (Khinchin 1957). For convenience as usual we shall consider the terms in the expression associated with the missing digits as zero.

Remark.—The total entropy of a number represented by finite decimal, i.e. according to our convention of an infinity of zeros coming after the finite places, is zero, since probability of occurrence of the digit 0 is 1 and for every other of digit the probability of occurrence is zero. Similarly, the entropy of a number with an infinite succession of any one particular digit is zero.

4. Standard Numbers

The probability distribution of the occurrence of digits in finite places in the decimal representation of a real number x in an r-ary scale is given by

$$ {}_rP(x, n) = \pi_r p_i^{n_i} \quad \ldots \quad (2) $$

where

$$ \Sigma n_i = n ; \quad \ldots \quad (3) $$

n_i is the frequency of the actual occurrence of the ith digit and ${}_rp_i$ is its probability in the decimal representation.

Definition.—A real number x is said to be a standard number if its probability distribution of occurrence of digits in first n finite places (for sufficiently large n) in the decimal representation of x is given by

$$_{r}P(x, n) = e^{-n_{r}H(x)} \quad \ldots \quad (4)$$

where $_{r}H(x)$ is expressed in terms of logarithms to the base e.

Proposition 1.—The set of numbers which are not standard is of measure zero.

PROOF.—The occurrence of a particular digit may be looked upon as an independent random event. Then, as a consequence of the central statistical theorem of Glivenko-Cantelli (Loeve 1960), the probability distribution converges in probability to the right-hand side of the relation (4) (Dutta 1966). Consequently, the set of numbers which are not standard are of measure zero.

5. Maximum Entropy and Simply Normal Number

From Information Theory it is known that for fixed r, $_{r}H(x)$ is maximum when the events are equally likely, i.e. when

$$_{r}p_0 = {_{r}p_1} = \ldots = {_{r}p_{r-1}}.$$

Now since $\sum_{i=0}^{r-1} {_{r}p_i} = 1$, therefore $_{r}p_i = \frac{1}{r}$, where i runs from 0 to $r-1$.

Definition.—In the decimal representation of a real number, a number is said to be simply normal in the scale of r if the probability of occurrence of each digit in the representation is $\frac{1}{r}$.

Hence, from the above discussions we get the following proposition.

Proposition 2.—The entropy of a number is maximum if and only if it is a simply normal number.

Note:

$$_{r}H(x) = -\Sigma_{r}p_i \log {_{r}p_i} = -\Sigma \frac{1}{r} \log \frac{1}{r}$$
$$= \log r$$

when

$$p_i = \frac{1}{r}.$$

It can be easily seen that the probability of the occurrence of a particular digit in the finite places in a decimal representation is $\frac{1}{r}$ and that it follows a binomial distribution. Therefore the number of all whose digits in an r-ary decimal representation have the same probability $\frac{1}{r}$ will have the maximum entropy.

Proposition 3.—All simply normal numbers are standard numbers in the scale r.

PROOF.—Let x be a simply normal number. Let us consider the first n digits of the decimal representation of x

$$\therefore \quad {}_rP(x, n) = \frac{1}{r^n}$$

$$_rH(x) = \log r$$

$$\therefore \quad r = e^{{}_rH(x)}$$

$$\therefore \quad {}_rP(x, n) = r^{-n} = e^{-n_rH(x)}.$$

Therefore x is a standard number. Hence the proposition.

Proposition 4.—Almost all standard numbers are simply normal numbers.

PROOF.—Let x be any standard number. Then we have

$$_rP(x, n) = e^{-n_rH(x)}.$$

Now $_rH(x)$ is a function of $_rp_i$'s ($_rp_i$ is the probability of occurrence of ith digit in the first n places of the decimal representation of x). If we expand $_rH(x)$ by Taylor's theorem about the values $_rp_i = \frac{1}{r}$, then we find that

$$_rH(x) = {}_rH_{\max}(x) + \sum_{i,j} \frac{\partial^2{}_rH}{\partial_rp_i\,\partial_rp_j}\left(p_i - \frac{1}{r}\right)\left(p_j - \frac{1}{r}\right) + \text{higher order terms,}$$

where the values of the partial derivatives are calculated for the values of p_i's equal to $\frac{1}{r}$.

Again,

$$_rH_{\max}(x) = \log r.$$

Therefore we can write

$$_rP(x, n) = C \,.\, r^{-n} e^{-\frac{1}{1/n}\sum_{i,j} \frac{\partial^2{}_rH}{\partial_rp_i\partial_rp_j}\left(p_i - \frac{1}{r}\right)\left(p_j - \frac{1}{r}\right)} + \text{higher order terms where}$$

C is a normalizing factor.

