Eq. (33) gives the solution of the problem which has been formulated. They permit the determination, for large n values, of the upper limit for the greatest amount, $x_m(n, d)$, of code combinations for which any two are different from each other by not less than a given number of elements.

Shown in Fig. 5 is a graph of $\phi(\alpha)$ constructed according to (33). Using this graph it is easy to determine the z/n ratio for a given d/n.

An analysis of (33) shows that the function 2^z, the upper limit of the desired $x_m(n, d)$ function, approaches 2^n asymptotically for any constant d and n increasing without limit. If, conversely, n remains constant and d increases from 1 to n, then the 2^z function decreases monotonically from 2^n to 1. If n and d simultaneously increase such that their ratio $\alpha = d/n$ is constant, then $\phi(\alpha) =$ const and the exponent in the 2^z function will increase, approximately, directly proportionally to n.

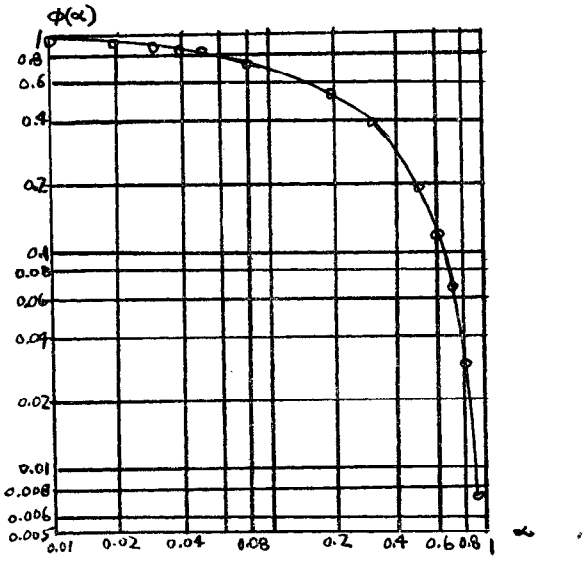

Fig. 5—Graph of $\phi(\alpha)$.

Two Inequalities Implied by Unique Decipherability*

BROCKWAY McMILLAN†

Summary—Consider a list of b words, each word being a string of letters from a given fixed alphabet of a letters. If every string of words drawn from this list, when written out in letters without additional space marks to separate the words, is uniquely decipherable, then

$$a^{-l_1} + a^{-l_2} + \cdots + a^{-l_b} \leq 1, \qquad (1)$$

where l_i, $1 \leq i \leq b$, is the length of the ith word in the list. This result extends a remark of J. L. Doob, who derived the same inequality for lists of a more restricted kind. A consequence of (1) and work of Shannon is that this more restricted kind of list suffices in the search for codes with specified amounts of redundancy.

Discussion

LET A be an alphabet of a letters. A finite nonvacuous sequence or string of letters of A will be called a *word* over A. The number of letters in a word will be called its *length*, or length over A. Consider a finite nonvacuous list B of words over A. This list B will be called *separable* if, whenever a string of words of B is written out in letters, without space marks between the words, the resulting string of letters is uniquely decipherable into the original string of words. A condition necessary and sufficient for separability is given by Sardinas and Patterson.[1]

The list of common English words, considered as words over the alphabet of 26 Latin letters, is not separable, as the sequences "together" and "to get her" show. If each English word is considered as ending with a space mark, so that the words are over an alphabet of 27 letters, then this list of words is separable.

A strong sufficient condition for separability of B is that no word of B appears as the initial string of letters in a longer word of B. For convenience, call this property of B *irreducibility*. The list of common English words terminated by spaces is irreducible. The list (1, 10, 100) of words over the binary alphabet is separable but not irreducible.

In an oral discussion,[2] Doob recently observed that the inequality (1) holds when B is an irreducible list. The main result of this note is a proof of (1) when B is merely separable. A strong converse result is also given: A construction used by Shannon shows that if $l_1, l_2, \cdots l_b$ are integers satisfying (1), then there exists an irreducible list B of b words whose lengths over A are respectively $l_1, l_2, \cdots l_b$. It follows as a corollary that if $l_1, l_2, \cdots, l_b$ are the lengths over A of a separable list of words, they are also the lengths over A of an irreducible list. This rather algebraic conclusion results, as will be seen, from largely analytic arguments.

* Manuscript received by the PGIT, April 6, 1956.

† Bell Telephone Labs., New York, N. Y.

[1] A. A. Sardinas and G. W. Patterson, "A necessary and sufficient condition for unique decomposition of encoded messages," 1953 IRE Convention Record, pt. 8, pp. 104–108.

[2] Comments given upon papers presented before a session on information theory at the summer meeting of the Inst. of Mathematical Statistics, Ann Arbor, Mich., August 30, 1955.

