KYBERNETIKA — VOLUME 13 (1977), NUMBER 5

On the Identification of a Subclass of Finite State Channels and their Capacity

BOHUMIL PÁTEK

The paper deals with finite state channels with stationary distribution over the set of states. The capacity of a more general class of channels is defined and a converse to the coding theorem proved. In case the channel is a finite state type a coding theorem is proved. Finally the possibility of identification of such a channel is discussed.

I. INTRODUCTION

In what follows Z, X, Y are finite sets, alphabet of a source and input and output alphabets of a channel, respectively. The notation $s \in Z^L$, $s = z_1 z_2 \ldots z_L$ and $(u, v) \in \in (X \times Y)^N$, $(u, v) = (x_1 x_2 \ldots x_N, y_1 y_2 \ldots y_N)$ is used. We consider a block code as a transformation φ of Z^L into X^N. The rate of a block code is defined as

$$R = \frac{L}{N} \log |Z|$$

($|A|$ denotes the cardinality of the set A). N is called block length. Let, further, S_L be a probability measure on Z^L and $P_N(v \mid u)$ the transition probability of $v \in Y^N$ given $u \in X^N$. The joint probability on $(X \times Y)^N$ is then given by $Q_N(u)\, P_N(v \mid u)$, where $Q_N = S_L \varphi^{-1}$.

The decoding rule used is the maximum-likelihood one, i.e. a transformation ψ of Y^N into Z^L fulfilling the restriction

$$\psi(v) = s \Rightarrow P_N(v \mid \varphi(s)) \geqq P_N(v \mid \varphi(s')) \quad \text{for all} \quad s' \neq s\,.$$

The average error probability over the sequence of L digits is defined as

$$\langle P_e \rangle = \frac{1}{L} \sum_{l=1}^{N} \sum_{\substack{s,s' \in Z^L \\ z_l = z \\ z_l' = z'}} \sum_{\substack{z,z' \in Z \\ z \neq z'}} S_L(s)\, P_N(\psi^{-1}(s') \mid \varphi(s))\,.$$

Independently of block code φ and decoding rule ψ, $\langle P_e \rangle$ satisfies by [1], expression (4.3.20),

$$(1) \qquad \langle P_e \rangle \log (|Z| - 1) + \mathscr{H}(\langle P_e \rangle) \geqq \frac{1}{L} H(S_L) - \frac{1}{L} I(Q_N \,.\, P_N) ,$$

where $H(S_L)$ is the entropy of the source L-tuples $s \in Z^L$ and $I(Q_N \,.\, P_N)$ is the average mutual information between the input N-tuples $u \in X^N$ and the output N-tuples $v \in Y^N$ having the joint probability distribution generated by Q_N and P_N. $\mathscr{H}$ is defined as

$$\mathscr{H}(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log (1 - \alpha) , \quad 0 \leqq \alpha \leqq 1 .$$

If $s \in Z^L$ enters the encoder, then the probability of decoding error is given by

$$P_e(s) = \sum_{v \notin \psi^{-1}(s)} P(v \mid \varphi(s)) .$$

For $|X| \geqq 2$ there exists a code such that for all $s \in Z^L$

$$(2) \qquad P_e(s) \leqq 2(2 \mid Z^L - 1)^\varrho \sum_{v \in Y^N} \{ \sum_{u \in X^N} Q_N(u) \, P_N(v \mid u)^{1/(1+\varrho)} \}^{1+\varrho} ,$$

where ϱ is arbitrarily chosen from the interval $\langle 0, 1 \rangle$, ([1], Theorem 5.6.1 and Corollary 2 of Theorem 5.6.2).

