91. *On the Channel Capacity of a State Machine**)

By Yatsuka NAKAMURA

Faculty of Engineering, the Shinshu University, Nagano

(Comm. by Kinjirô KUNUGI, M. J. A., June 12, 1973)

1. Introduction. In the information theory, the calculation of the channel capacity is not easy in general. The calculation method is known for only memoryless channels and several particular cases.

Finite automata are seen as information channels in various way (for example, see [1]). A kind of such automata, called a permutation machine, is described as a stationary 0-memory Markovian channel ([7]). In this paper, we shall give the method to get the capacity of such channels.

The author is indebted to Professor H. Umegaki for his advice and encouragement in preparing this paper.

2. Permutation channel. Let $\mathcal{A}=\{S, X, \tau\}$ be a state machine, i.e., (i) S is a non-empty finite set of states, (ii) X is a non-empty finite set of input letters, and (iii) τ is a mapping from $S\times X$ to S, called a transition function. A state machine can be represented by a finite directed graph G, where states correspond to vertices and transitions to directed edges indexed by elements in X. Terminology of the graph theory used here refers to Ore [5]. A directed edge (s_1, s_2) indexed by x is denoted by $s_1 \xrightarrow{x} s_2$, which exists if and only if $\tau(s_1, x)=s_2$. For such graph, let us assume the following property: (A) For every vertex s and every input letter x, there exists one and only one directed edge of the form $s_1 \xrightarrow{x} s$ for some state s_1, i.e., for every input letter x, a mapping $\tau(\cdot, x)$, which is from S onto S, is a permutation on S. A state machine, a graph of which satisfies the condition (A), is called a permutation machine (cf. [3] p. 195).

Let $X^I (I=\{0, \pm 1, \pm 2, \cdots\})$ be an alphabet space, where the state space X is a set of input letters of a permutation machine. And S^I be another alphabet space, where S is a set of states of the machine. For any sequence $s_i s_{i+1} \cdots s_{i+l}$ of states in S, we define an information channel ν by

$$\nu_{\bar{x}}(s_i s_{i+1} \cdots s_{i+l}) = \frac{1}{N} q^{x_{i+1}}_{s_i s_{i+1}} q^{x_{i+2}}_{s_{i+1} s_{i+2}} \cdots q^{x_{i+l}}_{s_{i+l-1} s_{i+l}} \qquad (\bar{x} \in X^I)$$

where

---

*) This work is partially supported by the Sakkokai Foundation.

$$q_{ss'}^{x}=\begin{cases}1\cdots\text{if } \tau(s,x)=s'\\ 0\cdots\text{else}\end{cases},$$

and $N=\mathrm{Card}\,(S)$. Then ν can be extended to a stationary channel from the input space X^I to the output space S^I, as ν is a kind of Markovian channels ([7]). Let us call such channel ν, which is derived from a permutation machine, a permutation channel. An input source, an output source and compound source are probability measures on X^I, S^I, $(X\times S)^I$ respectively. A compound source $r(.)=r(.\,;p,\nu)$ derived from p and ν is defined by (cf. [2]) $r(C)=\int_{X^I}\nu_{\bar{x}}(C_{\bar{x}})p(d\bar{x})$ where $C_{\bar{x}}$ is an $\bar{x}$-section of a measurable set C in $(X\times C)^I$. And $q(B)=r(X\times B)$.

Theorem 1. *Let ν be a permutation channel, and p be a Bernoulli input source, i.e., $p(x_1\cdots x_n)=p(x_1)p(x_2)\cdots p(x_n)$ for every $x_1,\cdots,x_n$. Then an output source q and a compound source r derived from the source p and the channel ν, are Markovian.*

Proof. $r((x_1s_1)\cdots(x_ns_n))=p(x_1\cdots x_n)\nu_x(s_1\cdots s_n)$

$$=p(x_1\cdots x_n)\frac{1}{N}q_{s_1s_2}^{x_2}q_{s_2s_3}^{x_3}\cdots q_{s_{n-1}s_n}^{x_n}$$

$$=p(x_1\cdots x_{n-1})p(x_n)\frac{1}{N}q_{s_1s_2}^{x_2}q_{s_2s_3}^{x_3}\cdots q_{s_{n-1}s_n}^{x_n}.$$

