KŌDAI MATH. SEM. REP.
22 (1970), 159—165

A NON-ERGODIC COMPOUND SOURCE WITH A MIXING INPUT SOURCE AND AN ADLER ERGODIC CHANNEL

BY YATSUKA NAKAMURA

1. Introduction.

Adler raised in [1] three kinds of channels which were called SM, WM and ERG channels respectively, and showed that the first two of them are ergodic but in general the rest one is not. At the end of the paper, he mentioned the problem; "Does ERG channel make the compound source ergodic when the input source is mixing?"

We will give here a negative answer to the problem by the method developed in [3]. The idea is derived from the fact that a channel is ergodic iff it is an extremal point in all channels. This was proved in [3] and [4] independently.

The author wishes to express his thanks to Professor H. Umegaki for his various encouragements and advices in preparing this paper.

2. Notation and definitions.

Let $(X, \mathfrak{X}, S)$ and $(Y, \mathfrak{Y}, T)$ be measurable spaces with measurable transformations S and T which act on X and Y respectively. A *stationary channel* 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:

(C 1) for each fixed $x\in X$, $\nu_x(\cdot)$ is a probability measure on $\mathfrak{Y}$;

(C 2) for each fixed $B\in\mathfrak{Y}$, $\nu_\cdot(B)$ is an $\mathfrak{X}$-measurable function on X;

(C 3) $\nu_{Sx}(B)=\nu_x(T^{-1}B)$ for all $x\in X$ and $B\in\mathfrak{Y}$.

Denote by Π a non empty set of S-invariant probability measures (we assume their existence) on $\mathfrak{X}$, and by Γ a set of all stationary channels from X to Y. For every $p\in\Pi$ and $\nu\in\Gamma$ we can construct a T-invariant probability measure q on $\mathcal{Y}$ and an $S\times T$-invariant probability measure r on $\mathcal{X}\times\mathcal{Y}$ as follows;

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

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

where C_x is a section of C determined by $x\in X$. If it is necessary we write as

Received November 21, 1969.

$q=q(p, \nu)$ and $r=r(p, \nu)$ since the probability measures q and r depend on $p\in\Pi$ and $\nu\in\Gamma$.

The channels $\nu^1\in\Gamma$ 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 probability measures on $\mathcal{X}\times\mathcal{Y}$ for every measure $p\in\Pi$, and we write

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

Denote by $\Pi_e\subset\Pi$ the set of all ergodic measures in Π, where the ergodicity is mean by one with respect to S.

Definition 2. 1. A channel $\nu\in\Gamma$ is said to be *ergodic* iff for all $p\in\Pi_e$, $r=r(p, \nu)$ is ergodic with respect to the transformation $S\times T$. While, a channel $\nu\in\Gamma$ is said to be *Adler-ergodic* (*ERG*, say), if it satisfies

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}[\nu_x(T^{-n}E\cap F)-\nu_x(T^nE)\nu_x(F)]=0 \qquad \Pi\text{-a.e.} \tag{2. 1}$$

for all $E, F\in\mathcal{Y}$, where Π-a.e. means p-a.e. for all $p\in\Pi$.

Definition 2. 2. For a channel $\nu\in\Gamma$ a channel $\nu^*\in\Gamma$ satisfying the following is called an *average* of ν.

(A 1) $\nu^*{}_{Sx}(B)=\nu^*_x(B)$ for all $B\in\mathcal{Y}$ and $x\in X$,

(A 2) $\int_G \nu_x(B)p(dx)=\int_G \nu^*_x(B)p(dx)$

for all $B\in\mathcal{Y}$ and S-invariant[1] $G\in\mathcal{X}$.

3. Average of channel.

In this section we assume the average of all channels in Γ.

Lemma 3. 1. *For all* $B\in\mathcal{Y}$,

$$\nu_x{}^*(B)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\nu_{S^nx}(B) \qquad \Pi\text{-a.e.} \tag{3. 1}$$

Proof. The condition (A 2) implies

$$\nu_x{}^*(B)=E_p(\nu_x(B)|\mathcal{G}) \qquad p\text{-a.e.} \tag{3. 2}$$

for all $p\in\Pi$, where $\mathcal{G}$ is a σ-subfield of all S-invariant $\mathcal{X}$-measurable sets and the right hand side of (3. 2) is the conditional expectation of a function $\nu_x(B)$ of x with respect to the σ-subfield $\mathcal{G}$ over probability measure space $(X, \mathcal{X}, p)$ with $p\in\Pi$. Hence by the individual ergodic theorem, (3. 1) follows from (3. 2).

