KŌDAI MATH. SEM. REP.
21 (1969), 133—150

MEASURE-THEORETIC CONSTRUCTION FOR INFORMATION THEORY

BY YATSUKA NAKAMURA

1. Introduction.

In the measure theoretic viewpoints, the information theory originated by Shannon [13] can be divided into a couple of basic parts, that is, the one is concerned to information source and the other is concerned to information channel. Kolmogorov [10a, b] and Sinai [14] gave the concept of entropy of measure preserving transformation, modifying the method of information source, and they classified certain dynamical systems which belong to the same spectral type.

Halmos introduced a measure theoretic construction of information source in his lecture note [7]. Under such a measure theoretic form, we can apply the theory to both the classifications of dynamical systems and the composition of the concrete information theory constructed on alphabet spaces. In this paper, we shall study a measure theoretic construction of information channel. For this purpose, main themes are devoted to define channels, between two abstract measurable spaces, and ergodic or stationary capacities of such channels, and to find conditions under which these two capacities coincide.

At first, an integral representation of entropy function will be done for the latter intention, namely to find the conditions for coincidence of the capacities. Parthasarathy [12] and Jacobs [9a] proved that the representation is possible when entropy is defined on alphabet space, and Umegaki [17a] showed that it is also possible even when the space is a compact totally disconnected topological space. Their constructions are available for the case of the abstract dynamical system, reducing to the special cases by certain mappings (see [9b]). But the method employed here needs only some simple calculations of entropy, and some knowledges of the ergodic theorem and the martingale convergence theorem.

Seconderly necessary and sufficient conditions for ergodicity of channels will be researched. Hinchin believed in his paper [8] that finite memory channels are ergodic, who gave the first mathematical and systematical construction to discrete information theory originated by Shannon. But, Takano [15] pointed out that finite memory channels are not always ergodic and it needs a concept so called "M-dependence", in addition to the assumption of finite memory, for ergodicity of channels.

Adler [1] showed that "weakly mixing" and "strongly mixing" channels in

Received July 23, 1968.

his sense are always ergodic, and M-dependent channels are strongly mixing, (so the assumption of finite memory is needless). But the necessary and sufficient conditions for ergodicity of channels were not known. (cf. Billingsley [2], p. 161) Recently Umegaki [17c] showed some necessary sufficient conditions independently from the author, by beautiful functional analysis methods. In this paper we also give some conditions by pure measure theoretic methods.

Lastly a condition of "completeness for ergodicity of a system $(X, \mathfrak{X}, \Pi)$" will be studied when X is a completely regular topological space. Alphabet spaces (even if alphabets are countable) satisfy the topological assumption, so our theory is applicable to the information theory with countable alphabets.

The author expresses his sincere thanks to Professor Umegaki for many instructive suggestions and advices in the course of preparing the present paper.

2. Notations and Preliminaries.

In this section, we shall refer to the several notations and functions and fundamental notions relative to the amounts of information and entropy which were given and formulated by Halmos [7].

Let $(X, \mathfrak{X}, p)$ be a probability measure space, and $\mathcal{A}$ be any measurable finite partition of X, or equivalently subfield of finite elements of $\mathfrak{X}$. Then *information of* $\mathcal{A}$ is defined by

$$I(\mathcal{A}) = -\sum_{A} \chi_A \log p(A), \tag{2.1}$$

where the sum is taken all over the atoms of $\mathcal{A}$, and χ_A is a characteristic function of A. If $\mathcal{C}$ is any subfield of $\mathfrak{X}$, then *conditional information of* $\mathcal{A}$ *relative to* $\mathcal{C}$ is defined by

$$I(\mathcal{A} \mid \mathcal{C}) = -\sum_{A} \chi_A \log p(A \mid \mathcal{C}), \tag{2.2}$$

where $p(A \mid \mathcal{C})$ is a conditional probability of A relative to $\mathcal{C}$, and A also moves on atoms of $\mathcal{A}$. If S is a measure preserving transformation on X, then

$$I(\mathcal{A} \mid \mathcal{C})S = I(S^{-1}\mathcal{A} \mid S^{-1}\mathcal{C}) \qquad \text{a.e.} \tag{2.3}$$

If $\mathcal{A} \subset \mathcal{C}$, then $I(\mathcal{A} \mid \mathcal{C}) = 0$, and if $\mathcal{A} \subset \mathcal{B}$, then

$$I(\mathcal{A} \mid \mathcal{C}) \leqq I(\mathcal{B} \mid \mathcal{C}) \qquad \text{a.e.} \tag{2.4}$$

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be subfields such that $\mathcal{B}$ are finite, then

$$I(\mathcal{A} \vee \mathcal{B} \mid \mathcal{C}) = I(\mathcal{B} \mid \mathcal{C}) + I(\mathcal{A} \mid \mathcal{B} \vee \mathcal{C}) \qquad \text{a.e.,} \tag{2.5}$$

and similarly let $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_n$ be a finite sequence of finite subfields, then

$$I\left(\bigvee_{i=1}^{n} \mathcal{B}_i \,\middle|\, \mathcal{C}\right) = I(\mathcal{B}_1 \mid \mathcal{C}) + \sum_{k=2}^{n} I\left(\mathcal{B}_k \,\middle|\, \bigvee_{i=1}^{k-1} \mathcal{B}_i \vee \mathcal{C}\right) \qquad \text{a.e.} \tag{2.6}$$

The conditional entropy of a finite field $\mathcal{A}$ relative to a subfield $\mathcal{C}$ is defined by

(2. 7) $$H(\mathcal{A}\mid\mathcal{C})=\int I(\mathcal{A}\mid\mathcal{C})dp$$

which is equal to

(2. 8) $$-\int\sum_{A}p(A\mid\mathcal{C})\log p(A\mid\mathcal{C})dp.$$

The entropy of $\mathcal{A}$ is defined by

(2. 9) $$H(\mathcal{A})=\int I(\mathcal{A})dp$$

which is equal to

(2. 10) $$-\sum_{A}p(A)\log p(A).$$

If $\mathcal{B}\subset\mathcal{C}$, then

(2. 11) $$H(\mathcal{A}\mid\mathcal{C})\leqq H(\mathcal{A}\mid\mathcal{B}).$$

Moreover under the same assumption as (2. 5),

(2. 12) $$H\left(\bigvee_{i=1}^{n}\mathcal{B}_i\mid\mathcal{C}\right)=H(\mathcal{B}_1\mid\mathcal{C})+\sum_{k=2}^{n}H\left(\mathcal{B}_k\middle|\bigvee_{i=1}^{k-1}\mathcal{B}_i\vee\mathcal{C}\right).$$

The entropy of a measure preserving transformation S relative to a finite subfield $\mathcal{A}$ is defined by

(2. 13) $$h(\mathcal{A},S)=H\left(\mathcal{A}\middle|\bigvee_{i=1}^{\infty}S^{-i}\mathcal{A}\right),$$

which is equal to

(2. 14) $$\lim_{n}\frac{1}{n}H\left(\bigvee_{i=0}^{n-1}S^{-i}\mathcal{A}\right),$$

where the limit always exists. *The entropy of a measure preserving transformation S* is defined by

(2. 15) $$h(S)=\sup_{\mathcal{A}}h(\mathcal{A},S),$$

where the supremum is taken over all finite subfields of $\mathcal{X}$.

