The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 1027–1033

A NOTE ON THE EXISTENCE OF THE WEAK CAPACITY FOR CHANNELS WITH ARBITRARILY VARYING CHANNEL PROBABILITY FUNCTIONS AND ITS RELATION TO SHANNON'S ZERO ERROR CAPACITY[1]

By R. Ahlswede

Ohio State University

0. Introduction. In [7] Kiefer and Wolfowitz stated the coding problem for channels with arbitrarily varying channel probability functions (a.v.ch.)—including cases with side information—for pure codes and maximal errors, and gave necessary and sufficient conditions for the channel rate to be positive. (A detailed discussion about the different coding problems which arise in the theory of a.v.ch. by using different code concepts and different error concepts is given in [2].)

It was undecided for a long time whether the coding theorem and the weak converse or also the strong converse of the coding theorem hold. (Compare [5] page 566.) If the coding theorem and weak converse hold for a channel then we say that this channel has a *weak* capacity, and if also the strong converse holds then we say that this channel has a *strong* capacity.

In Section 1 we prove that the a.v.ch. considered in [7] have a weak capacity. In case of output alphabets of length 2 an *explicit* formula for the strong capacity is even known [3]. One would like to have this sharper result for general finite output alphabets, but, while awaiting a solution of this difficult problem, it would be of interest to know that at least the weak capacity exists. The disadvantage of our method is obviously that it does not lead to a formula for the capacity. However, already for stationary semicontinuous compound channels a reasonable formula for the weak capacity is unknown [6]. Moreover, in Section 2 we shall show that Shannon's zero-error-capacity problem for the discrete memoryless channel (d.m.c.) is equivalent to finding an explicit formula for the capacity of a special a.v.ch. defined by a set of stochastic 0–1 matrices. Therefore an explicit formula for the weak capacity of an a.v.ch. would imply the solution of Shannon's problem, which is known to be of a graph theoretical nature and very difficult (cp. [8], [4]). The close relation between the two problems might give some hope of finding explicit formulas for the error-capacity of an a.v.ch. also in other cases of an alphabet length for which the zero-error capacity is known.

Our method for proving the existence of the weak capacity for an a.v.ch. applies to infinite alphabets and other channels than a.v.ch. (Corollary 1 and Corollary 2 in Section 1).

1. The existence of the weak capacity for a.v.ch. Let $X^t = \{1, \cdots, a\}$, $Y^t = \{1, \cdots, b\}$ for $t = 1, 2, \cdots$ and let $\mathscr{L} = \{w(\cdot|\cdot|s) | s \in S\}$ be a set of stochastic

Received February 20, 1969; revised January 5, 1970.

[1] Research of the author supported by the National Science Foundation under Grant Contract No. GP-9464 to the Ohio State University.

matrices with a rows and b columns. 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.

For every n-sequence $s_n = (s^1, \cdots s^n) \in \prod_1^n S$ we define a discrete memoryless channel $P(\cdot|\cdot|s_n)$ by $P(y_n|x_n|s_n) = \prod_{t=1}^n w(y^t|x^t|s^t)$ for every $x_n = (x^1, \cdots x^n) \in X_n$ and every $y_n = (y^1, \cdots y_n) \in Y_n$. Consider now the class of channels $\mathscr{L}_n = \{P(\cdot|\cdot|s_n) \mid s_n \in S_n\}$.

Suppose now sender and receiver want to communicate without knowing which channel of the channels in $\mathscr{L}_n$ actually governs the transmission. (cf. [7], Section 2.)

(1) A-code (n, N, λ) for the present problem (when the channel probability function varies arbitrarily) is a system $\{(u_i, A_i) \mid i = 1, \cdots, N\}$, where $u_i \in X_n$, $A_i \subset Y_n$, $A_i \cap A_j = \emptyset$ for $i \neq j$ and $P(A_i|u_i|s_n) \geqq i - \lambda$ for all $i = 1, \cdots, N$ and all $s_n \in S_n$.

(2) A number C is called the weak capacity (of the a.v.ch.), if (a) for any $\delta > 0$ and $\lambda (0 < \lambda < 1)$ there exists a code $(n, 2^{n(C-\delta)}, \lambda)$ for all sufficiently large n, (b) for any $\delta > 0$ there exists a $\lambda = \lambda(\delta)$ such that for all sufficiently large n there does not exist a code $(n, 2^{n(C+\delta)}, \lambda)$.