1) That is, $S^{-1}G=G$.

LEMMA 3. 2. *For a $p \in \Pi_e$, we assume $r(p, \nu)$ is not a direct product measure on $X \times Y$, i.e. $r = r(p, \nu)$ is not expressed by $r = p \times q$ $(q = q(p, \nu))$. Then*

$$r(p, \nu) \neq r(p, \nu^*). \tag{3. 3}$$

Proof. By (3. 1) together with the ergodicity of p, for $A \in \mathfrak{X}$

$$\int_A \nu_x^*(B) p(dx) = \int_A \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \nu_{S^n x}(B) p(dx)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \nu_{S^n x}(B) \chi_A(x) p(dx)$$

$$= \int \nu_x(B) p(dx) \int \chi_A(x) p(dx) = p(A) q(B).$$

By the assumption, we can choose $A \in \mathfrak{X}$ and $B \in \mathfrak{Y}$ as

$$r(A \times B) \neq p(A) q(B),$$

and the above chain of equalities shows (3. 3).

THEOREM 3. 1. *Let $\nu^1, \nu^2 \in \Gamma$ be ERG channels admitting the averages ν^{1*} and ν^{2*} respectively. If*

$$\nu^{1*} \equiv \nu^{2*} (\Pi) \tag{3. 4}$$

then $\nu^0 = (1/2)\nu^1 + (1/2)\nu^2$[2] *is also an ERG channel.*

Proof. Since the channels ν^1 and ν^2 satisfy (2. 1) and since the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \nu_x^i(T^{-n}E) \nu_x^i(F) = \left[\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \nu_{S^n x}^i(E)\right] \nu_x^i(F) \qquad (i=1, 2) \tag{3. 5}$$

always exists p-a.e. $x \in X$ for all $p \in \Pi$, the formula (2. 1) for $\nu = \nu^i$ becomes

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \nu_x^i(T^{-n}E \cap F) = \nu_x^{i*}(E) \nu_x^i(F) \qquad \Pi\text{-a.e.} \tag{3. 6}$$

Hence

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} [\nu_x^0(C \cap T^{-n}D) - \nu_x^0(C) \nu_x^0(T^{-n}D)]$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Big[\frac{1}{2} \nu_x^1(C \cap T^{-n}D) + \frac{1}{2} \nu_x^2(C \cap T^{-n}D)$$

$$- \frac{1}{4} \{\nu_x^1(C) + \nu_x^2(C)\}\{\nu_x^1(T^{-n}D) + \nu_x^2(T^{-n}D)\}\Big]$$

2) $\nu_x^0(B) = (1/2)\nu_x^1(B) + (1/2)\nu_x^2(B)$ $(x \in X,\ B \in \mathfrak{Y})$ satisfies the conditions (C 1)~(C 3), and so $\nu^0 \in \Gamma$.

and

$$=\frac{1}{2}\nu_x^{1*}(D)\nu_x^1(C)+\frac{1}{2}\nu_x^{2*}(D)\nu_x^2(C)$$

$$-\frac{1}{4}\{\nu_x^1(C)+\nu_x^2(C)\}\{\nu_x^{1*}(D)+\nu_x^{2*}(D)\}$$

$$=\frac{1}{4}\{\nu_x^{1*}(D)-\nu_x^{2*}(D)\}\{\nu_x^1(C)-\nu_x^2(C)\} \qquad \Pi\text{-a.e.}$$

Therefore (3. 4)[3] implies that (3. 7) equals to 0 Π-a.e. Q.E.D.

