Reports of the Academy of Sciences of the USSR

MATHEMATICS

1966

Abstract

Full Text

Reports of the Academy of Sciences of the USSR
1966. Volume 171, No. 1

UDC 519.92

MATHEMATICS

R. E. KRICHEVSKII

THE BLOCK LENGTH NECESSARY FOR OBTAINING A GIVEN REDUNDANCY

(Presented by Academician S. L. Sobolev on 22 I 1966)

Consider a source generating ν mutually independent letters with probabilities $p_1, \dots, p_\nu$, $p_1 + \dots + p_\nu = 1$, $\nu \geqslant 2$. These letters will be called the letters of the input alphabet. We shall divide the sequence of letters generated by the source into words (blocks) of length n and encode these blocks by words of an output two-letter alphabet.* Let the average number of letters of the output alphabet per one input letter under optimal coding be denoted by $\bar{l}(n)$; the difference

$$A(n) = \bar{l}(n) - H, \qquad \text{where} \qquad H = -p_1 \log p_1 - \dots - p_\nu \log p_\nu,$$

will be called the redundancy of coding.** As Shannon' s remarkable theorem shows ([1,2]), the redundancy of decodable coding is always nonnegative and can be made arbitrarily close to zero as n increases. The number n—the block length—characterizes the complexity of the coding. Therefore it is of interest to find the block length necessary for attaining a given redundancy. The present note is devoted to the solution of this problem. The problem just mentioned is equivalent to the problem of finding the average length of an optimal code for an alphabet whose letters are the n-letter words of the input alphabet with the corresponding probabilities. Although an optimal code can be constructed by Huffman' s method ([3]), this method does not make it possible to estimate a priori the average length of the code obtained. The method proposed by Shannon ([1]) gives, for the quantity $nA(n)$, the estimate $0 \leqslant nA(n) < 1$.

In the present note it is shown that if at least one of the numbers $\log_2 p_i/p_j$ $(i, j = 1, \dots, \nu)$ is irrational, then

$$\lim_{n\to\infty} nA(n) = \log_2 \log_2 e - \frac{1}{2},$$

whereas in the contrary case

$$\underline{\lim_{n\to\infty}}\, nA(n) \neq \overline{\lim_{n\to\infty}}\, nA(n).$$

Further,

$$\lim_{n\to\infty} nA(n) = 0$$

if and only if all the numbers $\log_2 p_i/p_j$ $(i, j = 1, \dots, \nu)$ are integers. Denote by $n(A)$ the minimum block length for which the redundancy is less than or equal to A. We have $n(A) = O(1/A)$ if not all the numbers $\log_2 p_i/p_j$ are integers; in the contrary case $n(A) = o(1/A)$; more precisely,

$$n(A) \leqslant c/\sqrt{A}, \qquad c = \text{const}.$$

For the case of rational $\log_2 p_i/p_j$ a formula is given that differs from $nA(n)$ by an arbitrarily small amount as $n \to \infty$.*** Apparently, this formula gives satisfactory accuracy already for small n.

The probability of a word u will be denoted by $p(u)$, and the length of the code word assigned to u by $l(u)$. We shall call a decodable code optimal if it minimizes $\sum_u p(u)l(u)$.

* The consideration of an output alphabet containing more than two letters involves no new difficulties.

** $A(n)$ is the absolute redundancy. In the literature the relative redundancy $A(n)/H$ is sometimes called redundancy.

*** Here a formula is given only for $\nu = 2$; the formula for arbitrary ν is derived in a completely analogous way, but it is more cumbersome.

Lemma. Let $p(u_0) = p_0,\ p(u_1) = \dots = p(u_t) = p',\ p' \leq p_0$. Then there exists an optimal code such that: a) if $l(u_j) < l(u_0) + [\log_2 p_0/p']$, $j = 1, \dots, t$, then

$$t < 2^{[\log_2 p_0/p']}; \tag{1}$$

b) if $l(u_j) > l(u_0) + [\log_2 p_0/p'] + 1$, $j = 1, \dots, t$, then

$$t < 2^{[\log_2 p_0/p']+1}. \tag{2}$$

Proof. We shall prove assertion a); assertion b) is proved analogously.

