ISRAEL JOURNAL OF MATHEMATICS, Vol. 29, Nos. 2-3, 1978

THE INFORMATION COCYCLE AND ε-BOUNDED CODES

BY
WILLIAM PARRY

ABSTRACT

We shall prove here that Bowen's bounded codes lead to a cocycle-coboundary equation which can be exploited in various ways: through central limit theorems, through the related information variance or through a certain group invariant. Another result which emerges is that it is impossible to boundedly code two Markov automorphisms when one is of maximal type and the other is not. The functions which appear in the above cited cocycle-coboundary equation may belong to various L^p spaces. We devote a section to this problem. Finally we show that the information cocycle associated with small smooth partitions of a C^2 Anosov diffeomorphism preserving a smooth probability is, in a sense, canonical.

0. Introduction

In [1] Bowen introduced the idea of bounded codes between finite partitions and between two measure preserving transformations (or stationary finite state stochastic processes.) Using this concept he was able to prove that small smooth partitions of C^2 Anosov diffeomorphisms (preserving a smooth probability) are weak-Bernoulli. He was also able to show that certain Bernoulli automorphisms with the same entropy cannot boundedly code each other. (To do this the central limit theorem (or rather the variance of the central limiting distribution) was used as an invariant.) A similar idea was used in [3] by Fellgett and the present author to show that various measure preserving transformations are not "regularly isomorphic". The principal aid to the results of [3] was a cocycle-coboundary equation.

We shall prove here that bounded codes also lead to the same cocyle-coboundary equation which can be exploited in various ways: through central limit theorems, through the closely related information variance or through a certain group invariant. Another result which emerges is that it is impossible to

Received January 23, 1977

boundedly code two Markov automorphisms when one is of "maximal type" and the other is not.

In the cocycle-coboundary equation it is of some importance to know to which L^p spaces the various functions occurring belong. We devote a section to this problem. Finally we take up a result of Bowen's and show that small smooth partitions (for C^2 Anosov diffeomorphisms with a smooth probability) lead to canonical information cocyles in the sense that any two differ by a coboundary (of a function which lies in L^p for all $1 \leqq p < \infty$).

1. Bounded and ε-bounded codes

Throughout $(X, \mathcal{B}, m)$ will denote a (Lebesgue) probability space. The distance between two finite ordered partitions with the same number of elements is given by

$$d_0(\alpha, \beta) = \sum_{i=1}^{k} m(A_i \Delta B_i), \quad \alpha = A_1, \cdots, A_k, \quad \beta = B_1, \cdots, B_k.$$

If α is a finite or countable partition we write $\hat{\alpha}$ for the σ-algebra generated by α. The (unsymmetric) distance between a finite partition α and a sub-σ-algebra $\mathcal{C}$ is defined as

$$d(\alpha, \mathcal{C}) = \inf\{d_0(\alpha, \beta) : \beta \subset \mathcal{C}, \alpha, \beta \text{ given any order}\}.$$

If $\{A_i\}$, $\{B_i\}$ $i \in I$ are subsets of α, β then

$$\begin{aligned} m\Big(\bigcup_{i\in I} A_i \Delta \bigcup_{i \in I} B_i\Big) &= \sum_{i\in I} m(A_i) + \sum_{i \in I} m(B_i) - 2m\Big(\bigcup_{i\in I} A_i \cap \bigcup_{i\in I} B_i\Big) \\ &\leqq \sum_{i\in I} m(A_i) + \sum_{i\in I} m(B_i) - 2\sum_{i\in I} m(A_i \cap B_i) \\ &= \sum_{i \in I} m(A_i \Delta B_i). \end{aligned}$$

From this it is easy to see that

$$d(\alpha, \mathcal{C}) = \sup\{d(\alpha', \mathcal{C}) : \hat{\alpha}' \subset \hat{\alpha}\}.$$

Quite generally, for σ-algebras $\mathcal{A}, \mathcal{C}$ contained in $\mathcal{B}$ we define

$$\begin{aligned} d(\mathcal{A}, \mathcal{C}) &= \sup\{d(\alpha, \mathcal{C}) : \alpha \subset \mathcal{A}\} \\ &= \sup_{\alpha \subset \mathcal{A}} \inf_{\beta \subset \mathcal{C}} d_0(\alpha, \beta). \end{aligned}$$

If $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ are three sub-σ-algebras then we have the following triangle inequality:

$$d(\mathscr{A}_1,\mathscr{A}_3)\leqq d(\mathscr{A}_1,\mathscr{A}_2)+d(\mathscr{A}_2,\mathscr{A}_3).$$

To see this let α_1, α_2, α_3 be finite ordered partitions with the same number of elements, in $\mathscr{A}_1$, $\mathscr{A}_2$, $\mathscr{A}_3$ respectively. The following inequalities are proved sequentially:

$$d_0(\alpha_1,\alpha_3)\leqq d_0(\alpha_1,\alpha_2)+d_0(\alpha_2,\alpha_3),$$

$$d(\alpha_1,\mathscr{A}_3)\leqq d_0(\alpha_1,\alpha_2)+d(\alpha_2,\mathscr{A}_3)\leqq d_0(\alpha_1,\alpha_2)+d(\mathscr{A}_2,\mathscr{A}_3),$$

$$d(\alpha_1,\mathscr{A}_3)\leqq d(\alpha_1,\mathscr{A}_2)+d(\mathscr{A}_2,\mathscr{A}_3),$$

$$d(\mathscr{A}_1,\mathscr{A}_3)\leqq d(\mathscr{A}_1,\mathscr{A}_2)+d(\mathscr{A}_2,\mathscr{A}_3).$$

We note that

$$d(\mathscr{A},\mathscr{C})=0 \text{ if and only if } \mathscr{A}\subset\mathscr{C}.$$

Together with the triangle inequality this implies

$$d(\mathscr{A}_1,\mathscr{A}_3)\leqq d(\mathscr{A}_2,\mathscr{A}_3) \text{ when } \mathscr{A}_1\subset\mathscr{A}_2,$$

$$d(\mathscr{A}_1,\mathscr{A}_3)\leqq d(\mathscr{A}_1,\mathscr{A}_2) \text{ when } \mathscr{A}_2\subset\mathscr{A}_3.$$