The finite state channel ([1], Section 4.6) is a channel with a generally infinite memory. It is characterized, apart of input and output sequences, by a sequence of states, which are assumed to be taken from a finite set. If the states are numbered by $1, 2, \ldots, m$, the channel may be described statistically by an ensemble $\mathscr{M}$ containing $|X| \times |Y|$ transition probability matrices $\mathbf{M}(y \mid x)$ of order $m \times m$. Any ensemble $\mathscr{M}$ describes a finite state channel iff

a) $\mathbf{M}_{ij}(y \mid x) \geqq 0$ for all $\mathbf{M} \in \mathscr{M}$

b) $\sum_{y \in Y} \sum_{j=1}^{m} \mathbf{M}_{ij}(y \mid x) = 1$ for all $x \in X$, $i = 1, 2, \ldots, |X|$

obviously, the $\mathbf{M}_{ij}(y \mid x)$'s coincide with the one-dimensional transition probabilities in [1]. For any $(u, v) = (u'u'', v'v'')$ the transition matrix $\mathbf{M}(v \mid u)$ is given by a recursive relation

$$\mathbf{M}(v \mid u) = \mathbf{M}(v' \mid u') \, \mathbf{M}(v'' \mid u'') .$$

If the initial distribution over the set of states is represented by a vector $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$, $\sum_{k=1}^{m} \pi_k = 1$ then the transition probability $P_N(v \mid u)$ for any N is given by

$$(3) \qquad P_N^\pi(v \mid u) = \pi \, \mathbf{M}(v \mid u) \, \eta ,$$

where η is a column vector with all components equal to one. (If for some i $\pi_i = 1$ we use the notation $P_N^i(v \mid u)$.)

For a finite state channel we define the lower capacity

$$\mathbf{C}_l = \lim_{N \to \infty} \frac{1}{N} \max_{Q_N} \min_{i} I(Q_N P_N^i)$$

which appears in a coding theorem ([1], Theorem 5.9.2) and the upper capacity.

$$\mathbf{C}_u = \lim_{N \to \infty} \frac{1}{N} \max_{Q_N} \max_{i} (I(Q_N P_N^i)$$

which appears in a converse of the coding theorem ([1], Theorem 4.6.2). This approach provides the upper and the lower bounds of the probability of error independent of the state sequence. Nevertheless, in case $\mathbf{C}_l$ and $\mathbf{C}_u$ are not equal, none of them can be taken for a true capacity.

II. CAPACITY

We shall be concerned with a communication channel whose statistical description is built by a system P of transition probabilities P_N satisfying for all N and for all $(u, v) \in (X \times Y)^N$ the conditions

(4) a) $\sum_{y \in Y} P_{N+1}(vy \mid ux) = P_N(v \mid u)$,

b) $\sum_{y \in Y} P_{N+1}(yv \mid xu) = P_N(v \mid u)$.

For such a channel we define the capacity as

$$\mathbf{C} = \lim_{N \to \infty} \frac{1}{N} \mathbf{C}_N ,$$

where

$$\mathbf{C}_N = \max_{Q_N} I(Q_N P_N) .$$

Before proving the plausibility of the definition we give here without mentioning the proof the following statement, (see [1], Appendix 4A).

Lemma 1. If $\{a_n\}_{n=1}^{\infty}$ is a bounded sequence of real numbers and for all $1 \leqq n < N$ is

$$a_N \geqq \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n}$$

then

$$\lim_{N \to \infty} a_N = \sup_N a_N .$$

Lemma 2.

$$\mathbf{C} = \sup_N \mathbf{C}_N \, .$$

Proof. Let us take $N = n + l$, $(u, v) = (u_1 u_2, v_1 v_2)$, $(u_1, v_1) \in (X \times Y)^n$, $(u_2, v_2) \in \in (X \times Y)^l$ and $Q_N(u) = Q_n(u_1)\, Q_l(u_2)$, where Q_n and Q_l permit to achieve $\mathbf{C}_n$ and $\mathbf{C}_l$, respectively. Then, using (4), we can write

$$N\mathbf{C}_N \geqq n\mathbf{C}_n + l\mathbf{C}_l - \sum_{u \in X^n} \sum_{v \in Y^n} Q_N(u)\, P_N(v \mid u) \, .$$

$$. \log \frac{\left(\sum_{u \in X^n} Q_N(u)\, P_N(v \mid u)\right) P_n(v_1 \mid u_1)\, P_l(v_2 \mid u_2)}{P_N(v \mid u) \left(\sum_{u_1 \in X^n} Q_n(u_1)\, P_n(v_1 \mid u_1)\right) \left(\sum_{u_2 \in X^l} Q_l(u_2)\, P_l(v_2 \mid u_2)\right)} \, .$$

The third expression on the right hand side can be upperbounded by zero using the inequality $\log z \leqq z - 1$, $(z > 0)$. In this way we have

$$N\mathbf{C}_N \geqq n\mathbf{C}_n + l\mathbf{C}_l \, .$$

Lemma 1 completes the proof.

