Z. Wahrscheinlichkeitstheorie verw. Geb. 15, 186–194 (1970)

The Capacity of a Channel with Arbitrarily Varying Channel Probability Functions and Binary Output Alphabet*

R. AHLSWEDE and J. WOLFOWITZ

Summary. Let $X=\{1,\ldots,a\}$ be the "input alphabet" and $Y=\{1,2\}$ be the "output alphabet". Let $X^t=X$ and $Y^t=Y$ for $t=1,2,\ldots$, $X_n=\prod_{t=1}^{n} X^t$ and $Y_n=\prod_{t=1}^{n} Y^t$. Let S be any set, $\mathscr{C}=\{w(\cdot|\cdot|s)|s\in S\}$ be a set of $(a\times 2)$ stochastic matrices $w(\cdot|\cdot|s)$, and $S^t=S$, $t=1,\ldots,n$. For every $s_n=(s^1,\ldots,s^n)\in\prod_{t=1}^{n} S^t$ define $P(\cdot|\cdot|s_n)$ by $P(y_n|x_n|s_n)=\prod_{t=1}^{n} w(y^t|x^t|s^t)$ for every $x_n=(x^1,\ldots,x^n)\in X_n$ and every $y_n=(y^1,\ldots,y^n)\in Y_n$. Consider the channel $\mathscr{C}_n=\{P(\cdot|\cdot|s_n)|s_n\in S_n\}$ with matrices $w(\cdot|\cdot|s)$ varying arbitrarily from letter to letter. The authors determine the capacity of this channel when a) neither sender nor receiver knows s_n, b) the sender knows s_n but the receiver does not, and c) the receiver knows s_n but the sender does not.

1. Introduction

Let $X=\{1,\ldots,a\}$ be the "input alphabet" and $Y=\{1,2\}$ be the "output alphabet" of the channels we shall study below. Results for $a>2$ will not appear until later sections. Hence, to simplify matters, we assume henceforth that $a=2$ unless the contrary is explicitly stated. (The case $a=1$ is trivial.) Let $X^t=X$ and $Y^t=Y$ for $t=1,2,\ldots$. By $X_n=\prod_{t=1}^{n} X^t$ we denote the set of input n-sequences (words of length n) and by $Y_n=\prod_{t=1}^{n} Y^t$ we denote the set of output n-sequences.

Let S be any set, and let $\mathscr{C}=\{w(\cdot|\cdot|s)|s\in S\}$ be a set of $(a\times 2)$ stochastic matrices $w(\cdot|\cdot|s)$. We shall refer to a $w(\cdot|\cdot|s)$ in the sequel either as a "matrix" or as a "channel probability function" (c.p.f.). Let $S^t=S$, $t=1,\ldots,n$. For every n-sequence $s_n=(s^1,\ldots,s^n)\in\prod_{t=1}^{n} S^t$ we define $P(\cdot|\cdot|s_n)$ by

$$P(y_n|x_n|s_n)=\prod_{t=1}^{n} w(y^t|x^t|s^t) \tag{1.1}$$

for every $x_n=(x^1,\ldots,x^n)\in X_n$ and every $y_n=(y^1,\ldots,y^n)\in Y_n$.

Now consider the channel

$$\mathscr{C}_n=\{P(\cdot|\cdot|s_n)|s_n\in S_n\}. \tag{1.2}$$

Suppose that sender and receiver want to communicate over the channel $\mathscr{C}_n$ without knowing which channel n-sequence s_n will govern the transmission of any

* Research of both authors supported by the U.S. Air Force under Grant AF-AFOSR-68-1472 to Cornell University.

word (input n-sequence). A code (n, N, λ) is a system

$$\{(u_1, A_1), \ldots, (u_N, A_N)\} \tag{1.3}$$

where the message sequence $u_i \in X_n$, $A_i \subset Y_n$, $i=1, \ldots, N$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and

$$P(A_i | u_i | s_n) \geqq 1-\lambda, \quad i=1, \ldots, N, \text{ and all } s_n \in S_n. \tag{1.4}$$

A number C is called the capacity of the channel if, for any $\varepsilon > 0$ and any λ, $0 < \lambda < 1$, the following is true for all n sufficiently large: There exists a code $(n, 2^{n(C-\varepsilon)}, \lambda)$ and there does not exist a code $(n, 2^{n(C+\varepsilon)}, \lambda)$.