If $\mathscr{A}_n\uparrow\mathscr{A}$ (i.e. $\mathscr{A}_1\subset\mathscr{A}_2\subset\cdots$ and $\bigcup_n\mathscr{A}_n$ generates $\mathscr{A}$) *then* $d(\mathscr{A}_n,\mathscr{C})\uparrow d(\mathscr{A},\mathscr{C})$. To see this, suppose $d<d(\mathscr{A},\mathscr{C})$; then there exists $\alpha\subset\mathscr{A}$ such that for all $\beta\subset\mathscr{C}$ (with the same number of elements as α) $d<d_0(\alpha,\beta)$. Given $d'>0$ we can choose n large enough so that there will exist $\alpha'\subset\mathscr{A}_n$ with $d_0(\alpha,\alpha')<d'$ and then

$$d_0(\alpha',\beta)\geqq d_0(\alpha,\beta)-d_0(\alpha,\alpha')>d-d' \quad \text{for all } \beta\subset\mathscr{C}.$$

Hence $d(\mathscr{A}_n,\mathscr{C})\geqq d-d'$, and $\lim_{n\to\infty}d(\mathscr{A}_n,\mathscr{C})\geqq d-d'$. Since d' is arbitrary we have $\lim_{n\to\infty}d(\mathscr{A}_n,\mathscr{C})\geqq d(\mathscr{A},\mathscr{C})$. The reverse inequality is obvious.

If γ is a finite partition with $\gamma\subset\mathscr{C}$ then

$$\begin{aligned}
d(\mathscr{A}\vee\hat{\gamma},\mathscr{C}) &= \sup\{d(\alpha,\mathscr{C}):\alpha\subset\mathscr{A}\vee\hat{\gamma}\}\\
&\leqq \sup\{d(\alpha\vee\gamma,\mathscr{C}):\alpha\subset\mathscr{A}\}\\
&= \sup\left\{\inf_\beta d_0(\alpha\vee\gamma,\beta):\alpha\subset\mathscr{A}\right\} \quad (\operatorname{card}\alpha\vee\gamma=\operatorname{card}\beta)\\
&\leqq \sup\left\{\inf_\beta d_0(\alpha\vee\gamma,\beta\vee\gamma):\alpha\subset\mathscr{A}\right\} \quad (\operatorname{card}\alpha=\operatorname{card}\beta)\\
&= \sup\left\{\inf_\beta d_0(\alpha,\beta):\alpha\subset\mathscr{A}\right\}\\
&= d(\mathscr{A},\mathscr{C})
\end{aligned}$$

and obviously $d(\mathscr{A},\mathscr{C}) \leqq d(\mathscr{A} \vee \hat{\gamma},\mathscr{C})$. Hence $d(\mathscr{A},\mathscr{C}) = d(\mathscr{A} \vee \hat{\gamma},\mathscr{C})$. By taking limits for a sequence $\hat{\gamma}_n \uparrow \mathscr{C}$ we have

$$d(\mathscr{A} \vee \mathscr{C},\mathscr{C}) = d(\mathscr{A},\mathscr{C}).$$

Let T be an automorphism (invertible measure preserving transformation) of $(X,\mathscr{B},m)$. If α,β are two finite partitions then β *ε-boundedly codes* α, with respect to T, if there exists k such that $d(\alpha_0^n,\beta_{-k}^{n+k}) \leqq \varepsilon$ for all $n = 0,1,\cdots$. Here we use the notation $\beta_{-k}^{l} = \bigvee_{i=-k}^{l} T^{-i}\beta$; we shall also write β^- (more correctly $\hat{\beta}^-$) for $\bigvee_{i=0}^{\infty} T^{-i}\hat{\beta}$ and β_T ($\hat{\beta}_T$) for $\bigvee_{i=-\infty}^{\infty} T^{-i}\hat{\beta}$.

If β ε-boundedly codes α then

$$d(\alpha_0^n, T^k\beta^-) = d\left(\alpha_0^n, \bigvee_{i=-k}^{\infty} T^{-i}\beta\right) \leqq \varepsilon \quad \text{for } n = 0,1,\cdots$$

and hence $d(\alpha^-, T^k\beta^-) \leqq \varepsilon$ (for some k).

According to Bowen [1], β *boundedly codes* α if for every $\varepsilon > 0$, β ε-boundedly codes α.

If α is a finite partition and $\mathscr{C}$ is a sub-σ-algebra then

$$I(\alpha \mid \mathscr{C}) = -\sum_{A \in \alpha} \chi_A \log m(A \mid \mathscr{C})$$

and more generally if $\mathscr{A}$ is another sub-σ-algebra *the information of $\mathscr{A}$ with respect to (or given) $\mathscr{C}$* is defined unambiguously (a.e.) as $I(\mathscr{A} \mid \mathscr{C}) = \lim_{n\to\infty} I(\alpha_n \mid \mathscr{C})$ when $\hat{\alpha}_n \uparrow \mathscr{A}$. The entropy of $\mathscr{A}$ with respect to $\mathscr{C}$ is $H(\mathscr{A} \mid \mathscr{C}) = \int I(\mathscr{A} \mid \mathscr{C})dm$. When $\mathscr{C}$ is the trivial σ-algebra $\mathscr{N}$ consisting of sets of measure zero and one we write $I(\mathscr{A} \mid \mathscr{N}) = I(\mathscr{A})$, $H(\mathscr{A} \mid \mathscr{N}) = H(\mathscr{A})$.

If α is a finite partition

$$I_T(\alpha) = I_T(\alpha^-) = I(\alpha \mid T^{-1}\alpha^-) = I(\alpha^- \mid T^{-1}\alpha^-)$$

is called the *information cocycle* of α or of α^-. (More generally if $\mathscr{A}$ is a sub-σ-algebra with $T^{-1}\mathscr{A} \subset \mathscr{A}$ then $I_T(\mathscr{A}) = I(\mathscr{A} \mid T^{-1}\mathscr{A})$ is the information cocycle of $\mathscr{A}$.)

The *entropy of T with respect to a finite partition* α is defined as $h(T,\alpha) = \lim_{n\to\infty}(1/n)\cdot H(\alpha_0^n) = \int I_T(\alpha)dm$.

