INFORMATION AND CONTROL 16, 268-284 (1970)

Infinite-Alphabet Channels and the Method of Codes of a Fixed Composition

UDO AUGUSTIN*

Mathematisches Institut, Universität Erlangen, W. Germany

A proof of the strong converse of the coding theorem for stationary infinite-alphabet channels without memory fulfilling a certain supposition on finite coverings is presented. The proof indicates to which point the method of fixed composition codes can be used for infinite-alphabet channels. The special supposition for the proof of the strong converse (though not the most general one; compare for this: Augustin [1]) is of technical relevance and is satisfied in all cases of practical interest.

1. THE CHANNEL

Let $[X_0, (Y_0, F_0), P_0]$ be given, where X_0, Y_0 are nonempty sets (input and output alphabet, respectively), F_0 a σ field in Y_0 and $P_0 = P_0(\ , \ )$ a real function on $X_0 \times F_0$ s.t. $P_0(x_0, \ )$ is a probability on (Y_0, F_0) for each $x_0 \in X_0$. Furthermore, let $[X_v, (Y_v, F_v), P_v]$ $(v = 1, 2,...)$ be copies of $[X_0, (Y_0, F_0), P_0]$, $X_{[1,t]} := \prod_{v=1}^{t} X_v$, $Y_{[1,t]} := \prod_{v=1}^{t} Y_v$ Cartesian products of the X_v and Y_v, respectively, $F_{[1,t]} := \prod_{v=1}^{t} F_v$ the product σ field on $Y_{[1,t]}$ and let $P_{[1,t]} = P_{[1,t]}(\ , \ )$ be the real function on $X_{[1,t]} \times F_{[1,t]}$ determined by: $P_{[1,t]}(x, \ )$ is a probability on $(Y_{[1,t]}, F_{[1,t]})$ for each $x = (x_1, ..., x_t) \in X_{[1,t]}$ and satisfies $P_{[1,t]}(x, E) = P_1(x_1, E_1) \cdots P_t(x_t, E_t)$ for each E of the form $E = E_1 \times \cdots \times E_t \in F_{[1,t]}$ $(E_v \in F_v, \ 1 \leqslant v \leqslant t)$. We call $[1, t]$ the time (time interval of t discrete time points, $v = 1, 2,..., t$). $\{[X_{[1,t]}, (Y_{[1,t]}, F_{[1,t]}), P_{[1,t]}]\}_{t=1,2,...}$ is called a stationary channel without memory for discrete time.

An ϵ code $(0 < \epsilon < 1)$ of length N for $[1, t]$ for this channel is a sequence $\{(x^i, E^i) : 1 \leqslant i \leqslant N\}$ $(x^i \in X_{[1,t]}, E^i \in F_{[1,t]})$, where the E^i $(1 \leqslant i \leqslant N)$ are pairwise disjoint sets of $Y_{[1,t]}$ and $P_{[1,t]}(x^i, E^i) > 1 - \epsilon$. Set

$$N_t(\epsilon) := \sup\{N \textit{ natural: there exists an } \epsilon\textit{-code of length } N \textit{ for } [1, t]\}.$$

* This paper was supported by the Deutsche Forschungsgemeinschaft.

Let

$$M_v := \{P_v(x_v\,,\ ) : x_v \in X_v\} \qquad (0 \leqslant v \leqslant t)$$

and

$$M_{[1,t]} := M_1 \times \cdots \times M_t := \{p_1 \times \cdots \times p_t \text{ on } F_{[1,t]} : p_v \in M_v\}.$$

We may assume without loss of generality that there is a 1 : 1 correspondence between x_v and $P_v(x_v\,,\ )$ (as far as one is interested in the length of codes alone) because the receiver can separate different input words only if their transition probabilities are different. Hence, we may identify X_v and M_v $(0 \leqslant v \leqslant t)$, $X_{[1,t]}$ and $M_{[1,t]}$, respectively.

For any probability a_0 with finite support on X_0 define

$$q_0(\ ) := \int_{X_0} da_0 P_0(x_0\,,\ ),$$

$$I(a_0\,, P_0) := \int_{X_0} da_0 \left[\int_{Y_0} dP_0(x_0\,,\ ) \ln \frac{dP_0(x_0\,,\ )}{dq_0}\right],$$

$$C := C(M_0) := \sup\{I(a_0\,, P_0) : a_0 \text{ has finite support} \subseteq X_0 = M_0\},$$

$$C(M_0{}') := \sup\{I(a_0\,, P_0) : a_0 \text{ has finite support} \subseteq M_0{}'(\subseteq M_0)\}.$$

$C = C(M_0)$ is called the capacity of the channel.

Obviously, the following holds:

$$\lim_{\epsilon\to 0} \limsup_{t\to\infty} (1/t) \ln N_t(\epsilon) = \lim_{\epsilon\to 0} \liminf_{t\to\infty} (1/t) \ln N_t(\epsilon) = C.$$

$\lim_{\epsilon\to 0} \liminf_{t\to\infty}(1/t) \ln N_t(\epsilon) \geqslant C$ follows from the coding theorem for finite alphabet channels by exhausting X_0 by finite alphabets; $\lim_{\epsilon\to 0} \limsup_{t\to\infty}(1/t) \ln N_t(\epsilon) \leqslant C$ follows from the estimates used for the weak converses of the coding theorem (Wolfowitz [2], p. 17, 100).