Arrange the words to be encoded in decreasing order of their probabilities; to each code assign the vector whose i-th coordinate is equal to $l(u_i)$. Order the optimal codes lexicographically*. We shall prove that the code which is lexicographically minimal among the optimal ones satisfies the lemma.

As is known ([2]), for a code to be decipherable it is necessary and sufficient that Kraft' s inequality be satisfied,

$$\sum \frac{1}{2^{l(u)}} \leq 1,$$

where the sum is taken over all words u. We shall call the sum standing on the left-hand side of this inequality the Kraft sum. If assertion a) of the lemma were not fulfilled for this code, then we could find $t_1 = 2^{[\log_2 p_0/p']}$ words for which inequality (1) is true. These words, together with u_0, contribute to the Kraft sum

$$\sum_{j=1}^{t_1} \frac{1}{2^{l(u_j)}} + \frac{1}{2^{l(u_0)}};$$

to the average code length they contribute

$$p_0 l(u_0) + p' \sum_{j=1}^{t_1} l(u_j).$$

Consider a new code for which $l'(u_0) = l(u_0) - 1,\ l'(u_j) = l(u_j) + 1,\ j = 1, \dots, t_1$, where $l'(u_j)$ is the length of the new code of the word $u_j,\ j = 0, 1, \dots, t_1$. As is easily verified,

$$\sum_{j=0}^{t_1} \frac{1}{2^{l(u_j)}} \geq \sum_{j=0}^{t_1} \frac{1}{2^{l'(u_j)}}, \tag{3}$$

$$p' \sum_{j=1}^{t_1} l(u_j) + p_0 l(u_0) \geq p' \sum_{j=1}^{t_1} l'(u_j) + p_0 l'(u_0). \tag{4}$$

In view of (3), the new code will again be decipherable. If the inequality in (4) were strict, this would contradict the optimality of the original code. Thus (4) becomes an equality, i.e. the new code is decipherable and optimal. However, contrary to the condition, it is lexicographically smaller than the original one. The lemma is proved.

Further we shall use the circumstance that

$$p(u) = p_1^{k_1(u)} \dots p_\nu^{k_\nu(u)},$$

where $k_i(u)$ is the number of occurrences of the i-th letter of the input alphabet in the word u, and

$$k_1(u) + \dots + k_\nu(u) = n.$$

Theorem 1. For every n there can be found an optimal code and such a word u_0 that almost all words of length n satisfy the inequality

$$[\log_2 p(u_0)/p(u)] \leq l(u) - l(u_0) \leq \log_2[p(u_0)/p(u)] + 1. \tag{5}$$

Here the expression "almost all" is used in the sense that the total probability of the words not satisfying (5) tends to zero as n increases.

In the proof of Theorem 1 a lemma is used: as u_0 one may take a word u for which

$$k_i(u) = np_i + a\sqrt{n}, \quad i = 1, \dots, \nu - 1 \quad (p_1 \geq \dots \geq p_\nu),$$

where the number a is sufficiently large.

* We shall say that $A \leq B$ if, in the vector corresponding to the code A, the i-th coordinate is smaller than the i-th coordinate of the vector corresponding to the code B, and all preceding coordinates are equal ($i = 1, 2, \dots$).

In Theorem 2 the discussion concerns a binary input alphabet; for the general case the result (more cumbersome) is obtained by the same method. For $v = 2$, $p_1 + p_2 = 1$; we shall assume that $p_1 > p_2$.

Theorem 2. Let $v = 2$, and let $\log_2 p_1/p_2$ be rational and equal to an irreducible fraction with denominator r, $r \geqslant 1$. Then

$$nA(n) = \frac{1-r}{2r} - \log_2 r(2^{1/r} - 1) + \frac{1}{r(2^{1/r} - 1)} + \delta - \frac{2^\delta}{r(2^{1/r} - 1)} + \varepsilon_n, \tag{6}$$

where δ is the distance from the number $-n \log_2 p_1$ to the nearest larger number of the form $\log_2 r(2^{1/r} - 1) + b/r$, b an integer, $\varepsilon_n \to 0$ as $n \to \infty$.