THEOREM 3. 2. *If $\nu\in\Gamma$ is an ERG channel, then ν^* is also an ERG channel.*

Proof. From (3. 2) and the bounded convergence theorem for conditional expectation, the following equations are valid;

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\nu_x^*(C\cap T^{-n}D)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}E_p(\nu_x(C\cap T^{-n}D)|\mathcal{G})$$

$$=E_p(\nu_x^*(D)\nu_x(C)|\mathcal{G})=\nu_x^*(D)E_p(\nu_x(C)|\mathcal{G})$$

$$=\nu_x^*(D)\nu_x^*(C)=\left\{\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\nu_x^*(T^{-n}D)\right\}\nu_x^*(C) \qquad p\text{-a.e.}$$

for all $p\in\Pi$. Q.E.D.

4. Main result.

In the preceding section, we have discussed the property of the average. In the following, we will show an existence theorem of ν^* under a topological condition over the output Y.

THEOREM 4. 1. *Let Y be a compact metric space, $\mathcal{Y}$ be the σ-field of Borel sets in Y and T be a homeomorphism acting on Y. Given the system $(X, \mathfrak{X}, \Pi, S)$ in any way, let Γ be the set of all stationary channels from X to Y. Then for every channel $\nu\in\Gamma$ there exists an average $\nu^*\in\Gamma$ of ν, that is, ν and ν^* are related by the conditions* (A 1) *and* (A 2).

Proof. Let $C(Y)$ be a Banach space of all continuous functions on Y, $\hat{C}(Y)$ be its conjugate space and $\hat{C}(Y)_1^+$ be a positive part of the unit sphere of $C(Y)$. Putting

$$\hat{\nu}_x(f)=\int f(y)\nu_x(dy) \qquad \text{for} \quad f\in C(Y),$$

3) By Th 2. 1 of [3], $\nu^{1*}\equiv\nu^{2*}(\Pi)$ iff $\nu_x^{1*}(D)=\nu_x^{2*}(D)$ Π-a.e. for all $D\in\mathfrak{Y}$.

then for all $x \in X$, $\hat{\nu}_x(\cdot) \in \hat{C}(Y)_1^+$. Let D be a countable dense subspace (scalar multiple is of rational complex) in $C(Y)$. Put

$$(4.1) \qquad F_x(f) = \lim_N \frac{1}{N} \sum_{n=0}^{N-1} \hat{\nu}_{S^n x}(f) \qquad \text{for} \quad f \in D.$$

By the ergodic theorem the right hand side of (4.1) exists Π-a.e. Let X_D be the subset in $\mathfrak{X}$ on which (4.1) exists for all $f \in D$, then X_D is an S- invariant set satisfying $p(X \backslash X_D) = 0$ for all $p \in \Pi$. Since

$$(4.2) \qquad |F_x(f)| = \left| \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \hat{\nu}_{S^n x}(f) \right|$$

$$\leqq \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int |f(y)| \nu_{S^n x}(dy) \leqq \|f\|,$$

$F_x(\cdot)$, for each $x \in X_D$, is uniquely extended to a bounded linear functional $\tilde{F}_x(\cdot)$ on $C(Y)$. From the definition of $F_x(\cdot)$

$$F_{Sx}(f) = F_x(f) \qquad \text{for all } x \in X_D \text{ and } f \in D,$$

which implies

$$(4.3) \qquad \tilde{F}_{Sx}(f) = \tilde{F}_x(f) \qquad \text{for all } x \in X_D \text{ and } f \in C(Y).$$

By (4.1) $\tilde{F}_.(f)$ is an $\mathfrak{X}$-measurable function on X_D for each $f \in D$, and for every $f \in C(Y)$ there exists a sequence $\{f_n\}$ in D which converges uniformly to f. Therefore the equality

$$\tilde{F}_x(f) = \lim_{n \to \infty} F_x(f_n) \qquad \text{for all} \quad x \in X_D$$

implies the measurability of $\tilde{F}_.(f)$ on X_D. By the Riesz-Markov-Kakutani's theorem, there exists measures $q_x(\cdot)$ $(x \in X_D)$ on $\mathcal{Y}$ such that

$$\tilde{F}_x(f) = \int f(y) q_x(dy) \qquad \text{for all} \quad f \in C(Y).$$

As in the paper [4], denote $B(Y)$ as all bounded Borel functions on Y. Then a class

$$\left\{ g \in B(Y);\ \int g(y) q_0(dy) \text{ is } \mathfrak{X}\text{-measurable on } X_D \right\}$$

contains all functions in $C(Y)$, and is clearly a monoton class. And so the class just coincides to $B(Y)$. Let us choose an arbitrary T-invariant probability measure q_0 on $\mathcal{Y}$ and put

$$\nu_x^*(B) = \begin{cases} q_x(B) & \text{if} \quad x \in X_D, \\ q_0(B) & \text{if} \quad x \in X \backslash X_D \end{cases}$$

for all $B \in \mathcal{Y}$.