3. Integral Representation of Entropy.

The following is a reformation of a theorem of Tulcea [16], which is a key point for our integral representation of entropy, and the proof is similar to that of

Tulcea. $(X, \mathfrak{X}, P)$ and S are the same as in § 2.

THEOREM 3.1.[1] *Let $\mathcal{G}$ be a subfield of $\mathfrak{X}$ with the property*

$$S^{-1}\mathcal{G}=\mathcal{G},^{2)} \tag{3.1}$$

then the sequence of functions

$$f_n(x)=\frac{1}{n}I\left(\bigvee_{i=0}^{n-1}S^{-i}\mathcal{A}\mid\mathcal{G}\right) \tag{3.2}$$

converges to some S-invariant function $\hat{g}(x)$ in L^1-mean and in almost everywhere sense. Moreover if $\mathcal{G}\subset\bigvee_{i=1}^{\infty}S^{-i}\mathcal{A}$, then

$$h(\mathcal{A}, S)=\int\hat{g}(x)dp. \tag{3.3}$$

Proof. Putting

$$g_0(x)=I(\mathcal{A}\mid\mathcal{G})(x)$$

and

$$g_n(x)=I\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{n}S^{-i}\mathcal{A}\vee\mathcal{G}\right)(x),\qquad n=1,2,\cdots,$$

then the function $f_n(x)$ defined by (3.2) is expressed by

$$f_n(x)=\frac{1}{n}\sum_{l=0}^{n-1}g_l(S^{n-l-1}x),$$

because

$$I\left(\bigvee_{i=0}^{n-1}S^{-i}\mathcal{A}\mid\mathcal{G}\right)=I\left(\bigvee_{j=1}^{n}S^{-(n-j)}\mathcal{A}\mid\mathcal{G}\right)$$

$$=I(S^{-(n-1)}\mathcal{A}\mid\mathcal{G})+\sum_{k=2}^{n}I\left(S^{-(n-k)}\mathcal{A}\,\middle|\,\bigvee_{j=1}^{k-1}S^{-(n-j)}\mathcal{A}\vee\mathcal{G}\right)\qquad\text{by (2.6)},$$

$$=I(S^{-(n-1)}\mathcal{A}\mid S^{-(n-1)}\mathcal{G})+\sum_{k=2}^{n}I\left(S^{-(n-k)}\mathcal{A}\,\middle|\,\bigvee_{j=1}^{k-1}S^{-(n-j)}\mathcal{A}\vee S^{-(n-j)}\mathcal{G}\right)\qquad\text{by (3.1)},$$

$$=I(\mathcal{A}\mid\mathcal{G})S^{n-1}+\sum_{l=1}^{n-1}I\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{l}S^{-i}\mathcal{A}\vee\mathcal{G}\right)S^{n-l-1}\qquad\text{by (2.3)},$$

$$=\sum_{l=0}^{n-1}g_l(S^{n-l-1}x).$$

By the martingale convergence theorem, for any atom $A\in\mathcal{A}$,

1) If $\mathcal{G}$ is trivial, i.e. $\mathcal{G}=2=[\phi, X]$, then Theorem 3.1 is just the McMillan's theorem.

2) It means $\{S^{-1}E; E\in\mathcal{G}\}=\{F; F\in\mathcal{G}\}$.

$$p\left(A\left|\bigvee_{i=1}^{n} S^{-i}\mathcal{A}\vee\mathcal{Q}\right.\right)\to p\left(A\left|\bigvee_{i=1}^{\infty} S^{-i}\mathcal{A}\vee\mathcal{Q}\right.\right)\qquad \text{a.e. as } n\to\infty.$$

So because of the continuity of $\log t$ on $(0,\infty)$,

$$g_n(x)=I\left(\mathcal{A}\left|\bigvee_{i=1}^{n} S^{-i}\mathcal{A}\vee\mathcal{Q}\right.\right)\to I\left(\mathcal{A}\left|\bigvee_{i=1}^{\infty} S^{-i}\mathcal{A}\vee\mathcal{Q}\right.\right)=g(x)\qquad \text{say, a.e. as } n\to\infty,$$

and by (2. 11),

$$\int g(x)dp=H\left(\mathcal{A}\left|\bigvee_{i=1}^{\infty} S^{-i}\mathcal{A}\vee\mathcal{Q}\right.\right)\leqq H(\mathcal{A}),$$

hence $g(x)$ is integrable. By the ergodic theorem, $(1/n)\sum_{l=0}^{n-1} g(S^{n-k-l}x)$ converges a.e. to an S-invariant integrable function $\hat{g}(x)$. Putting

$$E_k=\left\{x;\max_{1\leqq j<k} g_j(x)\leqq\lambda<g_k(x)\right\}$$

and

$$F_k^{(i)}=\left\{x;\max_{1\leqq j<k} f_j^{(i)}(x)\leqq\lambda<f_k^{(i)}(x)\right\}$$

where $f_j^{(i)}(x)=-\log p(A_i\mid\bigvee_{l=1}^{j} S^{-l}\mathcal{A}\vee\mathcal{Q})$ and A_i is an atom in $\mathcal{A}$, then

$$p(E_k)=\sum_i p(A_i\cap E_k)=\sum_i p(A_i\cap F_k^{(i)}).$$

Since the set $F_k^{(i)}$ is $\bigvee_{l=1}^{k} S^{-1}\mathcal{A}\vee\mathcal{Q}$-measurable,

$$p(A_i\cap F_k^{(i)})=\int_{F_k^{(i)}} p\left(A_i\left|\bigvee_{l=1}^{k} S^{-l}\mathcal{A}\vee\mathcal{Q}\right.\right)dp=\int_{F_k^{(i)}} e^{-f_k^{(i)}(x)}dp\leqq e^{-\lambda}p(F_k^{(i)}).$$