Suppose for this moment that the source is a stationary stochastic process, i.e. the system S of probabilities S_L fulfills for all L and $s \in Z^L$ the compatibility condition

$$\sum_{z \in Z} S_{L+1}(zs) = S_L(s) \, .$$

Then $(1/L)\, H(S_L)$ is nonincreasing with L ([1], Theorem 3.5. 1) and thus the following limit exists

$$H_\infty(S) = \lim_{L \to \infty} \frac{1}{L} H(S_L) \, .$$

Theorem 1. If the source is a stationary stochastic process and the channel satisfies conditions a) and b)*), then

$$\langle P_e \rangle \log (|Z| - 1) + \mathscr{H}(\langle P_e \rangle) \geqq H_\infty(S) - \frac{\log |Z|}{R} \mathbf{C} \, .$$

This statement provides a converse of the coding theorem. In fact, if S_N's are equiprobable distributions, $H_\infty(S)$ reaches its maximum $\log |Z|$ and the right hand side of the inequality is positive for any $R > \mathbf{C}$ and thus prevents $\langle P_e \rangle$ from being arbitrarily small.

Let us now pay our attention to a finite state channel with stationary distribution vector over its states π, i.e.

$$\pi \sum_{y \in Y} \mathbf{M}(y \mid x) = \pi \quad \text{for all} \quad x \in X \, .$$

*) Let us remark that such a channel need not be a finite state one.

In this case P_N^π's satisfy conditions a) and b) and thus they represent a channel belonging to the subclass we are restricted to (in what follows index π will be omitted).

Theorem 2. For a finite state channel with stationary distribution over its states $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$ let us denote

$$\alpha = \min_{\substack{k \\ \pi_k \neq 0}} \pi_k ,$$

$$E_N(\varrho, Q_N) = -\frac{1}{N} \log \sum_{v \in Y^N} \Big[\sum_{u \in X^N} Q_N(u)\, P_N(v \mid u)^{1/(1+\varrho)} \Big]^{1+\varrho} ,$$

$$F_N(\varrho) = \frac{\log \alpha}{N} + \max_{Q_N} E_N(\varrho, Q_N) ,$$

$$F_\infty(\varrho) = \lim_{N \to \infty} F_N(\varrho) ,$$

$$E_r(R) = \max_{0 \leqq \varrho \leqq 1} \big[F_\infty(\varrho) - \varrho R \big] .$$

Then for an arbitrary $\varepsilon > 0$ there exists a positive integer $N(\varepsilon)$ such that for any $N \geqq N(\varepsilon)$ and L there exists a code $\varphi_{N,R}$ for which $P_e(s) \leqq \exp\left[-N(E_r(R) - \varepsilon)\right]$ for all $s \in Z^L$. If $R < \mathbf{C}$ then the error exponent $E_r(R)$ is positive and decreasing with R.

In order to prove the theorem the folloving lemmas are needed:

Lemma 3.

$$P_{n+l}(v_1 v_2 \mid u_1 u_2) \leqq \frac{1}{\alpha}\, P_n(v_1 \mid u_1)\, P_l(v_2 \mid u_2)$$

for all n, l and $(u_1, v_1) \in (X \times Y)^n$, $(u_2, v_2) \in (X, Y)^l$.

Proof. a) At first let us suppose that π_k is positive for all k. Then

$$P_N(v_1 v_2 \mid u_1 u_2) = \sum_{k=1}^m \sum_{j=1}^m \sum_{i=1}^m \pi_k\, \mathbf{M}_{kj}(v_1 \mid u_1)\, \mathbf{M}_{ji}(v_2 \mid u_2) \leqq$$

$$\leqq \sum_{k=1}^m \sum_{j=1}^m \sum_{i=1}^m \pi_k\, \mathbf{M}_{kj}(v_1 \mid u_1) \frac{1}{\alpha}\, \pi_j\, \mathbf{M}_{ji}(v_2 \mid u_2) =$$

