CHAPTER I: INTRODUCTION

1. MOTIVATION

The problems discussed in these notes are motivated by the classical isomorphism problem of ergodic theory, which has received much attention since the introduction (by Kolmogorov) of entropy into the subject, and especially since Ornstein's celebrated solution for Bernoulli automorphisms. According to Kolmogorov, two Bernoulli automorphisms S, T (shifts on product spaces $(X, m) = \prod_{-\infty}^{\infty} (X_0, m_0)$ and $(Y, p) = \prod_{-\infty}^{\infty} (Y_0, p_0)$, respectively, where (X_0, m_0) and (Y_0, p_0) are finite probability spaces) are isomorphic only if their entropies $(-\sum_{i\in X_0} m_0(i) \log m_0(i)$ and $-\sum_{j\in Y_0} p_0(j) \log p_0(j))$ coincide. Ornstein's result ([O.1], [O.2]) complements this statement: if two Bernoulli automorphisms have the same entropy then they are isomorphic. This means that there is an essentially invertible measure-preserving transformation ϕ from almost all of X onto almost all of Y satisfying $\phi S = T\phi$ a.e. Writing $\phi(x) = (\phi_n(x))$ in terms of component functions ϕ_n, we see that $\phi_n(x) = \phi_0(S^n x)$ and ϕ is determined by $\phi_0 : X \to Y_0$. The present, $\phi_0(x)$, may depend on the entire past and future of x; this is not satisfactory from the point of view of communication and coding.

We have indicated that isomorphisms, which preserve only the most basic structures of measure theory, may be too weak for applications. It is therefore appropriate to ask for measure-preserving transformations which respect structures other than just measure e.g. a sub-σ-algebra which represents the 'past' or a partition (finite or countable) which represents the 'state space' of a stationary stochastic process. Also, finite state processes have natural topologies so that it is natural to consider continuous maps and homeomorphisms between them. These considerations lead us to the concepts of regular isomorphism, quasi-regular isomorphism, finitary code, block code, finite equivalence, almost topological conjugacy and topological conjugacy. A more detailed preview of the ideas involved in these will be given in the final section of this chapter.

In many of the above classifications an information function can be used as an invariant. Indeed, the question of the extent to which information, as opposed to entropy, can be used in classification problems is a principal motivation for many of the results presented in these notes.

2. BASIC DEFINITIONS AND CONVENTIONS

As is usual in measure theory, all measure theoretic objects which are equal almost everywhere are identified throughout these notes. Thus the qualification 'almost everywhere' is often omitted. We always require sets and functions to be measurable, even when this is not explicitly specified. $(X, \mathcal{B}, m)$ or a similar triple always denotes a probability space. If $\mathcal{A} \subset \mathcal{B}$ is a sub-σ-algebra and $\mathcal{A}_n$ is a sequence of sub-σ-algebras such that $\bigcup_n \mathcal{A}_n$ generates $\mathcal{A}$ then we write $\mathcal{A}_n \uparrow \mathcal{A}$.

For probability spaces $(X_i, \mathcal{B}_i, m_i)$ $(i = 1, 2)$, a measure-preserving transformation ϕ from almost all of X_1 onto almost all of X_2 is called a <u>homomorphism</u> and in keeping with the above convention we write $\phi : X_1 \to X_2$. Measure-preserving means $\phi^{-1}\mathcal{B}_2 \subset \mathcal{B}_1$ and $m_1(\phi^{-1}B) = m_2(B)$ for all $B \in \mathcal{B}_2$. If $\phi : X_1 \to X_2$ is a homomorphism, $(X_2, \mathcal{B}_2, m_2)$ is called a <u>homomorphic image</u> or <u>factor</u> of $(X_1, \mathcal{B}_1, m_1)$ and $(X_1, \mathcal{B}_1, m_1)$ an <u>extension</u> of $(X_2, \mathcal{B}_2, m_2)$. ϕ is called an <u>isomorphism</u> if there is a homomorphism $\psi : X_2 \to X_1$ such that $\psi\phi = \mathrm{id}_{X_1}$ and $\phi\psi = \mathrm{id}_{X_2}$ (a.e.). A homomorphism (isomorphism) from a probability space to itself is called an <u>endomorphism</u> (<u>automorphism</u>). An endomorphism T of $(X, \mathcal{B}, m)$ is <u>ergodic</u> if $T^{-1}B = B$, $B \in \mathcal{B}$ implies $mB = 0$ or 1.