Let r be the number of atoms in $\mathcal{A}$, then

$$\sum_k p(E_k)\leqq\sum_i e^{-\lambda}\sum_k p(F_k^{(i)})\leqq re^{-\lambda},$$

because $F_k^{(i)}\cap F_{k'}^{(i)}=\phi$ for $k\neq k'$. It follows that

$$p\left\{x;\sup_k g_k(x)>\lambda\right\}\leqq re^{-\lambda}\to 0\qquad \text{as } \lambda\to\infty,$$

which shows $\sup g_k(x)$ is integrable. And so

$$G_N(x)=\sup_{j\geqq N}|g_j(x)-g(x)|$$

is also integrable. Hence the Cesàro mean exists a.e., say $\hat{G}_N(x)$:

$$\hat{G}_N(x)=\lim_n\frac{1}{n}\sum_{k=0}^{n-1} G_N(S^{n-k-1}x)\qquad \text{a.e.}$$

and by the monotone convergence theorem,

$$E(G_N)=E(\hat{G}_N)\downarrow 0 \qquad \text{as} \quad N\to\infty.$$

Since $\hat{G}_N(x)$ is decreasing, $\lim_N \hat{G}_N(x)=0$ a.e.. Moreover

$$\begin{aligned}&\overline{\lim_n}\,|f_n(x)-g(x)|\\ &\leqq\overline{\lim_n}\left[\left|\frac{1}{n}\sum_{k=0}^{n-1}\left\{g_k(S^{n-k-1}x)-g(S^{n-k-1}x)\right\}\right|+\left|\frac{1}{n}\sum_{k=0}^{n-1}g(S^{n-k-1}x)-g(x)\right|\right]\\ &\leqq\overline{\lim_n}\left\{\frac{1}{n}\sum_{k=0}^{n-1}G_N(S^{n-k-1}x)+\left|\frac{1}{n}\sum_{k=0}^{n-1}g(S^{n-k-1}x)-g(x)\right|\right\}\\ &=\hat{G}_N(x)\to 0 \qquad \text{a.e. as } N\to\infty,\end{aligned}$$

which implies that $f_n(x)$ converges to $\hat{g}(x)$ in the both of the a.e. sense and the L^1-mean sense.[3)]

If we assume $\mathcal{G}\subset\bigvee_{i=1}^{\infty}S^{-i}\mathcal{A}$, then

$$\lim_k H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{k}S^{-i}\mathcal{A}\vee\mathcal{G}\right)=H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{\infty}S^{-i}\mathcal{A}\vee\mathcal{G}\right)=H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{\infty}S^{-i}\mathcal{A}\right)=h(\mathcal{A},S).$$

Hence

$$\begin{aligned}\int\hat{g}(x)dp&=\lim_n\int f_n(x)dp=\lim_n\frac{1}{n}H\left(\bigvee_{i=0}^{n-1}S^{-i}\mathcal{A}\mid\mathcal{G}\right)\\ &=\lim_n\frac{1}{n}\left\{H(\mathcal{A}\mid\mathcal{G})+\sum_{k=1}^{n-1}H\left(\mathcal{A}\,\middle|\,\bigvee_{i=1}^{k}S^{-i}\mathcal{A}\vee\mathcal{G}\right)\right\}=h(\mathcal{A},S).\end{aligned}$$

Q.E.D.

Let $(X,\mathfrak{X},S)$ be a measurable space with measurable transformation S and Π be a class of some S-invariant probability measures on $\mathfrak{X}$. (We assume that Π is not empty.) If we fix a p in Π, then we can consider $(X,\mathfrak{X},p,S)$ being a probability measure space with measure preserving transformation S. Over this space, we can also construct the entropy $h(\mathcal{A},S)$, of S relative to a finite partition $\mathcal{A}\subset\mathfrak{X}$, which depends on $p\in\Pi$. Hence it should be denoted by

$$h_p(\mathcal{A},S)=h(\mathcal{A},S).$$

Now we prove the following

Theorem 3. 2. *There exists an* S*-invariant* $\mathfrak{X}$*-measurable function* $h(x)$ *on* X, *which does not depend on* $p\in\Pi$, *and for every* $p\in\Pi$

3) The L^1-mean convergence of $f_n(x)$ is similarly proved as the McMillan's theorem. (See, e.g. [7], p. 28)

$$h_p(\mathcal{A}, S) = \int h(x) p(dx).$$

Proof. If we put,

$$\mathcal{G} = \left\{ B \in \bigvee_{n=1}^{\infty} S^{-n}\mathcal{A};\, S^{-1}B = B \right\}$$

then the preceding theorem is applicable; since $\mathcal{G}$ is a σ-subfield of $\mathcal{X}$, $S^{-1}\mathcal{G} = \mathcal{G}$ and moreover, $\mathcal{G} \subset \vee_{n=1}^{\infty} S^{-n}\mathcal{A}$. Hence, with the notations in the theorem,

$$h_p(\mathcal{A}, S) = \int \hat{g}(x) dp = \int \lim_n f_n(x) dp.$$

But $\hat{g}(x)$ and $f_n(x)$ depend on $p \in \Pi$. Now, for any atom $A \in \vee_{i=0}^{n-1} S^{-i}\mathcal{A}$

$$p(A \mid \mathcal{G}) = \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_A(S^i x) \qquad \text{a.e.} \tag{3. 4}$$

since the right hand side of (3. 4) is $\mathcal{G}$-measurable, and

$$\chi_B(x) = \chi_{S^{-i}B}(x) = \chi_B(S^i x) \qquad \text{for any } B \in \mathcal{G}$$

implies

$$\int_B \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_A(S^i x) p(dx) = \int \chi_B(x) \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_A(S^i x) p(dx)$$

$$= \int \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_B(x) \chi_A(S^i x) p(dx) = \int \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_{A \cap B}(S^i x) p(dx) = p(A \cap B),$$

where the last equality follows from $\chi_{A \cap B}(x)$ being an integrable function and from the ergodic theorem. Putting

$$f_A(x) = \overline{\lim_k} \frac{1}{k} \sum_{i=1}^{k} \chi_A(S^i x),$$

(3. 4) implies

$$f_n(x) = -\frac{1}{n} \sum_{A \in \vee_{i=0}^{n-1} S^{-i}\mathcal{A}} \chi_A(x) \log p(A \mid \mathcal{G}) = -\frac{1}{n} \sum_{A \in \vee_{i=0}^{n-1} S^{-i}\mathcal{A}} \chi_A(x) \log f_A(x) \qquad \text{a.e.}$$

But $f_n(x)$ converges to $\hat{g}(x)$ a.e., therefore

$$\hat{g}(x) = -\overline{\lim_n} \frac{1}{n} \sum \chi_A(x) \log f_A(x) \qquad \text{a.e.} \tag{3. 5}$$

We write $h(x)$ the right hand side of (3. 5). Then the function $h(x)$ is defined universally over X and does not depend on $p \in \Pi$, and

$$h_p(\mathcal{A}, S)=\int \hat{g}(x)dp=\int h(x)dp.$$

Evidently $\hat{g}(x)$ is an S-invariant function with mod p, hence $h(x)$ is also an S-invariant function with mod p. If we make non-essential alterations on a set $\{x; h(x)\neq h(Sx)\}$, then $h(x)$ becomes strictily S-invariant. Q.E.D.