As functionals on the set of transition functions, $N_t(\epsilon)$ and $C(M_0)$ have the following geometrical properties: Let $\mathrm{co}(M)$ denote the convex hull for a set M of probabilities. If $p \in \mathrm{co}(M_{[1,t]})$ and $p(E) > 1 - \epsilon$, then there exists $p' \in M_{[1,t]}$ with $p'(E) > 1 - \epsilon$. Hence,

$$N_t(\epsilon) = N_t(\epsilon, M_{[1,t]}) = N_t(\epsilon, \mathrm{co}(M_1) \times \cdots \times \mathrm{co}(M_t)).$$

This, together with the above behavior of $(1/t) \ln N_t(\epsilon)$, yields

$$C(M_0) = C(\mathrm{co}(M_0)).$$

2. The Strong Converse

We are now going to prove the so-called strong converse of the coding theorem, i.e.,

$$\limsup_{t\to\infty}(1/t)\ln N_t(\epsilon) \leqslant C \qquad (0 < \epsilon < 1),$$

under the following general supposition on M_0 to be analyzed in Section 3:

(C1). *For each $\eta > 0$ there is a natural number $K_{(\eta)}$ s.t. M_0 is union of $K_{(\eta)}$ subsets M_0^j $[1 \leqslant j \leqslant K_{(\eta)}]$ satisfying $C(M_0^j) < \eta$.*

This supposition admits a generalization of the method of fixed composition codes (method of π sequences) that f.i. is used by J. H. B. Kemperman for a proof of the strong converse of the coding theorem for finite alphabet channels (compare f.i., Wolfowitz [2], p. 121). We will stick to the ideas of Kemperman's proof as much as possible.

It is known from the theory of semicontinuous channels that the integrals below are well-defined and finite. Therefore, we will not make any additional remarks on that within the proofs.

Lemma. *$C(M_0) < \infty$ if (C1) holds.*

Proof. Let M_0 be union of the M_0^j with $C(M_0^j) < \eta$ $(1 \leqslant j \leqslant K_{(\eta)})$ and a_0 be any probability on M_0 with finite support. There exists a representation,

$$a_0 = \sum_{j=1}^{K_{(\eta)}} c^j b_0^j \qquad \left(c^j \geqslant 0, \sum_{j=1}^{K_{(\eta)}} c^j = 1\right),$$

of a_0 where b_0^j is a probability with finite support $\subseteq M_0^j$ $(1 \leqslant j \leqslant K_{(\eta)})$. Set

$$q_0(\ ) = \int_{X_0} da_0 P_0(x_0,\ )$$

as in Section 1 and

$$q_0^j(\ ) := \int_{X_0} db_0^j P_0(x_0,\ ).$$

We have

$$I(a_0, P_0) = \sum_{j=1}^{K_{(\eta)}} c^j \int_{X_0} db_0^j \left[\int_{Y_0} dP_0(x_0, \ ) \ln \frac{dP_0(x_0, \ )}{dq_0}\right]$$

$$= \sum_{j=1}^{K_{(\eta)}} c^j \int_{X_0} db_0^j \left[\int_{Y_0} dP_0(x_0, \ ) \ln \frac{dP_0(x_0, \ )}{dq_0^j}\right]$$

$$+ \sum_{j=1}^{K_{(\eta)}} c^j \int_{Y_0} dq_0^j \ln \frac{dq_0^j}{dq_0}$$

$$= \sum_{j=1}^{K_{(\eta)}} c^j I(b_0^j, P_0) + \sum_{j=1}^{K_{(\eta)}} c^j \int_{Y_0} dq_0^j \ln \frac{dq_0^j}{dq_0}$$

with

$$I(b_0^j, P_0) \leqslant C(M_0^j) < \eta \qquad (1 \leqslant j \leqslant K_{(\eta)}).$$

Set

$$\lambda := \frac{1}{K_{(\eta)}} \sum_{j=1}^{K_{(\eta)}} q_0^j.$$

Then

$$\int_{Y_0} dq_0^j \ln \frac{dq_0^j}{d\lambda} \leqslant \ln K_{(\eta)}$$

and

$$\int_{Y_0} dq_0 \ln \frac{dq_0}{d\lambda} \geqslant 0$$

(the latter follows by means of Jensen's inequality). Hence,

$$\sum_{j=1}^{K_{(\eta)}} c^j \int_{Y_0} dq_0^j \ln \frac{dq_0^j}{dq_0} = \sum_{j=1}^{K_{(\eta)}} c^j \int_{Y_0} dq_0^j \ln \frac{dq_0^j}{d\lambda} - \int_{Y_0} dq_0 \ln \frac{dq_0}{d\lambda} \leqslant \ln K_{(\eta)} .$$

This yields

$$\sup_{a_0} I(a_0, P_0) = C(M_0) \leqslant \eta + \ln K_{(\eta)} .$$

In the case that $C(M_0) < \infty$ holds $\liminf_{t\to\infty}(1/t) \ln N_t(\epsilon) \geqslant C(M_0)$ $(0 < \epsilon < 1)$ is equivalent to the following: For each ϵ $(0 < \epsilon < 1)$ and $\delta > 0$ there is $t_0(\epsilon, \delta)$ s.t. $\ln N_t(\epsilon) > t[C(M_0) - \delta]$ for all $t > t_0$. [If

$C(M_0) = \infty$ then $N_t(\epsilon)$ grows faster than exponentially with t (see, Augustin [1], p. 32)].