$$= \frac{1}{\alpha} \sum_{j=1}^m \Big(\sum_{k=1}^m \pi_k\, \mathbf{M}_{kj}(v_1 \mid u_1) \Big) \Big(\sum_{i=1}^m \pi_j\, \mathbf{M}_{ji}(v_2 \mid u_2) \Big) \leqq \frac{1}{\alpha}\, P_n(v_1 \mid u_1)\, P_l(v_2 \mid u_2) .$$

b) If $\pi_k = 0$ for some $k = k_1, k_2, \ldots, k_r$, $r < m$ then the stationarity of π implies

$$0 = \pi_k = \sum_{i=1}^n \pi_i\, \mathbf{M}_{ik}(v \mid u) \quad \text{for all} \quad (u, v) .$$

Hence if $\pi_i \neq 0$ then $\mathbf{M}_{ik}(v \mid u) = 0$ for all (u, v) and $k = k_1, k_2, \ldots, k_r$. It means the columns $k_1, k_2, \ldots, k_r$ do not contribute to the value of elements in rows corresponding to nonvanishing π_k's. Therefore omitting zero components in π and striking off k_i-th rows and columns, $i = 1, 2, \ldots, r$, we obtain a finite state channel which fulfills the assumption in a), generates the same $P_N(v \mid u)$ and has the same α.

Lemma 4.

$$\lim_{N \to \infty} F_N(\varrho) = \sup_N F_N(\varrho) . \tag{6}$$

In addition for $0 \leqq \varrho \leqq 1$ it converges uniformly and the limit $F_\infty(\varrho)$ is uniformly continuous.

Proof. Let us choose $N = n + l$, $(u, v) = (u_1 u_2, v_1 v_2)$, $(u_1, v_1) \in (X \times Y)^n$, $(u_2, v_2) \in (X \times Y)^l$ and $Q_N(u) = Q_n(u_1)\, Q_l(u_2)$ where distributions Q_n and Q_l permit to achieve maxima of $E_n(\varrho, Q_n)$ and $E_l(\varrho, Q_l)$ respectively. Then

$$F_N(\varrho) \geqq \frac{\log \alpha}{N} + E_N(\varrho, Q_N)$$

and

$$\exp\left[-N\, F_N(\varrho)\right] \leqq \frac{1}{\alpha} \sum_{v_1 \in Y^n} \sum_{v_2 \in Y^l} \Big\{ \sum_{u_1 \in X^n} \sum_{u_2 \in X^l} Q_n(u_1)\, Q_l(u_2)\, P_N(v_1 v_2 \mid u_1 u_2)^{1/(1+\varrho)} \Big\}^{1+\varrho} .$$

Using Lemma 3 and taking logarithms we obtain

$$F_N(\varrho) \geqq \frac{n}{N} F_n(\varrho) + \frac{l}{N} F_l(\varrho) .$$

In order to be able to apply Lemma 1 we show that $F_N(\varrho)$ is bounded uniformly with respect to N.

From Hölder's inequality it follows that $\{ \sum_{u \in X^N} Q_N(u)\, P_N(v \mid u)^{1/(1+\varrho)} \}^{1+\varrho}$ is a non-increasing function of ϱ and so

$$\frac{\partial E(\varrho, Q_N)}{\partial \varrho}$$

is nonnegative. Furthermore, if $\varrho_3 = \lambda \varrho_1 + (1 - \lambda)\, \varrho_2$ then also

$$\lambda \frac{1 + \varrho_1}{1 + \varrho_3} + (1 - \lambda) \frac{1 + \varrho_2}{1 + \varrho_3} = 1$$

and from Hölder's inequality we obtain

$$\Big\{ \sum_{u \in X^N} Q_N(u)\, P(v \mid u)^{1/(1+\varrho_3)} \Big\}^{1+\varrho_3} = \Big\{ \sum_{u \in X^N} \big(Q_N(u)^{(1+\varrho_1)/(1+\varrho_3)}\, P_N(v \mid u)^{\lambda/(1+\varrho_3)}\big) \,.$$

$$. \big(Q_N(u)^{(1-\lambda)(1+\varrho_2)/(1+\varrho_3)}\, P_N(v \mid u)^{(1-\lambda)/(1+\varrho_3)}\big) \Big\} \leqq \big(\sum_{u \in X^N} Q_N(u)\, P_N(v \mid u)^{1/(1+\varrho_1)} \big)^{\lambda(1+\varrho_1)} \,.$$

$$. \big(\sum_{u \in X^N} Q_N(u)\, P_N(v \mid u)^{1/(1+\varrho_2)} \big)^{(1-\lambda)(1+\varrho_2)} .$$