The *entropy of T* is $h(T) = \sup_\alpha h(T,\alpha)$, where the supremum is taken over all finite partitions α. If α is a *finite generator* (i.e. α is finite and $\hat{\alpha}_T = \mathscr{B}$) then $h(T) = h(T,\alpha)$.

We shall assume familiarity with the basic properties of information and entropy which may be found in [10]. The basic identity which will be used frequently is

$$I(\mathscr{A}_1 \vee \mathscr{A}_2 \mid \mathscr{C}) = I(\mathscr{A}_1 \mid \mathscr{C}) + I(\mathscr{A}_2 \mid \mathscr{C} \vee \mathscr{A}_1).$$

2. The information cocycle

LEMMA 1. *If $\mathscr{C} \supset \mathscr{A}$ are sub-σ-algebras such that $d(\mathscr{C}, \mathscr{A}) < \frac{1}{2}$ then $I(\mathscr{C} \mid \mathscr{A})$ is finite on a set of positive measure. More precisely if n satisfies $d(\mathscr{C}, \mathscr{A})/(1 - e^{-n}) < \frac{1}{2}$ then*

$$m\{x : I(\mathscr{C} \mid \mathscr{A}) > n\} \leqq d(\mathscr{C}, \mathscr{A})\left(1 + \frac{1}{1 - e^{-n}}\right).$$

PROOF. Let $d(\mathscr{C}, \mathscr{A})/(1 - e^{-n}) < \frac{1}{2}$ and let $\beta \subset \mathscr{C}$.

$$\begin{aligned}
\{x : I(\beta \mid \mathscr{A}) > n\} &= \bigcup_{B \in \beta} B \cap \{-\log m(B \mid \mathscr{A}) > n\} \\
&= \bigcup_{B \in \beta} B \cap \{m(B \mid \mathscr{A}) < e^{-n}\} \\
&\subset \bigcup_i A_i \cap \{m(B_i \mid \mathscr{A}) < e^{-n}\} \\
&\quad \cup \bigcup_i (B_i - A_i) \cap \{x : m(B_i \mid \mathscr{A}) < e^{-n}\}
\end{aligned}$$

where $\alpha = \{A_i\}$ is chosen (and ordered) ($\alpha \subset \mathscr{A}$) so that $\Sigma_i m(A_i \Delta B_i) \leqq d(\mathscr{C}, \mathscr{A})$. The latter union has measure at most $\Sigma_i m(B_i - A_i) \leqq d(\mathscr{C}, \mathscr{A})$ whereas

$$\begin{aligned}
A_i \cap \{x : m(B_i \mid \mathscr{A}) < e^{-n}\} &= A_i \cap \{x : E(\chi_{B_i} - \chi_{A_i} \mid \mathscr{A}) + 1 < e^{-n}\} \\
&\subset A_i \cap \{x : E(|\chi_{B_i} - \chi_{A_i}| \mid \mathscr{A}) > 1 - e^{-n}\}.
\end{aligned}$$

Hence

$$mA_i \cap \{x : m(B_i \mid \mathscr{A}) < e^{-n}\} \leqq \frac{1}{1 - e^{-n}} \int E(\chi_{B_i \Delta A_i} \mid \mathscr{A}) dm$$

and

$$m \bigcup_i A_i \cap \{x : m(B_i \mid \mathscr{A}) < e^{-n}\} \leqq \frac{1}{1 - e^{-n}} \sum_i m(B_i \Delta A_i) \leqq \frac{d(\mathscr{C}, \mathscr{A})}{1 - e^{-n}} < \frac{1}{2}.$$

We conclude that

$$m\{x : I(\beta \mid \mathscr{A}) > n\} \leqq d(\mathscr{C}, \mathscr{A})\left(1 + \frac{1}{1 - e^{-n}}\right).$$

Now let $\beta_n \subset \mathscr{C}$ be a sequence of finite partitions with $\hat{\beta}_n \uparrow \mathscr{C}$ then

$$m\{x : I(\mathscr{C} \mid \mathscr{A}) > n\} \leqq d(\mathscr{C}, \mathscr{A})\left(1 + \frac{1}{1 - e^{-n}}\right) < 1.$$

A real valued function of the form $fT - f$ is called a *coboundary* (with respect to T); $fT - f$ is the coboundary of the function f. Two functions which differ by a coboundary are said to be *cohomologous.*

THEOREM 1.[†] *Let T be an ergodic automorphism of $(X, \mathscr{B}, m)$. If α, β are finite partitions such that β ε-boundedly codes α for some $0 \leqq \varepsilon < \frac{1}{2}$ then $I(\alpha^-/\beta^-)$ is finite a.e. and $I_T(\alpha \vee \beta)$, $I_T(\beta)$ are cohomologous.*

The elements of the partition β^- consist (mod 0) *of countable unions of elements of $\alpha^- \vee \beta^-$.*

PROOF. The last statement means that after the deletion of a suitable set of measure zero the description is valid with no qualification; the statement follows from the finiteness a.e. of $I(\alpha^-/\beta^-) = I(\alpha^- \vee \beta^-/\beta^-)$ (cf. [10]).

We have seen that $d(\alpha^-, T^k\beta^-) \leqq \varepsilon$ for some $k \geqq 0$ and hence $d(\alpha^- \vee T^k\beta^-, T^k\beta^-) \leqq \varepsilon$. From Lemma 1, $I(\alpha^- \vee \beta^-/T^k\beta^-)$ is finite on a set of positive measure.