THEOREM (Strong Converse of the Coding Theorem). *Suppose* M_0 *fulfills* (C1). *Then for each* ϵ $(0 < \epsilon < 1)$ *and* $\delta > 0$ *there exists* $t_0(\epsilon, \delta)$, *s.t.*

$$\ln N_t(\epsilon) < t[C(M_0) + \delta] \quad \textit{for all} \quad t > t_0 .$$

Proof. Let $\{(x^i, E^i) : 1 \leqslant i \leqslant \bar{N}\}$ be an ϵ-code of length $\bar{N}$ for $[1, t]$; let

$$M_0 = \bigcup_{j=1}^{K_{(\eta)}} M_0{}^j, \qquad \text{with} \quad C(M_0{}^j) < \eta \quad [1 \leqslant j \leqslant K_{(\eta)}];$$

let $M_v{}^j$ denote the copy of $M_0{}^j$ in M_v, and $M_{j1}^{j_1} \times \cdots \times M_t^{j_t}$ the set

$$\{p_1 \times \cdots \times p_t \text{ on } (Y_{[1,t]}, F_{[1,t]}) : p_v \in M_v{}^v\}.$$

Furthermore, let $W(j_1, ..., j_t) := \{P_{[1,t]}(x, \ ):$ for some permutation πx of the letters of x holds $P_{[1,t]}(\pi x, \ ) \in M_t^{j_1} \times \cdots \times M_1^{j_t}\}$. There are at most $(t+1)^{K(\eta)}$ different classes $W(j_1, ..., j_t) \subseteq M_{[1,t]}$ given by the above covering of M_0. Therefore, some $W = W(j_1, ..., j_t)$ contains at least $N \geqslant \bar{N}(t+1)^{-K(\eta)}$ of the $P_{[1,t]}(x^i, \ )$ $(1 \leqslant i \leqslant \bar{N})$. Renumerating the x^i we have $P_{[1,t]}(x^i, \ ) \in W$ $(1 \leqslant i \leqslant N)$.

Let π^i be a permutation for the letters of x^i s.t.

$$P_{[1,t]}(\pi^i x^i, \ ) \in M_1^{j_1} \times \cdots \times M_t^{j_t} \qquad (1 \leqslant i \leqslant N).$$

$$P_{[1,t]}(\pi^i x^i, \pi^i E^i) = P_{[1,t]}(x^i, E^i) > 1 - \epsilon \qquad (1 \leqslant i \leqslant N).$$

Set

$$p^i := P_{[1,t]}(\pi^i x^i, \ ).$$

p^i is a product probability

$$p_1{}^i \times \cdots \times p_t{}^i := p^i.$$

Furthermore, let p_{0v}^i be the canonical image of $p_v{}^i$ in M_0,

$$q_0 := \frac{1}{t} \sum_{v=1}^{t} \frac{1}{N} \sum_{i=1}^{N} p_{0v}^i ,$$

q_v the image of q_0 in M_v and

$$q := q_1 \times \cdots \times q_t .$$

Finally, set

$$A_v := \{k \in [1, t] : M_0^{j_v} = M_0^{j_k}\},$$

$$h_{0v} := \frac{1}{|A_v|} \sum_{k \in A_v} \frac{1}{N} \sum_{i=1}^{N} p_{0k}^i \qquad [\in \operatorname{co}(M_0)],$$

$$h := h_1 \times \cdots \times h_t ,$$

where $h_v \in \operatorname{co}(M_v)$ is the image of h_{0v} .

We have for any real R^i, S^i

$$\begin{aligned} 1 - \epsilon - p^i \left\{\frac{dp^i}{dh} \geqslant \exp R^i\right\} - p^i \left\{\frac{dh}{dq} \geqslant \exp S^i\right\} & \\ \leqslant p^i(\pi^i E^i) - p^i \left\{\frac{dp^i}{dh} \geqslant \exp R^i\right\} - p^i \left\{\frac{dh}{dq} \geqslant \exp S^i\right\} & \\ \leqslant p^i\left(\pi^i E^i \cap \left\{\frac{dp^i}{dh} < \exp R^i\right\} \cap \left\{\frac{dh}{dq} < \exp S^i\right\}\right) & \\ \leqslant \exp[R^i + S^i] \cdot q(\pi^i E^i) = \exp[R^i + S^i] \cdot q(E^i) & \\ (1 \leqslant i \leqslant N). & \end{aligned}$$

Moreover,

$$p^i \left\{\ln \frac{dp^i}{dh} \geqslant \tilde{R}^i\right\} \leqslant \frac{1-\epsilon}{4}$$

for

$$\tilde{R}^i := \frac{4}{1-\epsilon} \int dp^i \ln^+\left(\frac{dp^i}{dh}\right),$$

and

$$p^i \left\{\ln \frac{dh}{dq} \geqslant \tilde{S}^i\right\} \leqslant \frac{1-\epsilon}{4}$$

for

$$\tilde{S}^i := \int dp^i \ln \frac{dh}{dq} + \left(\frac{4}{1-\epsilon}\right)^{1/2} \left[\int dp^i \left(\ln \frac{dh}{dq} - \int dp^i \ln \frac{dh}{dq}\right)^2\right]^{1/2}.$$

Therefore,

$$\begin{aligned}\frac{1}{N} &\geqslant \frac{1}{N}\sum_{i=1}^{N} q(E^i)\\ &\geqslant \frac{1}{N}\sum_{i=1}^{N} \exp[-\tilde{S}^i - \tilde{R}^i]\\ &\quad \cdot \left(1 - \epsilon - p^i\left\{\ln\frac{dp^i}{dh} \geqslant \tilde{R}^i\right\} - p^i\left\{\ln\frac{dh}{dq} \geqslant \tilde{S}^i\right\}\right)\\ &\geqslant \frac{1}{N}\sum_{i=1}^{N} \exp[-\tilde{S}^i - \tilde{R}^i]\,\frac{1-\epsilon}{2}\\ &\geqslant \frac{1-\epsilon}{2}\exp\left[-\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{S}^i\right) - \left(\frac{1}{N}\sum_{i=1}^{N}\tilde{R}^i\right)\right]\end{aligned}$$

(Jensen's inequality).

