INFORMATION THEORY FOR MATHEMATICIANS[1]

By J. Wolfowitz

Cornell University

A more descriptive term for information theory and one preferred by the present writer is "the theory of coding of messages." In this expository note we will describe briefly some basic concepts of this theory when transmission is through a "noisy channel" (noise = chance errors). We shall assume that both the transmitting alphabet and the receiving alphabet consist of two symbols, 0 and 1, say. This represents no loss in generality because the extension to any other alphabet, say one of twenty-six symbols, is immediate and presents no difficulty at all.

The fundamental paper of the theory is [1]; other important papers are [2], [3], [4], and [5]. The papers most easily intelligible to the mathematician are probably [3], [4], [7], and [8]. The latter three deal with the subject matter of the present paper; [4] and [7] may each be read without any prior reading, and [8] is a sequel of [7]. In the present paper we describe four theorems proved in [7] and [8] and their relation to prior results.

Suppose that a person has a vocabulary of S words, any of which he may want to transmit, in any frequency and in any order, over some channel. We emphasize that we do not assume anything about the frequency with which particular words are transmitted, nor that the words to be transmitted are selected by any random process; in this respect our treatment differs from most of those in the literature.

Let the words be numbered in some fixed but arbitrary manner. Then transmitting a word is equivalent to transmitting one of the integers $1, 2, \cdots, S$. Let $s = \log S$ (all logarithms in this paper are to the base 2). Then there are S sequences of s elements each[2], each element either 0 or 1. If there is no noise, i.e., error of transmission, then, to transmit any word one has only to transmit the appropriate sequence of s zeros or ones.

If there is noise then this is clearly not enough, for the transmitted sequence will usually be incorrectly received. What is needed is that the received sequence, which will usually be a moderately garbled version of the transmitted sequence, should still be different from the moderately garbled version of any other transmitted sequence, so that one can infer what sequence it is that has been transmitted. But this requires that the sequences to be sent be not too similar in some reasonable sense, lest they be confused in transmission. Hence one must employ sequences of length greater than s, and *not all* such sequences (so that "neighboring" sequences be not sent). All these remarks will now be made precise.

Let the integer m ($\geqq 0$) be the "memory". A sequence of n (respectively $(n - m)$,

Received October 31, 1957.

[1] Rietz lecture delivered (under a different title) at the Atlantic City meeting of the Institute of Mathematical Statistics on September 10, 1957, by invitation of the Council of the Institute. Work under contract with the Office of Naval Research.

[2] Obviously, if s is not an integer one should replace it by the smallest integer $\geqq s$.