The channel described above has been called (see [2]) a channel with arbitrarily varying c.p.f.'s, which we abbreviate thus: a.v.ch. The study of a.v.ch. was initiated in [2]. The authors of [2] did not limit themselves to the case where Y has only two elements, and obtained various partial results. We limit ourselves here to the case $|Y|=2$, but for the problems we treat our results go considerably beyond [2] and give the capacity of the particular channels studied. The problems where $|Y|>2$ seem to be unamenable to our methods.

Theorem 1 can easily be improved by using sharper estimates of the maximal code length for discrete memoryless channels.

2. Preliminary Lemmas

Once again we remind the reader that $a=2$ until the contrary is explicitly stated. In particular, $a=2$ in Sections 2 and 3.

We shall consider first a special case of a.v.ch. from which the general case can be easily derived. Consider two matrices w, w'. We denote the i-th row vector in w by i and the i-th row vector in w' by i'. We represent these vectors as points in E^2. Let the matrices w, w' be such that their representation is given by the following Fig. 1, in which the abscissa is the first coordinate of a vector:

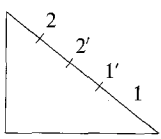

Fig. 1

If we define $w(\cdot|\cdot|1) = w(\cdot|\cdot)$ and $w(\cdot|\cdot|2) = w'(\cdot|\cdot)$, then $P(\cdot|\cdot|s_n)$ can be defined as in (1.1).

We say that the code (1.3) is a *strict* maximum likelihood code (s.m.l.c.) with respect to $P(\cdot|\cdot|s_n^*)$, where $s_n^* = (2, \ldots, 2)$, if, for given $u_1, \ldots, u_N$,

$$A_i = \{y_n | y_n \in Y_n \text{ and } P(y_n | u_i | s_n^*) > P(y_n | u_j | s_n^*) \text{ for } j \neq i\} \tag{2.1}$$

for $i=1, \ldots, N$. (Cf. [4], 7.3.1.)

Define

$${}_1A_i^t = \{y_n | y_n \in A_i \text{ and } y^t = u_i^t\},$$

$${}_2A_i^t = \{y_n | y_n \in A_i \text{ and } y^t \neq u_i^t\},$$

$${}_1A_i^{*t} = \{(y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n) | \text{ there exists } y^t \text{ such that } (y^1, \ldots, y^n) \in {}_1A_i^t\},$$

$${}_2A_i^{*t} = \{(y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n) | \text{ there exists } y^t \text{ such that } (y^1, \ldots, y^n) \in {}_2A_i^t\}.$$

Lemma 1. *With w and w' as in the figure, if $\{(u_i, A_i) | i=1, \ldots, N\}$ is a s.m.l.c. with respect to $P(\cdot|\cdot|s_n^*)$, then*

$$(2.2) \qquad {}_1A_i^{*t} \supset {}_2A_i^{*t}, \quad i=1, \ldots, N;\ t=1, \ldots, n,$$

and

$$(2.3) \qquad P(A_i|u_i|s_n) \geqq P(A_i|u_i|s_n^*)$$

for $i=1, \ldots, N$ and all $s_n \in S_n$.

Proof. Suppose first that $u_i^t=1$.

Let $(y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n) \in {}_2A_i^{*t}$ and $(y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n) \notin {}_1A_i^t$. This could occur for only one of two reasons:

(a) there exists u_j, $j \neq i$, such that

$$(2.4) \quad P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_j|s_n^*\big) > P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_i|s_n^*\big)$$

or

(b) $(y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n) \notin \bigcup_{i=1}^{n} A_i$, and there exists a $k \neq i$ such that

$$(2.5) \quad P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_k|s_n^*\big) = P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_i|s_n^*\big).$$

Whatever the situation may be, we can find a $j \neq i$, such that

$$(2.6) \quad P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_j|s_n^*\big) \geqq P\big((y^1, \ldots, y^{t-1}, u_i^t, y^{t+1}, \ldots, y^n)|u_i|s_n^*\big).$$

In case $w(\cdot|2|2) \equiv w(\cdot|1|2)$, $N=1$ and the lemma holds. We can therefore assume that $w(1|2|2) \neq w(1|1|2)$. This implies that $w(1|1|2)>0$ (see Fig. 1).