We now estimate $(1/N)\cdot\sum_{i=1}^{N}\tilde{R}^i$:

$$\frac{1-\epsilon}{2}\frac{1}{N}\sum_{i=1}^{N}\tilde{R}^i = \frac{1}{N}\sum_{i=1}^{N}\int dp^i \ln^+\left(\frac{dp^i}{dh}\right) \leqslant \frac{1}{N}\sum_{i=1}^{N}\int dp^i \ln\frac{dp^i}{dh} + e^{-1}$$

because

$$\int dp^i \left|\ln^-\left(\frac{dp^i}{dh}\right)\right| = \int dh \left|z^-\left(\frac{dp^i}{dh}\right)\right| \leqslant e^{-1},$$

where $z(u) := u \ln u$ (and $-u \ln u \geqslant e^{-1}$ for $u > 0$).

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\int dp^i \ln\frac{dp^i}{dh} &= \sum_{v=1}^{t}\frac{1}{N}\sum_{i=1}^{N}\int dp_{0v}^i \ln\frac{dp_{0v}^i}{dh_{0v}}\\ &= \sum_{v=1}^{t}\frac{1}{|A_v|}\sum_{k\in A_v}\frac{1}{N}\sum_{i=1}^{N}\int dp_{0k}^i \ln\frac{dp_{0k}^i}{dh_{0k}} \leqslant \sum_{v=1}^{t} C(M_0^{j_v}) < t\eta.\end{aligned}$$

Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{R}^i \leqslant \frac{4}{1-\epsilon}(t\eta + e^{-1}).$$

Next, we estimate $(1/N) \cdot \sum_{i=1}^{N} \tilde{S}^i$:

$$\frac{1}{N}\sum_{i=1}^{N}\int dp^i \ln\frac{dh}{dq} = \sum_{v=1}^{t}\frac{1}{N}\sum_{i=1}^{N}\int dp_v{}^i \ln\frac{dh_v}{dq_v}$$

$$= \sum_{v=1}^{t}\frac{1}{|A_v|}\sum_{k\in A_v}\frac{1}{N}\sum_{i=1}^{N}\int dp_{0k}^i \ln\frac{dh_{0k}}{dq_0}$$

$$= t\left(\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln\frac{dh_{0v}}{dq_0}\right);$$

furthermore (using Jensen's inequality twice),

$$\frac{1}{N}\sum_{i=1}^{N}\left[\int dp^i\left(\ln\frac{dh}{dq}-\int dp^i\ln\frac{dh}{dq}\right)^2\right]^{1/2}$$

$$= t^{1/2}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{t}\sum_{v=1}^{t}\int dp_{0v}^i\ln^2\left(\frac{dh_{0v}}{dq_0}\right)-\frac{1}{t}\sum_{v=1}^{t}\left(\int dp_{0v}^i\ln\frac{dh_{0v}}{dq_0}\right)^2\right]^{1/2}$$

$$\leqslant t^{1/2}\left[\frac{1}{t}\sum_{v=1}^{t}\frac{1}{N}\sum_{i=1}^{N}\int dp_{0v}^i\ln^2\left(\frac{dh_{0v}}{dq_0}\right)-\left(\frac{1}{t}\sum_{v=1}^{t}\frac{1}{N}\sum_{i=1}^{N}\int dp_{0v}^i\ln\frac{dh_{0v}}{dq_0}\right)^2\right]^{1/2}$$

$$= t^{1/2}\left[\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln^2\left(\frac{dh_{0v}}{dq_0}\right)-\left(\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln\frac{dh_{0v}}{dq_0}\right)^2\right]^{1/2}.$$

Together these estimates yield

$$\frac{1}{\bar{N}}(t+1)^{K_{(\eta)}} \geqslant \frac{1}{N} \geqslant \exp\left[-\frac{1}{N}\sum_{i=1}^{N}\tilde{S}^i-\frac{4}{1-\epsilon}(t\eta+e^{-1})+\ln\frac{1-\epsilon}{2}\right]$$

and

$$\ln\bar{N} \leqslant t\left(\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln\frac{dh_{0v}}{dq_0}\right)$$

$$+\left(\frac{4t}{1-\epsilon}\right)^{1/2}\left[\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln^2\left(\frac{dh_{0v}}{dq_0}\right)-\left(\frac{1}{t}\sum_{v=1}^{t}\int dh_{0v}\ln\frac{dh_{0v}}{dq_0}\right)^2\right]^{1/2}$$

$$+\frac{4}{1-\epsilon}(t\eta+e^{-1})+\ln\frac{2}{1-\epsilon}+K_{(\eta)}\ln(t+1).$$

Finally, observe $q_0 = (1/t) \cdot \sum_{v=1}^{t} h_{0v}$ and

$$\frac{1}{t} \sum_{v=1}^{t} \int dh_{0v} \ln^2 \left(\frac{dh_{0v}}{dq_0} \right) \leqslant 4e^{-2} + \ln^2 t$$

(the latter with a similar discussion as in Wolfowitz [2], p. 123) then it follows

$$\ln \bar{N} \leqslant t \left(\frac{1}{t} \sum_{v=1}^{t} \int dh_{0v} \ln \frac{dh_{0v}}{d\left(\frac{1}{t} \sum_{k=1}^{t} h_{0k}\right)} \right) + \frac{4}{1-\epsilon} (t^{1/2} \ln t + t\eta)$$

$$+ K_{(\eta)} \ln(t+1) + \frac{4}{1-\epsilon}.$$

Now let $\eta = \eta(t) \xrightarrow{t \to \infty} 0$ in such a way that $K_{(\eta)} = K_{[\eta(t)]} \leqslant O(t^{1/2})$. Then with

$$\frac{1}{t} \sum_{v=1}^{t} \int dh_{0v} \ln \frac{dh_{0v}}{d\left(\frac{1}{t} \sum_{k=1}^{t} h_{0k}\right)} \leqslant C[\mathrm{co}(M_0)] = C(M_0),$$

we obtain

$$\ln N_t(\epsilon) \leqslant tC(M_0) + \frac{4}{1-\epsilon} o(t),$$

where $o(t)$ does not depend on ϵ.