We note that

$$I(\alpha^-/\beta^-) = I(T^{-k}\alpha^-/\beta^-) + I(\alpha^{k-1}/T^{-k}\alpha^- \vee \beta^-)$$

and since the latter function is integrable and $I(T^{-k}\alpha^-/\beta^-) = I(\alpha^-/T^k\beta^-) \circ T^k = I(\alpha^- \vee \beta^-/T^k\beta^-) \circ T^k$ we see that $I(\alpha^- \mid \beta^-)$ is finite on a set of positive measure.

$$\begin{aligned} I(\alpha^- \vee \beta^- \mid T^{-1}\beta^-) &= I(\beta \mid T^{-1}\beta^-) + I(\alpha^- \mid \beta^-) \\ &= I(\alpha^- \vee T^{-1}\alpha^- \vee \beta^- \mid T^{-1}\beta^-) \\ &= I(T^{-1}\alpha^- \mid T^{-1}\beta^-) + I(\alpha^- \vee \beta^- \mid T^{-1}\alpha^- \vee T^{-1}\beta^-) \\ &= I(\alpha^- \mid \beta^-) \circ T + I(\alpha \vee \beta \mid T^{-1}(\alpha^- \vee \beta^-)). \end{aligned}$$

$I(\beta \mid T^{-1}\beta^-)$ and $I(\alpha \vee \beta \mid T^{-1}\alpha^- \vee T^{-1}\beta^-)$ are integrable (in fact they belong to $L^p(X)$ for every $1 \leqq p < \infty$, as we shall see) and therefore $I(\alpha^- \mid \beta^-)$ is finite on a T invariant set of positive measure. Since T is ergodic, $I(\alpha^- \mid \beta^-)$ is finite a.e. Hence

$$I(\beta \mid T^{-1}\beta^-) - I(\alpha \vee \beta \mid T^{-1}\alpha^- \vee T^{-1}\beta^-) = I(\alpha^- \mid \beta^-) \circ T - I(\alpha^- \mid \beta^-)$$

i.e. $I_T(\beta)$, $I_T(\alpha \vee \beta)$ are cohomologous.

[†] The author has subsequently discovered that here and in the remainder of the paper the upper bound $\frac{1}{2}$ may be replaced by 2.

COROLLARY. *If T is ergodic and if α, β ε-boundedly code each other (for some $0 \leqq \varepsilon < \frac{1}{2}$) then their information cocycles are cohomologous.*

3. Central limits and other invariants

We now show that the central limiting distribution of

$$F_n(\alpha) = \frac{1}{\sqrt{n}}(I_T(\alpha) + \cdots + I_T(\alpha) \circ T^{n-1} - nh(T, \alpha)),$$

if it exists, is independent of α for all those α which ε-boundedly code each other $(0 \leqq \varepsilon < \frac{1}{2})$.

The corresponding statement for

$$G_n(\alpha) = \frac{1}{\sqrt{n}}(I(\alpha^n) - nh(T, \alpha))$$

was proved by Bowen [1] for partitions which boundedly code each other. We include a proof of Bowen's result also.

First note that if α, β ε-boundedly code each other where $0 \leqq \varepsilon < \frac{1}{2}$ then

$$I_T(\alpha) + \cdots + I_T(\alpha) \circ T^{n-1} - (I_T(\beta) + \cdots + I_T(\beta) \circ T^{n-1}) = f \circ T^n - f$$

for some finite valued f and $(fT^n - f)/\theta_n \to 0$ in measure whenever $\theta_n \to \infty$. Hence

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_T(\alpha) \circ T^i = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_T(\beta) \circ T^i$$

and $\int I_T(\alpha) dm = \int I_T(\beta) dm$. In other words $h(T, \alpha) = h(T, \beta)$.

It is now clear that $F_n(\alpha) - F_n(\beta) = (fT^n - f)/\sqrt{n}$ which tends to zero in measure. This is enough to ensure that $m\{x : F_n(\alpha) > t\}$ converges for all t if and only if $m\{x : F_n(\beta) > t\}$ converges for all t. The limiting function will necessarily be the same.

As for Bowen's result (when α, β ε-boundedly code each other for *all* $\varepsilon > 0$), let $0 < \varepsilon < \frac{1}{2}$, then

$$\begin{aligned}
\lim_{n \to \infty} m\left\{x : \frac{I(\alpha_0^n) - I(\beta_0^n)}{\sqrt{n}} > t\right\} &= \lim_{n \to \infty} m\left\{x : \frac{I(\alpha_0^n) - I(\beta_{-k}^{n+k})}{\sqrt{n}} > t\right\} \\
&\leqq \lim_{n \to \infty} m\left\{x : \frac{I(\alpha_0^n \vee \beta_{-k}^{n+k}) - I(\beta_{-k}^{n+k})}{\sqrt{n}} > t\right\} \\
&= \lim_{n \to \infty} m\left\{x : \frac{1}{\sqrt{n}} I(\alpha_0^n \mid \beta_{-k}^{n+k}) > t\right\} \\
&\leqq \lim_{n \to \infty} \varepsilon\left(1 + \frac{1}{1 - e^{-t\sqrt{n}}}\right) = \varepsilon.
\end{aligned}$$

Since ε can be chosen arbitrarily small

$$\lim_{n\to\infty} m\left\{x : \frac{I(\alpha_0^n) - I(\beta_0^n)}{\sqrt{n}} > t\right\} = 0.$$

Interchanging α and β we see that $(I(\alpha_0^n) - I(\beta_0^n))/\sqrt{n} \to 0$ in measure. This is enough to show that $m\{x : G_n(\alpha) > t\}$ converges for all t if and only if $m\{x : G_n(\beta) > t\}$ converges for all t and that the limiting function will be the same.

In many cases the limiting distribution of $F_n(\alpha)$ is the same as the limiting distribution of $G_n(\alpha)$. In fact the difference between $F_n(\alpha)$ and $G_n(\alpha)$ is

$$\frac{1}{\sqrt{n}}(I(\alpha_0^n) - I(\alpha_0^n \mid T^{-(n+1)}\alpha^-))$$

or

$$\lim_{k\to\infty} \frac{1}{\sqrt{n}}[(I(\alpha_0^n) - I(\alpha_0^n \mid T^{-(n+1)}\alpha^k))].$$

On cylinders $(x_0, \cdots, x_{n+k})$ the quantity inside the square brackets is

$$-\log\left(\frac{m(x_0, \cdots, x_n)m(x_{n+1}, \cdots, x_{n+k+1})}{m(x_0, \cdots, x_n, \cdots, x_{n+k+1})}\right).$$

When $m(x_0, \cdots, x_n)m(x_{n+1}, \cdots, x_{n+k+1})/m(x_0, \cdots, x_n, \cdots, x_{n+k+1})$ is bounded from above and below, as in the case of finite state Markov chains, for example, it is clear that $F_n(\alpha) - G_n(\alpha)$ will converge to zero in measure.

