

LA-4822

C. 1

The Notion of Complexity

For Reference

Not to be taken from this room



los alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

This report expresses the opinions of the author or authors and does not necessarily reflect the opinions or views of the Los Alamos Scientific Laboratory.

Printed in the United States of America. Available from
National Technical Information Service
U. S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy \$3.00; Microfiche \$0.95

LA-4822

UC-32

ISSUED: December 1971



Los Alamos
scientific laboratory
of the University of California
LOS ALAMOS, NEW MEXICO 87544

The Notion of Complexity

by

W. A. Beyer
M. L. Stein
S. M. Ulam



THE NOTION OF COMPLEXITY

by

W. A. Beyer, M. L. Stein, and S. M. Ulam

ABSTRACT

The notion of the arithmetic complexity $|n|$ of an integer n is defined in terms of the minimum number of additions, multiplications, and exponentiations required to combine 1's to form n . The value of $|n|$ is calculated for $n < 2^{10}$. n is called complicated if $|n| > |n_1|$ for every $n_1 < n$. Of the first 19 complicated numbers, 14 are prime. A conjecture about a relation between complexity and entropy is proposed. Some computations are presented to support this conjecture.

I. INTRODUCTION

In this report we discuss notions of complexity in some algebraic structures. These notions are also applicable to more general combinatorial situations that perhaps lack any algebraic pattern in the classical sense. We concentrate on a few special cases for which we studied and calculated a special notion of complexity. Essentially, we examined a special notion of complexity for ordinary integers with a little excursion on such a notion for integers modulo a prime.

The notion of complexity, in our view, is separate, though associated with the idea of the amount of information or entropy of a system. We mention briefly a possible axiomatic approach to defining a real number called complexity for elements of a set or of a class on which certain operations are performed. These could be binary operations; our set could be a set of integers, and the operations could be addition, multiplication, and exponentiation, for example. It is this case that was examined on a computing machine and to which most of this report is devoted.

Another case would be a class of subsets of a given set, with allowed operations being the Boolean operations of union and intersection or

union and complementation. One could add other operations, for example, the direct product of sets and also projection. This would correspond to allowing quantifiers in our theory. One can study a notion of complexity for vectors in a countable space or even in the continuum. An important study would be that of a relative complexity; that is to say, complexity of elements or "expressions" when the complexity of certain symbols is normalized to 1. In what has been sometimes called "speculation" on constants in physical theories, for example, the whole art seems to depend on the success of attempts to define some known important numbers, e.g., the dimensionless ratios

$$M_{\text{proton}}/M_{\text{electron}} = 1836.11\dots$$

and

$$e^2/hc = 137.1\dots$$

by use of only a few artificially introduced constants which should be as "simple" as possible. (cf. the attempts by Eddington¹ and some very recent ones by Good² and Wyler.³)

Considered "genetically," a mathematical theory resembles a tree in that one obtains from a given number of symbols corresponding to "variables"

and from a number of allowed operations, expressions that elongate by branching. The simplifications and abbreviations may then reduce the length of the expressions.

One could try to define complexity in a mathematical structure by postulating certain of its properties, somewhat like postulating properties of a measure.

Let the structure, S , consist of elements x , y , ... It may be finite or infinite. We have in the set S a number of, say, binary operations R_1 , R_2 , ... R_n . We want to assign a number $c(x) \geq 0$ to each element x of S and to each R_i ($i = 1 \dots n$) so that the following properties should hold.

- a. If $z = R_i(x, y)$, then $c(z) = c(R_i(x, y)) \leq c(x) + c(y) + c(R_i)$ $i = 1 \dots n$.
- b. For each element z , if $z = R_j(x, y)$, we should have for one case at least, $c(z) = c(x) + c(y) + c(R_j)$.
- c. $z(x_0) = z(x_1) = \dots = z(x_n)$ for some pre-assigned elements $x_0 \dots x_n$.

Needless to say, one can define analogous desiderata for the case in which the operations are more general than binary ones.