3. The Condition (C1)

We analyze how the condition (C1) is related to compactness conditions and to moment conditions. Furthermore, we show that (C1) is of interest for a computational cutdown for decoding procedures. This cutdown is related to approximations of the channel by finite input alphabet subchannels.

Let G be the class of convex functions φ on R^+ which have a representation $\varphi(u) = \int_0^u m(u)\, du$ with a nonnegative continuous and strictly increasing function $m(u)$ satisfying $m(0) = 0$, $m(u) \to \infty$ $(u \to \infty)$. It has been shown (in Augustin [1], Section 9) in another context that (C1) is equivalent to the following condition:

(C2). (a) *M_0 is conditionally variational norm compact and*

(b) *there is $\varphi \in G$ such that*

$$K^{\varphi}(M_0) := \sup_{a_0} \left(\int_{X_0} da_0(x_0) \int_{Y_0} dP_0(x_0\,,\ )\, \varphi \left\{ \ln^+ \left[\frac{dP_0(x_0\,,\ )}{dq_0} \right] \right\} \right) < \infty,$$

where $q_0 = \int da_0(x_0)\, P_0(x_0\,,\ )$.

Observe that part (b) of condition (C2) alone is of the same type as the requirement that $C(M_0) < \infty$ holds.

We are now going to use (C2) to describe the behavior of the channel with respect to approximations by finite input alphabet subchannels. The index 0 of M_0 is dropped for brevity.

LEMMA. *Let $p^1,..., p^n$ and $\bar{p}^1,..., \bar{p}^n$, respectively, be probabilities in M and let $a^j \geqslant 0$ be real numbers with $\sum_{j=1} a^j = 1$. Set $q := \sum a^j p^j$, $\bar{q} := \sum a^j \bar{p}^j$, $\tilde{q} := \frac{1}{2}q + \frac{1}{2}\bar{q}$. If*

$$\| p^j - \bar{p}^j \| \leqslant \gamma \leqslant 2 \qquad (1 \leqslant j \leqslant n)$$

holds, then

$$A := \int d\tilde{q} \left| \sum a^j \frac{dp^j}{d\tilde{q}} \ln \frac{dp^j}{dq} - \sum a^j \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{d\bar{p}^j}{d\bar{q}} \right| \leqslant B^{\varphi}(\gamma, M),$$

where

$$B^{\varphi}(\gamma, M) := \gamma^{1/3}[C(M) + 27] + \frac{2}{m(\gamma^{-1/3})} K^{\varphi}(M),$$

for any $\varphi \in G$ with m being the derivative of φ. ($0 \leqslant B^{\varphi}(\gamma, M) \leqslant \infty$ for every $\varphi \in G$.)

Proof. We set $g^j := \frac{1}{2}p^j + \frac{1}{2}\bar{p}^j$ $(1 \leqslant j \leqslant n)$, $z(u) := u \ln u$ and $\tilde{m}(u)$ for the inverse function of $m(\tilde{m}(m(u)) = u)$. For any $u_1\,, u_2 \geqslant 0$ holds

$$u_1 \cdot u_2 \leqslant u_1 \tilde{m}(u_1) + \varphi(u_2)$$

and

$$u_1 \cdot u_2 \leqslant u_1 \tilde{m}(K \cdot u_1) + \frac{1}{K} \cdot \varphi(u_2) \qquad (K > 0).$$

This can easily be seen by a simple geometric argument and it is also a familiar step for the derivation of generalized Hölder inequalities. The inequality of the lemma may be considered as a generalized Hölder inequality.

We estimate A:

$$A \leqslant A_1 + A_2,$$

where

$$A_1 := \int d\tilde{q} \left| \sum a^j \frac{dp^j}{d\tilde{q}} \ln \frac{dp^j}{d\tilde{q}} - \sum a^j \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{d\bar{p}^j}{d\tilde{q}} \right|$$

and

$$A_2 := \int d\tilde{q} \left| \frac{dq}{d\tilde{q}} \ln \frac{dq}{d\tilde{q}} - \frac{d\bar{q}}{d\tilde{q}} \ln \frac{d\bar{q}}{d\tilde{q}} \right| = \int d\tilde{q} \left| z\left(\frac{dq}{d\tilde{q}}\right) - z\left(\frac{d\bar{q}}{d\tilde{q}}\right) \right|$$

$$\leqslant \int d\tilde{q} \left| z\left(\frac{dq}{d\tilde{q}}\right) \right| + \int d\tilde{q} \left| z\left(\frac{d\bar{q}}{d\tilde{q}}\right) \right|.$$

Observe $0 \leqslant dq/d\tilde{q} \leqslant 2$, $0 \leqslant d\bar{q}/d\tilde{q} \leqslant 2$ and $|z(u)| \leqslant |u - 1| \cdot z(2)$ $(0 \leqslant u \leqslant 2)$. Hence, $A_2 \leqslant 2 \cdot z(2)(\| q - \tilde{q} \| + \| \bar{q} - \tilde{q} \|)$. But $\| q - \tilde{q} \| = \| \bar{q} - \tilde{q} \| = \frac{1}{2} \cdot \| q - \bar{q} \| \leqslant \frac{1}{2} \sum a^j \| p^j - \bar{p}^j \| \leqslant \frac{1}{2}\gamma$. Therefore,

$$A_2 \leqslant \gamma \cdot 4 \ln 2.$$

$$A_1 \leqslant A_3 + A_4,$$

where

$$A_3 := \int d\tilde{q} \left| \sum a^j \frac{dp^j}{d\tilde{q}} \ln \frac{dp^j}{dg^j} - \sum a^j \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{d\bar{p}^j}{dg^j} \right|$$

and

$$A_4 := \int d\tilde{q} \left| \sum a^j \frac{dp^j}{d\tilde{q}} \ln \frac{dg^j}{d\tilde{q}} - \sum a^j \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{dg^j}{d\tilde{q}} \right|.$$

$$A_3 \leqslant \sum a^j \int d\tilde{q} \left| \frac{dp^j}{d\tilde{q}} \ln \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{d\bar{p}^j}{dg^j} \right|$$

$$= \sum a^j \int dg^j \left| z\left(\frac{dp^j}{dg^j}\right) - z\left(\frac{d\bar{p}^j}{dg^j}\right) \right| \leqslant \gamma \cdot 4 \cdot \ln 2,$$

where the last inequality sign is obtained similarly as for A_2.