(3) C is called the strong capacity if (a) holds and (b) is replaced by (b′) for any $\delta > 0$ and any $\lambda (0 < \lambda < 1)$ there does not exist a code $(n, 2^{n(C+\delta)}, \lambda)$ for all sufficiently large n. (a), (b′) imply (a), (b).

(4) Let $N(n, \lambda)$ be the maximal length of a (n, N, λ)-code for $\mathscr{L}_n$.

(5) The entropy of a probability vector $\pi = (\pi_1, \cdots \pi_c)$ is defined to be $H(\pi) = -\sum_{i=1}^c \pi_i \log \pi_i$.

(6) Denote the rate for the probability vector π on X and cpf $w(\cdot|\cdot|s)$ by $R(\pi, w(\cdot|\cdot|s)) = H(\pi'(s)) - \sum_{i=1}^a \pi_i H(w(\cdot|\cdot|s))$, where $\pi'(s) = \pi \cdot w(\cdot|\cdot|s)$.

(7) In case $a = b = d$ we call the stochastic matrix $(w(i|j))\ i = 1, \cdots, d; j = 1, \cdots, d$ d-ary symmetric, if

$$\begin{aligned} w(j \mid i) &= 1 - \varepsilon \qquad &\text{for} \quad i = j \\ &= \varepsilon(d-1)^{-1} \qquad &\text{for} \quad i \neq j. \end{aligned}$$

(8) The discrete memoryless channel $P(\cdot|\cdot)$ defined by $P(y_n|x_n) = \prod_{t=1}^n w(y^t|x^t)$ for every $x_n = (x^1, \cdots x^n) \in X_n$, $y_n = (y^1, \cdots y^n) \in Y_n$, and all $n = 1, 2 \cdots$ is called a d-ary symmetric channel.

(9) The a.v.ch. $(\mathscr{L}_n)$ $n = 1, 2, \cdots$ is called d-ary symmetric if for some $\varepsilon(\frac{1}{2} > \varepsilon > 0)$ $\mathscr{L} = \{w(\cdot|\cdot) \mid w(i|i) \geqq 1 - \varepsilon \text{ for } i = 1, \cdots, d\}$. Using a suitable index-set S we can write $\mathscr{L} = \{w(\cdot|\cdot|s) \mid s \in S\}$.

(10) We say $\{(u_i, A_i) \mid u_i \in X_n, A_i \subset Y_n \text{ for } i = 1, \cdots N, A_i \cap A_j = \emptyset \text{ for } i \neq j\}$ is a *strict* maximum likelihood code (s.m.l.c.) with respect to $P(\cdot|\cdot)$, if $A_i = \{y_n \mid y_n \in Y_n$ and $P(y_n|u_i) > P(y_n|u_j)$ for $j \neq i\}$ for $i = 1, \cdots, N$. (cf. [10] 7.3.1.)

(11) As usual the Hamming distance between n-sequences is defined as $h(x_n, y_n)$ = number of components in which x_n and y_n are different. After these preparations we can state

LEMMA 1. *Let* $\mathscr{L}_n$ *be a d-ary symmetric* a.v.ch. *and let* $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ *be a* s.m.l.c. *with respect to the* d.m.c. $P(\cdot \mid \cdot)$, *given by*

$$w(j \mid i) = 1-2\varepsilon \quad \text{for} \quad i=j$$
$$= 2\varepsilon(d-1)^{-1} \quad \text{for} \quad i \neq j.$$

Then

(12) $P(A_i \mid u_i \mid s_n) \geqq P(A_i \mid u_i)$ *for* $i = 1, \cdots, N$ *and all* $P(\cdot \mid \cdot \mid s_n) \in \mathscr{L}_n$.

PROOF. Notice first that $y_n \in A_i$ if and only if

(13) $$h(y_n, u_i) < h(y_n, u_j) \qquad \text{for} \quad j \neq i.$$

(14) we can assume for symmetry reasons without loss of generality that the tth component of u_i is 1.

Define

$${}_jA_i^t = \{y_n \mid y_n \in A_i \text{ and } y^t = j\} \quad \text{and}$$
$${}_jA_i^{*t} = \{(y^1, \cdots, y^{t-1}, y^{t+1}, \cdots, y^n) \mid (y^1, \cdots, y^{t-1}, j, y^{t+1}, \cdots, y^n) \in {}_jA_i^t\}$$

for $i = 1, \cdots, N$ and $j = 1, \cdots, d$.