In any case it is clear that central limiting distributions are invariants of the relationships implied by ε-bounded codes ($0 \leqq \varepsilon < \frac{1}{2}$) and bounded codes. In the most important cases central limiting distributions are Gaussian so that the only invariant to be extracted is the variance. In many cases (cf. [4]) this will be

$$\sigma^2(T, \alpha) = \lim_{n\to\infty} \frac{1}{n}\int (I_T(\alpha) + \cdots + I_T(\alpha)\circ T^{n-1} - nh(T, \alpha))^2 dm$$

which was introduced in [3] as the *information variance.* (See [8] for a computation of this quantity when T is a finite state Markov chain and α is the canonical partition.)

Fellgett and the author showed that the Meshalkin examples [5] of Bernoulli automorphisms based on $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$ were not "regularly isomorphic"; Bowen proved that they do not boundedly code each other. $\sigma^2(T, \alpha)$ was (essentially) the invariant used in both cases.

Another invariant of the relation $I_T(\alpha) = I_T(\beta) + fT - f$ is the group (cf. [7])

$$\Lambda(T,\alpha)=\{(a,b)\in R\times R: e^{2\pi i(a+bI_T(\alpha))}=F\circ T/F$$

for some measurable $F: X\to K=\{z:|z|=1\}\}$.

In other words if α,β ε-boundedly code each other $(0\leqq\varepsilon<\frac{1}{2})$ then $\Lambda(T,\alpha)=\Lambda(T,\beta)$. $\Lambda(T,\alpha)$ is readily computable when T is a finite state Markov automorphism and α is the canonical (state) partition, since functions F appearing in the definition of $\Lambda(T,\alpha)$ are necessarily functions of one variable, i.e. $F(x)=F(x_0)$, where x_0 is the zero co-ordinate of x.

Two automorphisms T_1, T_2 with preferred partitions α_1,α_2 such that α_1,α_2 are generators (for T_1, T_2 respectively) are said to be *ε-bounded* (*bounded*) *equivalent or isomorphic* if there is an isomorphism ϕ $(\phi T_1=T_2\phi)$ such that $\alpha_1,\phi^{-1}\alpha_2$ ε-boundedly (boundedly) code each other.

Clearly these relationships are particularly relevant to finite state stationary processes. Again using the Λ invariant, it is easy to show that the Meshalkin examples are not ε-bounded equivalent $(0\leqq\varepsilon<\frac{1}{2})$. Implicitly the work of [7] shows that no two of the Markov automorphisms

$$\begin{pmatrix}pq\\pq\end{pmatrix},\ \begin{pmatrix}pq\\qp\end{pmatrix},\ \begin{pmatrix}qp\\pq\end{pmatrix}\quad(p\neq q)$$

are ε-bounded equivalent $(0\leqq\varepsilon<\frac{1}{2})$. (The invariant $\sigma^2(T,\alpha)$ is not sharp enough to distinguish these.)

THEOREM 2. *Let T_1, T_2 be Markov automorphisms (based on irreducible finite stochastic matrices) with T_2 of "maximal type". If T_1, T_2 are ε-bounded equivalent $(0\leqq\varepsilon<\frac{1}{2})$ then T_1 is of maximal type. More generally, if T_2 is only assumed to be an automorphism whose information cocycle is cohomologous to a constant, then T_1 is of maximal type.*

PROOF. Let T be defined by the stochastic matrix $P=\{P(i,j)\}$. T_2 is of maximal type means that its stochastic matrix has the form

$$\left\{\frac{\tau(i,j)}{\beta}\frac{\lambda_j}{\lambda_i}\right\},$$

where β is the maximum eigenvalue of a 0-1 irreducible matrix τ and $\Sigma_j\tau(i,j)\lambda_j=\beta\lambda_i$.

In this case it is easy to see that I_{T_2} is cohomologous to the constant $\log\beta$. More generally, supposing I_{T_2} is cohomologous to $\log\beta$ and T_1, T_2 ε-boundedly code each other $(0\leqq\varepsilon<\frac{1}{2})$, then I_{T_1} is cohomologous to $\log\beta$. But I_{T_1} is cohomologous to $-\log P(x_0,x_1)$ and therefore $-\log P(x_0,x_1)=\log\beta+fT_1-f$

for some finite valued f. Since $-\log P(x_0, x_1) - \log \beta$ is a function of the zero and first co-ordinates only it follows (cf, [7]) that $e^{2\pi i r f}$ is a function of the zero co-ordinate only for each real r. Hence f is a function of the zero co-ordinate only, i.e. $f(x) = f(x_0)$. Consequently $p(i,j) = (e^{-f(j)}/\beta e^{-f(i)})\sigma(i,j)$ where

$$\sigma(i,j) = \begin{cases} 1 & \text{when } p(i,j) > 0, \\ 0 & \text{when } p(i,j) = 0. \end{cases}$$

We see that p has the required form for T_1 to be of maximal type since β must be the maximal eigenvalue of the matrix σ.

COROLLARY. *Natural extensions of "β-transformations" are ε-bounded equivalent to Markov autormorphisms ($0 \leqq \varepsilon < \frac{1}{2}$) only when the latter are of maximal type.*

PROOF. For the definition of "β-transformations" cf. [9], [6]. These transformations have information cocycles which are cohomologous to constants.

4. The Lebesgue class of information functions

The basic equation we have been investigating reads

$$I(\alpha \mid T^{-1}\alpha^-) = I(\beta \mid T^{-1}\beta^-) + I(\alpha^- \mid \beta^-)\circ T - I(\alpha^- \mid \beta^-)$$

when α, β are finite partitions with $\hat{\alpha} \supset \hat{\beta}$ and when α, β ε-boundedly code each other ($0 < \varepsilon < \frac{1}{2}$). For some purposes it is clearly desirable to know what kind of functions appear here; more specifically, to which L^p spaces do they belong?

The following result is a simple extension of an estimate due to Chung [2] for the case $p = 1$.