Suppose first that $u_j^t=1$. Multiplying both sides of (2.6) by $\dfrac{w(2|1|2)}{w(1|1|2)} \geqq 0$ we obtain

$$(2.7) \quad P\big((y^1, \ldots, y^{t-1}, 2, y^{t+1}, \ldots, y^n)|u_j|s_n^*\big) \geqq P\big((y^1, \ldots, y^{t-1}, 2, y^{t+1}, \ldots, y^n)|u_i|s_n^*\big).$$

This contradicts the fact that $(y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n) \in {}_2A_i^{*t}$ and proves (2.2) in this case.

Suppose now that $u_j^t=2$. It follows from $w(1|2|2) \leqq w(1|1|2)$, $w(1|1|2)>0$, and (2.6) that

$$(2.8) \quad \begin{aligned} &P\big((y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n)|(u_j^1, \ldots, u_j^{t-1}, u_j^{t+1}, \ldots, u_j^n)|s_n^*\big) \\ &\geqq P\big((y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n)|(u_i^1, \ldots, u_i^{t-1}, u_i^{t+1}, \ldots, u_i^n)|s_n^*\big). \end{aligned}$$

However, $w(2|2|2) \geqq w(2|1|2)$ and (2.8) imply that

$$(2.9) \quad P\big((y^1, \ldots, y^{t-1}, 2, y^{t+1}, \ldots, y^n)|u_j|s_n^*\big) \geqq P\big((y^1, \ldots, y^{t-1}, 2, y^{t+1}, \ldots, y^n)|u_i|s_n^*\big),$$

which also contradicts the fact that $(y^1, \ldots, y^{t-1}, y^{t+1}, \ldots, y^n) \in {}_2A_i^{*t}$. This proves (2.2) when $u_i^t=1$. The proof when $u_i^t=2$ is (symmetrically) the same.

We now prove (2.3) inductively. Assume that (2.3) holds for $s_n' \in S_n$. We shall show that (2.3) then holds for s_n, where s_n is obtained from s_n' by changing the element 2 in the k-th component of s_n' to a 1.

Define $B(i, k)$ by

$$ {}_1A_i^{*k} = {}_2A_i^{*k} \cup B(i,k), \qquad {}_2A_i^{*k} \cap B(i,k) = \emptyset. $$

This definition is possible because of (2.2). Let

$$ P\big(B(i,k)|(u_i^1, \ldots, u_i^{k-1}, u_i^{k+1}, \ldots, u_i^n)|s_n'\big) = a_1', $$
$$ P\big({}_2A_i^{*k}|(u_i^1, \ldots, u_i^{k-1}, u_i^{k+1}, \ldots, u_i^n)|s_n'\big) = a_2'. $$

If $u_i^k = 1$, then

$$ P(A_i|u_i|s_n') = w(1|1|2)\, a_1' + a_2' $$

and if $u_i^k = 2$, then

$$ P(A_i|u_i|s_n') = w(2|2|2)\, a_1' + a_2'. $$

If now in the k-th component of s_n' we replace $w(\cdot|\cdot|2)$ by $w(\cdot|\cdot|1)$, then in both cases we get $P(A_i|u_i|s_n) \geqq P(A_i|u_i|s_n')$, because $w(1|1|1) \geqq w(1|1|2)$ and $w(2|2|1) \geqq w(2|2|2)$. This completes the proof of the lemma.

We now need the following definitions:

(2.10) The entropy of a probability vector $\pi = (\pi_1, \ldots, \pi_c)$ is defined to be

$$ H(\pi) = -\sum_{i=1}^{c} \pi_i \log \pi_i. $$

(2.11) The rate for the probability vector π on X and c.p.f. $w(\cdot|\cdot|s)$ is $R\big(\pi, w(\cdot|\cdot|s)\big) = H\big(\pi'(s)\big) - \sum_i \pi_i H\big(w(\cdot|\cdot|s)\big)$, where $\pi'(s) = \pi \cdot w(\cdot|\cdot|s)$.