A probability space which is isomorphic to a subinterval of $[0, 1]$ (with Lebesgue measure) together with a countable number of atoms is called a <u>Lebesgue space.</u> All of our spaces will be Lebesgue. This is not a very restrictive condition; for instance a complete separable metric space together with a complete Borel probability always defines a Lebesgue space. The main reasons for imposing this condition are that Lebesgue spaces are separable and that the following holds (see [R.1]):

Let $(X_1, \mathcal{B}_1, m_1)$ and $(X_2, \mathcal{B}_2, m_2)$ be Lebesgue spaces. Every σ-algebraic homomorphism $\Phi : \mathcal{B}_2 \to \mathcal{B}_1$ $(\Phi(B^C) = (\Phi B)^C$, $\Phi(\bigcup_n B_n) = \bigcup_n \Phi(B_n)$, $m_1(\Phi B) = m_2 B$ for all $B, B_n \in \mathcal{B}_2$ and $\Phi(X_2) = X_1)$ is realised, essentially uniquely, by a homomorphism $\phi : X_1 \to X_2$ in the sense that $\Phi B = \phi^{-1} B$ for all $B \in \mathcal{B}_2$.

For $i = 1, 2$ let T_i be an endomorphism of $(X_i, \mathcal{B}_i, m_i)$. A homomorphism $\phi : X_1 \to X_2$ satisfying $\phi T_1 = T_2 \phi$ is called a homomorphism of T_1 to T_2; T_2 is then a factor of T_1 and T_1 an extension of T_2. (We write $T_1 \overset{\phi}{\to} T_2$.) If such a homomorphism is an isomorphism, then T_1 and T_2 are said to be isomorphic or conjugate.

Let T be an endomorphism of $(X, \mathcal{B}, m)$. There is a 1-1 correspondence between T-invariant sub-σ-algebras $\mathcal{A}$ $(T^{-1}\mathcal{A} \subset \mathcal{A} \subset \mathcal{B})$ and factors of T: If $T' : (X', \mathcal{B}', m') \to (X', \mathcal{B}', m')$ is a factor of T by ϕ then $T^{-1}(\phi^{-1}\mathcal{B}') \subset \phi^{-1}\mathcal{B}' \subset \mathcal{B}$ and, on the other hand, if $T^{-1}\mathcal{A} \subset \mathcal{A} \subset \mathcal{B}$ then there exists a factor (by some ϕ) $T' : (X', \mathcal{B}', m') \to (X', \mathcal{B}', m')$, unique up to isomorphism and called the factor with respect to $\mathcal{A}$, such that $\mathcal{A} = \phi^{-1}\mathcal{B}'$. Moreover there is an automorphism $\hat{T}$ of a (Lebesgue) space $(\hat{X}, \hat{\mathcal{B}}, \hat{m})$, unique up to isomorphism and called the natural extension of T, such that $\hat{T}$ is an extension of T (by ϕ, say) and $\hat{T}^n \phi^{-1}\mathcal{B} \uparrow \hat{\mathcal{B}}$ (see [R. 3]).

If T is an automorphism of $(X, \mathcal{B}, m)$, an invariant sub-σ-algebra $\mathcal{A}$ with $T^n\mathcal{A} \uparrow \mathcal{B}$ is called exhaustive. An automorphism of a Lebesgue space together with a preferred exhaustive sub-σ-algebra is called a process.