Using the same inequality once more yields

$$\sum_{v\in Y^N}\Big\{\sum_{u\in X^N} Q_N(u)\,P_N(v \mid u)^{1/(1+\varrho_3)}\Big\}^{1+\varrho_3} \leqq \Big\{\sum_{v\in X^N}\Big(\sum_{u\in X^N} Q_N(u)\,P_N(v \mid u)^{1/(1+\varrho_1)}\Big)^{1+\varrho_1}\Big\}^{\lambda}\,.$$
$$\cdot\Big\{\sum_{v\in Y^N}\Big(\sum_{u\in X^N} Q_N(u)\,P_N(v \mid u)^{1/(1+\varrho_2)}\Big)^{1+\varrho_2}\Big\}^{1-\lambda}$$

and hence

$$E_N(\varrho_3, Q_N) \geqq \lambda\, E_N(\varrho_1, Q_N) + (1-\lambda)\, E_N(\varrho_2, Q_N)$$

which establishes that $E_N(\varrho, Q_N)$ is convex $\cap$ in $\varrho \in \langle 0, 1\rangle$. Carrying out the differentiation we have

$$\left.\frac{\partial E_N(\varrho, Q_N)}{\partial \varrho}\right|_{\varrho=0} = \frac{1}{N} I(Q_N \cdot P_N)\,. \tag{7}$$

Due to the monotonicity and convexity of $E_N(\varrho, Q_N)$ and equation (7) there si

$$0 \leqq \frac{\partial E_N(\varrho, Q_N)}{\partial \varrho} \leqq \log |X|$$

and for $0 \leqq \varrho_1 \leqq \varrho_2 \leqq 1$ then

$$0 \leqq F_N(\varrho_2) - F_N(\varrho_1) \leqq (\varrho_2 - \varrho_1) \log |X|\,.$$

Hence it follows that for $0 \leqq \varrho \leqq 1$, $F_N(\varrho)$ is bounded uniformly with respect to N. Thus using Lemma 1 we have (6). The uniform convergence and uniform continuity follow from the bounded slope of $F_N(\varrho)$ for each N.

Proof of the theorem. We shall proceed analogically to [1], Section 5.9. According to (2) one can find a code such that for all $0 \leqq \varrho \leqq 1$

$$P_e(s) \leqq 2(2|Z|^L - 1)^{\varrho} \min_{Q_N} \sum_{v\in Y^N}\Big\{\sum_{u\in X^N} Q_N(u)\,P_N(v \mid u)^{1/(1+\varrho)}\Big\}^{1+\varrho} \leqq$$
$$\leqq 4\alpha |Z|^{L\varrho} \frac{1}{\alpha} \min_{Q_N} \sum_{v\in Y^N}\Big\{\sum_{u\in X^N} Q_N(u)\,P_N(v \mid u)^{1/(1+\varrho)}\Big\}^{1+\varrho} \leqq$$
$$\leqq \exp\left[-N(-\varrho R + F_N(\varrho) - \log\frac{4\alpha}{N}\right].$$

Lemma 4 confirms the existence of $N(\varepsilon)$ such that for $N \geqq N(\varepsilon)$ and all $0 \leqq \varrho \leqq 1$

$$F_\infty(\varrho) - F_N(\varrho) + \frac{\log 4\alpha}{N} \leqq \varepsilon\,.$$

This implies

$$P_e(s) \leqq \exp\left[-N(\varrho R + F_\infty(\varrho) - \varepsilon)\right].$$

Finally maximizing $F_\infty(\varrho) - \varrho R$ over all $0 \leqq \varrho \leqq 1$ we obtain (5).