We note that if all the numbers $\log_2 p_i/p_j$ $(i, j = 1, \dots, v)$ are integers, then formula (6) will be valid for any v. If, moreover, $p_1 = \dots = p_v$, then formula (6) coincides with the formula derived by S. S. Kislitsyn [6] for the redundancy of the Huffman code for an alphabet of equiprobable letters.

Proof. In the proof, in addition to Theorem 1, the equality

$$\sum_{k \equiv c (\mathrm{mod}\, r)} C_n^k p_1^k p_2^{n-k} = \frac{1}{r} + \varepsilon', \qquad c = 0, \dots, r-1, \tag{7}$$

is used, where ε' can be taken arbitrarily small for $n > n_0(\varepsilon')$. Equality (7) is proved by means of the Moivre–Laplace theorem [4].

As u_0 one may here take the most probable word, $p(u_0) = p_1^n$; define $l(u_0)$ so that the inequality $0 \leqslant \log_2 r(2^{1/r} - 1) + l(u_0) + \log_2 p(u_0) + 1 \leqslant 1$ is satisfied; put

$$k = [r(\log_2 r(2^{1/r} - 1) + l(u_0) + \log_2 p(u_0) + 1)], \tag{8}$$

$$\theta = 2^{l_0 + \log_2 p(u_0) + 1 - k/r} - \frac{1}{r(2^{1/r} - 1)}. \tag{9}$$

Let

$$l(u) = l(u_0) + [\log_2 p(u_0)/p(u)], \tag{10}$$

if $\{\log_2 p(u_0)/p(u)\} = d/r,\ d = 0, 1, \dots, k$.

Formula (10) also defines $l(u)$ for words u for which $\{\log_2 p(u_0)/p(u)\} = (k+1)/r$ and whose total probability is equal to $\theta + \varepsilon'_n$, $\varepsilon'_n \to 0$ as $n \to \infty$. For all remaining words

$$l(u) = l(u_0) + [\log_2 p(u_0)/p(u)] + 1. \tag{11}$$

It can be shown that the code thus defined is decipherable and, as n grows, differs arbitrarily little from the optimal one.

Theorem 3. Let at least one of the numbers $\log_2 p_i/p_j$ be irrational, $i, j = 1, \dots, v$. Then

$$\lim_{n\to\infty} nA(n) = \log_2 \log_2 e - \frac{1}{2}.$$

In the proof one uses Theorem 2 and an inequality estimating the accuracy of approximation of an irrational number by rational numbers [5].

We give several corollaries of Theorems 2 and 3.

Corollary 1. Let all the numbers $\log_2 p_i/p_j$, $i, j = 1, \dots, v$, be rational. Then $\overline{\lim}_{n\to\infty} nA(n) \neq \underline{\lim}_{n\to\infty} nA(n)$. For $v = 2$

$$\overline{\lim_{n\to\infty}}\, nA(n) = \frac{1-r}{2r} + \frac{1}{r(2^{1/r} - 1)} - \log_2 \ln 2 - \frac{1}{\ln 2} =$$

$$= \log_2 \log_2 e - \frac{1}{2} + \frac{5 \ln 2}{24 r^2} + O\left(\frac{1}{r^3}\right), \tag{12}$$

$$\underline{\lim_{n\to\infty}}\, nA(n) = \frac{1-r}{2r} - \log_2 r(2^{1/r} - 1) = \log_2 \log_2 e - \frac{1}{2} + \frac{\ln 2}{12 r^2} + O\left(\frac{1}{r^3}\right), \tag{13}$$

where r is the denominator of the irreducible fraction $\log_2 p_1/p_2$, $r \geqslant 1$.

Corollary 2. a) *If not all the numbers* $\log_2 p_i/p_j$ $(i, j = 1, \dots, \nu)$ *are integers, then*

$$\lim_{n\to\infty} nA(n) > 0,$$

i.e. $n(A) = O(1/A)$; *b) if all the numbers* $\log_2 p_i/p_j$ $(i, j = 1, \dots, \nu)$ *are integers, then*

$$\lim_{n\to\infty} nA(n) = 0, \qquad n(A) < c/\sqrt{A}.$$

Recall that $n(A)$ is the smallest block length necessary to attain the prescribed redundancy.