3. PROCESSES

Let X_0 be a finite or countable set, and let $X = \prod_{n=-\infty}^{\infty} X_n$ where $X_n = X_0$ for all $n \in \mathbb{Z}$. The shift T is defined $(Tx)_n = x_{n+1}$ for $x = (x_n) \in X$. A cylinder is a set of the form

$$[i_0, i_1, \ldots, i_l]^k = \{x = (x_n) \in X : x_k = i_0, x_{k+1} = i_1, \ldots, x_{k+l} = i_l\}$$

where $k, l \in \mathbb{Z}$ and $l \geq 0$. If $k = 0$, we sometimes omit this superscript. Let $\mathcal{B}$ be the σ-algebra generated by all cylinder sets. Clearly $T^{-1}\mathcal{B} = \mathcal{B}$. If m is a T-invariant probability on $(X, \mathcal{B})$ then $(X, \mathcal{B}, m, T)$, or simply T, is called a countable or finite state process according to the cardinality of X_0, the state space.

Consider an automorphism T of a Lebesgue space $(X, \mathcal{B}, m)$. A finite or countable (measurable) partition α is called a generator if $\bigcup_{n=-\infty}^{\infty} T^n\alpha$ generates $\mathcal{B}$. If $\mathcal{B}$ is even generated by $\bigcup_{n=0}^{\infty} T^{-n}\alpha$, α is called a strong generator. The state space X_0 of a countable or finite state process defines a generator called the state partition: $\{[i] : i \in X_0\}$. In fact there is no significant difference between automorphisms with specified generators and (countable or finite) state processes, as the exercise below shows. This rests on the fact that we are restricting attention to Lebesgue spaces, since we need the following: If $(\mathcal{P}_n)$ is a sequence of partitions of a Lebesgue space $(X, \mathcal{B}, m)$ such that $\bigcup_n \mathcal{P}_n$ generates $\mathcal{B}$, then there is a null set N such that for all $x, y \in X - N$ there exist n and $B \in \mathcal{P}_n$ with $x \in B$, $y \in X - B$ (see [R. 1]).

Exercise. Let T be an automorphism of a Lebesgue space $(X, \mathcal{B}, m)$ and let $\alpha = \{A_1, A_2, \ldots\}$ be a finite or countable generator. Let $X_0 = \{1, 2, \ldots\}$ have the same cardinality as α and define $X' = \prod_{-\infty}^{\infty} X_0$. Define a probability μ on X' by

$$\mu[i_0, i_1, \ldots, i_l]^k = m(A_{i_0} \cap T^{-1}A_{i_1} \cap \ldots \cap T^{-l}A_{i_l}).$$

Note that the shift S preserves μ, and prove that S and T are isomorphic.

Given a countable or finite state process $(X, \mathcal{B}, m, T)$, let $\mathcal{A}$ be the 'past' sub-σ-algebra generated by the cylinders $[i_0, \ldots, i_l]^k$ with $k \geq 0$. Then $\mathcal{A}$ is T-invariant and exhaustive. $\mathcal{A}$ is called the standard past, and $(X, \mathcal{B}, m, T)$ is understood to have $\mathcal{A}$ as its preferred exhaustive sub-σ-algebra.

4. MARKOV CHAINS

In this section we list some definitions and facts concerning non-negative matrices and Markov chains. The basic references are [S'] and [F].

Let A be a non-negative $k \times k$ matrix. A is irreducible if for each pair (i, j), $1 \leq i, j \leq k$, we can find $n \geq 1$ such that the product A^n has $A^n(i, j) > 0$. The period of (a state) i $(1 \leq i \leq k)$ is the highest common factor of $n \geq 1$ with $A^n(i, i) > 0$. If A is irreducible, this is independent of the state chosen and is called the period of A. A is called aperiodic if (it is irreducible and) it has period 1; this is equivalent to requiring $A^n > 0$ for some $n \geq 1$.