And so,

$$\begin{aligned}&r((x_ns_n)\,|\,(x_1s_1)\cdots(x_{n-1}s_{n-1}))\\ &\quad=r((x_1s_1)\cdots(x_ns_n))/r((x_1s_1)\cdots(x_{n-1}s_{n-1}))\\ &\quad=p(x_n)q_{s_{n-1}s_n}^{x_n}=r((x_ns_n)\,|\,(x_{n-1},s_{n-1}))\end{aligned}$$

for all possible sequence $(s_1\cdots s_{n-1})$, which shows that the source r is Markovian. For the output source,

$$\begin{aligned}q(s_1\cdots s_n)&=\sum_{x_1\cdots x_n}p(x_1\cdots x_n)\nu_x(s_1\cdots s_n)\\ &=\sum_{x_1\cdots x_n}p(x_1\cdots x_n)\frac{1}{N}q_{s_1s_2}^{x_2}q_{s_2s_3}^{x_3}\cdots q_{s_{n-1}s_n}^{x_n}\\ &=\sum_{x_n\in X_{s_{n-1}s_n}}\sum_{x_1\cdots x_{n-1}}p(x_1\cdots x_{n-1})p(x_n)\frac{1}{N}q_{s_1s_2}^{x_2}\cdots q_{s_{n-2}s_{n-1}}^{x_{n-1}}\end{aligned}$$

where $X_{s_{n-1}s_n}=\{x\in X:\ \tau(s_{n-1},x)=s_n\}$, hence we get

$$\begin{aligned}q(s_n\,|\,s_1\cdots s_{n-1})&=q(s_1\cdots s_n)/q(s_1\cdots s_{n-1})\\ &=\sum_{x_n\in X_{s_{n-1}s_n}}p(x_n)=q(s_n\,|\,s_{n-1}).\end{aligned}$$

Q.E.D.

3. The Capacity of permutation channels. The entropy of a source p is defined by

$$h_p=\lim_n\left\{-\frac{1}{n}\sum_{x_1\cdots x_n}p(x_1\cdots x_n)\log p(x_1\cdots x_n)\right\}.$$

Then the transmission rate R_p is $R_p=h_p+h_q-h_r$, where q and r are the output and compound sources derived from p and a channel ν respectively. The stationary capacity C_s of the channel ν is defined by

$C_s=\sup_p R_p$ where p moves on all input sources. And the ergodic capacity C_e is defined by $C_e=\sup_{p\in\Pi'} R_p$ where $\Pi'=\{p: p$ is an input source and $r(.)=r(.\,;p,\nu)$ is ergodic$\}$.

Theorem 2. *If ν is a permutation channel, then*

$$C_s=\sup \frac{1}{N}\left\{-\sum_{s_1s_2}\left(\sum_{x_i\in Ms_1s_2} p(x_i)\right)\log\left(\sum_{x_i\in Ms_1s_2} p(x_i)\right)\right\} \tag{1}$$

where the supremum is taken all over k-dimensional probability vectors $(p_1p_2\cdots p_k)$ $(p_i=p(x_i),\ k=\text{Card}\,(X))$ *and* $Ms_1s_2=\{x: q^x_{s_1s_2}>0\}$.

Proof. The following chain of formulae is valid;

$$\begin{aligned}R_p&=h_p+h_q-h_r\\&=h_q-\lim_n \frac{1}{n}\left\{\sum_{x_1\cdots x_n} p(x_1\cdots x_n)\cdot\sum_{s_1\cdots s_n}\nu_x(s_1\cdots s_n)\log\nu_x(s_1\cdots s_n)\right\}\\&=h_q-\lim_n\frac{1}{n}\left\{\sum_{x_1\cdots x_n}p(x_1\cdots x_n)\cdot\sum_{s_1\cdots s_n}\frac{1}{N}q^{x_2}_{s_1s_2}\cdots q^{x_n}_{s_{n-1}s_n}\log\frac{1}{N}q^{x_2}_{s_1s_2}\cdots q^{x_n}_{s_{n-1}s_n}\right\}\\&=h_q-\lim_n\frac{1}{n}\left\{\sum_{x_1}\sum_{s_1}\frac{1}{N}\log\frac{1}{N}\right\}=h_q.\end{aligned}$$

And we know that ([2]),

$$h_q=\lim_n H(S_2|S_{-n}\cdots S_1)\le H(S_2|S_1). \tag{2}$$

The right side of (2) is equal to

$$-\sum_{s_1} q(s_1)\{q(s_2|s_1)\log q(s_2|s_1)\}.$$

But $q(s_2|s_1)=\sum_{x_1x_2}p(x_1x_2)q^{x_2}_{s_1s_2}=\sum_{x_2\in Ms_1s_2}p(x_2)$ and $q(s_1)=\sum_{x_1}p(x_1)\frac{1}{N}=\frac{1}{N}$, therefore

$$R_p=h_q\le\text{the right hand side of (1).}$$

The equality is actually achieved by some Bernoulli probability measure determined by a k-dimensional probability vector $(\tilde{p}_1\tilde{p}_2\cdots\tilde{p}_k)$, as the output source q becomes Markovian and the equality holds in (2).