4. Classification of Channels.

Let us consider two measurable spaces $(X, \mathfrak{X})$, $(Y, \mathfrak{Y})$. A *channel* ν from X to Y is a real valued function $\nu_x(B)$ on $X\times\mathfrak{Y}$ $(x\in X, B\in\mathfrak{Y})$ which satisfies the following conditions:

(i) If we fix $x\in X$, then $\nu_x(\cdot)$ is a probobility measure on $\mathfrak{Y}$;

(ii) If we fix $B\in\mathfrak{Y}$, then $\nu_.(B)$ is an $\mathfrak{X}$-measurable function on X.

Let S and T be measurable transformations on X and Y respectively. A channel ν is called *stationary* iff

(iii) $$\nu_{Sx}(B)=\nu_x(T^{-1}B) \qquad \text{for all } x\in X \text{ and } B\in\mathfrak{Y}.$$

We put Γ as a set of all stationary channels from X to Y. As in § 3, Π is a set of some S-invariant probability measures on $\mathfrak{X}$. Then for every $p\in\Pi$ we can construct a T-invariant probability measure q on $\mathfrak{Y}$ and a $S\times T$-invariant probability measure r on $\mathfrak{X}\times\mathfrak{Y}$ as follows:

$$q(B)=\int \nu_x(B)p(dx) \qquad \text{for every } B\in\mathfrak{Y},$$

$$r(C)=\int \nu_x(C_x)p(dx) \qquad \text{for every } C\in\mathfrak{X}\times\mathfrak{Y},$$

where C_x is a section of C with $x\in X$. Obviously q and r depend on a probability $p\in\Pi$ and a channel $\nu\in\Gamma$, therefore sometimes we write as

$$q=q(p,\nu), \qquad r=r(p,\nu).$$

DEFINITION 4.1. $\nu^1\in\Gamma$ and $\nu^2\in\Gamma$ are *equivalent with* mod Π iff $r^1=r(p,\nu^1)$ and $r^2=r(p,\nu^2)$ coincide as probabilities on $\mathfrak{X}\times\mathfrak{Y}$ for every measure $p\in\Pi$. In this case we write

$$\nu^1\equiv\nu^2\ (\Pi).$$

Let $\Pi_e\subset\Pi$ be a set of all ergodic measures in Π with respect to S. Then we can introduce an equivalence relation with mod Π_e in Γ.

DEFINITION 4.2. A system $(X, \mathfrak{X}, \Pi)$ is *complete for ergodicity* iff $p(A)>0$ $(p\in\Pi, A\in\mathfrak{X})$ implies $P_e(A)>0$ for some $P_e\in\Pi_e$.

THEOREM 4.1. *If the system* $(X, \mathfrak{X}, \Pi)$ *is complete for ergodicity, then for every* ν^1 *and* ν^2 *in* Γ *the following conditions are equivalent to each other:*

1) $\nu^1 \equiv \nu^2\ (\Pi)$,

1′) $\nu^1 \equiv \nu^2\ (\Pi_e)$,

2) $\nu_x^1(C_x) = \nu_x^2(C_x)$ *a.e.* Π *for every* $C \in \mathcal{X} \times \mathcal{Y}$,

2′) $\nu_x^1(C_x) = \nu_x^2(C_x)$ *a.e.* Π_e *for every* $C \in \mathcal{X} \times \mathcal{Y}$,

3) $\nu_x^1(B) = \nu_x^2(B)$ *a.e.* Π *for every* $B \in \mathcal{Y}$,

3′) $\nu_x^1(B) = \nu_x^2(B)$ *a.e.* Π_e *for every* $B \in \mathcal{Y}$,

where a.e. Π *(or* Π_e*) means that it is true a.e. for every* p *in* Π *(or* Π_e*).*

Proof. 1) $\Rightarrow$ 1′), 2) $\Rightarrow$ 3), 3) $\Rightarrow$ 3′) are obvious.

1′) $\Rightarrow$ 2′): Assume 2′) is not true, then

$$p_e\{x;\ \nu_x^1(C_x) \neq \nu_x^2(C_x)\} > 0$$

for some $C \in \mathcal{X} \times \mathcal{Y}$ and $p_e \in \Pi_e$. Suppose now

$$p_e\{x;\ \nu_x^1(C_x) > \nu_x^2(C_x)\} > 0,$$

then

$$\int_D \nu_x^1(C_x)p(dx) > \int_D \nu_x^2(C_x)p(dx)$$

where

$$D = \{x;\ \nu_x^1(C_x) > \nu_x^2(C_x)\}.$$

As

$$\nu_x([C \cap (D \times Y)]_x) = \chi_D(x)\nu_x(C_x),$$

it follows

$$r^1(C \cap (D \times Y)) > r^2(C \cap (D \times Y)),$$

which contradicts 1′), and the contradiction of the other case

$$p_e\{x;\ \nu_x^1(C_x) < \nu_x^2(C_x)\} > 0$$

follows from the same manner.

2′) $\Rightarrow$ 2): If 2) is not true, then

$$p\{x;\ \nu_x^1(C_x) \neq \nu_x^2(C_x)\} > 0 \qquad \text{for some } p \in \Pi \text{ and } C \in \mathcal{X} \times \mathcal{Y}.$$

Then, as $(X, \mathcal{X}, \Pi)$ is complete for ergodicity,

$$p_e\{x;\ \nu_x^1(C_x) \neq \nu_x^2(C_x)\} > 0 \qquad \text{for some } p_e \in \Pi_e.$$

3′) $\Rightarrow$ 3): Same as the above proof.

3) $\Rightarrow$ 1): Choose $A\in\mathcal{X}$ and $p\in\Pi$ arbitrarily, and integrate on A the each term in 3):

$$r^1(A\times B)=\int_A \nu_x^1(B)p(dx)=\int_A \nu_x^2(B)p(dx)=r^2(A\times B),$$

which shows that r^1 and r^2 coincide on $\mathcal{X}\times\mathcal{Y}$. Q.E.D.

REMARK. If the system $(X,\mathcal{X},\Pi)$ is not complete for ergodicity then only the following implications hold: 1) $\Leftrightarrow$ 2) $\Leftrightarrow$ 3) $\Rightarrow$ 1$'$) $\Leftrightarrow$ 2$'$) $\Leftrightarrow$ 3$'$).

5. Ergodicity of Channel.

$(X,\mathcal{X})$, $(Y,\mathcal{Y})$, S, T, Π and Γ are same as in the preceding section. Now we give a new definition:

DEFINITION 5.1. Channel $\nu\in\Gamma$ is called *ergodic* iff the ergodicity of $p_e\in\Pi_e$ always implies the ergodicity of $r=r(p_e,\nu)$.

THEOREM 5.1. *The following five conditions are equivalent to each other for any* $\nu\in\Gamma$.