A non-negative matrix is stochastic if all its row sums equal 1. A probability vector is a strictly positive vector with sum 1. Given an irreducible stochastic $k \times k$ matrix P, there is a unique probability vector p with $pP = p$. Hence we obtain a finite state process $(X, \mathcal{B}, m, T)$ by taking $X = \prod_{-\infty}^{\infty} \{1, 2, \ldots, k\}$ and defining

Cambridge University Press
978-0-521-28794-4 - Classification Problems in Ergodic Theory
William Parry and Selim Tuncel
Excerpt
More information

$$m[i_0, i_1, \ldots, i_l]^n = p(i_0)\,P(i_0, i_1)\,P(i_1, i_2)\,\ldots\,P(i_{l-1}, i_l).$$

$(X, \mathcal{B}, m, T)$, together with its state partition, is called the Markov chain defined by P. The (shift) automorphism T and the measure m are also called Markov. If P has identical rows and $pP = p$ as above, the unique probability vector p is easily seen to be the one giving the rows of P so that m is the product measure obtained from the measure on $\{1, \ldots, k\}$ which assigns $p(i)$ to i. In this case the process, shift and measure are called Bernoulli.

If P is an irreducible stochastic $k \times k$ matrix of period t and p is the unique probability vector with $pP = p$ then for each pair (i, j), $1 \leq i, j \leq k$, there exists $0 \leq t_0 \leq t-1$ such that $\lim_{n\to\infty} P^{t_0+nt}(i, j) = tp(j)$ and $P^m(i, j) = 0$ if m is not of the form $t_0 + nt$. This shows that all Markov automorphisms are ergodic. The convergence of $P^{t_0+nt}(i, j)$ is exponentially fast.

We will need:

Perron-Frobenius Theorem [S']. Let A be a non-negative irreducible matrix with period t. Then

(i) there is a positive eigenvalue β with a corresponding strictly positive eigenvector;

(ii) β is a simple eigenvalue (i.e. it is a simple root of the characteristic equation of A);

(iii) $\beta\omega^i$, $i = 0, 1, \ldots, t-1$ are eigenvalues where ω is a primitive t^{th} root of 1, and for all other eigenvalues α, $|\alpha| < \beta$;

(iv) if $r = (r_1, \ldots, r_k)$ is a strictly positive vector then

$$\min_{1\leq i\leq k} \left\{\frac{(Ar)_i}{r_i}\right\} \leq \beta \leq \max_{1\leq i\leq k} \left\{\frac{(Ar)_i}{r_i}\right\}$$

with equality on either side implying equality throughout. In particular $Ar = \alpha r$, r strictly positive, implies that $\alpha = \beta$.

5. REDUCED PROCESSES AND TOPOLOGICAL MARKOV CHAINS

We shall consider all finite state shift spaces, whenever necessary, as topologized in the following natural manner: If $X = \prod_{-\infty}^{\infty} X_0$ where X_0 is a finite set, we give X_0 the discrete topology and X the product topology. Then X is compact and metrizable and the shift is a homeomorphism. The closed-open subsets of X are the cylinder sets and finite unions of these. Cylinders form a base for the topology of X and so, generate the Borel σ-algebra. Moreover, X is zero-dimensional (it has a base consisting of open-closed sets).