Let us now suppose that $\mathbf{C}_N - R = \delta > 0$. By a simple substitution we ascertain that $E_N(0, Q_N) = 0$. Because of (7) and the continuity of

$$\frac{\partial E_N(\varrho, Q_N)}{\partial \varrho}$$

there must exist ϱ_0 such that for all $0 \leqq \varrho \leqq \varrho_0$

$$\max_{Q_N} E_N(\varrho, Q_N) \geqq \varrho \left(\mathbf{C}_N - \frac{\delta}{2} \right)$$

and therefore

$$F_N(\varrho) \geqq \varrho R + \frac{\varrho \delta}{2} + \frac{\log \alpha}{N}.$$

$F_N(\varrho) - \varrho R$ is thus positive for sufficiently large N and so behaves $F_\infty(\varrho) - \varrho R$. Because of vanishing for $\varrho = 0$, $F_\infty(\varrho) - \varrho R$ achieves its maximum in the interval $0 \leqq \varrho \leqq 1$ for some positive ϱ and then $F_\infty(\varrho) - \varrho R$ is a decreasing function of R. This makes the proof complete.

Theorem 2 together with Theorem 1 show the plausibility of $\mathbf{C}$ as capacity. In case the communication channel is a finite state type the upper and the lower capacities always exist but $\mathbf{C}$ is definable only if π is stationary. For different stationary π's, if any, $\mathbf{C}$ may also differ (we shall use an upper index). Nevertheless, the following inequality holds always.

Theorem 3. If π is a stationary distribution over the states of a finite state channel then

$$\mathbf{C}_u \geqq \mathbf{C}^\pi \geqq \mathbf{C}_l \,. \tag{8}$$

Proof. $\mathbf{C}_N^\pi$ could be written in the form

$$N\mathbf{C}_N^\pi = \max_{Q_N} \sum_{k=1} \sum_{u \in X^N} \sum_{v \in Y^N} Q_N(u)\, P_N^k(v \mid u)\, \pi_k \left(\log \frac{P_N^k(v \mid u)}{\sum_{u \in X^N} Q_N(u)\, P_N^k(v \mid u)} + \right.$$
$$\left. + \log \frac{\sum_{i=1}^m P_N^i(v \mid u)\, \pi_i}{P_N^k(v \mid u)} + \log \frac{\sum_{u \in X^N} Q_N(u)\, P_N^k(v \mid u)}{\sum_{i=1}^m \sum_{u \in X^N} Q_N(u)\, P_N^i(v \mid u)\, \pi_i} \right).$$

Upper and lower bounding every summand we obtain

$$N\mathbf{C}_N^\pi \leqq N\mathbf{C}_u + 0 + \log m$$
$$N\mathbf{C}_N^\pi \geqq N\mathbf{C}_l - \log m - 0 \,.$$

Limiting $N \to \infty$ yields the statement being investigated.

Remark. In inequalities (8) the sign of equality need not be valid. Let us for instance consider two different finite state channels given by transition matrices $\mathbf{M}^1(y \mid x)$, $\mathbf{M}^2(y \mid x)$ and having stationary distributions over the set of states π_1, π_2 and capacities $\mathbf{C}_1 = \mathbf{C}_{u,1} = \mathbf{C}_{l,1}$, $\mathbf{C}_2 = \mathbf{C}_{u,2} = \mathbf{C}_{l,2}$, $\mathbf{C}_1 \neq \mathbf{C}_2$ respectively. (For an example of such a channel see [1], Figure 4.6.1.) If we compose another channel taking

$$\mathbf{M}(y \mid x) = \begin{pmatrix} \mathbf{M}^1(y \mid x) & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^2(y \mid x) \end{pmatrix}, \quad \pi = \begin{pmatrix} \omega \pi_1 \\ (1 - \omega)\, \pi_2 \end{pmatrix}, \quad 0 \leqq \omega \leqq 1 ,$$

then it is obviously a finite state channel with stationary distribution π over its states and has a capacity $\mathbf{C}^\pi$ satisfying

$$\mathbf{C}_l = \min(\mathbf{C}_1, \mathbf{C}_2) < \mathbf{C}^\pi < \max(\mathbf{C}_1, \mathbf{C}_2) = \mathbf{C}_u \quad \text{for all} \quad \omega \in (0, 1) .$$

III. ON THE POSSIBILITY OF CHANNEL IDENTIFICATION

Let us consider the class of channels introduced in Section II by conditions a) and b). In order to be able to estimate by input-output experiments the set P of transition probabilities of such a channel it is reasonable to assume that P is fully determined by a finite set of transition probabilities $\{P_{l_k}(v_k \mid u_k)\}_{k=1}^h$.