Or

$$r^n{}_rP(x, n) = C \,.\, e^{-\frac{1}{1/n}\sum_{i,j} \frac{\partial^2{}_rH}{\partial_rp_i\partial_rp_j}\left(p_i - \frac{1}{r}\right)\left(p_j - \frac{1}{r}\right)} + \cdots$$

Now $r^n{}_rP(x, n)$ is the total number of the different possibilities of the occurrence of the digits in the first n places of the standard numbers with $_rp_i$'s as the probability of occurrence of ith digit. Let us denote $r^n{}_rP(x, n)$ by $N_r(p_0, p_1, \ldots p_{r-1})$.

Now as $n \to \infty$ we get (Gelfond and Schilow 1960)

$$N_r(p_0, \ldots, p_{r-1}) = C\delta\left(\vec{P} - \vec{P}_0\right)$$

where $\vec{P}_0 = \left(\frac{1}{r}, \frac{1}{r}, \ldots \frac{1}{r}\right)$, $\vec{p} = (p_0, \ldots, p_{r-1})$ and δ is the r-dimensional Dirac δ-function (Dirac 1943).

Now, if both sides are integrated for all values of $\vec{p}$'s then the r.h.s. is equal to C and the l.h.s. denotes the measure of all standard numbers. Hence, C denotes the measure of all standard numbers as

$$\int_{R-d} \delta\left(\vec{p}-\vec{p}_0\right) \Delta \vec{p} = 0$$

where R denotes the entire range of integration over $\vec{p}$ and d is any arbitrary small domain about $\vec{p}_0$; so, the measure of all standard numbers which are not simply normal numbers is zero.

Proposition 5.—Almost all real numbers in the range 0 to 1 are simply normal numbers.

PROOF.—From Proposition 1 we find that the set of all non-standard numbers is of measure zero. Also from Proposition 4, we get that the measure of all standard numbers which are not simply normal numbers is zero. Hence the proposition.

Proposition 6.—Almost all real numbers are simply normal numbers.

PROOF.—The entire set of real numbers is the union of enumerable sets of unit intervals in each of which the set of numbers which are not simply normal is of measure zero and so is the case of the set of all real numbers. Hence the proposition.

6. NORMAL NUMBERS

Definition (A).—A real number x is said to be normal in the scale of r if $x, rx, r^2x, \ldots$ are respectively simply normal in the scale $r, r^2, r^3, \ldots$ (Hardy and Wright 1960; Niven 1956).

Definition (B).—If a real number x is simply normal in all the scales $r, r^2, r^3, \ldots$ then it is normal in the scale r.

Proposition 7 (S. S. Pillai's Theorem).—The two definitions (A) and (B) are equivalent.

PROOF.—

$${}_{r^k}H(r^{k-1}x) = r^k H(r^{k-1}) + {}_{r^k}H(x) = {}_{r^k}H(x).$$

So x is simply normal in the scale of r^k if $r^{k-1}x$ is so. Hence the proposition.

Note: The equivalence of the definitions (A) and (B) was first proved by S. S. Pillai (1940). Here we have proved Pillai's theorem simply by using the well-known theorem of entropy.

Proposition 8.—Almost all real numbers in the interval (0, 1) are normal numbers to a given scale r.

PROOF.—The set of normal numbers is the intersection of the sets of simply normal numbers in the scale of $r, r^2, \ldots, r^x, \ldots$

Let us denote it by $\mathcal{S}_{N,r}$ and the set of simply normal numbers in the scale r^j by S_{N,r^j}

$$\mathcal{S}_{N,r} = \bigcap_{j=1,2,\ldots} S_{N,r^j}$$

$$\therefore \quad {}_c\mathcal{S}_{N,r} = \bigcup_j {}_cS_{N,r^j}$$

where c in the subscript to the left indicates the operation of complementation. As each of ${}_cS_{N,r^j}$ is of measure zero, therefore ${}_c\mathcal{S}_{N,r}$ is of measure zero, i.e. almost all numbers are normal.

Proposition 9.—Almost all real numbers are normal to a given scale r.

The proof is similar to that for Proposition 6.

7. Absolutely Normal Numbers

Definition.—A real number is said to be absolutely normal number if it is a simply normal number to every scale r (Niven 1956).

Remark.—Obviously, all absolutely normal numbers are normal numbers, but the converse appears to be an open question still.

Proposition 10.—Almost all real numbers are absolutely normal in the interval (0, 1).

PROOF.—Let $\mathcal{S}_N$ be the set of absolutely normal numbers and $S_{N,r}$ that of normal numbers in the scale r.