(2.12) $N(n, \lambda)$ is the maximal length of a (N, n, λ)-code for $\mathscr{C}_n$.

(2.13) For every fixed $i \in X$, $T(i)$ denotes the minimal closed convex system of probability distributions on Y which contains all distributions $\{w(\cdot|i|s) \,|\, s \in S\}$.

(2.14) The set of $(a \times 2)$ stochastic matrices

$$ \bar{\mathscr{C}} = \Big\{ \big(w(j|i)\big)_{\substack{i=1,\ldots,a \\ j=1,2}} \,\Big|\, w(\cdot|i) \in T(i),\ i = 1, \ldots, a \Big\} $$

is called the row convex closure of the set $\mathscr{C}$.

We shall need

Lemma 2. *Let $\{(u_i, A_i) \,|\, i = 1, \ldots, N\}$ be a code with average error $\bar{\lambda}$ for a single channel n-sequence. There exists a subcode of length $N/2$ with maximal error $\lambda = 2\bar{\lambda}$.* (See [4], Lemma 3.1.1.)

Lemma 3. *An (n, N, λ) code for $\mathscr{C}_n$ is an (n, N, λ) code for $\bar{\mathscr{C}}_n$, and conversely.*

Proof. Denote by Σ the σ-field of all subsets of S, and by Σ_n the σ-field of all subsets of S_n.

$$ P(A_i|u_i|s_n) \geqq 1 - \lambda \qquad \text{for all } s_n \in S_n $$

implies that

$$ \int_{S_n} d\, q_{u_i}(s_n)\, P(A_i|u_i|s_n) \geqq 1 - \lambda $$

for all probability distributions q_{u_i} on (S_n, Σ_n) and all $i=1, \ldots, N$. Any element $P(\cdot|\cdot)$ of $\overline{\mathscr{C}}$ can be approximated row-wise arbitrarily closely by expressions of the form

$$\int_S d\, q_x(s)\, P(\cdot|x|s), \qquad x \in X$$

where q_x is concentrated on finitely many points. Any element $P_n(\cdot|\cdot)$ of $\overline{\mathscr{C}}_n$ can be approximated row-wise arbitrarily closely by expressions of the form

$$\int_{S_n} d\, q^*_{u_i}(s_n)\, P(\cdot|u_i|s_n), \qquad i=1, \ldots, N,$$

where $q^*_{u_i}$ is a distribution on S_n which is concentrated on finitely many points and which is a product of suitable distributions q_{a_j}, where $j=1, \ldots, n$ and $u_i=(a_1, \ldots, a_n)$. This proves the first part of the lemma. The converse is obvious.

Lemma 4.

$$\max_{\pi} \min_{w \in \overline{\mathscr{C}}} R(\pi, w) = \min_{w \in \overline{\mathscr{C}}} \max_{\pi} R(\pi, w).$$

Proof. It is known that $R(\pi, w)$ is concave in π for each w and convex in w for each π. $\overline{\mathscr{C}}$ and $\{\pi\}$ are normcompact convex sets and $R(\pi, w)$ is normcontinuous in both variables. Therefore the minimax theorem ([6]) is applicable and the desired result follows.

This lemma is due to Stiglitz [7]. His proof was given here because it is so brief.

3. The Capacity when $a=2$

We shall now prove

Theorem 1. *Define*

$$C = \max_{\pi} \inf_{w \in \overline{\mathscr{C}}} R(\pi, w).$$

For every λ, $0<\lambda<1$, *the following estimates hold:*

a) $N(n, \lambda) > e^{Cn - k(\lambda)\sqrt{n}}$,

b) $N(n, \lambda) < e^{Cn + k(\lambda)\sqrt{n}}$

where $k(\lambda)$ *is a known function of* λ *and* $n=1, 2, \ldots$.

