ENTROPY AND CONJUGACY

By Thomas A. Brown[1]

The RAND Corporation

0. Introduction. The notion of using entropy, a concept from information theory, to define a conjugacy invariant for measure-preserving transformations is due to Kolmogorov [2], [3]; it has been exploited by Sinai [5] and Rokhlin [4]. Halmos [1] has given a good treatment in English, in which he develops the entropy from its information theoretic origins. It is possible, however, to develop the theory of entropy as a conjugacy invariant of measure-preserving transformation in an elementary way without even mentioning information theory, and also without invoking powerful theorems such as McMillan's theorem or the martingale theorem of Doob, and that is what we shall do in the present paper.

1. Partitions. Let X be a measure space with measure m such that $m(X) = 1$. We define a *partition* $\mathfrak{A}$ to be a finite family of disjoint measurable sets which cover X. If every set in a partition $\mathfrak{A}$ is the union of sets in another partition $\mathfrak{B}$, then we call $\mathfrak{B}$ a *refinement* of $\mathfrak{A}$, and write $\mathfrak{A} \subset \mathfrak{B}$. The least common refinement of $\mathfrak{A}$ and $\mathfrak{B}$ is denoted by $\mathfrak{A} \vee \mathfrak{B}$. Clearly, $\mathfrak{A} \vee \mathfrak{B} = \{A \cap B \mid A \in \mathfrak{A}, B \in \mathfrak{B}\}$. If S is a measure-preserving transformation on X, then if we define $S\mathfrak{A} = \{SA \mid A \in \mathfrak{A}\}$, it follows that $S^{-1}(\mathfrak{A} \vee \mathfrak{B}) = S^{-1}\mathfrak{A} \vee S^{-1}\mathfrak{B}$.

We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *independent* if, for every $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, $m(A \cap B) = m(A)m(B)$.

2. The function L. Consider the continuous real-valued function L defined as follows:

$$L(t) = -t \log t \qquad 0 < t \leqq 1$$

$$= 0 \qquad t = 0.$$

The only properties of the function L that we shall use in most of the subsequent discussion are the following:

(a) The function L is continuous at zero, and $L(0) = 0$.

(b) The function L is *concave* in the sense that, if a, b, and $a + b$ are in the domain of L, then $L(a) + L(b) \geqq L(a + b)$.

Note that this is not quite the same thing as being concave downward in the sense of elementary calculus. However, a function (such as L) which has value zero at zero and negative second derivative throughout the unit interval (and

Received August 29, 1961; revised July 1, 1962.

[1] Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors.

thus is concave downward in the sense of elementary calculus) will be concave in our sense.

(c) The function L is *additive* in the sense that, if $\{a_i \mid i = 1, 2, \cdots, n\}$ and $\{b_j \mid j = 1, 2, \cdots, m\}$ are two sets of nonnegative real numbers such that $\sum_{i=1}^{n} a_i = 1$ and $\sum_{j=1}^{m} b_j = 1$, then $\sum_{i=1}^{n} L(a_i) + \sum_{j=1}^{m} L(b_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} L(a_i b_j)$. To prove that L has this property, one has merely to write out the right-hand sum, and use the fact that $\log(ab) = \log a + \log b$.

3. Entropy of a partition. Given any real-valued function f defined on the interval [0, 1], we can define the *mean entropy with respect to f* of a partition $\mathfrak{A}$ as follows: $\bar{H}_f(\mathfrak{A}) = \sum_{A \in \mathfrak{A}} f(m(A))$.

In case $f = L$, we call this simply the *mean entropy* of $\mathfrak{A}$, and write $\bar{H}(\mathfrak{A}) = \bar{H}_L(\mathfrak{A})$. Many of the results which follow are true for the mean entropy with respect to any concave f or with respect to any additive f. We shall indicate these results with a "con" or an "add" as the case may be.

LEMMA 1. ("*con*"): $\mathfrak{A} \subset \mathfrak{B}$ *implies* $\bar{H}(\mathfrak{A}) \leqq \bar{H}(\mathfrak{B})$.