It follows from (13), (14) that

(15) $${}_1A_i^{*t} \supset {}_jA_i^{*t} \qquad \text{for all} \quad j = 1, \cdots, d.$$

In order to have a unique description we write $P(\cdot \mid \cdot \mid s_n)$ with $s_n = (1, \cdots, 1)$ instead of $P(\cdot \mid \cdot)$. We shall now prove the lemma iteratively. Suppose (12) holds for s_n', we shall show that (12) holds for s_n^*, where s_n^* can be produced by changing the tth component of s_n' from a 1 to s^{*t}. Because $w(1 \mid 1 \mid s^{*t}) \geqq w(1 \mid 1 \mid 1) + \varepsilon$ and because of (14) and (15) $P(A_i \mid u_i \mid s_n^*) \geqq P(A_i \mid u_i \mid s_n')$ for $i = 1, \cdots, N$. This completes the proof. (Similar arguments were used in [3].)

THEOREM 1. *The capacity of the d-ary symmetric* a.v.ch. *is*

$$\geqq \max_\pi R(\pi, w(\cdot \mid \cdot)) = \log d + (1-2\varepsilon)\log(1-2\varepsilon) + 2\varepsilon \log 2\varepsilon/(d-1)$$
$$\geqq (1-2\varepsilon)\log d + (1-2\varepsilon)\log(1-2\varepsilon) + 2\varepsilon \log 2\varepsilon.$$

PROOF. We have to verify (a). Using Shannon's random coding method we achieve a code with average error $\bar{\lambda}$ of the desired length for channel $P(\cdot \mid \cdot)$. Without essential loss in the code length we can reduce this code to a code with maximal error $\lambda = 2\bar{\lambda}$ (Lemma 3.1.1 [10]). By Lemma 1 this code is even a λ-code for all channels in $\mathscr{L}_n$. This proves (a). An elementary calculation gives the formula for $\max_\pi R(\pi, w(\cdot \mid \cdot))$. Theorem 1 plays a basic role in the proof for the main result of this section:

THEOREM 2. *The weak capacity for an* a.v.ch. *exists.*

PROOF. Define $C^+(\lambda) = \limsup_{n \to \infty} n^{-1} \log N(n, \lambda)$ and $C = \inf_{0<\lambda<1} C^+(\lambda)$.

$C^+(\lambda)$ and therefore C exist, because $\log N(n, \lambda) \leqq n \log b$. For $\delta > 0$ and $\varepsilon (0 < \varepsilon < 1)$ there exists a $k = k(\delta, \varepsilon)$ such that $k^{-1} \log N(k, \varepsilon) \geqq C - \frac{1}{4}\delta$. Let $\{u_i, A_i) \mid i = 1, \cdots, N(k, \varepsilon)\}$ be a $(k, N(k, \varepsilon), \varepsilon)$-code for $\mathscr{L}_k$.

We use $\{u_i \mid i = 1, \cdots, N(k, \varepsilon)\}$ as input alphabet and $\{A_i \mid i = 1, \cdots, N(k, \varepsilon)\}$ as output alphabet of the a.v.ch. determined by the class of stochastic matrices $\mathscr{L}' = \{(w'(\cdot \mid \cdot) \mid w'(j \mid i) = P(A_j | u_i | s_k) \text{ for some } P(\cdot \mid \cdot | s_k) \in \mathscr{L}_k, i, j = 1, \cdots, N(k, \varepsilon)\}$.

$\mathscr{L}'$ contains by definition only matrices with

$$w'(i \mid i) \geqq 1 - \varepsilon \qquad \text{for} \quad i = 1, \cdots, N(k, \varepsilon). \tag{16}$$

Let $\mathscr{L}^* \supset \mathscr{L}'$ be the set of all stochastic matrices satisfying (16). $\mathscr{L}^*$ determines a d-ary symmetric a.v.ch. with $d = N(k, \varepsilon)$. Let $N^*(t, \lambda)$ be the maximal length of a (t, N, λ)-code for $\mathscr{L}_t^*$.

(17) A code (t, N, λ) for $\mathscr{L}_t^*$ is a code (tk, N, λ) for $\mathscr{L}_{tk}$.