Proof. Let w' be such that $\max_{\pi} R(\pi, w') = \inf_{w \in \overline{\mathscr{C}}} \max_{\pi} R(\pi, w)$. It follows from Lemma 3 that a λ-code for $\mathscr{C}_n$ is also a λ-code for the d.m.c. determined by w'. Therefore statement b) is a consequence of the strong converse for the d.m.c. ([3], [4]). We can assume without loss of generality (w.l.o.g.) that w' has a representation

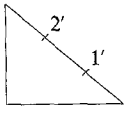

Fig. 2

Choose any $\bar{w} \in \mathscr{C}$. Then $\bar{w}(1|1) \geqq w'(1|1)$, $\bar{w}(2|2) \geqq w'(2|2)$, because otherwise we could, by convex combinations, produce a matrix $w^* \in \overline{\mathscr{C}}$ with $\max_{\pi} R(\pi, w^*) <$

$\max_{\pi} R(\pi, w')$. We therefore have the representation

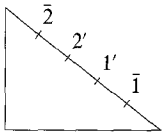

Fig. 3

Now let $w \in \bar{\bar{\mathscr{C}}}$ be such that

$$w(1|1) \geqq w^{**}(1|1), \qquad w(2|2) \geqq w^{**}(2|2)$$

for all $w^{**} \in \bar{\bar{\mathscr{C}}}$. Obviously $\bar{\bar{\mathscr{C}}} =$ row convex closure of $\{w, w'\}$. Again by Lemma 3 it is sufficient to prove a) for $\mathscr{C} = \{w, w'\}$.

It follows from Shannon's random coding theorem ([5], [4]) that we can find a s.m.l.c. for the d.m.c. w' with average error $\bar{\lambda} = \lambda/2$ and length $N(n, \lambda) > e^{Cn - k(\lambda)\sqrt{n}}$.

By Lemma 2 there exists a subcode with length $N(n, \lambda)/2$ and maximal error λ.

Application of Lemma 1 completes the proof.

An examination of the proof of Theorem 1 shows the following:

$$T(1) \cap T(2) \neq \emptyset \Leftrightarrow C = 0. \tag{3.1}$$

This was proved in [2], Theorem 1. When $C = 0$ then, in the next to the last diagram, $1'$ and $2'$ coincide, and conversely.

(3.2) *C is the capacity of the channel with $|S| = 1$ whose single matrix has, as its i-th row, $i = 1, 2$, the point of $T(i)$ closest to $T(i')$, $i' \neq i$.*

4. Extension of Theorem 1 to $a > 2$

Theorem 2. *Theorem 1 holds verbatim when $a > 2$.*

For each pair $i \neq i'$ let $C(i, i')$ be the capacity of the matrix whose rows are the point of $T(i)$ nearest to $T(i')$ and the point of $T(i')$ nearest to $T(i)$. (When $T(i) \cap T(i') \neq \emptyset$ these points may not be uniquely defined, but $C(i, i') = 0$ anyhow.)

Theorem 3. *Under the conditions of Theorem 2 we have*

$$C = \max_{i \neq i'} C(i, i'). \tag{4.1}$$

From (4.1) we easily obtain

$$C > 0 \Leftrightarrow T(i) \cap T(i') = \emptyset \quad \text{for some pair } i, i'. \tag{4.2}$$

((4.2) is a special case for $b = 2$ of Theorem 1 of [2].)

Before proving Theorems 2 and 3 we shall need

Lemma 5. *Let M be an $(a \times b)$ stochastic matrix. Suppose all rows are convex linear combinations of two (extreme) rows. Let M' be the $(2 \times b)$ matrix of these rows. Then the capacity of M equals the capacity of M'.*

This lemma must be present, explicitly or implicitly, in one of Shannon's papers. It is easily proved from the expression for the capacity of an individual channel given, e.g., in Theorem 3.1.1 of [4].

We now prove Theorems 2 and 3. In order to define a convenient terminology let us say that, in the next to the last diagram (which appears in the proof of Theorem 1), the point 2′ is to the left of the point 1′, the point 1′ is to the right of 2′, and 1′ (resp. 2′) is the right (resp. left) end of the interval [1′, 2′]. If two points coincide, each is to be to the left and right of the other. Let z (resp. z') be the farthest to the right (resp., to the left) of the left (resp. right) ends of the intervals $T(i)$, $i=1, \ldots, a$. Without loss of generality we assume that z is the left end of $T(1)$ and z' is the right end of $T(2)$.