PROOF. Each A in $\mathfrak{A}$ is the union of elements of $\mathfrak{B}$. Say $A_j = \bigcup_{i=1}^{n} B_{j,i}$. Thus $m(A_j) = \sum_{i=1}^{n} m(B_{j,i})$, and so by concavity it follows that $L(m(A_j)) \leqq \sum_{i=1}^{n} L(m(B_{j,i}))$. Thus

$$\bar{H}(\mathfrak{A}) = \sum_{j=1}^{m} L(m(A_j)) \leqq \sum_{j=1}^{m} \sum_{i=1}^{n} L(m(B_{j,i})) = \bar{H}(\mathfrak{B}).$$

LEMMA 2. ("*add*"): *If* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are independent, then* $\bar{H}(\mathfrak{A} \vee \mathfrak{B}) = \bar{H}(\mathfrak{A}) + \bar{H}(\mathfrak{B})$.

PROOF. Simply look at the definitions and apply the additive property of L.

LEMMA 3. *If* T *is a measure-preserving transformation on* X, *and is a partition, then* $T^{-1}\mathfrak{A}$ *is also a partition, and* $\bar{H}(T^{-1}\mathfrak{A}) = \bar{H}(\mathfrak{A})$.

PROOF. For any A in $\mathfrak{A}$, $m(T^{-1}A) = m(A)$ and the result follows.

4. Entropy of a transformation with respect to a partition. Let T be a measure-preserving transformation on X, and define the *entropy* $h(\mathfrak{A}, T)$ of T (with respect to the partition $\mathfrak{A}$) by the equation

$$h(\mathfrak{A}, T) = \limsup_{n \to \infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right).$$

We could also define the *entropy with respect to f* $h_f(\mathfrak{A}, T)$ by the equation

$$h_f(\mathfrak{A}, T) = \limsup_{n \to \infty} \frac{1}{n} \bar{H}_f\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right).$$

Obviously, $h(\mathfrak{A}, T) = h_L(\mathfrak{A}, T)$ by definition.

LEMMA 4. *If* S *is a measure-preserving transformation which commutes with* T, *then* $h(S^{-1}\mathfrak{A}, T) = h(\mathfrak{A}, T)$.

PROOF.

$$h(S^{-1}\mathfrak{A}, T) = \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}S^{-1}\mathfrak{A}\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} S^{-1}T^{-i}\mathfrak{A}\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(S^{-1}\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right) = h(\mathfrak{A}, T).$$

LEMMA 5. ("*con*"): *If* $\mathfrak{A} \subset \mathfrak{B}$, *then* $h(\mathfrak{A}, T) \leqq h(\mathfrak{B}, T)$.

PROOF. If $\mathfrak{A} \subset \mathfrak{B}$, then clearly $\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A} \subset \bigvee_{i=0}^{n-1} T^{-i}\mathfrak{B}$, and thus by Lemma 1 the result follows.

LEMMA 6. ("*con*"): *If* $\mathfrak{A} \subset \bigvee_{i=-N}^{N} T^{-i}\mathfrak{B}$ *for some* N, *and* T *is invertible, then* $h(\mathfrak{A}, T) \leqq h(\mathfrak{B}, T)$.

PROOF. Clearly, $\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A} \subset \bigvee_{i=-N}^{N+n-1} T^{-i}\mathfrak{B}$, and it follows by Lemma 1 that $\bar{H}(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}) \leqq \bar{H}(\bigvee_{i=-n}^{N+n-1} T^{-i}\mathfrak{B})$. Thus

$$h(\mathfrak{A}, T) = \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n+2N} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right)$$
$$\leqq \limsup_{n\to\infty} \frac{1}{n+2N} \bar{H}\left(\bigvee_{i=0}^{2N+n-1} T^{-i}T^{N}\mathfrak{B}\right)$$
$$= h(T^{N}\mathfrak{B}, T) = h(\mathfrak{B}, T).$$

LEMMA 7. ("*con*"): *If* k *is a positive integer, then* $h(\mathfrak{A}, T^k) \leqq kh(\mathfrak{A}, T)$.

PROOF. Let $\mathfrak{B} = \bigvee_{i=0}^{k-1} T^{-i}\mathfrak{A}$. Clearly, $\mathfrak{A} \subset \mathfrak{B}$, so

$$h(\mathfrak{A}, T^k) \leqq h(\mathfrak{B}, T^k) = \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} (T^k)^{-i}\mathfrak{B}\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-ki}\left(\bigvee_{j=0}^{k-1} T^{-j}\mathfrak{A}\right)\right)$$
$$= \limsup_{n\to\infty} \frac{1}{n} \bar{H}\left(\bigvee_{i=0}^{nk-1} T^{-i}\mathfrak{A}\right)$$
$$= k \limsup_{n\to\infty} \frac{1}{nk} \bar{H}\left(\bigvee_{i=0}^{nk-1} T^{-i}\mathfrak{A}\right)$$
$$\leqq kh(\mathfrak{A}, T).$$

The last inequality follows from the fact that the superior limit of a subsequence is always less than or equal to the superior limit of the sequence.