1) ν *is ergodic.*
2) *If* $C\in\mathcal{X}\times\mathcal{Y}$ *and* $(S\times T)^{-1}C=C$ *then* $\nu_x(C_x)=0$ *or* 1 *a.e.* Π_e.
3) *If* $\nu\equiv\alpha\nu^1+(1-\alpha)\nu^2\,(\Pi_e)$ *for some* $\nu^1,\nu^2\in\Gamma$ *and* $0<\alpha<1$, *then* $\nu\equiv\nu^1\equiv\nu^2(\Pi_e)$.
4) *If* $\mathcal{X}_0$ *and* $\mathcal{Y}_0$ *are any semi-rings generating* $\mathcal{X}$ *and* $\mathcal{Y}$ *respectively, then*

$$\lim_N \frac{1}{N}\sum_{n=0}^{N-1}\int_{S^{-n}A\cap B}[\nu_x(T^{-n}C\cap D)-\nu_x(T^{-n}C)\nu_x(D)]p(dx)=0$$

for every $A,B\in\mathcal{X}_0$, $C,D\in\mathcal{Y}_0$ *and* $p\in\Pi$.

5) $\nu'_x\ll\nu_x$ *a.e.* Π_e *implies* $\nu\equiv\nu'(\Pi_e)$ *for any* $\nu'\in\Gamma$.

Let us give some explanations to the above: In 3), $\alpha\nu_x^1(B)+(1-\alpha)\nu_x^2(B)$ $(x\in X, B\in\mathcal{Y})$ is also a stationary channel, so Γ is a convex set and 3) means that ν is ergodic iff ν is an extremal point in Γ classified by the equivalence relation of mod Π_e. In 5), $\nu'_x\ll\nu_x$ a.e. Π_e means that there exists a set $D\in\mathcal{X}$ such that $p_e(D)=1$ for any $p_e\in\Pi_e$ and the measure ν'_x over $(Y,\mathcal{Y})$ is absolutely continuous with respect to the measure ν_x over $(Y,\mathcal{Y})$ for every $x\in D$. The condition 4) is a reformation of Adler [1]. The equivalences between 1), 3) and 4) are independently proved by Umegaki [17c] for the case $[X,Y]$ being a pair of compact Hausdorff spaces with a pair of homeomorphisms on each X and Y.

Proof. 1) $\Rightarrow$ 2): Suppose 1) true, then for every $p_e\in\Pi_e$, $r=r(p_e,\nu)$ is ergodic with a measure preserving transformation $S\times T$. Therefore $(S\times T)^{-1}C=C$ implies $r(C)=0$ or 1, i.e.

$$\int\nu_x(C_x)p_e(dx)=0 \quad\text{or } 1.$$

If this is 0, then $\nu_x(C_x)=0$ a.e. p_e, and on the other case, $\nu_x(C_x)=1$ a.e. p_e, hence

$$\nu_x(C_x)=0 \quad \text{or } 1 \text{ a.e. } \Pi_e.$$

2)$\Rightarrow$1): Choose $C\in\mathfrak{X}\times\mathfrak{Y}$ with $(S\times T)^{-1}C=C$ and $p_e\in\Pi_e$, then the $\mathfrak{X}$-measurable sets

$$D_0=\{x;\nu_x(C_x)=0\} \qquad \text{and} \qquad D_1=\{x;\nu_x(C_x)=1\}$$

are S-invariant with mod p_e. Hence $p_e(D_0)$ and $p_e(D_1)$ are 0 or 1 from the ergodicity of p_e. Thus, the ergodicity of r follows from

$$r(C)=\int\nu_x(C_x)p_e(dx)=0 \quad \text{or } 1.$$

1)$\Rightarrow$3): Suppose 3) is false, then for some $\nu^1,\nu^2\in\Gamma$, $\nu^1\not\equiv\nu^2(\Pi_e)$ and for α, $0<\alpha<1$,

$$\nu\equiv\alpha\nu^1+(1-\alpha)\nu^2\ (\Pi_e).$$

Hence for some $p_e\in\Pi_e$, $r^1=r(p_e,\nu^1)$ and $r^2=r(p_e,\nu^2)$ are not identical and for every $C\in\mathfrak{X}\times\mathfrak{Y}$,

$$\int\nu_x(C_x)p_e(dx)=\int\alpha\nu_x^1(C_x)p_e(dx)+\int(1-\alpha)\nu_x^2(C_x)p_e(dx),$$

that is,

$$r(C)=\alpha r^1(C)+(1-\alpha)r^2(C),$$

which shows that r can be written by a linear combination of the different measures r^1 and r^2. Consequently r is not ergodic.

3)$\Rightarrow$2): Suppose 2) is false. Then for some $p_e\in\Pi_e$ and $C\in\mathfrak{X}\times\mathfrak{Y}$ with $(S\times T)^{-1}C=C$, $X'=\{x;\nu_x(C_x)\neq 0,1\}$ is not of p_e-measure null. Now we define new channels $\bar\nu^1,\bar\nu^2\in\Gamma$ by

$$\bar\nu_x^1(D)=\begin{cases}\nu_x(D\cap C_x)/\nu_x(C_x) & \text{if } x\in X',\\ \nu_x(D) & \text{if } x\notin X',\end{cases}$$

$$\bar\nu_x^2(D)=\begin{cases}\nu_x(D\cap(Y\backslash C_x))/\nu_x(Y\backslash C_x) & \text{if } x\in X',\\ \nu_x(D) & \text{if } x\notin X',\end{cases}$$

where $D\in\mathfrak{Y}$. These are stationary, because $C_x=T^{-1}C_{Sx}$ implies

$$\nu_x(C_x)=\nu_x(T^{-1}C_{Sx})=\nu_{Sx}(C_{Sx}),$$

hence $S^{-1}X'=X'$ and

$$\nu_{Sx}(D\cap C_{Sx})=\nu_x(T^{-1}D\cap T^{-1}C_{Sx})=\nu_x(T^{-1}D\cap C_x).$$

Moreover $\bar\nu_x^1(C_x)=1$ and $\bar\nu_x^2(C_x)=0$ for all $x\in X'$, and so

$$\bar{\nu}^1 \doteqdot \bar{\nu}^2 (\Pi_e).$$

Then we can see easily,

$$\nu_x(D) = \nu_x(C_x)\bar{\nu}_x^1(D) + \{1-\nu_x(C_x)\}\bar{\nu}_x^2(D).$$

Putting

$$A = \left\{x;\, \nu_x(C_x) \geqq \frac{1}{2}\right\} \qquad \text{and} \qquad B = \left\{x;\, \nu_x(C_x) < \frac{1}{2}\right\},$$

then A and B are $\mathfrak{X}$-measurable. Finally we define channels ν^1 and ν^2 as follows:

$$\nu_x^1(D) = \begin{cases} \bar{\nu}_x^1(D) & \text{if} \quad x \in A, \\ \{1-2\nu_x(C_x)\}\bar{\nu}_x^2(D) + 2\nu_x(C_x)\bar{\nu}_x^1(D) & \text{if} \quad x \in B, \end{cases}$$

$$\nu_x^2(D) = \begin{cases} \{2\nu_x(C_x)-1\}\bar{\nu}_x^1(D) + \{2-2\nu_x(C_x)\}\bar{\nu}_x^2(D) & \text{if} \quad x \in A, \\ \bar{\nu}_x^2(D) & \text{if} \quad x \in B, \end{cases}$$