In case b), the minimum of $nA(n)$ over $n \le N$ is attained at n which is the denominator of a convergent for $-\log_2 p_1$. For such an n, from (6) we obtain, using properties of continued fractions, the assertion of part b) of Corollary 2. As one example of the application of assertion b), consider the case $\nu = 10$, $p_1 = \cdots = p_{10} = 0.1$, which corresponds to coding decimal digits by binary symbols. Uniform coding of n-digit decimal numbers gives the redundancy

$$nA'(n) = 1 - \{n \log_2 10\} = \delta.$$

Under optimal coding, from (6) we obtain*

$$nA(n) = 1 + \delta - 2^{\delta}.$$

Table 1

n	$r = 1,\ p_1 = 4/5$		$r = 2,\ p_1 = \sqrt{2}/(\sqrt{2}+1)$	
	$A(n)$	$\overline{A}(n)$	$A(n)$	$\overline{A}(n)$
1	0.03	0.8	0.021	0.028
2	0.06	0.04	0.0213	0.0207
3	0.006	0.003	0.01153	0.01146
4	0.0189	0.0184	0.00967	0.00969
5	0.01599	0.01593	0.00772	0.00776
6	0.00332	0.00340	0.00580	0.00582
7	0.00983	0.00984	0.00592	0.00593
8	0.01030	0.01031	0.00372	0.00388
9	0.00321	0.00317	0.00677	0.00678

Choosing as n the denominators of convergents for $\log_2 10$, we obtain

$$n(A) \le c/\sqrt{A}.$$

We can also obtain uniform (with respect to the distribution of probabilities of letters of the input alphabet) estimates of the quantity $nA(n)$.

Corollary 3. *If not all the numbers* $\log_2 p_i/p_j$ $(i,j = 1,\dots,\nu)$ *are integers, then*[**]

$$nA(n) \gtrsim c_1 > 0. \tag{14}$$

The inequality

$$nA(n) \lesssim c_2 < 1 \tag{15}$$

always holds.

Here c_1 and c_2 are constants depending only on ν. Their values for $\nu = 2$ can be obtained by taking the maximum of expression (12) over $r \geq 1$ and the minimum of (13) over $r \geq 2$. We also note that the function $A(n)$ is not monotone.

In conclusion, Table 1 is given, in which the values $A(n)$ (found by the Huffman method (3)) are compared with the approximate values $\overline{A}(n)$ found by formula (6). The calculations were made for $\nu = 2$.

Institute of Mathematics
Siberian Branch of the Academy of Sciences of the USSR

Received
18 I 1966

REFERENCES

1. C. Shannon, in: *Works on Information Theory and Cybernetics*, IL, 1963.
2. R. Fano, *Transmission of Information*, Mir, 1965.
3. D. Huffman, *Cybernetic Collection*, No. 3, IL, 1961.
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, Mir, 1964.
5. A. Ya. Khinchin, *Continued Fractions*, Moscow, 1961.
6. S. S. Kislytsin, *Uspekhi Mat. Nauk*, 20, no. 6 (126), (1965).

[*] For example, the following code will be optimal: if the numbers $0,1,\dots,k$ are coded by $l = [n\log_2 10]$ symbols of their binary representation, and the numbers $k+1,\dots,10^n-1$ by $l+1$ symbols of their binary representation, where $k = 2^{[n\log_2 10]} - 10^n$, and, for $n = 1$, we obtain the following optimal code for decimal digits: $0 \leftrightarrow 000$, $1 \leftrightarrow 001$, $\cdots$, $5 \leftrightarrow 101$, $6 \leftrightarrow 110$, $7 \leftrightarrow 0111$, $\cdots$, $10 \leftrightarrow 1010$.

[**] Here $a(n) \gtrsim b(n)$ means that the inequality holds starting from some n.

Note: Figure translations are in progress. See original paper for figures.