LEMMA 8. (*"con"*): *If T is an invertible measure-preserving transformation, then $h(\mathfrak{A}, T^k) \leqq |k|\, h(\mathfrak{A}, T)$ for all integers k.*

PROOF. We have already proved the result for positive k. If $k = 0$, it is natural and usual to take $T^0 = I$, the identity transformation.

Thus $\bar{H}(\bigvee_{i=0}^{n-1} I^{-i}\mathfrak{A}) = \bar{H}(\mathfrak{A})$, so $h(\mathfrak{A}, I) = \limsup_{n\to\infty} \frac{1}{n}\bar{H}(\mathfrak{A}) = 0$. If $k = -1$, note that

$$\bar{H}\left(\bigvee_{i=0}^{n-1} T^{+i}\mathfrak{A}\right) = \bar{H}\left(T^{n-1}\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right) = \bar{H}\left(\bigvee_{i=0}^{n-1} T^{-i}\mathfrak{A}\right);$$

thus $h(\mathfrak{A}, T^{-1}) = h(\mathfrak{A}, T)$, and the desired result follows.

5. Relative entropy. A *σ-field* (or simply *field*) is a collection of measurable sets closed under complementation and countable unions. A *minimal set* in a field is a set which contains no other set in the field except the empty set ϕ. The collection of all minimal sets of a finite field is clearly a partition. Conversely, the collection of unions of sets from a given partition (together with the set ϕ) form a finite field. Thus there is a natural one to one correspondence between the set of all finite fields over X and the set of all partitions of X.

Given a measurable set A and a field $\mathfrak{B}$, let $P(A/\mathfrak{B})$ denote the essentially unique function on X which is measurable with respect to $\mathfrak{B}$, and such that $\int_B P(A/\mathfrak{B}) = m(A \cap B)$ for all $B \in \mathfrak{B}$.

Thus if $B = \{X, \phi\}$, $P(A/\mathfrak{B})$ is the constant $m(A)$. If $\mathfrak{B}$ is finite, then let $\mathfrak{B}'$ denote the corresponding partition, and we have

$$P(A/\mathfrak{B}) = \sum_{B\in\mathfrak{B}'} [m(A \cap B)/m(B)]\chi(B)$$

where $\chi(B)$ is the characteristic function of B. Note that if A is a set in the field $\mathfrak{B}$, then $P(A/\mathfrak{B}) = \chi(A)$. Now we define the *mean entropy of a partition $\mathfrak{A}$ with respect to a field $\mathfrak{B}$ as follows*:

$$\bar{H}(\mathfrak{A}/\mathfrak{B}) = \sum_{A\in\mathfrak{A}} \int_X L(P(A/\mathfrak{B}))$$

If $\mathfrak{B} = (X, \phi)$, we simply write this as $\bar{H}(\mathfrak{A})$. Clearly this definition agrees with the one given in Section 3. Since $\chi(A) = 0$ or 1 at each point in X, and since $L(0) = L(1) = 0$, it follows that if every set in $\mathfrak{A}$ is in the field $\mathfrak{B}$, then $\bar{H}(\mathfrak{A}/\mathfrak{B}) = 0$. The following computation, which depends strongly on the special nature of the function L, is self-explanatory:

$$\begin{aligned}
\bar{H}(\mathfrak{A} \vee \mathfrak{B}') &= \sum_{A\in\mathfrak{A},B\in\mathfrak{B}'} L(m(A \cap B)) \\
&= \sum_{A\in\mathfrak{A},B\in\mathfrak{B}'} L(m(A \cap B)) - \sum_{B\in\mathfrak{B}'} L(m(B)) + \sum_{B\in\mathfrak{B}'} L(m(B) \\
&= -\sum_{A\in\mathfrak{A},B\in\mathfrak{B}'} m(A \cap B)[\log (m(A \cap B)) - \log (m(B))] + \bar{H}(\mathfrak{B}') \\
&= \sum_{A\in\mathfrak{A},B\in\mathfrak{B}'} L(m(A \cap B)/m(B))m(B) + \bar{H}(\mathfrak{B}') \\
&= \bar{H}(\mathfrak{A}/\mathfrak{B}) + \bar{H}(\mathfrak{B}').
\end{aligned}$$