PROPOSITION 1. *If α is a countable partition and if $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \cdots$ is an increasing sequence of σ-algebras then for $1 \leqq p < \infty$, p an integer,*

$$\int \sup_n (I(\alpha \mid \mathcal{B}_n))^p \leqq p! \sum_{A \in \alpha} m(A)\left[1 + |\log m(A)| + \cdots + \frac{|\log m(A)|^p}{p!}\right].$$

PROOF. Let $E_t^p = \{x : \sup_n |I(\alpha \mid \mathcal{B}_n)|^p > t\}$ then $\int \sup_n (I(\alpha \mid \mathcal{B}_n))^p dm = \int_0^\infty m(E_t^p)dt$. Using Chung's estimate for $p = 1$, $m(E_t^1) \leqq \sum_{A \in \alpha} \min(m(A), e^{-t})$, we have

$$m(E_t^p) = m(E_{t^{1/p}}^1) \leqq \sum_{A \in \alpha} \min(m(A), e^{-t^{1/p}}).$$

Hence

$$\int_0^\infty m(E_t^p)dt = \sum_{A\in\alpha}\int_0^\infty \min(m(A), e^{-t^{1/p}})dt$$

$$= \sum_{A\in\alpha}\left(m(A)|\log m(A)|^p + \int_{|\log m(A)|^p}^\infty e^{-t^{1/p}}dt\right).$$

$$= \sum_{A\in\alpha}\left(m(A)|\log m(A)| + \int_{|\log m(A)|}^\infty pe^{-u}u^{p-1}du\right).$$

If $J_p = \int_{|\log m(A)|}^\infty u^p e^{-u}du$ then a recurrence relation shows that

$$pJ_{p-1} = p!\,e^{-|\log m(A)|}\left(1 + |\log m(A)| + \cdots + \frac{|\log m(A)|^{p-1}}{(p-1)!}\right).$$

Hence

$$\int_0^\infty m(E_t^p)dt \leqq \sum_{A\in\alpha}(mA\,|\log m(A)|^p + pJ_{p-1})$$

$$= \sum_{A\in\alpha} p!\,m(A)\left|1 + |\log m(A)| + \cdots + \frac{|\log m(A)|^p}{p!}\right|.$$

Corollary. *If $I(\alpha)\in L^p(X)$ then $\sup_n I(\alpha\mid\mathcal{B}_n)\in L^p(X)$. If α is finite then $\sup_n I(\alpha\mid\mathcal{B}_n)\in L^p(X)$ for all $1\leqq p<\infty$ and, in particular, $I(\alpha\mid T^{-1}\alpha^-)\in L^p(X)$ for all $1\leqq p<\infty$.*

The estimate is not good enough to imply that $I(\alpha/T^{-1}\alpha^-)$ is bounded and, in fact, it can happen that $I(\alpha/T^{-1}\alpha^-)$ is unbounded when α is finite, as the following example shows.

Let $p_1, p_2, \cdots$ be a sequence of positive numbers with $\sum_{i=1}^\infty p_i = 1$, $s = \sum_{i=1}^\infty ip_i < \infty$ and $p_n/p_{n+m}\to 1$ for each m, e.g. $p_n = 1/Kn^3$, $K = \sum_{n=1}^\infty 1/n^3$. Let T be the Markov automorphism based on the stochastic matrix

$$\begin{pmatrix} p_1 & p_2 & p_3 & \cdots \\ 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \cdots \\ \cdot & \cdot\cdot & \cdots\cdot & \end{pmatrix}$$

The stationary initial probabilities are $\lambda_n = \sum_{i=n}^\infty p_i/s$. Let β be the canonical (countable) partition $\beta = [1],[2],\cdots$ where $[i] = \{x : x_0 = i\}$ and let $\alpha = [1],[1]^c$. $T^{-1}[1] = [2]\cup[1,1]$ and hence $[1]^c\cup T^{-1}[1] = [2]$; $T^{-2}[2] = [2]\cup[1,2]$ and hence $[1]^c\cup T^{-1}[2] = [3]$, etc. We conclude that $\alpha^-\geqq\beta\geqq\alpha$, i.e. $\alpha^- = \beta^-$.

Since $I(\alpha/T^{-1}\alpha^-) = I(\alpha^-/T^{-1}\alpha^-) = I(\beta^-/T^{-1}\beta^-) = I(\beta/T^{-1}\beta^-) = I(\beta/T^{-1}\beta)$ (β is the Markov canonical partition) it suffices to show that $I(\beta/T^{-1}\beta)$ is unbounded. It will then follow that the two set partition α has the property that $I(\alpha/T^{-1}\alpha^-)$ is unbounded but belongs to $L^p(X)$ for all $1 \leqq p < \infty$. But $I(\alpha/T^{-1}\alpha^-) = I(\beta/T^{-1}\beta)$ has the value $\log(\lambda_n/\lambda_1 p_n)$ on the cylinder $[1, n] = \{x : x_0 = 1, x_1 = n\}$ and $\lambda_n/\lambda_1 p_n = (\Sigma_{i=n}^{\infty} p_i)/p_n \to \infty$ since $p_{n+m}/p_n \to 1$ for each m.

We proceed now to an estimate for $\int I(\alpha \mid \beta)^p dm$ when $1 \leqq p < \infty$ and α, β are two ordered partitions with $d_0(\alpha, \beta) = d < (1 - e^{-p})^2$.

Let $\alpha = (A_1, A_2, \cdots, A_k)$ and $\beta = (B_1, B_2, \cdots, B_k)$. Evidently

$$\int I(\alpha \mid \beta)^p dm = \sum_{i,j} m(A_i \cap B_j) \left| \log \frac{m(A_i \cap B_j)}{m} \right|^p$$

$$= S_1 + S_2,$$

where

$$S_1 = \sum_{i \neq j} m(A_i \cap B_j) \left| \log \frac{m(A_i \cap B_j)}{m} \right|^p,$$

$$S_2 = \sum_{i} m(A_i \cap B_j) \left| \log \frac{m(A_i \cap B_j)}{m} \right|.$$