Then the above arguments show that ν^* is a stationary channel from X to Y. It follows from (4.1) and the ergodic theorem that

$$\iint f(y)\nu_x(dy)p(dx)=\iint f(y)\nu_x{}^*(dy)p(dx) \tag{4. 4}$$

for all $f\in D$ and S- invariant measures p on $\mathfrak{X}$, and by the approximation (4. 4) holds for all $f\in C(Y)$ and such p. As

$$p_G(A)=p(A\cap G)/p(G) \qquad (A\in\mathfrak{X})$$

is an S-invariant measure on $\mathfrak{X}$ for all $G\in\mathfrak{X}$ such that $S^{-1}G=G$ and $p(G)>0$,

$$\begin{aligned}\int_G \nu_x(B)p(dx)&=p(G)\int\nu_x(B)p_G(dx)\\ &=p(G)\int\nu_x{}^*(B)p_G(dx)=\int_G \nu_x{}^*(B)p(dx).\end{aligned} \tag{4. 5}$$

(4. 5) is valid even if $p(G)=0$, and so ν^* is a required average. Q.E.D.

The following theorem is our ultimate object.

THEOREM 4. 2. *There exists an ERG channel ν^0 and a mixing measure p_0 such that $r_0=r(p_0, \nu_0)$ is not ergodic.*

Proof. Let $X=\{0, 1\}^I$ be alphabet spaces of two alphabets, i.e. countable direct product over $I=(0, \pm 1, \pm 2, \cdots)$ of $\{0, 1\}$'s; let $\mathfrak{X}$ be the induced σ-algebra generated by the cylinder sets in X, and S be the shift transformation on X. We can assume X being compact metric space with Tychonov-product topology, and $\mathfrak{X}$ Borel fields of this topology. And S can be seen as homeomorphism on X. Let Π be all S-invariant probability measures on $\mathfrak{X}$. In order to construct the channel, take input as X and output Y same one; $X=Y$. (*E*.g., these are referred to [4].) The transformation T is taken as shift on Y which is same as S on X.

Now ν is a memoryless channel from X to Y determined by a stochastic matrix

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}. \tag{4. 6}$$

Then the channel ν is clearly ERC. (See [2] p. 153.)

By Theorem 4. 1, there exists an average ν^* of ν, and which is also ERG by Theorem 3. 2. Then $\nu^*=\nu^{**}$ together with Theorem 3. 1 shows that a channel

$$\nu^0=\frac{1}{2}\nu+\frac{1}{2}\nu^* \tag{4. 7}$$

is also ERG. By the form of the matrix (4. 6), for a Bernoulli-probability p_0 on X determined by a stochastic probability (1/2, 1/2), $r=r(p_0, \nu)$ is not a direct product measure. Hence Lemma 3.2 implies

$$r=r(p_0, \nu)\neq r(p_0, \nu^*)=r^*, \tag{4.8}$$

while (4.7) implies

$$r^0=r(p_0, \nu^0)=\frac{1}{2}r+\frac{1}{2}r^*. \tag{4.9}$$

Hence (4.8) and (4.9) show that r^0 is not ergodic. On the other hand, p_0 is mixing and ν^0 is ERG.

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, Inc. (1964).

[3] Nakamura, Y., A measure-theoretic construction for information theory. Kōdai Math. Sem. Rep. **21** (1969), 133–150.

[4] Umegaki, H., Representations and extremal properties of averaging operators and their application to information channels. Journ. Math. Analysis and Appl. **24** (1968), 41–73.

Department of Mathematics,
Tokyo Institute of Technology.