In the sequel we shall assume that this is achieved by the existence of a linea-recurrence relation of the form

c) $P_{n+l}(vv' \mid uu') = \sum_{k=1}^{h} P_{n+l_k}(vv_k \mid uu_k)\, b_k(v' \mid u')$,

where $(u, v) \in (X \times Y)^n$, $(u', v') \in \in (X \times Y)^l$ and $\{(u_k, v_k)\}_{k=1}^l$ is a fixed set of input-output sequences of finite lengths l_k, respectively. (This recurrence relation is an analogy to that one used in [3], for linearly dependent processes.)

Conditions a) of Section II and c) imply that a compound sequence matrix $\mathbf{H}$ defined as

$$\mathbf{H}_{ij} = P_{l_i + l_j}(v_i v_j' \mid u_i u_j') ,$$

where $\{(u_i, v_i)\}_{i=1}^h$, $\{(u_j', v_j')\}_{j=1}^h$ are arbitrary sets of input-output sequences, has always the rank less then h. The maximum among the ranks of all possible compound sequence matrices that can be formed by taking various sets of (u_i, v_i) and (u_j', v_j') will be called the rank of the channel denoted by r.

For the channel of finite rank, Carlyle's procedure ([2], Chapter I.C) for finding a pseudostochastic sequential machine generating P is applicable. This procedure is based on the following facts:

1) For an arbitrary r-dimensional vector α, $\sum_{i=1}^{r} \alpha_i = 1$, there exists an ensemble $\mathscr{A}$ containing $|X| \times |Y|$ square matrices $\mathbf{A}(y \mid x)$ of order r which generates P,

i.e. for any $(u, v) = (x_1, \ldots, x_N, y_1, \ldots, y_N)$

$$P_N(v \mid u) = \alpha\, \mathbf{A}(v \mid u)\, \eta\,,$$

where

$$\mathbf{A}(v, u) = \mathbf{A}(y_1 \mid x_1) \ldots \mathbf{A}(y_N \mid x_N)\,.$$

Moreover

$$\sum_{y \in Y} \sum_{j=1}^{r} \mathbf{A}_{ij}(y \mid x) = 1 \quad \text{for all} \quad x \in X \quad \text{and} \quad i = 1, 2, \ldots, r\,.$$

$(\mathscr{A}, \alpha)$ will be called generator of P.

2) The knowledge of $P_N(v \mid u)$ for all $(u, v) \in (X \times Y)^N$, $N = 1, 2, \ldots, 2r - 1$ suffices for all $\mathbf{A}(y \mid x)$ to be evaluated as

$$\mathbf{A}(y \mid x) = \mathbf{Q}^{-1} \mathbf{H}(y \mid x)\, \mathbf{H}^{-1} \mathbf{Q}\,,$$

where $\mathbf{H}$ is given by $\{(u_i \mid v_i)\}_{i=1}^{r}$, $\{(u'_j, v'_j)\}_{j=1}^{r}$ for which u_1, v_1, u'_1, v'_1 are empty sequences (such a choise can be always done), $\mathbf{H}(v \mid u)$ is defined by the formula

$$\mathbf{H}_{ij}(v \mid u) = P_{l_i + l_j + 1}(v_i y v'_j \mid u_i x u'_j)$$

and $\mathbf{Q}$ is an arbitrary regular matrix satisfying the restrictions

$$\mathbf{Q}_{1k} = \alpha_k$$

$$\sum_{j=1}^{r} \mathbf{Q}_{ij} = \mathbf{H}_{i1} \quad \text{for} \quad i = 1, 2, \ldots, r\,.$$

3) If $(\mathscr{A}, \alpha)$ and $(\mathscr{A}', \alpha')$ are two generators of the same P then there exists a regular matrix $\mathbf{G}$, $\sum_{j=1}^{r} \mathbf{G}_{ij} = 1$ for $i = 1, 2, \ldots, r$ such that $\alpha = \alpha' \mathbf{G}$ and $\mathbf{A}(y \mid x) = \mathbf{G}^{-1} \mathbf{A}'(y \mid x)\, \mathbf{G}$ for all $(x, y) \in X \times Y$.