Q.E.D.

A channel ν is ergodic if and only if ergodicity of input source implies ergodicity of the compound source r. As well known ([4], [6]), ergodicity of a channel implies $C_s=C_e$. We can deduce easily that permutation channels are not ergodic in general, even if the associated graph is connected, but we get the following:

Theorem 3. *Let G be a graph representing a permutation machine $\mathcal{A}$. If G is (weakly) connected, then for the permutation channel ν constructed from $\mathcal{A}$, the stationary capacity and the ergodic capacity consist, i.e., $C_s=C_e$. (Weak connectedness implies strong connectedness in this case.)*

Proof. For the Bernoulli probability measure p determined by the vector $(\tilde{p}_1\cdots\tilde{p}_k)$ defined in the proof of the previous theorem, the compound source $r(.)=r(.;p,\nu)$ is Markovian by Theorem 1. Let us prove that this Markov chain is irreducible. The Markov chain r is stationary as both the input source p and the channel are stationary, so it suffices to prove that the graph H which represents transients of the Markov chain r, is weakly connected. The graph H can be constructed easily using the graph G by the following method:

Put $H=X\times G$, where an edge $b\to a$ for $a=(x,s)$, $b=(x',s')\in X\times G$ exists if and only if the edge (s,s') indexed by x exists in G, i.e., $\tau(s',x)=s$, Actually,

$$r((xs)\,|\,(x's'))=\tilde{p}(x)q^{x}_{s's}>0$$

if and only if $q^{x}_{s's}>0$ which is equivalent to $\tau(s',x)=s$. (We can assume $\tilde{p}(x)>0$.)

For any state s and any letters x,x', the vertices (x,s) and (x',s) are weakly connected as for any x'', the edges $(x,s)\to(x'',s)$ and $(x',s)\to(x'',s')$ exist, where $s'=\tau(s,x'')$. The notation of weakly connectedness is $(x,s)\sim(x',s)$.

Now let us prove that $(x_1,s_1)\sim(x_2,s_2)$ for every s_1,s_2,x_1,x_2. The states s_1 and s_2 are strongly connected in G, hence there exists a sequence of input letters x_3,x_4,x_m and a sequence of indexed directed edges

$$s_1\xrightarrow{x_3}s_{i_3},\ s_{i_3}\xrightarrow{x_4}s_{i_4},\ \cdots,\ s_{i_{m-1}}\xrightarrow{x_m}s_2.$$

Then for any x a sequence of edges $(x,s_1)\to(x_3,s_{i_3}),(x_3,s_{i_3})\to(x_4,s_{i_4}),\cdots,(x_{m-1},s_{i_{m-1}})\to(x_m,s_2)$ exists, therefore putting $x=x_1$ we get $(x_1,s_1)\sim(x_m,s_2)\sim(x_2,s_2)$, which implies that H is weakly connected and r is ergodic.

References

[1] Cerny, J.: Approximation in the space of information channels. Information and Control, **16**, 384–395 (1970).

[2] Feinstein, A.: Foundations of Information Theory. MacGraw-Hill (1958).

[3] Hartmanis, J., and Stearns, R. E.: Algebraic Structure Theory of Sequential Machines. Prentice-Hall (1966).

[4] Jacobs, K.: Ergodic decomposition of the Kolmogorov-Sinai invariant. Ergodic theory (edited by F. B. Wright). Proc. of International Symposium, Tulane Univ. Oct. 1961, Acad. Press, pp. 173–190 (1963).

[5] Ore, O.: Theory of Graphs. Coll. Publ., Amer. Math. Soc., **38** (1962).

[6] Parthasarathy, K. R.: On the integral representation of the rate of transmission of a stationary channel. Ill. J. Math., **5**, 299–305 (1961).

[7] Nakamura, Y.: On finite-memory Markovian channels (to appear).