Suppose first that z is to the left of z'. Then obviously $C=0$. According to Theorem 1 of [2], the capacity of the channel is zero. Hence Theorems 2 and 3 are valid in this case.

Suppose now that z is to the right of z'. From Lemma 5 it follows that C is the capacity of the (2×2) matrix with rows z and z'. Using only the letters 1 and 2 of the input alphabet we see that the capacity of the channel is at least C. Thus the proof of Theorems 2 and 3 will be complete when we prove the converse part.

Suppose that the message sequences u_i, $i=1, \ldots, N$, consist only of 1's and 2's. Then the converse is obvious (or follows from Theorem 1). Suppose now all the other input letters are also used in the u_i. Since we require the error of decoding to be no greater than λ for *every* word u_i and *every* channel n-sequence, we can picture the situation as if some malevolent being, to be called, say, the "jammer", could choose the c.p.f. for each letter *after* he knows the letter being sent. It follows from Lemma 3 that he can achieve that the point on the diagram which corresponds to the letter being sent lies in the interval $[z, z']$. The desired converse now follows from Lemma 5. This completes the proof of Theorems 2 and 3.

Let $m_1, \ldots, m_a$ (resp., $m_1', \ldots, m_a'$) be the left (resp., the right) end points of $T(1), \ldots, T(a)$, respectively. Let μ (resp., μ') be that one of m_i (resp., m_i'), $i=1, \ldots, a$, which is farthest to the right (resp., to the left). A convenient way of computing C is given by the following, which we state as a theorem for ease of reference:

Theorem 3′. *If μ' is to the right of μ, then $C=0$. Otherwise C is the capacity of the matrix with rows μ and μ'.*

5. The Case where the Sender but not the Receiver Knows the c.p.f. for Each Letter, and $a \geqq 2$

We now study the case described in the title of this section. We assume that the sender knows the c.p.f. for each letter in advance of sending that letter. We also assume that he knows all the preceding c.p.f.'s, but does *not* know any future c.p.f.'s.

Using the method of proof of Theorem 4.9.1 of [4] (see also the proof of Theorem 4.8.1 of [4]) one can show that the capacity of our channel is unaltered if we limit ourselves to codes where the sender chooses the next letter to be sent solely on the basis of the c.p.f. which will govern the transmission of *this* letter and not on the basis of preceding c.p.f.'s. (The fundamental reason for this is that the channel is memoryless.) Henceforth we limit ourselves to such codes.

To make the proof easier to follow we start with the case $|S|=2$, and then remove this limitation. Denote the two matrices in $\mathscr{C}$ by A and B, say. We can describe the codes for the present channel by the following device: The sender's (input)

alphabet is to consist of a^2 pairs (i,j), $i,j=1,\ldots,a$. The "letter" (i,j) means that, when the sender knows that A (resp. B) will govern the transmission of the letter, he sends the letter i (resp., j). By this simple device we have reduced the problem to that treated in Theorem 2.

We have already seen that Theorem 3 or Theorem 3′ implies that the expression C of Theorem 1 is a function only of the closed convex sets $T(i)$, $i=1,\ldots,a$. We may therefore write

$$C=C\big(T(1),\ldots,T(a)\big). \tag{5.1}$$

It is clear (e.g., from Theorem 3′) that the right member of (5.1) is well defined even if the number of sets T is infinite, provided that, in the computations implied by Theorems 3 and 3′, we replace the operation "max" by "sup".

