AN UPPER BOUND ON THE RATE OF TRANSMISSION OF MESSAGES[1]

BY

J. WOLFOWITZ

This paper is a sequel to an earlier one entitled *The Coding of messages subject to chance errors* (Illinois Journal of Mathematics, vol. 1 (1957), pp. 591–606), and should be regarded as the ninth section of the latter. The previous notation, definitions, and list of references remain in force, except that it is convenient to replace $p(\alpha)$ by $p(1 \mid \alpha)$ and $(1 - p(\alpha))$ by $p(0 \mid \alpha)$. The purpose of the present paper is to state and prove Theorem 4, which, for any memory m, gives an upper bound on the length of an error correcting code for which the probability of transmitting any word incorrectly is $\leqq \lambda$, $0 \leqq \lambda < 1$. In Theorem 2 of the earlier paper we gave such an upper bound for the case $m = 0$. It will be shown at the end of the proof of Theorem 4 that the latter implies Theorem 2. The discussion after the proof will show how it is possible greatly to simplify the computation of the constant J_0 in the upper bound.

Let the 2^{m+1} α-sequences be numbered, in some fixed manner, from 1 to 2^{m+1}. Any x-sequence can be written as a sequence of $(n - m)$ α-sequences, namely, the α-sequence of its first $(m + 1)$ elements, followed by the α-sequence of its 2^{nd}, 3^{rd}, $\cdots$, $(m + 2)^{\text{nd}}$ elements, etc. Replacing each α-sequence by its serial number, we obtain that any x-sequence can be written as a sequence of $(n - m)$ elements, each element one of $1, \cdots, 2^{m+1}$. Henceforth we consider each x-sequence to be written in this manner. Not every sequence of $(n - m)$ elements, each of which is one of $1, \cdots, 2^{m+1}$, is an x-sequence. For example, if $m = 1$, the serial number of the α-sequence (0, 0) cannot, in an x-sequence, be followed by the serial number of the α-sequence (1, 0). Let i, j be, respectively, the serial numbers of α-sequences which are related so that the α-sequence i can be followed, in an x-sequence, by the α-sequence j; we shall say then that j can be a successor of i.

Let A be a subset of the integers $1, \cdots, 2^{m+1}$, and $M(A)$ an irreducible ([6], p. 318) stationary transition matrix of a Markov chain whose states are the set A; say $M(A) = \{t_{ij}\}$, $i, j \in A$. We shall always require that $t_{ij} = 0$ if j cannot be a successor of i. Let T_i be the stationary absolute probability of the i^{th} state, i.e., the T_i, $i \in A$, are the unique ([6], p. 325) solutions of the system

$$T_i = \textstyle\sum_{j \in A} T_j t_{ji} \qquad i \in A \tag{9.1}$$

which satisfy $\sum_{i \in A} T_i = 1$. Define (for $M(A)$)

$$u(i) = \textstyle\sum_{j \in A} T_j p(i \mid j), \qquad i = 0, 1 \tag{9.2}$$

Received July 11, 1957.

[1] Research under contract with the Office of Naval Research.

$$H_1(M(A)) = -\textstyle\sum_{i=0}^{1} u(i) \log u(i), \tag{9.3}$$

$$H_2(M(A)) = -\textstyle\sum_{j=0}^{1} \sum_{i\epsilon A} T_i\, p(j \mid i) \log p(j \mid i) \tag{9.4}$$

(recall the definition of $H_X(Y)$ in Section 4), and

$$J_0 = \sup_{A,M(A)} [H_1(M(A)) - H_2(M(A))]. \tag{9.5}$$

Our object will now be to prove the following

Theorem 4. *Let* λ, $0 \leqq \lambda < 1$, *be any given number. There exists a* $K'' > 0$ *such that, for any* n, *any code with the property that the probability of transmitting any word incorrectly is* $\leqq \lambda$, *cannot have a length greater than* $2^{nJ_0+K''n^{1/2}}$.

Let x be any x-sequence (written as a sequence of α-sequences). Let t'_{ij} be the number of times the sequence i is followed by the sequence j in x; for this purpose we count the first element (sequence) in x as if it followed the last element. Let A be the totality of sequences from among $1, \cdots, 2^{m+1}$ which occur in x. Write $n' = n - m$ and, for $i, j \epsilon A$,

$$n'T_i = \textstyle\sum_{j\epsilon A} t'_{ij}, \qquad t_{ij} = t'_{ij}/n'T_i.$$

Clearly, $\{t_{ij}\}$ is a stochastic matrix. From its construction it is irreducible. Moreover, the quantities $n'T_i$ satisfy, for $i \epsilon A$,

$$n'T_i = \textstyle\sum_{j\epsilon A} n'T_j\, t_{ji},$$

which is the same system as (9.1). Since $\sum T_i = 1$ it follows that the T_i are the stationary absolute probabilities of the matrix $\{t_{ij}\}$. The sequence x will be said to be a member of the domain of $\{t_{ij}\}$.