where $D \in \mathcal{Y}$. Then these satisfy the conditions (i) and (ii) of channel (§ 4). Moreover the stationarity of ν^1 and ν^2 follows from $S^{-1}A = A$, $S^{-1}B = B$, the stationarity of $\bar{\nu}^1$ and $\bar{\nu}^2$, and the S-invariantness of $\nu_x(C_x)$. We see easily

$$\nu \equiv \frac{1}{2}\nu^1 + \frac{1}{2}\nu^2 \ (\Pi_e)$$

and for all $x \in X'$,

$$\{2\nu_x(C_x)-2\}\{\bar{\nu}_x^1(C_x) - \bar{\nu}_x^2(C_x)\} \neq 0 \qquad \text{and} \qquad \nu_x(C_x)\{\bar{\nu}_x^1(C_x) - \bar{\nu}_x^2(C_x)\} \neq 0,$$

hence

$$\bar{\nu}_x^1(C_x) \neq \{2\nu_x(C_x)-1\}\bar{\nu}_x^1(D) + \{2-2\nu_x(C_x)\}\bar{\nu}_x^2(D) \qquad \text{for} \quad x \in X'$$

and

$$\{1-2\nu_x(C_x)\}\bar{\nu}_x^2(D) + 2\nu_x(C_x)\bar{\nu}_x^1(D) \neq \bar{\nu}_x^2(D) \qquad \text{for} \quad x \in X',$$

which imply $\nu^1 \doteqdot \nu^2 (\Pi_e)$.

1) $\Rightarrow$ 4): For any $p_e \in \Pi_e$, $A, B \in \mathfrak{X}_0$ and $C, D \in \mathcal{Y}_0$,

$$\frac{1}{N}\sum_{n=0}^{N-1}\int_{S^{-n}A\cap B}\{\nu_x(T^{-n}C\cap D) - \nu_x(T^{-n}C)\nu_x(D)\}p_e(dx)$$

$$(5.1) \quad = \frac{1}{N}\sum_{n=0}^{N-1}\left\{\int_{S^{-n}A\cap B}\nu_x(T^{-n}C\cap D)p_e(dx) - \int_A \nu_x(C)p_e(dx)\int_B \nu_x(D)p_e(dx)\right\}$$

$$+ \frac{1}{N}\sum_{n=0}^{N-1}\left\{\int_A \nu_x(C)p_e(dx)\int_B \nu_x(D)p_e(dx) - \int_{S^{-n}A\cap B}\nu_x(T^{-n}C)\nu_x(D)p_e(dx)\right\}$$

$$=\frac{1}{N}\sum_{n=0}^{N-1}\{r((S^{-n}A\times T^{-n}C)\cap(B\times D))-r(A\times C)r(B\times D)\}$$

$$(5.2)\qquad +\frac{1}{N}\sum_{n=0}^{N-1}\left\{\int\chi_A(x)\nu_x(C)p_e(dx)\int\chi_B(x)\nu_x(D)p_e(dx)\right.$$

$$\left.-\int\chi_A(S^n x)\nu_{S^n x}(C)\chi_B(x)\nu_x(D)p_e(dx)\right\}.$$

The first and second terms of the last hand side of (5. 2) converge to zero as $N\to\infty$ by the ergodicities of r and p_e.

4) $\Rightarrow$ 1): For any $p_e\in\Pi_e$, $A, B\in\mathfrak{X}_0$ and $C, D\in\mathcal{Y}_0$, similarly to the reformation of the formula (5. 1),

$$\frac{1}{N}\sum_{n=0}^{N-1}\{r((S^{-n}A\times T^{-n}C)\cap(B\times D))-r(A\times C)r(B\times D)\}$$

$$=\frac{1}{N}\sum_{n=0}^{N-1}\int_{S^{-n}A\cap B}\{\nu_x(T^{-n}C\cap D)-\nu_x(T^{-n}C)\nu_x(D)\}p_e(dx)$$

$$+\frac{1}{N}\sum_{n=0}^{N-1}\left\{\int\chi_A(S^n x)\nu_{S^n x}(C)\chi_B(x)\nu_x(D)p_e(dx)\right.$$

$$\left.-\int\chi_A(x)\nu_x(C)p_e(dx)\int\chi_B(x)\nu_x(D)p_e(dx)\right\}.$$

The first and second terms of the right hand side of the equation converge to zero as $N\to\infty$ by the assumption in 4) and by the ergodicity of p_e, which shows the ergodicity of $r=r(p_e,\nu)$.

5) $\Rightarrow$ 2): Suppose 2) is false. Then for some $p_e\in\Pi_e$ and $C\in\mathfrak{X}\times\mathcal{Y}$ with $(S\times T)^{-1}C=C$, $p\{x; 0<\nu_x(C_x)<1\}>0$. If we put

$$E=\{x; 0<\nu_x(C_x)<1\}$$

and define for $D\in\mathcal{Y}$,

$$\nu'_x(D)=\begin{cases}\nu_x(D\cap C_x)/\nu_x(C_x) & \text{if } x\in E,\\ \nu_x(D) & \text{if } x\notin E,\end{cases}$$

then $\nu'\in\Gamma$ and $\nu'\neq\nu\,(\Pi_e)$ are proved similarly in the proof of 3) $\Rightarrow$ 2). Moreover if $\nu_x(B)=0$ then $\nu'_x(B)=0$ for any $B\in\mathcal{Y}$, hence $\nu'_x\ll\nu_x$, which contradicts 5).

2), 3) $\Rightarrow$ 5): Suppose 5) false. Then there exists $\nu'\in\Gamma$ and $\nu'_x\ll\nu_x$ a.e. Π_e and $\nu'\neq\nu(\Pi_e)$. Assuming 2), for any $C\in\mathfrak{X}\times\mathcal{Y}$ with $(S\times T)^{-1}C=C$,

$$\nu_x(C_x)=0 \text{ or } 1 \quad \text{a.e. } \Pi_e.$$

Choose $D\subset\{x\in X; \nu'_x\ll\nu_x\}$, $D\in\mathfrak{X}$ and $p_e(D)=1$ for all $p_e\in\Pi_e$, then for every $x\in D$, $\nu_x(C_x)=0$ implies $\nu'_x(C_x)=0$ and $\nu_x(C_x)=1$ implies $\nu'_x(C_x)=1$. Consider now a channel

$$\nu''=\alpha\nu+(1-\alpha)\nu', \qquad 0<\alpha<1,$$

which satisfies

$$\nu''_x(C_x)=0 \text{ or } 1 \quad \text{a.e. } \Pi_e.$$

Hence $\nu''\in\Gamma$ is ergodic, which contradicts 3). Q.E.D.