The equation $\bar{H}(\mathfrak{A} \vee \mathfrak{B}') = \bar{H}(\mathfrak{A}/\mathfrak{B}) + \bar{H}(\mathfrak{B}')$ implies (by an argument similar to that used in Lemma 6) that for any integer k, $h(\mathfrak{A}, T) \leqq h(\mathfrak{B}', T) + \bar{H}(\mathfrak{A}/\bigvee_{i=-k}^{+k} T^i\mathfrak{B})$.

Now we come to the sticky question. Let $\mathfrak{C}_k = \bigvee_{i=-k}^{+k} T^i\mathfrak{B}$, and $\mathfrak{C} = \bigvee_{i=-\infty}^{+\infty} T^i\mathfrak{B}$, (i.e., the σ-field generated by all the $T^i\mathfrak{B}$). Suppose every element of $\mathfrak{A}$ is in the field $\mathfrak{C}$. Then we know that $\bar{H}(\mathfrak{A}/\mathfrak{C}) = 0$. Does it follow that by choosing k large, we can make $\bar{H}(\mathfrak{A}/\mathfrak{C}_k)$ as small as we please? We could invoke the deep and powerful martingale theorem of Doob at this point, but in fact it is not necessary to do so: we shall prove the result we need directly. Let A be some element of $\mathfrak{A}$ such that $m(A) \neq 0$. Since the union of all the $\mathfrak{C}_k$ is dense in $\mathfrak{C}$, for any small $\epsilon > 0$ we can find a set C in $\mathfrak{C}_k$, such that $m((A \cup C) - (A \cap C)) < \epsilon \min\{m(A), m(C), 1 - m(C)\}$.

Let C' denote the complement of C. Then

$$1 - [m(A \cap C)/m(C)] < \epsilon$$

and

$$[m(A \cap C')/m(C')] < \epsilon,$$

and therefore

$$\begin{aligned}\int_X L(P(A/\mathfrak{C}_k)) &= \sum_{B \varepsilon \mathfrak{C}_k'} - m(A \cap B) \log \frac{m(A \cap B)}{m(B)} \\ &\leqq -m(A \cap C) \log \frac{m(A \cap C)}{m(C)} - m(A \cap C') \log \frac{m(A \cap C')}{m(C')} \\ &< -\epsilon \log \epsilon .\end{aligned}$$

Repeating the process for each $A \varepsilon \mathfrak{A}$, and taking k sufficiently large, we can make $\bar{H}(\mathfrak{A}/\mathfrak{C}_k)$ as small as we please. Thus we obtain the following lemma.

LEMMA 9. (proved only for entropy defined with L): *If* $\mathfrak{A} \subset \bigvee_{i=-\infty}^{+\infty} T^i\mathfrak{B}$, *then* $h(\mathfrak{A}, T) \leqq h(\mathfrak{B}, T)$.

6. Entropy of a transformation. We define the *entropy* h of a measure-preserving transformation T by $h(T) = \sup h(\mathfrak{A}, T)$ where the supremum is taken over all partitions of X. Two measure-preserving transformations S and T are said to be *conjugate* if there exists an invertible measure-preserving transformation R such that $S = RTR^{-1}$. It is easy to see that, for any partition $\mathfrak{A}$,

$$h(\mathfrak{A}, RTR^{-1}) = h(R\mathfrak{A}, T).$$

It follows that $h(RTR^{-1}) = h(T)$; thus we say that the entropy h of a transformation is a conjugacy invariant.

LEMMA 10. (proved only for entropy defined with L): *If* T *is invertible, and* $\mathfrak{A}$ *has the property that for every partition* $\mathfrak{B}$, $\mathfrak{B} \subset \bigvee_{i=-\infty}^{+\infty} T^i\mathfrak{A}$, *then* $h(\mathfrak{A}, T) = h(T)$.

PROOF. Corollary to Lemma 9.

LEMMA 11. ("*con*"): *If* k *is a positive integer, then* $h(T^k) = kh(T)$.