Obviously, in the case to which our exercise is devoted, these postulates are satisfied. Moreover, they define the complexity uniquely if, as must be the case in general, the complexity was normalized for some elements. (In our case, we assume the complexity of the integer 0 to be equal to 1. We hope to study this notion more thoroughly for the more general case and also to perform experiments to determine complexity functions for the case in which S is a class of sets.) Ultimately, one would wish to discuss the complexity of genetic codes and biological organisms quantitatively.

("Integer" always means a positive integer.)

II. ARITHMETIC COMPLEXITY OF INTEGERS

The arithmetic complexity $|n|$ of an integer n is defined as the fewest number of operators: $+$, x , xx (addition, multiplication, and exponentiation) which combine 1's to form n . Thus, $|1| = 0$; $|2| = 1$ since $2 = 1 + 1$; and $|5| = 4$ since $5 = (1 + 1)xx(1 + 1) + 1$ and not fewer than four operators with 1's will form five. Obviously, for a and b integers, $|a + b|$, $|ab|$, and $|a^b|$ are each not more than

$|a| + |b| + 1$. For an infinity of integers n , the relation $|n + 1| = |n| + 1$ holds.

For the purpose of calculating the complexity of some integers, all correct formulas (up to some number of operators) involving $+$, x , xx , and the number 1 were enumerated using parenthesis-free notation on a computer. It required one hour of computer time to enumerate the integers with complexity ≤ 6 . Ralph Cooper made the following observation. Each correct formula involving n (> 0) operators is the composition of two formulas, one formula with n_1 operators and one formula with n_2 operators such that $n = n_1 + n_2 + 1$. One generates the integers of complexity n by first generating tables of integers of complexity $< n$. One partitions $n - 1$ into $n_1 + n_2$ in all ways and combines the integers of complexity n_1 with the integers of complexity n_2 to produce integers of complexity not larger than n . This method is considerably more efficient than the previous method. Table I lists the complexity of all integers $< 2^{10}$.

From the above construction, one sees that an upper bound $\ell^1(k)$ to $\ell(k)$, the number of integers of complexity k , is given by the solution of

$$\ell^1(k + 1) = \sum_{j=0}^k \ell^1(j) \ell^1(k - j),$$

with $\ell^1(0) = 1$. The solution to this equation is given by

$$\ell^1(k) = \frac{1}{k+1} \binom{2k}{k} 2^{-k},$$

which implies that

$$\ell(k) \leq \frac{2^k}{k\sqrt{\pi k}} + o\left(2^k k^{-5/2}\right).$$

Two additional forms of complexity have been considered and calculated.

- a. Complement complexity. To make complexity symmetric in 0's and 1's, we introduce a slightly different complexity, the complement complexity $\bar{K}(y|n)$. Define the complement operation C by $C(x|n) = 2^n - 1 - x$. $\bar{K}(y|n)$ is defined as the fewest operations of addition, multiplication, exponentiation, and complementation that combine 1's to form y . In the count of operations, the

first three are given the value 1 and the last is given the value zero. Thus $\bar{K}(y|n) = \bar{K}(2^n - 1 - y|n)$. Table II gives the values of $\bar{K}(y|n)$ for $y < 2^{10}$ and $n = 10$.

- b. Modulo a prime p complexity. In addition to the operations of +, x, and xx, the operation of mod_p is allowed and is defined by $\text{mod}_p(x) = x - p[x/p]$ where p is a fixed prime and [] denotes the greatest integer. Table III gives the modulo prime $p = 137$ complexity for integers < 137 . Table IV gives the modulo prime $p = 1009$ complexity for integers < 1009 .

III. COMPLICATED NUMBERS

One defines n to be a complicated number if $|n| > |n_1|$ for every $n_1 < n$. The complicated numbers $< 2^{10}$ are 1, 2, 3, 4, 5, 7, 11, 13, 21, 23, 41, 43, 71, 94, 139, 211, 215, 431, and 863. (Those underlined are also prime.) Obviously, there are an infinity of complicated numbers. We propose the following conjectures.

- There exists K such that all complicated numbers $K_1 > K$ are prime.
- Every sufficiently large integer n is the sum of $k < \log n$ complicated integers.
- There exists c such that every sufficiently large n satisfies $|n| < c + \sqrt{\log n}$.