The function $x^{\frac{1}{2}} |\log x|^p$ has a maximum at e^{-2p} (with value $e^{-p} 2^p p^p$), is increasing for $0 < x < e^{-p}$ and decreasing for $e^{-p} < x < 1$. Therefore

$$S_1 \leqq \sum_{i \neq j} m(A_i \cap B_j)^{\frac{1}{2}} m(B_j)^{\frac{1}{2}} e^{-p}(2p)^p$$

$$\leqq e^{-p}(2p)^p \left(\sum_{i \neq j} m(A_i \cap B_j) \right)^{1/2} \left(\sum_{i \neq j} m(B_j) \right)^{1/2}$$

$$\leqq e^{-p}(2p)^p d^{\frac{1}{2}} \cdot k^{\frac{1}{2}}.$$

In order to estimate S_2 we shall use the function $x |\log x|^p$, which has a maximum at e^{-p} (with value $e^{-p} p^p$) and which increases in the range $0 < x < e^{-p}$ and decreases in the range $e^{-p} < x < 1$. Let

$$I = \left\{ i : \frac{m(A_i \cap B_j)}{m(B_j)} > 1 - \sqrt{d} > e^{-p} \right\} \quad \text{and}$$

$$I' = \{1, 2, \cdots, k\} - I.$$

For $i \in I'$, $m(B_i) - m(A_i \cap B_i) > \sqrt{d}\, m(B_i)$ and therefore

$$d \geqq \sum_{i \in I'} m(B_i) - m(A_i \cap B_i) > \sqrt{d}\, m\left(\bigcup_{i \in I'} B_i \right),$$

i.e. $m(\bigcup_{i \in I'} B_i) < \sqrt{d}$.

$$
\begin{aligned}
S_2 &= \sum_{i\in I} m(A_i \cap B_i)\left|\log \frac{m(A_i \cap B_i)}{m}\right|^p + \sum_{i\in I} m(A_i \cap B_i)\left|\log \frac{m(A_i \cap B_i)}{m}\right|^p \\
&\leqq \sum_{i\in I} m(B_i)(1-\sqrt{d})|\log(1-\sqrt{d})|^p + \sum_{i\in I'} m(B_i)e^{-p}\cdot p^p \\
&\leqq (1-\sqrt{d})\left(\frac{\sqrt{d}}{1-\sqrt{d}}\right)^p + e^{-p}p^p\sqrt{d} \\
&= d^{p/2}(1-\sqrt{d})^{p-1} + e^{-p}p^p d^{\frac{1}{2}} \\
&\leqq (1+e^{-p}p^p)d^{\frac{1}{2}}.
\end{aligned}
$$

In conclusion we have

$$\int I(\alpha \mid \beta)^p dm \leqq d^{\frac{1}{2}}[1 + e^{-p}p^p + k^{\frac{1}{2}}e^{-p}(2p)^p].$$

For the next estimate we have to contend with the fact that $x\,|\log x\,|^p$ is not concave if $p > 1$. (It is for $p = 1$.) In fact, $x\,|\log x\,|^p$ is concave in the interval $0 < x < e^{-(p-1)}$ and convex in the interval $e^{-(p-1)} < x < 1$. The line $y = m_p(1-x)$ meets $y = x\,|\log x\,|^p$ tangentially at some point (x_0, y_0) with $e^{-p} \leqq x_0 \leqq e^{-(p-1)}$, so that the function $\phi_p(x) = x\,|\log x\,|^p$ $(x \leqq x_0)$, $\phi_p(x) = m_p(1-x)$ $(x \geqq x_0)$ is concave in $[0,1]$, with maximum value $e^{-p}p^p$ at e^{-p}.

PROPOSITION 2. *For $p \geqq 1$, there exist constants K_p, K'_p such that, if α is a partition with k elements and if $\mathscr{C}$ is a sub-σ-algebra with $d(\alpha, \mathscr{C}) \leqq (1-e^{-p})^2$, then*

$$\int I(\alpha \mid \mathscr{C})^p dm \leqq (K_p + k^{\frac{1}{2}}K'_p)d(\alpha \mid \mathscr{C})^{\frac{1}{2}}.$$

PROOF. Let $d(\alpha \mid \mathscr{C}) < d$ so that there exists $\beta \subset \mathscr{C}$ with $d_0(\alpha, \beta) < d$.

$$
\begin{aligned}
\int I(\alpha \mid \mathscr{C})^p dm &= \int \sum_{A\in\alpha} m(A \mid \mathscr{C})\log^p m(A \mid \mathscr{C})dm \\
&\leqq \int \sum_{A\in\alpha} \phi_p(m(A \mid \mathscr{C}))dm \\
&= \sum_{A\in\alpha} E(\phi_p(m(A \mid \mathscr{C})) \mid \beta)dm \\
&\leqq \sum_{A\in\alpha} \phi_p(m(A \mid \beta))dm \\
&= \sum_{A\in\alpha}\sum_{B\in\beta} m(B)\phi_p\left(\frac{m(A\cap B)}{m(B)}\right) \\
&= S_1 + S_2,
\end{aligned}
$$

where

$$S_1 = \sum_{i \neq j} m(B_j)\phi_p\left(\frac{m(A_i \cap B_j)}{m(B_j)}\right) \quad \text{and} \quad S_2 = \sum_i m(B_i)\phi_p\left(\frac{m(A_i \cap B_i)}{m(B_i)}\right).$$

$$S_1 = \sum_{i \neq j} m(A_i \cap B_j)^{\frac{1}{2}} m(B_j)^{\frac{1}{2}} \left(\frac{m(B_j)}{m(A_i \cap B_j)}\right)^{1/2} \phi_p\left(\frac{m(A_i \cap B_j)}{m(B_j)}\right)$$

$$\leqq \sum_{i \neq j} m(A_i \cap B_j)^{\frac{1}{2}} m(B_j)^{\frac{1}{2}} K'_p,$$

where K'_p is the maximum of $e^{-p}(2p)^p$ and $m_p(1-x)/x^{\frac{1}{2}}$ in the range $x \geqq e^{-p}$, i.e.

$$K'_p \leqq \max(e^{-p}(2p)^p, m_p e^{p/2}).$$

Therefore

$$S_1 \leqq K'_p\left(\sum_{i \neq j} m(A_i \cap B_j)\right)^{1/2}\left(\sum_{i \neq j} m(B_j)\right)^{1/2},$$

i.e. $S_1 \leqq K'_p k^{\frac{1}{2}} d^{\frac{1}{2}}$.