Now let

$$A=\begin{pmatrix} g_1 \\ \vdots \\ g_a \end{pmatrix}, \qquad B=\begin{pmatrix} h_1 \\ \vdots \\ h_a \end{pmatrix}.$$

Consider the following a^2 convex bodies, each determined by the two points exhibited:

$$T(g_i,h_j), \qquad i,j=1,\ldots,a.$$

It follows from Theorem 2 that we have proved that the capacity of the channel being discussed is

$$C\big(\{T(g_i,h_j),\ i,j=1,\ldots,a\}\big). \tag{5.2}$$

We now drop the restriction that $|S|=2$. Consider the totality $\mathscr{B}'=\{B'\}$ of sets B' such that each B' consists of *exactly* one row from each matrix $w(\cdot\,|\,\cdot\,|\,s)$, $s\in S$. Let B be the convex hull of B' and $\mathscr{B}=\{B\}$ be the totality of sets (intervals) B. The general case follows from the preceding remarks and the argument of Theorem 2 of [8]. We have thus proved

Theorem 4. *When the sender, but not the receiver, knows the c.p.f. being used for each letter of an a.v.ch., the capacity of the channel is* $C(\mathscr{B})$.

6. The Case where the Receiver but not the Sender Knows the c.p.f. for Each Letter, and $a\geqq 2$

We now study the case described in the title of this section. We introduce one change from the preceding channels: We assume that the jammer is allowed to choose each c.p.f. (for each letter) by a random process, i.e., the i-th c.p.f., $i=1,\ldots,n$, is chosen according to a probability distribution q_i on (S,Σ). The receiver knows the sequence $(q_1,q_2,\ldots,q_n)$ when he decodes the received n-sequence (i.e., decides which transmitted n-sequence u_i was sent). This assumption is very realistic in this case and in the case treated in Theorems 1 and 2. It was not made explicitly there because it is unnecessary; Lemma 3 essentially involves it.

We shall now prove

Theorem 5. *When the receiver (but not the sender) knows the sequence* $(q_1,\ldots,q_n)$, *the capacity of the channel is the same as that given in Theorem* 2, *i.e.,*

the capacity is the same as it would be if the receiver did not know the sequence $(q_1, \ldots, q_n)$.

Proof. Since the capacity cannot be less than that in Theorem 2, it remains only to prove the converse. The capacity of Theorem 2 is, by Lemma 4, the smallest of the capacities of the matrices in $\bar{\mathscr{C}}$. Clearly, the jammer can choose a q for each letter which will produce (row-wise) the matrix in $\bar{\mathscr{C}}$ whose capacity is smallest. This proves the desired result.

7. Miscellaneous Remarks

Theorems 2, 4, and 5 hold, with essentially the same proofs, when a is not finite. One uses the argument of Theorem 2 in Chapter II of [8].

Theorems 1 and 3 of [2] were proved for arbitrary but finite a and b and $|S| < \infty$. It is a consequence of Theorems 2 and 4 of the present paper and the argument of Theorem 2 in Chapter II of [8] that Theorems 1 and 3 of [2] hold for arbitrary (not necessarily finite) a and S, and $b = 2$. An examination of the proofs of Theorems 1 and 3 of [2] in the light of these latter results shows that the restriction (made in [2]) to $|S| < \infty$ was unnecessary, and that the proofs of [2] carry over verbatim to arbitrary S.

References

1. Ahlswede, R., Wolfowitz, J.: Correlated decoding for channels with arbitrarily varying channel probability functions. Inform. and Control **14**, 457–473 (1969).
2. Kiefer, J., Wolfowitz, J.: Channels with arbitrarily varying channel probability functions. Inform. and Control **5**, 44–54 (1962).
3. Wolfowitz, J.: The coding of messages subject to chance errors. Illinois Jour. Math. **1**, 591–606 (1957).
4. – Coding theorems of information theory. Berlin-Heidelberg-New York: Springer, first edition, 1961; second edition, 1964.
5. Shannon, C.E.: Certain results in coding theory for noisy channels. Inform. and Control **1**, 6–25 (1957).
6. Kakutani, S.: A generalization of Brouwer's fixed point theorem. Duke math. J. **8**, 457–458 (1941).
7. Stiglitz, I.G.: Coding for a class of unknown channels. IEEE Trans. Inform. Theory **IT-12**, 189–195 (1966).
8. Ahlswede, R.: Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. Z. Wahrscheinlichkeitstheorie verw. Geb. **10**, 1–42 (1968).

Professor R. Ahlswede
Ohio State University
Columbus, Ohio, USA

Professor J. Wolfowitz
University of Illinois
Dept. of Mathematics
Urbana, Ill. 61801, USA

(Received January 7, 1969)