We note here a fact which will be of importance later: Since the t'_{ij} are all integers, it follows that the number of matrices $\{t_{ij}\}$ which can be obtained in this manner is $O(n^a)$, where $a = 2^{m+2}$.

Let x be any x-sequence, and $M(A) = \{t_{ij}\}$ the matrix of whose domain x is a member. Let B' be any set of y-sequences generated by x such that $P\{Y(x) \epsilon B'\} > (1 - \lambda)/2$. Then it follows from (3.8) that there exists a positive constant K''_2 which does not depend on $M(A)$ or the sequence x, such that the number of sequences in B' is greater than

$$2^{n'H_2(M(A))-K''_2\,\delta_2 n'^{1/2}}.$$

Let y be any y-sequence generated by x. Then y contains at most

$$V_1 = n'u(1) + \delta_2\, n'^{1/2} \textstyle\sum_{i\epsilon A} [T_i\, p(1 \mid i) p(0 \mid i)]^{1/2}$$

elements one, and at most

$$V_0 = n'u(0) + \delta_2\, n'^{1/2} \textstyle\sum_{i\epsilon A} [T_i\, p(1 \mid i) p(0 \mid i)]^{1/2}$$

elements zero. We shall now obtain an upper bound on the number of y-sequences which contain at most V_1 elements one and V_0 elements zero. For this purpose suppose $Z = (Z_1, \cdots, Z_{n'})$ is a sequence of independent, identically distributed chance variables such that

$$P\{Z_1 = i\} = u(i), \qquad i = 0, 1.$$

If y contains at most V_1 ones and V_0 zeros, then

$$\begin{aligned} \log P\{Z = y\} &> -n' H_1(M(A)) \\ &\quad + \delta_2 n'^{1/2}\left(\sum [T_i\, p(1 \mid i) p(0 \mid i)]^{1/2}\right) \log (u(1)\, u(0)) \\ &> -n' H_1(M(A)) - K_1'' \delta_2\, n'^{1/2}, \end{aligned} \tag{9.6}$$

where K_1'' is a positive constant which does not depend on A or $M(A)$. It follows that the number of y-sequences generated by all x-sequences in the domain of $M(A)$ is less than

$$2^{n' H_1(M(A)) + K_1'' \delta_2 n'^{1/2}},$$

with K_1'' an absolute constant.

Now let $(x_1, A_1), \cdots, (x_w, A_w)$ be any code such that

(a) $x_1, \cdots, x_w$ are x-sequences which are all members of the domain of the same matrix $M(A)$,

(b) $P\{Y(x_i) \in A_i\} > (1 - \lambda)/2$,

(c) A_i, $i = 1, \cdots, w$, consists only of y-sequences generated by x_i (for some fixed, sufficiently large, δ_2).

Since $(x_1, A_1), \cdots, (x_w, A_w)$ is a code, the A_i are all disjoint. Hence the total number of y-sequences in $A_1 \cup A_2 \cup \cdots \cup A_w$ is at least

$$w \cdot 2^{n' H_2(M(A)) - K_2'' \delta_2 n'^{1/2}}.$$

However, the total number of y-sequences generated by all x in the domain of $M(A)$ is less than

$$2^{n' H_1(M(A)) + K_1'' \delta_2 n'^{1/2}}.$$

Hence

$$w < 2^{n' [H_1(M(A)) - H_2(M(A))] + \delta_2 (K_1'' + K_2'') n'^{1/2}}. \tag{9.7}$$

Now let $(x_1, A_1), \cdots, (x_N, A_N)$ be any code whatever such that

$$P\{Y(x_i) \in A_i\} \geqq 1 - \lambda, \qquad i = 1, \cdots, N.$$

Choose δ_2 so large that, for any x,

$$P\{Y(x) \text{ is a sequence generated by } x\} > 1 - (1 - \lambda)/2. \tag{9.8}$$