PROOF. Refer to the proof of Lemma 7. We have

$$h(\mathfrak{B}, T^k) = k \limsup_{n\to\infty} \frac{1}{nk} \bar{H}\left(\bigvee_{i=0}^{nk-1} T^{-i}\mathfrak{A}\right) \leqq kh(\mathfrak{A}, T).$$

Actually we may replace the inequality sign by equality, since $\bar{H}(\bigvee_{i=0}^{nk-1} T^{-i}\mathfrak{A})$ is a monotone increasing sequence, and thus, if n is large,

$$(1/nk - j)\bar{H}(\bigvee_{i=0}^{nk-1} T^{-i}\mathfrak{A}).$$

So we know that $h(\mathfrak{B}, T^k) = kh(\mathfrak{A}, T)$, and thus $h(T^k) \geqq kh(T)$. On the other hand, by Lemma 7, $h(\mathfrak{A}, T^k) \leqq kh(\mathfrak{A}, T)$, so $h(T^k) \leqq kh(T)$. Thus $h(T^k) = kh(T)$.

7. An application. Consider the set Y of all sequences $\{a_i\}$ (where i ranges over all integers, positive and negative) of integers mod k. We call a subset of Y a *cylinder* if it consists of all sequences which have specified entries at certain components. For example, the set of all sequences such that $a_0 = 1$ and $a_3 = 0$ is a cylinder. Assign the quantity $1/k^m$ to each cylinder which involves the specification of m components (thus the cylinder specified above is assigned $1/k^2$), and extend this function in the obvious way to the field generated by the collection of all cylinders. Thus we have converted Y into a measure space. Let T be the transformation which carries each sequence $\{a_i\}$ into the sequence whose ith component is a_{i-1}. The transformation T is clearly an invertible measure-preserving transformation. It is called the *k-shift*. Since the set Y can be considered as the set of points in the unit square by letting $\{a_i\}$ correspond to the point $(.a_0a_1a_2 \cdots, .a_{-1}a_{-2}a_{-3} \cdots)$, where the coordinates are in k-ary fractions, it follows that all the k-shifts can be thought of as invertible measure-preserving transformations of the unit square.

For a long time it was an unsolved problem whether or not the various k-shifts are conjugate to each other. The entropy invariant solved this problem. For let T be the k-shift. Let $\mathfrak{A}$ be the partition of Y into k equal subsets according to the quantity in the zero component. Then any other partition $\mathfrak{B}$ will be contained in $\bigvee_{i=-\infty}^{+\infty} T^i\mathfrak{A}$. Thus by Lemma 10, $h(T) = h(\mathfrak{A}, T)$. The partitions $\mathfrak{A}, T^{-1}\mathfrak{A}, \cdots, T^{-n}\mathfrak{A}$ are all independent, and thus

$$\bar{H}\left(\bigvee_{i=0}^{n} T^{-i}\mathfrak{A}\right) = \sum_{i=0}^{n} \bar{H}(T^{-i}\mathfrak{A}) = n\bar{H}(\mathfrak{A})$$

$$= n \sum_{i=1}^{k} -\frac{1}{k} \log \frac{1}{k} = -n \log \frac{1}{k} = n \log k.$$

It follows that

$$h(T) = \limsup_{n\to\infty} \frac{1}{n} \cdot n \log k = \log k.$$

So every different k-shift is in a different conjugacy class!

In fact, if T is a k_1-shift and S is a k_2-shift, with k_1 and k_2 relatively prime, it

is easy to show that no power of T can be conjugate to a power of S. For if T^m were conjugate to S^n, it would follow that $m \log k_1 = n \log k_2$, which implies $k_1^m = k_2^n$. Thus no power of the two-shift is conjugate to a power of the three-shift.

REFERENCES

[1] HALMOS, P. R. (1959). Entropy in ergodic theory, (mimeographed lecture notes). University of Chicago.

[2] KOLMOGOROV, A. N. (1958). A new metric invariant of transient dynamical systems and automorphisms in Lebesque spaces, *Dokl. Akad. Nauk SSSR* **119** 861–864.

[3] KOLMOGOROV, A. N. (1959). Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR* **124** 754–755.

[4] ROKHLIN, V. A. (1959). Entropy of metric automorphisms. *Dokl. Akad. Nauk SSSR* **124** 980–983.

[5] SINAI, YA (1959). On the concept of entropy for a dynamic system, *Dokl. Akad. Nauk SSSR* **124** 768–771.