IV. COMPLEXITY AND ENTROPY

Kolmogorov^{4,5} has introduced the notion of complexity of a finite string over a given alphabet. For simplicity, suppose the alphabet to be {0,1}. Let A be an algorithm that transforms finite binary sequences into binary sequences. By an algorithm is meant any of the various equivalent concepts used in logic. For a binary string x, one defines the complexity by

$$K_A(x) = \begin{cases} \min \ell(p) \\ A(p)=x \\ \infty \\ \text{if no } p \text{ exists such that } A(p) = x, \end{cases}$$

where $\ell(p)$ denotes the length of the binary string p. Analogously, one defines conditional complexity.

Let $A(p,x)$ be an algorithm defined from pairs of binary strings to binary strings. Put

$$K_A(y|x) = \begin{cases} \min \ell(p) \\ A(p,x)=y \\ \infty \\ \text{if no } p \text{ exists such that } A(p,x) = y. \end{cases}$$

$K_A(y|x)$ is called the conditional complexity of y with respect to x. Kolmogorov regards complexity as analogous to entropy. We make the following conjecture.

Conjecture. Let a discrete binary information source S in the sense of Shannon⁶ be given with entropy $H = -p \log p - (1-p) \log (1-p)$ where probability (0) = p and probability (1) = 1-p; $0 < p < 1$. Let $\{x_1, x_2, \dots, x_{2^n}\}$ be the set of all binary strings of length n arranged in order of decreasing probability. Let $k(n)$ be the least integer so that $\sum_{i=1}^{k(n)} \text{prob}(x_i) > r$ where $1/2 < r < 1$. Then asymptotically for large n,

$$H \approx \frac{1}{k(n)} \sum_{i=1}^{k(n)} K_A(x_i|n). \quad (1)$$

(In Eq. (1), K_A should be normalized so that when $p = 1/2$,

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} K_A(x_i|n) = 1.)$$

In other words, the most likely sequences from A have complexity approximately equal to the entropy of S.

In order to test the conjecture expressed in Eq. (1), we replaced $K_A(x_i|n)$ by $\lambda \bar{K}(y|n)$, where λ is selected so that when $p = 1/2$,

$$\frac{1}{k(n)} \sum_{i=1}^{k(n)} \lambda \bar{K}(x_i|n) = 1.$$

Graphs of $H_1 = -p \log p - (1-p) \log (1-p)$ and

$$H_2 = \frac{1}{k(n)} \sum_{i=1}^n \lambda \bar{K}(x_i|n)$$

when $n = 10$ and $r = .75$ are shown in Fig. 1

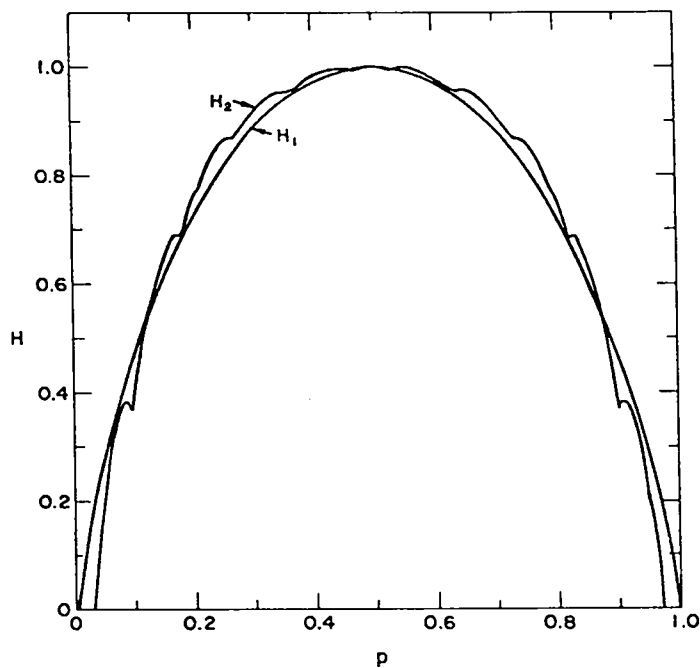


Fig. 1. Comparison of entropy $H_1 = - \sum p_i \log p_i$ and complement complexity H_2 as defined and discussed in text.