$$S_2 = \sum_{i \in I} m(B_i)\phi_p\left(\frac{m(A_i \cap B_i)}{m(B_i)}\right) + \sum_{i \in I'} m(B_i)\phi_p\left(\frac{m(A_i \cap B_i)}{m(B_i)}\right)$$

$$\leqq \sum_{i \in I} m(B_i)\phi_p(1-\sqrt{d}) + \sqrt{d}e^{-p}p^p$$

$$\leqq \phi_p(1-\sqrt{d}) + e^{-p}p^p\sqrt{d} \quad (\text{when } 1-\sqrt{d} \geqq e^{-p})$$

$$\leqq m_p\sqrt{d} + e^{-p}p^p\sqrt{d}$$

$$= d^{\frac{1}{2}}(m_p + e^{-p}p^p).$$

With $K_p = m_p + e^{-p}p^p$ we have $\int I(\alpha \mid \mathscr{C})dm \leqq (K_p + k^{\frac{1}{2}}K'_p)d^{\frac{1}{2}}$, i.e. if $d(\alpha \mid \mathscr{C}) \leqq (1-e^{-p})^2$ then

$$\int I(\alpha \mid \mathscr{C})^p dm \leqq (K_p + k^{\frac{1}{2}}K'_p)d(\alpha \mid \mathscr{C})^{\frac{1}{2}}.$$

5. Smooth partitions

Bowen [1] has shown that if T is a C^2 Anosov diffeomorphism preserving a smooth probability then for arbitrary smooth partitions α there exist constants C, $0 < \lambda < 1$ such that if β is a finite measurable partition whose elements have small enough diameters then $d(\alpha_0^n, \beta_{-k}^{n+k}) \leqq C \cdot \lambda^k$ for all n, $k = 0, 1, 2, \cdots$.

In particular small smooth partitions boundedly code each other.

THEOREM 3. *If T is an automorphism of the Lebesgue space $(X, \mathcal{B}, m)$ and if α, β are finite partitions for which there exist C, $0 < \lambda < 1$ such that $d(\alpha_0^n, \beta_{-k}^{n+k}) \leqq C \cdot \lambda^k$ for all n, $k = 0, 1, \cdots$ then $I(\alpha^-/\beta^-) \in L^p(X)$ for all $1 \leqq p < \infty$.*

PROOF. We shall only need $d(\alpha, T^k\beta^-) \leqq C \cdot \lambda^k$ for $k = 0, 1, 2, \cdots$. We note that $I(\alpha^n/\beta^-) = I(\alpha \mid \beta^-) + \cdots + I(\alpha/T^n\beta^-)T^n$ and therefore

$$\|I(\alpha^n/\beta^-)\|_p \leqq \|I(\alpha/\beta^-)\|_p + \cdots + \|I(\alpha/T^n\beta^-)\|_p$$

$$\leqq \sum_{k=0}^{\infty} \|I(\alpha \mid T^k\beta^-)\|_p.$$

For fixed α and p we have seen that $\|I(\alpha \mid T^k\beta^-)\|_p \leqq Kd(\alpha \mid T^k\beta^-)^{\frac{1}{2p}}$ for some constant K when $d(\alpha \mid T^k\beta^-) \leqq C\lambda^k$ is small enough.

The first few terms of the series are finite by Proposition 1 and the remainder are dominated by a series whose kth term is $K \cdot C^{\frac{1}{2p}}\lambda^{k/2p}$. Hence the sequence $\|I(\alpha^n/\beta^-)\|_p$ is bounded ($n = 0, 1, \cdots$).

The Martingale theorem ensures that $I(\alpha^n \mid \beta^-)$ increases to $I(\alpha^- \mid \beta^-)^p$. We conclude that $I(\alpha^- \mid \beta^-)^p$ is integrable, i.e. $I(\alpha^- \mid \beta^-) \in L^p(X)$.

The import of this result is the following: Bowen proved that for the case of a C^2 Anosov diffeomorphism which preserves a smooth probability, small smooth partitions boundedly code each other. If we consider such a partition α then "to some extent" α^- is independent of α, i.e. α^- is almost canonical. Precisely, if β is another small smooth partition then α^-, β^- are closely related through $\alpha^- \vee \beta^-$ in that α^- (or β^-) has for its elements sets which (mod 0) are countable unions of elements of $\alpha^- \vee \beta^-$. α^- would consist of sets which are *finite* (in fact bounded) unions of elements of $\alpha^- \vee \beta^-$ if $I(\alpha^-/\beta^-) \in L^\infty(X)$—we have proved the next best thing. Corresponding to the close relationship between α^- and β^- we have the close relationship between $I(\alpha/T^{-1}\alpha^-)$ and $I(\beta/T^{-1}\beta^-)$ given by

$$I(\alpha/T^{-1}\alpha^-) = I(\beta/T^{-1}\beta^-) + f \circ T - f$$

where all functions here belong to $L^p(X)$ for all $1 \leqq p < \infty$.

REFERENCES

1. R. Bowen, *Smooth partitions of Anosov diffeomorphisms are weak Bernoulli*, Israel J. Math. **21** (1975), 95–100.

2. K. L. Chung, *A note on the ergodic theorem of information theory*, Ann. Math. Statist. **32** (1961), 612–614.

3. R. Fellgett and W. Parry, *Endomorphisms of a Lebesgue space II*, Bull. London Math. Soc. **7** (1975), 151–158.

4. I. A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters–Noordhoff, 1969.

5. L. Meshalkin, *A case of isomorphism of Bernoulli schemes*, Dokl. Akad. Nauk SSSR **128** (1959), 41–44.

6. W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hungar. **11** (1960), 401–416.

7. W. Parry, *Endomorphisms of a Lebesgue space III*, Israel J. Math **21** (1975), 167–172.

8. W. Parry and K. Schmidt, *A note on cocycles of unitary representations*, Proc. Amer. Math. Soc. **55** (1976), 185–190.

9. A. Renyi, *Representation for real numbers and their ergodic properties*, Acta Math. Acad. Sci. Hungar. **8** (1957), 477–493.

10. V. A. Rohlin, *Lectures on the entropy theory of transformations with invariant measure*, Uspehi Mat. Nauk **22** (5) 137 (1967), 3–56 (translated as Russian Math. Surveys **22** (5) (1967), 1–52).

Mathematics Institute
University of Warwick
Coventry, England