$$A_4 \leqslant \sum a^j \int d\tilde{q} \left| \frac{dp^j}{d\tilde{q}} \ln \frac{dg^j}{d\tilde{q}} - \frac{d\bar{p}^j}{d\tilde{q}} \ln \frac{dg^j}{d\tilde{q}} \right|$$

$$\leqslant \sum a^j \int dg^j \left| \frac{dp^j}{dg^j} \ln \frac{dg^j}{d\tilde{q}} - \frac{d\bar{p}^j}{dg^j} \ln \frac{dg^j}{d\tilde{q}} \right|$$

$$\leqslant \sum a^j \int dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln \frac{dg^j}{d\tilde{q}} \right| = A_5 + A_6,$$

where

$$A_5 := \sum a^j \int dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^- \left(\frac{dg^j}{d\tilde{q}} \right) \right|,$$

$$A_6 := \sum a^j \int dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^+ \left(\frac{dg^j}{d\tilde{q}} \right) \right|.$$

To estimate A_5 set $B^j := \{dg^j/d\tilde{q} < \delta\}$, where $0 < \delta \leqslant 1/e$. We obtain

$$\begin{aligned}
&\sum a^j \int_{B^j} dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^- \left(\frac{dg^j}{d\tilde{q}} \right) \right| \\
&\quad = \sum a^j \int_{B^j} d\tilde{q} \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| z^- \left(\frac{dg^j}{d\tilde{q}} \right) \right| \\
&\quad \leqslant \sum a^j \tilde{q}(B^j)\, 2\, | z(\delta)| \leqslant 2\, | z(\delta)|
\end{aligned}$$

and

$$\begin{aligned}
&\sum a^j \int_{(\mathrm{compl} B^j)} dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^- \left(\frac{dg^j}{d\tilde{q}} \right) \right| \\
&\quad \leqslant \sum a^j \| p^j - \bar{p}^j \| \, | \ln \delta | \leqslant \gamma \ln \frac{1}{\delta}.
\end{aligned}$$

Setting $\delta = (1/2e)\gamma$ (then $\delta \leqslant 1/e$, because $\gamma \leqslant 2$) we have

$$\begin{aligned}
A_5 &\leqslant 2 \left| \frac{1}{2e} \gamma \ln \left(\frac{1}{2e} \gamma \right) \right| + \gamma \left| \ln \left(\frac{1}{2e} \gamma \right) \right| \leqslant 3\gamma + \frac{3}{2} \, | z(\gamma)| \\
&= 3\gamma + 3\gamma^{1/2} \, | z(\gamma^{1/2})| \leqslant 3(\gamma + \gamma^{1/2}).
\end{aligned}$$

To estimate A_6, set $D^j := \{| dp^j/dg^j - d\bar{p}^j/dg^j | < \delta'\}$, where $\delta' > 0$. We obtain

$$\begin{aligned}
&\sum a^j \int_{D^j} dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^+ \left(\frac{dg^j}{d\tilde{q}} \right) \right| \\
&\quad \leqslant \delta' \sum a^j \int_{D^j} dg^j \left| \ln \left(\frac{dg^j}{d\tilde{q}} \right) \right| \leqslant \delta' \left[C(M) + \frac{1}{e} \right]
\end{aligned}$$

because of $C(M) = C[\mathrm{co}(M)]$ and

$$\begin{aligned}
&\sum a^j \int_{(\mathrm{compl} D^j)} dg^j \left| \frac{dp^j}{dg^j} - \frac{d\bar{p}^j}{dg^j} \right| \left| \ln^+ \left(\frac{dg^j}{d\tilde{q}} \right) \right| \\
&\quad \leqslant \sum a^j g^j(\mathrm{compl}\, D^j)\, \tilde{m}(2K) + \frac{1}{K} \sum a^j \int dg^j \varphi \left[\ln^+ \left(\frac{dg^j}{d\tilde{q}} \right) \right]
\end{aligned}$$

by means of the inequality given at the beginning of the proof. Together with $g^j(\text{compl}\, D^j) \leqslant 1/\delta' \, \| p^j - \bar{p}^j \|$ (by Chebyshev's inequality) and

$$\sum a^j \int dg^j \varphi \left[\ln^+ \left(\frac{dg^j}{d\tilde{q}}\right)\right] = \sum a^j \int d\tilde{q} \, \frac{dg^j}{d\tilde{q}} \, \varphi \left[\ln^+ \left(\frac{dg^j}{d\tilde{q}}\right)\right]$$
$$\leqslant \frac{1}{2} \sum a^j \int d\tilde{q} \, \frac{dp^j}{d\tilde{q}} \, \varphi \left[\ln^+ \left(\frac{dp^j}{d\tilde{q}}\right)\right]$$
$$+ \frac{1}{2} \sum a^j \int d\tilde{q} \, \frac{d\bar{p}^j}{d\tilde{q}} \left[\ln^+ \left(\frac{d\bar{p}^j}{d\tilde{q}}\right)\right]$$