REMARK. M-dependent channel is defined as follows; A and B are finite (or countable) sets. We put

$$X=A^I=\{\omega=(\cdots,\omega_{-1},\omega_0,\omega_1,\cdots);\ \omega_i\in A\},$$

$$Y=B^I=\{\omega'=(\cdots,\omega'_{-1},\omega'_0,\omega'_1,\cdots);\ \omega'_i\in B\}$$

and $\mathfrak{X}=F_A$, $\mathfrak{Y}=F_B$ are Borel fields generated by rectangles in A^I and B^I respectively. We call such system (A^I, F_A) as alphabet space. S and T are shift transformations on A^I and B^I, i.e.

$$(S\omega)_n=\omega_{n+1}, \qquad (T\omega')_n=\omega'_{n+1}.$$

For a fixed integer $M>0$, a channel ν from A^I to B^I is called M-dependent iff

$$\nu_x([\omega'_s,\cdots,\omega'_t]\cap[\omega'_u,\cdots,\omega'_v])=\nu_x([\omega'_s,\cdots,\omega'_t])\nu_x([\omega'_u,\cdots,\omega'_v])$$

for every $x\in A^I$ whenever $u-t\geqq M$, where $[\omega'_s,\cdots,\omega'_t]$ means a rectangle of coordinates from s to t. If we write the rectangles in B^I as $\mathfrak{Y}_0$, then M-dependentness implies for any $C, D\in\mathfrak{Y}_0$,

$$\nu_x(C\cap T^{-n}D)=\nu_x(C)\nu_x(T^{-n}D)$$

for large n, and so the condition of Theorem 1, 4) is satisfied. Hence, if we consider Π being the set of all S-invariant probability Borel measures, M-dependent channel is always ergodic.

6. Capacity of Channel.

In this section, $(X,\mathfrak{X})$, $(Y,\mathfrak{Y})$, S, T, Π and Γ are same as in §§ 4, 5.

DEFINITION 6.1. *A transmission rate of a channel $\nu\in\Gamma$ with respect to a measure $p\in\Pi$* is defined by

$$R_p(\nu)=\sup_{\mathcal{A},\mathcal{B}}\{h_p(\mathcal{A},S)+h_q(\mathcal{B},T)-h_r(\mathcal{A}\times\mathcal{B},S\times T)\} \tag{6.1}$$

where the supremum is taken within all finite partitions $\mathcal{A}$ and $\mathcal{B}$ in $\mathfrak{X}$ and $\mathfrak{Y}$ respectively, and $q=q(p,\nu)$, $r=r(p,\nu)$.

REMARK. (1) Using (2.14), (2.13) and (2.11), we can see easily $R_p(\nu)\geqq 0$.
(2) In finite alphabet spaces (see § 5, Remark), it holds that

$$h(\mathcal{A},S)\leqq h(S)<+\infty,\qquad h(\mathcal{B},T)\leqq h(T)<+\infty,$$

$$h(\mathcal{A}\times\mathcal{B},S\times T)\leqq h(S\times T)<+\infty,$$

and[4)]

$$R_p(\nu)=h(S)+h(T)-h(S\times T).$$

The amount $R_p(\nu)$ is just a transmission rate in usual sense. (cf. Feinstein [6])

DEFINITION 6. 2.

$$C_s(\nu)=\sup_{p\in\Pi}R_p(\nu)\qquad\text{and}\qquad C_e(\nu)=\sup_{p\in\Pi'}R_p(\nu)$$

are called *stationary capacity* and *ergodic capacity* (of a channel $\nu\in\Gamma$) respectively, where

$$\Pi'=\{p\in\Pi;\ r=r(p,\nu)\text{ is ergodic with }S\times T\},$$

and if $\Pi'=\phi$ then we put $C_e(\nu)=0$, where Π is always assumed non-empty.

THEOREM 6. 1. *If a system $(X,\mathfrak{X},\Pi)$ is complete for ergodicity and a stationary channel $\nu\in\Gamma$ is ergodic, then*

$$C_s(\nu)=C_e(\nu).$$

Proof. Obviously $C_e(\nu)\leqq C_s(\nu)$. Now we assume $C_s(\nu)<+\infty$. For arbitrary $\varepsilon>0$ there exist a measure $p\in\Pi$ and finite partitions $\mathcal{A}$ and $\mathcal{B}$,

$$h_p(\mathcal{A},S)+h_q(\mathcal{B},T)-h_r(\mathcal{A}\times\mathcal{B},S\times T)>C_s(\nu)-\varepsilon, \tag{6. 2}$$

where $q=q(p,\nu)$ and $r=r(p,\nu)$. Then there exist measurable functions $h_1(x)$, $h_2(y)$ and $h_3(x,y)$ by Theorem 3. 2 such that

$$h_p(\mathcal{A},S)=\int_X h_1(x)p(dx),$$

$$h_q(\mathcal{B},T)=\int_Y h_2(y)q(dy)=\int_X\int_Y h_2(y)\nu_x(dy)p(dx),$$

$$h_r(\mathcal{A}\times\mathcal{B},S\times T)=\int_{X\times Y}h_3(x,y)r(dx,dy)=\int_X\int_Y h_3(x,y)\nu_x(dy)p(dx).$$

Let us denote

4) The right hand side of (6. 1) can be proved to be monotone increasing for refinements of finite partitions $\mathcal{A}$ and $\mathcal{B}$, by the formula (2. 1. 2) in [4]. And $h(\mathcal{A},S)$, $h(\mathcal{B},T)$ and $h(\mathcal{A}\times\mathcal{B},S\times T)$ increase and approximate $h(S)$, $h(T)$ and $h(S\times T)$ respectively as $\mathcal{A}$ and $\mathcal{B}$ being refined. So the formula is valid.

$$\tilde{h}(x)=h_1(x)+\int_Y h_2(y)\nu_y(dy)+\int_Y h_3(x,y)\nu_x(dy).$$

Then the left hand side of (6. 2) equals to

$$\int \tilde{h}(x)p(dx).$$

Since $\tilde{h}(x)$ is S-invariant, there exist simple functions

$$\tilde{h}_n(x)=\sum_{i=1}^{k_n}\lambda_i^{(n)}\chi_{D_i^{(n)}}(x), \qquad n=1,2,\cdots$$

satisfying $\tilde{h}_n(x)\uparrow\tilde{h}(x)$, where $\{D_i^{(n)}\}_{i=1}^{k_n}$ $(n=1,2,\cdots)$ is a sequence of measurable partitions with $S^{-1}D_i^{(n)}=D_i^{(n)}$. By virtue of the monotone convergence theorem,

$$\lim_n\int\tilde{h}_n(x)p(dx)=\int\tilde{h}(x)p(dx),$$

and hence by (6. 2)

$$\int\tilde{h}_n(x)p(dx)=\sum_{i=1}^{k_n}\lambda_i^{(n)}p(D_i^{(n)})>C_s(\nu)-\varepsilon \quad \text{for some } n.$$

Consequently

$$\lambda_{i_0}^{(n)}>C_s(\nu)-\varepsilon \qquad p(D_{i_0}^{(n)})>0 \quad \text{for some } i_0,$$

then there exists some $p_e\in\Pi_e$ and $p_e(D_{i_0}^{(n)})=1$, since $(X,\mathfrak{X},\Pi)$ is complete for ergodicity and $S^{-1}D_{i_0}^{(n)}=D_{i_0}^{(n)}$. Hence

$$R_{p_e}(\nu)\geqq\int\tilde{h}(x)p_e(dx)\geqq\int\tilde{h}_n(x)p_e(dx)=\lambda_{i_0}^{(n)}>C_s(\nu)-\varepsilon,$$

and by the ergodicity of ν,

$$r=r(p_e,\nu)\in\Pi',$$

which shows

$$C_e(\nu)\geqq C_s(\nu).$$

We can prove the inequality in the case $C_s(\nu)=+\infty$ similarly. Q.E.D.