As Doob observed, one interest of (1) is in the following application: Consider a universe of b events, respectively of probabilities $p_1, p_2, \cdots, p_b$. Suppose that a code is established in which each event is designated by a distinct word of B. If the results of a sequence of independent trials are recorded by writing the corresponding words of B in order without space marks, the expected number of letters written per trial is $p_1l_1 + p_2l_2 + \cdots + p_bl_b$. If (1) holds; *e.g.*, if B is separable, then one has a direct proof of the inequality

$$\sum_{i=1}^{b} p_i l_i \geq -\sum_{i=1}^{b} p_i \log_a p_i. \tag{2}$$

This inequality, of course, also follows from the basic theorems of Shannon.[3]

It follows from the corollary remark above that the set of values assumed by $\sum p_i l_i$ as B is varied over the class of separable lists coincides exactly with the set of values assumed when B is restricted to be irreducible. In particular, given a separable code, there exists an irreducible code for the same universe of events which has the same redundancy.

Proofs

Given a separable list B, let $l = \max l_i$ (for $1 \leq i \leq b$), and let n_r be the number of words of B which are of length r, $1 \leq r \leq l$. Then (1) reads

$$\sum_{r=1}^{l} n_r a^{-r} \leq 1. \tag{1'}$$

What will be shown is that the polynomial $Q(x) - 1$, where

$$Q(x) = \sum_{1}^{l} n_r x^r$$

has no zeros in the circle $| x | < a^{-1}$ in the complex plane. In particular then, $Q(x) - 1$ has no zeros in the interval $0 \leq x < a^{-1}$. Since $Q(x)$ is monotone and continuous for $x \geq 0$, and $Q(0) = 0$, (1') follows.

Let $N(k)$ be the number of distinct sequences of words of B each of which, written as a string of letters, is of total length k over A. Since B is separable, each of these $N(k)$ sequences of words gives a distinct sequence of k letters of A. Hence $0 \leq N(k) \leq a^k$. A simple comparison test then shows that for any complex x such that $| xa | < 1$ the infinite series $1 + N(1)x + N(2)x^2 + \cdots$ converges. This series, therefore, represents a function $F(x)$ analytic in $| xa | < 1$.

Suppose for a moment that $k > l$. Consider the $N(k)$ strings of k letters of A mentioned above: those which are decipherable into sequences of words of B. They can be partitioned into l subclasses C_r, $1 \leq r \leq l$, thus: Let C_r consist of all strings whose first word is of length r. Then if $r \neq s$, C_r and C_s cannot overlap; if a string were in both, it would be decipherable into two distinct sequences of words. Since there are n_r distinct words of length r, and $N(k - r)$ possible distinct subsequent strings of length $k - r$ which are decipherable into sequences of words of B, C_r contains exactly $n_rN(k - r)$ distinct strings of letters. Hence if $k > l$

$$N(k) = n_1N(k - 1) + n_2N(k - 2) + \cdots + n_lN(k - l). \tag{3}$$

It is easy to see that if one defines $N(0) = 1$, and $N(-k) = 0$ for $k > 0$, then (3) in fact holds for all $k \geq 1$.

Multiply (3) by x^k and sum from $k = 1$ to $k = \infty$. If $| xa | < 1$, one gets $F(x) - 1 = Q(x)F(x)$, or

$$F(x) = \frac{1}{1 - Q(x)}.$$

Since $F(x)$ is analytic inside $| x | < a^{-1}$, $Q(x) - 1$ cannot vanish in that circle. Hence (1').

Conversely, suppose that $l_1, l_2, \cdots, l_b$ are integers satisfying (1). So enumerate them that $l_1 \leq l_2 \leq \cdots \leq l_b$. It is easy to see that there is then an integer $k \geq 0$ such that

$$a^{-l_1} + a^{-l_2} + \cdots + (k + 1)a^{-l_b} = 1.$$

Let $q_i = a^{-l_i}$, $1 \leq i \leq b - 1$, and $q_i = a^{-l_b}$ for $b \leq i \leq b + k$. Then $q_1, q_2, \cdots, q_{b+k}$ qualify as an exhaustive list of probabilities. The construction of Shannon[3] for an efficient binary code to transmit messages drawn from a universe with the probabilities q_i can easily be extended to an alphabet A of a letters. This extension then describes the construction of an irreducible list of $b + k$ words, the first b of which have lengths $l_1, l_2, \cdots, l_b$. Deleting the last k words leaves the list irreducible, and with words of the desired length.

Finally to prove (2) from (1), it is necessary only to invoke the inequality

$$\log_a x \leq (x - 1) \log_a e, \tag{4}$$

which is well known, and easily proved by elementary calculus. Then if $p_1, \cdots, p_b$ are any nonnegative numbers which sum to 1,

$$\sum p_k \log_a \frac{1}{p_k} - \sum p_k l_k = \sum p_k \log_a \frac{a^{-l_k}}{p_k} \leq \sum p_k \left(\frac{a^{-l_k}}{p_k} - 1 \right) \log_a e = 0.$$

Since equality in (4) occurs only when $x = 1$, equality in (2) can occur only if each $p_k = a^{-l_k}$, *i.e.*, if and only if $l_k = -\log_a p_k$.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656; July–October, 1948.