(Jensen's inequality for the convex function $u \cdot \varphi(\ln^+ u)$ on R^+) and the latter being $\leqslant K^\varphi(M)$, we obtain for A_6 :

$$A_6 \leqslant \delta' \left[C(M) + \frac{1}{e}\right] + \frac{\gamma}{\delta'} \, \tilde{m}(K) + \frac{2}{K} \cdot K^\varphi(M) \qquad (\delta', K > 0).$$

We set $\delta' = \gamma^{1/3}$ and give $24\gamma^{1/3}$ as a crude upper estimate for $A_2 + A_3 + A_5$. Then

$$A \leqslant \gamma^{1/3} \left[C(M) + \frac{1}{e} + 24\right] + 2\gamma^{2/3} \tilde{m}(K) + \frac{2}{K} \cdot K^\varphi(M).$$

Finally, we set $\tilde{m}(K) = \gamma^{-1/3}$; hence, $K = m(\gamma^{-1/3})$. This yields

$$A \leqslant \gamma^{1/3} [C(M) + 27] + \frac{2}{m(\gamma^{-1/3})} \, K^\varphi(M). \qquad \text{Q.E.D.}$$

Set $\text{diam}(M') := \sup\{\| p - p' \| : p, p' \in M'\}$.

COROLLARY. (a) $B^\varphi(\gamma, M) = B^\varphi[\gamma, \text{co}(M)]$.

(b) *Let* $\text{diam}(M') \leqslant \gamma \leqslant 2$ *for* $M' \subseteq M$. *Then*

$$C(M') \leqslant B^\varphi(\gamma, M') \leqslant B^\varphi(\gamma, M).$$

Proof. (a) To obtain $K^\varphi(\gamma, M) = K^\varphi[\gamma, \text{co}(M)]$ check the last part of the proof of the lemma. Together with $C(M) = C[\text{co}(M)]$ follows $B^\varphi(\gamma, M) = B^\varphi[\gamma, \text{co}(M)]$.

(b) Take $p^1, \ldots, p^n \in M'$ and $\bar{p}^1 = \cdots = \bar{p}^n \in M'$ in the lemma. Then

$$\left| \sum a^j \int dp^j \ln \frac{dp^j}{dq} \right| \leqslant \sum a^j \int d\tilde{q} \left| \frac{dp^j}{d\tilde{q}} \ln \frac{dp^j}{dq} - 0 \right| \leqslant B^\varphi(\gamma, M').$$

$C(M')$ is supremum of expressions of the form $\sum a^j \int dp^j \ln(dp^j/dq)$ (which are nonnegative). Q.E.D.

COROLLARY. *Let $M' \subseteq M'' \subseteq M$ and suppose*

$$M'' \subseteq V_\gamma(M') := \{p \in M : \inf_{p' \in M'} \| p - p' \| < \gamma\} \qquad (\gamma \leqslant 2).$$

Then

$$C(M') \leqslant C(M'') \leqslant C(M') + B^\varphi(\gamma, M'').$$

Proof. Apply the lemma picking $p^j \in M''$ and $\bar{p}^j \in M'$ such that $\| p^j - \bar{p}^j \| < \gamma$. Q.E.D.

Pinsker's inequality implies (compare, Augustin [1], Section 9) that $\operatorname{diam}(M') \leqslant r \cdot C(M')^{1/2}$ holds with an absolute constant r. For more details on such inequalities and on optimum constants compare a forthcoming paper of J. H. B. Kemperman, "On the Optimum Rate of Transmitting Information."

It should be noticed, finally, that the following is wrong in almost all cases: Let $M' \subseteq M$. For each $\eta > 0$ there is $\gamma > 0$ such that $M' \subseteq M'' \subseteq M$ and $C(M'') \leqslant C(M') + \eta$ implies $M'' \subseteq V_\gamma(M')$. Detailed discussions of relations between topological convergence and information convergence have been given by Csiszar in [3].

The next theorem gives a prescription for the computational cutdown for coding.

THEOREM. *Suppose* $\sup_{0 \leqslant \gamma \leqslant 2} B^\varphi(\gamma, M) < \infty$ *for* M. *Let for a given* $\delta > 0$ γ_0 *be sufficiently small s.t.* $B^\varphi(\gamma_0, M) < \delta$ *holds. Furthermore, let* $M = \bigcup_{j=1}^{K_{(\gamma_0)}}$ *be any finite covering of* M *by subsets* M^j *with* $\operatorname{diam}(M^j) < \gamma_0$ $[1 \leqslant j \leqslant K_{(\gamma_0)}]$.

(a) *Then there are* $K_{(\gamma_0)}$ *numbers* $a^j \geqslant 0$ *with* $\sum_{j=1}^{K_{(\gamma_0)}} a^j = 1$ *such that*

$$\sum_{j=1}^{K_{(\gamma_0)}} a^j \int dp^j \ln \frac{dp^j}{dq} > C(M) - \delta$$

holds (where $q = \sum a^j p^j$*) for any choice of* $p^1 \in M^1, \ldots, p^{K(\gamma_0)} \in M^{K(\gamma_0)}$.

(b) *Choose any* $p^j \in M^j$ $[1 \leqslant j \leqslant K_{(\gamma_0)}]$. *Then take constants* a^j *s.t.* $\sum_{j=1}^{K_{(\gamma_0)}} a^j \int dp^j \ln(dp^j/dq)$ *is maximal. Then for arbitrary* $\bar{p}^j \in M^j$ $(1 \leqslant j \leqslant K_{(\gamma_0)})$ *holds*

$$\sum_{j=1}^{K_{(\gamma_0)}} a^j \int d\bar{p}^j \ln \frac{d\bar{p}^j}{d\bar{q}} > C(M) - 2\delta,$$

where $\bar{q} = \sum a^j \bar{p}^j$.