Now

$$\mathcal{S}_N = \bigcap_{r=1}^{\infty} S_{N,r} \quad \therefore \quad {}_cS_N = \bigcup_{r=1}^{\infty} {}_cS_{N,r}$$

${}_cS_N$, being a union of enumerable aggregate of sets of measure zero, is of measure zero. So, S_N is of measure zero. So, almost all real numbers are absolutely normal.

Proposition 11.—Almost all real numbers are absolutely normal.

The proof is similar to that for Proposition 6.

8. Entropy and Normalcy of Numbers

In Information Theory, the base of logarithm in the definition of entropy is arbitrary. In the mathematical theory of communication this is taken as 2 in the majority of problems. Now for the applications in number theory, particularly in the theory of decimal representations of numbers in the scale of r, we get the simplest expression for entropy if we take r, the scale of notation, as the base of logarithm. If we stick to this convention, the entropy of real numbers ranges from 0 to 1 and the property of simple normalcy for numbers is associated with that of having the value of entropy equal to one for given r.

Thus, the properties normalcy and of absolute normalcy are associated with the invariance of this property of the value of the entropy being one under the set of transformation of the scale of notations r in the set $\{r^k, k = 1, 2, \ldots\}$ and in the set $\{r', r' = 1, 2, \ldots\}$ respectively.

Now, we defined above the entropy of a real number. We may think of finding out the average entropy of the entire real number system. As for a given r, the distribution of real number with respect to entropy is a δ-function, so the average value of entropy will be unity (with base r). As the total measure of real number is infinite (of power c), so the total entropy will also be infinite.

Now, in theory of numbers we have the notions of simple normalcy and its invariance, i.e. normalcy and absolute normalcy, but we have no notion, like degree of normalcy, or a measure of normalcy. In the information-theoretic approach to theory of decimal representation of a number, we not merely get some notions like the definitions of entropy, etc., and some propositions like Propositions 6 and 8 greatly simplified but also we may introduce some notions like the degree of normalcy or, better, simple normalcy by the value of the entropy. Also defining neighbourhood about 1 and 0 suitably and conveniently as the coincidence interval of probability, we can define almost simply normal numbers or almost non-normal numbers. Moreover, attempts can be made to define other notions like rationality and irrationality, in terms of entropy.

These problems will be the subject-matter of future investigations.

Acknowledgements

The authors express their gratitude to National Professor S. N. Bose, F.R.S., for having permitted to include this work in his research scheme and for the keen interest shown in the problem and to Dr. A. C. Choudhury and Mr. B. C. Chatterjee for their interest and for constant encouragement and necessary permission to one of the authors (Sen) for completing this work.

References

Boltzmann, L. (1923). Vorlesungen über Gas-Theorie (3rd edn.). Verlag Johann Ambrosius Barth.

Cramer, H. (1949). Mathematical Methods of Statistics. Princeton University Press, Princeton.

Dirac, P. A. M. (1943). The Principles of Quantum Mechanics (3rd edn.). Oxford University Press.

Dutta, M. (1966). Maximum information-theoretic entropy estimation. *Sankhya*, A, **28**, 319.

Dutta, M., and Pal, S. (1963). Introduction to the Mathematical Theory of Probability and Statistics. World Press, Calcutta.

Gelfond, I. M., and Schilow, G. E. (1960). Verallgemeinerte Funktionen (Distributionen) 1. Veb Deutscher Verlag der Wissenchaften, Berlin.

Gibbs, J. W. (1902). Principles of Statistical Mechanics. (Reprinted by Dover Publications Inc., New York).

——— (1928). Collected Works. (Reprinted by Dover Publications Inc., New York).

Hardy, G. H., and Wright, E. M. (1960). An Introduction to the Theory of Numbers (4th edn.). Clarendon Press, Oxford.

Khinchin, A. I. (1957). Mathematical Foundations of Information Theory. Dover Publications Inc., New York.

Loeve, M. (1960). Probability Theory. D. Van Nostrand Co., New York.

Niven, I. (1956). Irrational Numbers. (Carus Mathematical Monographs). Mathematical Association of America, New York.

Pillai, S. S. (1940). A note on normal numbers. *Proc. Indian Acad. Sci.*, 12.

Planck, Max (1887). The Theory of Heat Radiation. (Reprinted by Dover Publications Inc., New York).

——— (1927). A Treatise on Thermodynamics. (Reprinted by Dover Publications Inc., New York, 1945).

Shannon, C. E. (1948). A mathematical theory of communications. *Bell Syst. Tech. J.*, **27**.

Zeisse, H. (1944). Thermodynamik. Verlag Von S. Hirzel, Leipzig.