The idea is now that it is possible to find codes for $\mathscr{L}_t^*$ which are long enough even for $\mathscr{L}_{tk}$ if only ε is sufficiently small. Notice that, with decreasing ε, $k = k(\delta, \varepsilon)$, and therefore also the alphabet length $d = N(k, \varepsilon)$, will in general increase. This, for instance, makes Theorem 1 of chapter II of [1] inapplicable.

However, as a consequence of Theorem 1 we have

$$N^*(t, \lambda) > \exp\{(1-2\varepsilon)\log N(k, \varepsilon) + (1-2\varepsilon)\log(1-2\varepsilon) + 2\varepsilon \log 2\varepsilon - \eta\}t$$

for t sufficiently large and therefore and because of (17)

$$\begin{aligned} 1/tk \log N(tk, \lambda) &\geqq 1/tk \log N^*(t, \lambda) \\ &\geqq 1/k\{(1-2\varepsilon)\log N(k, \varepsilon) + (1-2\varepsilon)\log(1-2\varepsilon) + 2\varepsilon\log 2\varepsilon - \eta\} \\ &\geqq \{(1-2\varepsilon)(C - \tfrac{1}{4}\delta) + (1-2\varepsilon)\log(1-2\varepsilon) + 2\varepsilon \log 2\varepsilon - \eta\} \\ &\geqq C - \tfrac{1}{2}\delta \end{aligned} \tag{18}$$

for ε, η sufficiently small and t sufficiently large. Every nonnegative integer can be written as $n = tk + r$, where $0 \leqq r < k$. Using $N(tk + r, \lambda) \geqq N(tk, \lambda)$ and $\lim_{t \to \infty} (tk + r)/tk = 1$ we get $N(n, \lambda) \geqq C - \delta$ for all n sufficiently large.

Corollary 1. *The weak capacity of an* a.v.ch. *with infinite input and output alphabet exists if and only if*

$$C = \inf_{0 < \lambda < 1} \limsup_{n \to \infty} n^{-1} \log N(n, \lambda)$$

is finite.

This follows immediately by observing that only the finiteness of C was used in the proof of Theorem 2 and from the definition of the weak capacity.

(19) For every positive integer t let $\mathscr{L}^t = \{w(\cdot \mid \cdot | s^t) \mid s^t \in S^t\}$ be a set of stochastic matrices and let $\mathscr{L}_n$ be defined as usual.

Corollary 2. *Let* $(\mathscr{L}_n)_{n=1,2}\ldots$ *be a channel system as defined under* (19) *satisfying* $S^t \supset S^{t+1}$ *for* $t = 1, 2, \cdots$, *then the weak capacity exists.*

Noticing that (18) still holds one can give the same proof as for Theorem 2.

Remark 1. One can generalize this result to channel systems with an independence structure which satisfies (18). The independence property also can be weakened. It is enough that (18) holds asymptotically.

REMARK 2. The a.v.ch. considered above can be more fully described as an a.v.ch. where neither the sender nor the receiver knows the individual channel which governs the transmission of a code word. (S^-, R^-). Kiefer and Wolfowitz also introduced in [7] an a.v.ch. where only the sender (S^+, R^-) or only the receiver (S^-, R^+) knows the individual channel which governs the transmission of a code word.

In the case (S^-, R^+) a code (n, N, λ) is a system

$$\{(u_i, A_i(s_n)) \mid i = 1, \cdots, N;\ s_n \in S_n\},$$

where $u_i \in X_n$, $A_i(s_n) \subset Y_n$, the $A_1(s_n), \cdots, A_N(s_n)$ are disjoint for every $s_n \in S_n$, and

$$P(A_i(s_n) | u_i | s_n) \geqq 1-\lambda \quad \text{for} \quad s_n \in S_n, \qquad i = 1, \cdots, N.$$

Again the weak capacity exists. To see this one proceeds as in the case (R^-, S^-). First one defines $C^+(\lambda) = \limsup_{n\to\infty} n^{-1} \log N(n, \lambda)$ and $C = \inf_{0<\lambda<1} C^+(\lambda)$.

For $\delta > 0$ and $\varepsilon (0 < \varepsilon < 1)$ there exists a $k = k(\delta, \varepsilon)$ such that $k^{-1} \log N(k, \varepsilon) \geqq C - \frac{1}{4}\delta$.

