INFINITE CODES FOR MEMORYLESS CHANNELS

By David Blackwell[1]

University of California, Berkeley

1. Introduction and summary. For a memoryless channel with finite input alphabet A, finite output alphabet B, and probability law $p(b \mid a)$, the *capacity* C is defined as the maximum over all probability distributions q on A of

$$\sum_{a\,b} q(a)p(b \mid a)\log_2(p(b \mid a)/\sum_a q(a)p(b \mid a)).$$

Shannon [1] has obtained the following result.

Exponential error bound. For any $C_0 < C$ there is a number $\rho < 1$ such that, for every positive integer N, there is a set $S \subset A^{(N)}$ with at least $2^{C_0 N}$ elements and a function g from $B^{(N)}$ to S, such that, for every $s = (a_1, \cdots, a_N) \varepsilon S$,

$$\sum p(b_1 \mid a_1)\cdots p(b_N \mid a_N) < 2\rho^N,$$

where the sum extends over all sequences $b_1, \cdots, b_N$ for which $g(b_1, \cdots, b_N) \neq s$.

Thus if the sender selects any $s \varepsilon S$ and places its letters $a_1, \cdots, a_N$ successively into the channel, and the receiver, on observing the resulting output sequence $b_1, \cdots, b_N$, decides that the input was $g(b_1, \cdots, b_N)$, the probability that he makes an error is less than $2\rho^N$, no matter what $s \varepsilon S$ was chosen. This result may be described as follows: it is possible to transmit at any rate $C_0 < C$, with arbitrarily small probability of error, by using block codes of sufficient length.

We wish to draw a slightly stronger conclusion, as follows. We imagine an infinite sequence $x = (x_1, x_2, \cdots)$ of 0's and 1's, which we are required to transmit across the channel. At time N, the sender will have observed the first $[C_0 N]$ coordinates of x, and will place the Nth input symbol in the channel. The receiver, having at this point observed the first N channel outputs, will estimate the first $M(N)$ coordinates of x. If $M(N)/C_0 N \to 1$ as $N \to \infty$ and if, for every x, all but a finite number of his estimates are correct (i.e., agree with x in every coordinate estimated) with probability 1, we shall say that the channel is being used at rate C_0. Our result is that, in this sense, a (memoryless) channel can be used at any rate $C_0 < C$.

The result stated below is exactly this result, for the special case $C_0 = 1$. The general case involves no new ideas, but only more notation, and we shall restrict attention to the case $C_0 = 1$. The function f_n of a code, as defined below, specifies the nth channel input symbol, as a function of the first n coordinates of x. The number $M(n)$ is the number of x coordinates to be estimated by the

Received March 27, 1959; revised June 1, 1959.

[1] This paper was prepared with the partial support of the Office of Naval Research (Nonr-222-53). This paper in whole or in part may be reproduced for any purpose of the United States Government.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

receiver after observing the first n output symbols, and the function g_n specifies the estimate.

We now state the result precisely.

For any finite set S, we denote by $S^{(N)}$ the set of all sequences $(s_1, \cdots, s_N)$, where $s_n \varepsilon S$ for $n = 1, 2, \cdots, N$. For a memoryless channel with finite input alphabet A, finite output alphabet B, an infinite code (for transmitting at rate 1) is defined as consisting of (a) a sequence $\{f_n\}$ of functions, where f_n maps $I^{(n)}$ into A, and I consists of the two elements 0 and 1, (b) a nondecreasing sequence $\{M(n)\}$ of positive integers such that $M(n)/n \rightarrow 1$ as $n \rightarrow \infty$, and (c) a sequence $\{g_n\}$ of functions, where $\{g_n\}$ maps $B^{(n)}$ into $I^{(M(n))}$.

An infinite sequence $x = (x_1, x_2, \cdots)$ of 0's and 1's, together with an infinite code, defines a sequence of independent output variables $y_1, y_2, \cdots$, with

$$\Pr\{y_n = b\} = p(b \mid f_n(x_1, \cdots, x_n)),$$

where $p(b \mid a)$ is the probability that the output symbol of the channel is b, given that the corresponding input symbol is a, and defines a sequence of estimated messages $t_1, t_2, \cdots$, where $t_n = g_n(y_1, \cdots, y_n)$. We shall say that the code is effective at x if, with probability 1,

$$t_n = (x_1, \cdots, x_{M(n)})$$

for all sufficiently large n, and shall say that the code is effective if it is effective for every x. The result of this note is the