7. Topological Argument and Application.

Let us consider a system $(X,\mathfrak{X},\Pi)$ and a transformation S on X, where X is a completely regular topological space, $\mathfrak{X}$ is a Borel field generated by open sets in X, S is a homeomorphism on X and Π is a class of all S-invariant inner regular[5] probability measures on $\mathscr{X}$. (We assume $\Pi \neq \phi$.)

THEOREM 7.1. *The above system* $(X, \mathscr{X}, \Pi)$ *is complete for ergodicity.*

Proof. Let $\check{X}$ be the Čech's compactification of X, and a homeomorphism $\check{S}$ on $\check{X}$ be an extension of S. (Such extension always exists.) Now for a measure $p \in \Pi$ and for every Borel set A in $\check{X}$ we write

$$\check{p}(A) = p(A \cap X),$$

where $A \cap X \in \mathscr{X}$ because a class $\{B; B \cap X \in \mathscr{X}\}$ contains all open sets in X and closed with countable union and complementation.

Then $\check{p}$ is an inner regular Borel measure, since compact sets in X are also compact in $\check{X}$.

Now we assume $p(C) > 0$ for some $C \in \mathscr{X}$, then, as p is inner regular, there exists some compact set K in C and $p(K) > 0$, which follows $\check{p}(K) > 0$. Then by the similar reason as Farrel ([5] p. 459, even if X is not a metric space), there exists some inner regular ergodic Borel measure $\check{p}_e$ and $\check{p}_e(K) > 0$.

Now we define

$$p_e(A) = \check{p}_e\left(A \cap \bigcup_{n=-\infty}^{\infty} S^{-n}K\right) \qquad \text{for all } A \in \mathscr{X},$$

where $A \cap \bigcup_{n=-\infty}^{\infty} S^{-n}K$ is a Borel set in $\check{X}$ since a class $\{A; A \cap \bigcup_{n=-\infty}^{\infty} S^{-n}K$ is a Borel set in $\check{X}\}$ contains $\mathscr{X}$.

Then p_e is an ergodic probability measure on $\mathscr{X}$, and inner regular as compact subsets of $\check{X}$, contained in X, is also compact in X. Thus $p(C) > 0$ implies $p_e(C) > 0$ for some $p_e \in \Pi_e$. Q.E.D.

Application: We will be able to construct information theory of countable alphabet by Hinchin's method, on countable alphabet spaces A^I and B^I. The countable alphabet space A^I, where A is a countable set and I is a set of integer, can be seen as Polish space (separable complete metric space) with Tychonoff's product topology, because a discrete space A is of course a Polish space and a countable product of Polish spaces is also a Polish space.

If we consider the system (A^I, F_A, Π, S), where F_A is a Borel field generated by open sets in A^I, S is a shift transformation, and Π is a class of all S-invariant Borel measures, (See, Remark in § 5) then the system satisfies the topological condition of Theorem 1, because every Borel measure is necessarily inner regular in Polish spaces. (See [11] p. 64)

Therefore, when we treat the capacities of ergodic channels from A^I to B^I (or to other spaces), we need not distinguish the ergodic capacity and the stationary capacity.

5) A finite measure μ is inner regular iff $\mu(E) = \sup_{K \subset E} \mu(K)$ where K is compact.

Bibliography

[1] ADLER, R. L., Ergodic and mixing properties of infinite memory channels. Proc. Amer. Math. Soc. **12** (1961), 924–930.

[2] BILLINGSLEY, P., Ergodic theory and information. John Wiley & Sons (1965).

[3] BREIMAN, L., The individual ergodic theorem of information theory. Ann. Math. Stat. **28** (1957), 809–811.

[4] DOBRUSHIN, R. L., General formulation of Shannon's main theorem in information theory. Amer. Math. Soc. Trans. (2) **33** (1963), 323–438. (English translation)

[5] FARREL, R. H., Representation of invariant measures. Ill. J. Math. **6** (1958), 175–180.

[6] FEINSTEIN, A., Foundations of information theory. McGraw-Hill, New York (1958).

[7] HALMOS, P. R., Entropy in ergodic theory. Mimeographed notes, The Univ. of Chicago (1959).

[8] HINCHIN, A. I., Mathematical foundations of information theory. Dover Publications, New York (1958). (English translation)

[9] JACOBS, K., (a) Über die Struktur der mittleren Entropie. Math. Zeitschr. **78** (1962), 33–43; (b) Ergodic decomposition of the Kolmogorov-Sinai Invariant, Ergodic theory (edited by F. B. Wright). Proc. of International Symposium, Tulane Univ. Oct. 1961, Acad. Press (1963), 173–190.

[10] KOLMOGOROV, A. N., (a) A new metric invariant of transitive dynamic systems and automorphisms in Lebesgue spaces. Dokl. Akad. Nauk, SSSR **119** (1958), 861–864 (in Russian); (b) On the entropy per unit time as a metric invariant of automorphisms. Ibid. **124** (1959), 754–755. (in Russian)

[11] NEVEU, J., Mathematical foundations of the calculus of probability. Holden-day, (1965). (English translation)

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

[13] SHANNON, C. E., A mathematical theory of communication. Bell System Tech. Journ. **27** (1948), 379–423, 623–656.

[14] SINAI, Ya. G., The concept of the entropy of a dynamical system. Dokl. Akad. Nauk SSSR **124** (1959), 768–771.

[15] TAKANO, K., On the basic theorems of information theory. Ann. Inst. Stat. Math., Tokyo **9** (1957), 53–77.

[16] TULCEA, A. I., Contributions to information theory for abstract alphabets. Arkiv för Math. **4** (1963), 235–247.

[17] UMEGAKI, H., (a) General treatment of alphabet-message space and integral representation of entropy. Kōdai Math. Sem. Rep. **16** (1964), 18–26; (b) A functional method for stationary channels. Ibid., 27–39; (c) Representations and extremal properties of averaging operators and their applications to information channels. To appear.

DEPARTMENT OF MATHEMATICS,
TOKYO INSTITUTE OF TECHNOLOGY.