Delete from each A_i those y-sequences not generated by x_i; call the residue A_i'. In view of (9.8) we have

$$P\{Y(x_i) \in A_i'\} > (1 - \lambda)/2. \tag{9.9}$$

The code $(x_1, A_1'), \cdots, (x_N, A_N')$ may be divided into subcodes according to the matrix to whose domain the x-sequences belong; the x-sequences of a subcode belong to the domain of the same matrix, two x-sequences which do not belong to the same subcode belong to the domains of different matrices. The number of such subcodes is $O(n^a)$, the number of matrices $M(A)$. The length of any one subcode is by (9.7) less than

$$2^{nJ_0+\delta_2(K_1''+K_2'')n^{1/2}}. \tag{9.10}$$

Hence

$$N < [O(n^a)]2^{nJ_0+\delta_2(K_1''+K_2'')n^{1/2}}, \tag{9.11}$$

from which Theorem 4 follows at once.

$H_1(M(A)) - H_2(M(A))$ is of course a continuous function of the T_i, and a function of $M(A)$ only through the T_i. It appeared in the course of the preceding proof that the T_i were the stationary absolute probabilities of an irreducible matrix because they were the proportions in which the various α-sequences appeared in some x-sequence. From the strong law of large numbers for Markov chains it follows that the converse is essentially true, i.e., that there exist x-sequences in which the proportion of the various elements is within ε ($\varepsilon > 0$ arbitrary) of the stationary absolute probabilities of any irreducible matrix of transition probabilities (for n sufficiently large).

To see that Theorem 4 implies Theorem 2 (when $m = 0$) we have only to note that then T_1 and T_0 may be any (positive) pair such that $T_1 + T_0 = 1$. This follows either from the fact that any such pair is the stationary absolute probability vector of an irreducible transition matrix with two states, or else from the fact that any such pair may be the vector of proportions of zeros and ones in an x-sequence.

Suppose $m = 1$. Let us number the α-sequences: (0, 0), 1; (0, 1), 2; (1, 0), 3; (1, 1), 4. Obviously, in any x-sequence, $n'T_2 = n'T_3$. There are clearly no restrictions on T_1 and T_4 except the trivial ones. We conclude that the totality of vectors (T_1, T_2, T_3, T_4) is the totality of vectors (b_1, b_2, b_2, b_3) with $b_1, b_2, b_3 > 0$, and $b_1 + 2b_2 + b_3 = 1$. There are also the obvious vectors which correspond to sets A which are proper subsets of $\{1, 2, 3, 4\}$.

The above characterization of the possible vectors $\{T_i\}$ for the case $m = 1$ greatly reduces the labor of computing J_0, so that it becomes of interest to do this for general m. This is easy to do, but it will be less burdensome for reader and writer to do it for the case $m = 2$; the procedure in the general case will be readily apparent.

We number the α-sequences (say):

(0 0 0)	1	(1 0 0)	5
(0 0 1)	2	(1 0 1)	6
(0 1 0)	3	(1 1 0)	7
(0 1 1)	4	(1 1 1)	8.

Only the following t_{ij} can be different from zero: $t_{11}, t_{12}, t_{23}, t_{24}, t_{35}, t_{36}, t_{47}, t_{48}, t_{51}, t_{52}, t_{63}, t_{64}, t_{75}, t_{76}, t_{87}, t_{88}$. Consider the system (9.1) for $A = \{1, \cdots, 8\}$, and add the equations in the proper pairs. We obtain

$$\begin{aligned} T_1 + T_2 &= T_1 + T_5 , \\ T_3 + T_4 &= T_2 + T_6 , \\ T_7 + T_8 &= T_4 + T_8 . \end{aligned}$$

Hence

$$T_2 = T_5 , \qquad T_4 = T_7 , \qquad T_6 = T_3 + T_4 - T_2 .$$

We have therefore that $(T_1 , \cdots , T_8)$ must be of the form

$$(b_1 , b_2 , b_3 , b_4 , b_2 , (b_3 + b_4 - b_2), b_4 , b_5) \tag{9.12}$$

with all elements positive and summing to unity. It is easy to verify that these necessary conditions are sufficient, i.e., that for any vector (9.12) there exists an 8×8 matrix $\{t_{ij}\}$ with all possible t_{ij} (enumerated earlier in this paragraph) positive (hence $\{t_{ij}\}$ is irreducible) such that this vector is a solution of the system (9.1). The vectors which correspond to the A which are proper subsets of $\{1, \cdots , 8\}$ are characterized similarly.

Added in proof. By using the methods of the present paper, the constant J_0 in the upper bound can in general be reduced when $m > 0$. A paper which describes this is in preparation.

CORNELL UNIVERSITY
ITHACA, NEW YORK