Let $\{(u_i, A_i(s_n) \mid i = 1, \cdots, N(k, \varepsilon)\}$ be a $(k, N(k, \varepsilon), \varepsilon)$-code. We use $\{u_i \mid i = 1, \cdots, N(k, \varepsilon)\}$ as input alphabet and $\{i \mid i = 1, \cdots, N(k, \varepsilon)\}$ as output alphabet of the a.v.ch. (R^-, S^-) determined by the class of stochastic matrices $\mathscr{L}' = \{w'(\cdot | \cdot | s') | w'(j | i | s') = P(A_j(s_n) | u_i | s_n) \text{ for } i, j = 1, \cdots, N(k, \varepsilon) \text{ and some } s_n \in S_n\}$.

That means the set $\{A_j(s_n) \mid s_n \in S_n\}$ will be identified with letter j. The receiver knowing s_n still knows how to decode. We can now proceed as in the proof for Theorem 2.

A similar argument applies in case (S^+, R^-) and we therefore omit it.

2. The relation to Shannon's zero error capacity. In [8] Shannon introduced the zero error capacity C_0 of a d.m.c. as the least upper bound of rates at which it is possible to transmit information with zero probability of error. More precisely:

(a) a code $(n, N, 0)$ is a system of pairs $\{(u_i, A_i) \mid i = 1, \cdots, N\}$, where $u_i \in X_n$, $A_i \subset Y_n$ for $i = 1, \cdots, N$ and $A_i \cap A_j = \emptyset$ for $i \neq j$; satisfying $P(A_i | u_i) = 1$ for $i = 1, \cdots, N$.

(b) $N(n, 0)$ = maximal length N for which a code $(n, N, 0)$ exists. Then the zero error capacity can be defined as

(c) $C_0 = \limsup_{n\to\infty} n^{-1} \log N(n, 0)$. C_0 exists because $N(n, 0) \leqq d^n$.

It is easy to see that $N(n_1 + n_2, 0) \geqq N(n_1, 0) \cdot N(n_2, 0)$. Therefore one can write $C_0 = \lim_{n\to\infty} n^{-1} \log N(n, 0)$.

Shannon proved that for alphabet length $d \leqq 4$ always $C_0 = \log N(1, 0)$, in other words $N(n, 0) = N(1, 0)^n$.

This multiplicativity formula fails for $d = 5$ in the case: $X = Y = \{0, 1, \cdots, 4\}$.

$$\begin{aligned} w(j | i) &> 0 \qquad \text{for} \quad j = i, i+1 \pmod 5; \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

$$N(1, 0) = 2, \quad N(2, 0) = 5. \quad \text{(cf. [8].)}$$

One would like to have a "reasonable" formula for C_0 which does not "depend on an infinite product space." Such a formula is unknown. (An answer as: for given d there exist a $k = k(d)$ such that $N(nk, 0) = (N(k, 0))^n$ could be considered "reasonable".)

(20) We say the a.v.ch. $(\mathscr{L}_n)$ $n = 1, 2, \cdots$ is of 0–1-type, if $\mathscr{L}$ is a subset of the finite set $\{w(\cdot|\cdot) \mid w(\cdot|\cdot)$ stochastic and $w(i|j) = 0$ or $1\}$.

The elements of $\mathscr{L}$ may be written as $w(\cdot|\cdot|s)$ where s is an element of a suitable finite index set S. We ask for a formula for the weak capacity in this special case.

Let $\{(u_i, A_i) \mid i = 1, \cdots, N\}$ be a (n, N, λ)-code for $\mathscr{L}_n$, then by definition

$$P_n(A_i|u_i|s_n) \geqq 1 - \lambda \qquad \text{for all} \quad i = 1, \cdots, N \quad \text{and} \quad s_n \in S_n. \tag{21}$$

However, $P_n(y_n|x_n|s_n) = 0$ or 1 implies

$$P_n(A_i|u_i|s_n) = 1 \qquad \text{for} \quad i = 1, \cdots, N \quad \text{and all} \quad s_n \in S_n. \tag{22}$$

We define $w^*(i|j) = |S|^{-1} \sum_{s \in S} w(i|j|s)$ and the d.m.c. $P^*(\cdot|\cdot)$ by $P^*(y_n|x_n) = \prod_1^n w^*(y^t|x^t)$ for all $y_n \in Y_n$, $x_n \in X_n$; $n = 1, 2, \cdots$.

We say $P^*(\cdot|\cdot)$ is the d.m.c. corresponding to the 0–1-type a.v.ch. $(\mathscr{L}_n)$ $n = 1, \cdots$.