4) If the channel is a finite state one with m states, then $m \geqq r$.

By the application of condition (4) to the generator obtained above we can write, denoting the i-th row of a matrix $\mathbf{B}$ by $\mathbf{B}_{i.}$

$$\alpha \sum_{y \in Y} \mathbf{A}(y \mid x) = \alpha \mathbf{Q}^{-1} \sum_{y \in Y} \mathbf{H}(y \mid x)\, \mathbf{H}^{-1} \mathbf{Q} = \sum_{y \in Y} \mathbf{H}_{1.}(y \mid x)\, \mathbf{H}^{-1} \mathbf{Q} =$$
$$= \mathbf{H}_{1.} \mathbf{H}^{-1} \mathbf{Q} = \mathbf{Q}_{1.} = \alpha\,.$$

The channel is then characterized by a generator $(\mathscr{A}, \alpha)$. Thus it appears formally to be a finite state channel with stationary distribution over the set of states up to the fact that the matrix elements need not be necessarily nonnegative. The statement 3) gives a motive for trying to obtain a true finite state representation by taking some $(\mathscr{A}, \alpha)$ and searching a suitable transformation matrix $\mathbf{G}$. The corresponding iterative method is to be found in [4]. In what follows we give an example of a channel whose generator cannot have nonnegative elements though the channel is a finite state one with stationary distribution over its set of states but in a number exceeding its rank. Consequently the question of identification of finite state channels with stationary distribution over the set of states is not completely answered in this way. Nevertheless the Carlyle's method gives a lower bound of the number of states of the channel and provides a simple algorithm by which transition probabilities $P_N(v \mid u)$ can be evaluated.

Let us consider a binary channel whose transition probabilities are evaluated in this way: for any $(u, v) = (x_1, \ldots, x_N, y_1, \ldots, y_N)$ we find $\bar{v} = \bar{y}_1 \bar{y}_2 \ldots \bar{y}_N$ where

$$\bar{y}_i = y_i x_i + (1 - y_i)(1 - x_i),$$

then

$$P_N(v \mid u) = P_N(\bar{v} \mid 11 \ldots 1) = \begin{cases} 0 & \text{if } \bar{v} = v'00v'' \\ \frac{1}{3} & \text{otherwise} \end{cases}$$

The rank r of the channel is 2 and one of the potential generators is

$$\mathbf{A}(1 \mid 1) = \mathbf{A}(0 \mid 0) = \begin{pmatrix} 0 & \frac{2}{3} \\ \frac{3}{2} & -1 \end{pmatrix}, \quad \mathbf{A}(0 \mid 1) = \mathbf{A}(1 \mid 0) = \begin{pmatrix} 1 & -\frac{2}{3} \\ \frac{3}{2} & -1 \end{pmatrix}$$

$$\pi = (1, 0).$$

If we notice that the trace of $\mathbf{A}(1 \mid 1)$ is negative we come, realizing the statement 3), to the conclusion that none of the generators has nonnegative elements. Nevertheless in this case we can deduce, inspecting the probability tree of $P_N(v \mid 11 \ldots 1)$, that the channel is a 3-state type with an equiprobable distribution over the set of states and transition matrices

$$\mathbf{M}(1 \mid 1) = \mathbf{M}(0 \mid 0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{M}(0 \mid 1) = \mathbf{M}(1 \mid 0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

(Received March 9, 1977.)

REFERENCES

[1] R. G. Gallager: Information Theory and Reliable Communication. Wiley, New York 1968.

[2] A. Paz: Introduction to Probabilistic Automata. Academic Press, New York 1971.

[3] T. L. Booth: Sequential Machines and Automata. Wiley, New York 1967.

[4] F. Bancilhon, M. Depeyrot: Idetification des automates nonfiables. Rev. fr. Automat. Informat. Rech. Operat. (juin 1974), *B-2*, 127.

Ing. Bohumil Pátek, Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 180 76 Praha 8. Czechoslovakia.