THEOREM: *For any memoryless channel with capacity $C > 1$, there is an effective code.*

2. Proof of the theorem. Choose a number D with $1 < D < C$, and let ρ be the number <1 which Shannon's exponential error bound associates with transmitting at rate D. Thus we can, for any positive integer R, transmit any $[DR]$ x-coordinates with R uses of the channel, with error probability at most $2\rho^R$. We shall divide the x-sequence into successive blocks, of length $R(1), R(2), \cdots$, where $\{R(k)\}$ is an appropriately chosen increasing sequence of positive integers. We may use the channel, during the time the $k + 1$st block of x-symbols is observed, to transmit up to $[DR(k + 1)]$ x-coordinates, among those received to date, with error probability at most $2\rho^{R(k+1)}$. We choose to transmit the kth block, containing $R(k)$ x-coordinates, and to repeat the first $Q(k)$ coordinates of x, where $\{Q(k)\}$ is a nondecreasing sequence of nonnegative integers such that

$$Q(k) + R(k) \leqq [DR(k + 1)],$$
$$Q(k) \leqq R(1) + \cdots + R(k - 1).$$

Since $\{R(k)\}$ is strictly increasing, $\sum_k \rho^{R(k)}$ converges, so that, with probability 1, only a finite number of errors will be committed. That is to say, the receiver, after observing the $k + 1$st block of output symbols, estimates the first $Q(k)$ x-symbols, say as $u(k)$, and the kth block of x-symbols, say as $v(k)$, and we have, with probability 1,

$$u(k) = c(k), \qquad v(k) = d(k)$$

for all sufficiently large k, where $c(k)$ denotes the first $Q(k)$ coordinates of x and $d(k)$ denotes the kth block of x-coordinates. After observing the $k + 1$st block of output symbols and making the estimates $u(k)$, $v(k)$, the receiver will have estimated each of the first $R(1) + \cdots + R(k) = T(k)$ coordinates of x at least once. He now forms an estimate $w(k)$ of the first $T(k)$ coordinates, using the latest estimate made on each coordinate. If

$$Q(k) = R(1) + \cdots + R(i - 1) + h, 0 \leqq h < R(i),$$

the estimate $w(k)$ is:

$$w(k) = (u(k), v^*(i), v(i + 1), \cdots, v(k)),$$

where $v^*(i)$ consists of the last $R(i) - h$ coordinates of $v(i)$. If $Q(k) \to \infty$ with k, so does i. Since, with probability 1, all $u(i)$, $v(i)$ for i sufficiently large are correct, we conclude that, with probability 1,

$$w(k) = (x_1, \cdots, x_{T(k)})$$

for all sufficiently large k. We have thus defined a sequence $\{w(k)\}$ of estimates, where $w(k)$ estimates the first $T(k)$ coordinates of x after $T(k + 1)$ outputs have been received, such that, with probability 1, all but a finite number of $w(k)$ are correct.

For $n < T(2)$, we define g_n arbitrarily; for $T(k + 1) \leqq n < T(k + 2)$, we define g_n as $w(k)$. Thus, for $T(k + 1) \leqq n < T(k + 2)$, we have $M(n) = T(k)$, and $M(n)/n \to 1$ as $n \to \infty$ if $T(k)/T(k + 2) \to 1$ as $k \to \infty$.

In summary, any two sequences $\{R(k)\}$, $\{Q(k)\}$ can be used to define an effective code, if

(1) $\{R(k)\}$ is a strictly increasing sequence of positive integers.

(2) $\{Q(k)\}$ is a nondecreasing sequence of nonnegative integers.

(3) $Q(k) + R(k) \leqq [DR(k + 1)]$.

(4) $Q(k) \leqq R(1) + \cdots + R(k - 1)$.

(5) $Q(k) \to \infty$ as $k \to \infty$.

(6) $(R(1) + \cdots + R(k))/(R(1) + \cdots + R(k + 2)) \to 1$ as $k \to \infty$.

The sequences $R(k) = k$, $Q(k) = [\min(1, D - 1)(k - 1)]$, for instance, satisfy (1) $\cdots$ (6).

This completes the proof.

It would be desirable to extend the theorem to finite-state channels. The method of this paper relies on Shannon's exponential error bounds, and such bounds are not yet known for general finite-state channels.

REFERENCE

[1] C. E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, Vol. 1 (1956), pp. 6–25.