Let now q be a uniformly distributed probability measure on S: $q = (|S|^{-1}, \cdots, |S|^{-1})$ and q_n its independent product on $S_n = \prod_1^n S$, then

$$P^*(A|x_n) = \sum_{s_n \in S_n} q_n(s_n) P(A|x_n|s_n) \tag{23}$$

for all $x_n \in X_n$, $A \subset Y_n$ and therefore by (22)

$$P^*(A_i|u_i) = 1 \qquad \text{for} \quad i = 1, \cdots, N. \tag{24}$$

On the other hand, let us assume that (24) holds for a code $\{(u_i, A_i) \mid i = 1, \cdots, N\}$, then also (22) must hold for this code, because $q_n(s_n) > 0$ for all $s_n \in S_n$ and $P(A_i|u_i|s_n) < 1$ for some (i, s_n) would violate (24).

THEOREM 3. *Let* $(\mathscr{L}_n)_{n=1,2,\cdots}$ *be an* a.v.ch. *of* 0–1-*type and* $P^*(\cdot|\cdot)$ *the corresponding* d.m.c., *then*

(i) *The zero-error capacity* C_0 *of* $P^*(\cdot|\cdot)$ *is equal to the strong capacity of* $(\mathscr{L}_n)_{n=1,2,\cdots}$.

(ii) *The zero-error capacity of an arbitrary* d.m.c. *is equal to the strong capacity of a suitable* a.v.ch. *of* 0–1-*type*.

PROOF. (i) follows immediately from the equivalence of (21) and (24). Shannon calls two matrices w and w' adjacent, if $w(i|j) > 0$ when and only when $w'(i|j) > 0$. It is easy to see that d.m.c.'s which correspond to adjacent matrices have the same zero error capacity (cf. [8] page 10). Choose now $\mathscr{L} = \{w(\cdot|\cdot|s) \mid w(\cdot|\cdot|s)$ stochastic, $w(i|j|s) = 0$ or 1 and $w(i|j|s) = 0$ if $w(i|j) = 0\}$. The corresponding w^* and w are adjacent.

(ii) follows now by means of (i).

Shannon raised the question of finding a connection between the zero error capacity of a d.m.c. and the error capacity of a suitable other d.m.c. Theorem 3 gives a connection with the capacity of an a.v.ch.

REMARK 3. Theorem 3 makes it possible to compare results for C_0 with results for the capacity of a.v.ch.

The estimate of Shannon ([8] Theorem 1) $C_0 \leqq \inf\{C(w') \mid w' \text{ adjacent to } w\}$ means in terms of a.v.ch. $C_0 \leqq C_D$ (as defined in [2] Section 3).

That C_0 is in general unequal to $\inf\{C(w') \mid w' \text{ adjacent to } w\}$ follows therefore also from Example 1 in Section 3 of [2], where

$$w_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad w_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

and $\mathscr{L} = \{w_1, w_2\}$. Thus

$$w^* = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix},$$

$C_0 = 0$ for w^*, but $C_D > 0$.

REFERENCES

[1] AHLSWEDE, R. (1968). Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **10** 1–42.

[2] AHLSWEDE, R. and WOLFOWITZ, J. (1969). Correlated decoding for channels with arbitrarily varying channel probability functions. *Information and Control* **14** 457–473.

[3] AHLSWEDE, R. and WOLFOWITZ, J. The capacity of channels with arbitrarily varying channel probability functions and binary output alphabets. (to appear)

[4] BERGE, C. (1958). *Théorie des graphes et ses applications*. Dunod, Paris.

[5] BLACKWELL, D., BREIMAN, L. and THOMASIAN, I. (1960). The capacities of certain channel classes under random coding. *Ann. Math. Statist.* **31** 558–567.

[6] KESTEN, H. (1961). Some remarks on the capacity of compound channels in the semicontinuous case. *Information and Control* **4** 169–184.

[7] KIEFER, J. and WOLFOWITZ, J. (1962). Channels with arbitrarily varying channel probability functions. *Information and Control* **5** 44–54.

[8] SHANNON, C. E. (1956). The zero error capacity of a noisy channel. *Trans. IRE P6IT-2* 8–19.

[9] WOLFOWITZ, J. (1957). The coding of messages subject to chance errors. *Illinois J. Math.* **1** 591–606.

[10] WOLFOWITZ, J. (1964). Coding Theorems of Information Theory (2nd ed.). Springer-Verlag, New York.