Proof. (a) For given $\delta' > 0$ there is a probability a on $X = M$ with finite support s.t.

$$I(a, P) = \int da(x) \left[\int dP(x, \ ) \ln \frac{dP(x, \ )}{d \int da(x) P(x, \ )} \right] > C(M) - \delta'.$$

a has a representation $a = \sum_{j=1}^{K(\gamma_0)} a^j b^j$, where $a^j \geqslant 0$, $\sum_{j=1}^{K(\gamma_0)} a^j = 1$ and b^j is a probability with support M^j. Then

$$\left| \int da(x) \left[\int dP(x, \ ) \ln \frac{dP(x, \ )}{d \int da(x) P(x, \ )} \right] - \sum a^j \int dp^j \ln \frac{dp^j}{dq} \right| < B^{\varphi}(\gamma_0, M)$$

for any choice of $p^j \in M^j$ $[1 \leqslant j \leqslant K_{(\gamma_0)}]$ according to the lemma. (a) follows from $\sup_a I(a, P) = C(M)$.

Part (b) is an immediate consequence of Part (a) and of the lemma. Q.E.D.

The last theorem can also be used to derive a device for the treatment of certain nonstationary channels by means of approximation by finitely many channels having finite input alphabets. Together with the strong converse of Section 2, the last theorem can be used to describe how good the nonstationary channel can be approximated by those finitely many channels. (For this kind of question the type of strong converse proved in Section 2 is just the right one.) Provided sender and receiver know the nonstationary behavior of the channel the prescriptions given by the last theorem can be used to determine simpler, almost optimal, input sources and to simplify coding.

4. On Weakening (C1)

It has to be considered as very different problems to try to prove coding theorems and converses in relation to coding effort and to try to prove coding theorems and converses without paying respect to coding effort.

In the first case manipulations of each of the transition probabilities from a certain set and manipulations of each of the transition probabilities of codes of maximal length are necessary.

In the second case only manipulations of averages of transition probabilities are necessary.

By sticking to the line of manipulating averages one can easily prove coding theorem and its strong converse for nonstationary infinite-alphabet channels without memory satisfying (C2)(b) alone, even together with speed estimates for the convergence in those theorems (compare, Augustin [1], where also necessary and sufficient conditions are given for the coding theorem and its strong converse for infinite alphabet nonstationary channels without memory). Furthermore, it is possible with this idea to transfer the known estimates for $\ln N_t(\epsilon)$ up to an order $o(t^{1/2})$ and the estimates for the error probability of codes to those channels.[1]

We finally give an example of a mathematical channel not satisfying (C2)(a) but (C2)(b) that has a very nice coding theorem and strong converse.

Let Y_0 be the interval $[0, 1] \subseteq R$, F_0 the natural σ field, λ the Lebesgue measure on $([0, 1], F_0)$ and $M_0 := \{p^j\}$ $(j = 1, 2, \ldots)$, where p^j is defined by means of the following density: For $j = 2i$ define

$$\frac{dp^{2i}}{d\lambda} := \begin{cases} 2 & \text{on } [(2k-1)\,2^{-i}, 2k2^{-i}] \quad \text{for} \quad k = 1, \ldots, 2^{i-1} \\ 0 & \text{otherwise} \end{cases},$$

for $j = 2i - 1$ define

$$\frac{dp^{2i-1}}{d\lambda} := 2 - \frac{dp^{2i}}{d\lambda}.$$

M_0 is not norm-totally bounded because $\| p^{j_1} - p^{j_2} \| \geqslant 1$ for $j_1 \neq j_2$. Hence, M_0 does not satisfy (C2)(a). However, it can easily be seen that M_0 satisfies (C2)(b).

Now consider the stationary channel without memory generated by M_0. (a) Surely $N_t(\epsilon) \geqslant 2^t$ because p^{2i} and p^{2i-1} have disjoint support. (b) Let $p \in M_{[1,t]}$, $E \in F_{[1,t]}$ and $p(E) > 1 - \epsilon$. Then $(\lambda \times \cdots \times \lambda)(E) > (1 - \epsilon)\,2^{-t}$, hence $N_t(\epsilon) < 2^t \cdot 1/(1 - \epsilon)$. Together we have $t(\ln 2) \leqslant \ln N_t(\epsilon) < t(\ln 2) - \ln(1 - \epsilon)$ with $\ln 2$ being the capacity of the channel.

It is possible to treat these kind of channels automatically together with the well-known practical channels if one tries to manipulate only averages of the transition probabilities. By accepting these remarks, one sees that it is not very good heuristics to talk about information distances (as is frequently done) when proving strong converses. One also sees that one should avoid the use of such unstable methods as the method of fixed-composition codes because these methods are related to the coding effort and to manipulations of each of the transition probabilities and to conditions as condition (C1).

When interested in the coding effort, however, one has to use even stronger conditions than (C1). To work with weak coverings instead of coverings

[1] Papers by the author on that matter have appeared the summer of 1969 in *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* and in the Hungarian journal *Studia.*

with norm balls in order to obtain prescriptions for coding would give very slow approximations only of the capacity by means of finite input alphabet subchannels. Therefore, the two problems of coding theorems and converses in relation to coding effort and without relation to the effort of coding should be treated separately.

RECEIVED: August 20, 1969

REFERENCES

1. U. AUGUSTIN, Gedächtnisfreie Kanäle für diskrete Zeit, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **6** (1966), 10–61.
2. J. WOLFOWITZ, "Coding Theorems of Information Theory," in "Ergebnisse der Mathematik und ihrer Grenzgebiete," 2nd. ed., Bd. 31, Springer-Verlag, Berlin/New York, 1964.
3. I. CSISZÁR, Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen, *Magy. Tud. Akad. Mat. Kut. Intez. Közlem.* **7** (1962), 137–158.