Suppose $(X, \mathcal{B}, m, T)$ is a finite state process. Then $\mathcal{B}$ is the Borel σ-algebra. Let X' be the support of m i.e. $X' = X - U$ where U is the largest open null subset of X. X' is T-invariant. We denote by $\mathcal{B}'$, m', T' the restrictions of $\mathcal{B}$, m, T to X'. X' is compact and zero-dimensional and T' is a homeomorphism. $(X', \mathcal{B}', m', T')$ is the <u>reduced process.</u> Measure theoretically there is no distinction between a finite state process and its reduced process and, since supporting measures are more convenient for topological considerations, we shall always assume that <u>finite state processes are reduced.</u>

Suppose A is a 0-1 irreducible $k \times k$ matrix. A defines a closed, shift-invariant subset X of $\prod_{-\infty}^{\infty}\{1, 2, \ldots, k\}$:

$$X = \{x = (x_n) : A(x_n, x_{n+1}) = 1 \text{ for all } n\} .$$

X, together with (the restriction of) the shift, is called the <u>topological Markov chain</u> or <u>subshift of finite type</u> defined by A. If P is a $k \times k$ (irreducible) stochastic matrix compatible with A (i.e. $P(i, j) = 0$ iff $A(i, j) = 0$), the Markov measure defined by P has X as its support. Our assumption that finite state processes are reduced means that we regard Markov measures as being defined on their supporting topological Markov chains. The period of a topological Markov chain is defined according to the period of its defining matrix A.

The 0-1 matrix A may be viewed as a matrix of transitions: we have k vertices and a transition from i to j is allowed iff $A(i, j) = 1$. The topological Markov chain given by A is the space of all doubly infinite sequences of (allowable) transitions. From this point of view, there is no reason to restrict ourselves to 0-1 matrices; given an irreducible non-negative integral $k \times k$ matrix A' we again have k vertices and $A'(i, j)$ specifies the number of paths from i to j. The topological Markov chain now consists of all doubly infinite sequences of directed paths, with the shift transformation. This, however, is no more general than the 0-1 case since we may index with the directed paths a 0-1 matrix (transition from path a to path b is allowed iff b starts at the terminal vertex of a) which gives an equivalent (topologically conjugate) topological Markov chain in the sense that there exists a homeomorphism between the two spaces which conjugates the shifts. We shall work with 0-1 matrices, resorting to general non-negative matrices only to save space in examples.

[A. M.] contains a beautiful exposition of topological Markov chains.

6. INFORMATION AND ENTROPY

In this section we give a brief review of the basics of information and entropy theory. Details may be found in [B], [R. 2] and [W. 1], but [P. 1] is perhaps the best reference for our purposes.

If α is a countable partition of the (Lebesgue) space $(X, \mathcal{B}, m)$ and $\mathcal{C} \subset \mathcal{B}$ is a sub-σ-algebra, the conditional information of α given $\mathcal{C}$ is

$$I(\alpha|\mathcal{C}) = -\sum_{A\in\alpha} \chi_A \log m(A|\mathcal{C}) .$$

All logarithms are to the base e. $H(\alpha|\mathcal{C}) = \int I(\alpha|\mathcal{C})\,dm$ is the conditional entropy of α given $\mathcal{C}$. If $\mathcal{C}$ is the trivial σ-algebra consisting of sets of measure 0 and 1, we have, respectively, the information and entropy of α, $I(\alpha)$ and $H(\alpha)$. Note that $I(\alpha|\mathcal{C}) \geq 0$ and that $H(\alpha|\mathcal{C}) = 0$ (or $I(\alpha|\mathcal{C}) = 0$) iff α consists of sets in $\mathcal{C}$.

For a countable partition α, we shall use the same symbol to denote the σ-algebra generated by the partition, the distinction being clear from the context. For partitions α, β we put $\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$. We use similar notation for the refinement of any number of partitions. When $\mathcal{C}_1, \mathcal{C}_2, \ldots$ is a sequence of σ-algebras, $\bigvee_{n=1}^{\infty} \mathcal{C}_n$ denotes the σ-algebra generated by their union.

The basic identities for information and entropy are:

$$I(\alpha \vee \beta|\gamma) = I(\alpha|\beta \vee \gamma) + I(\beta|\gamma)$$

$$H(\alpha \vee \beta|\gamma) = H(\alpha|\beta \vee \gamma) + H(\beta|\gamma)$$

for countable partitions α, β, γ. These are easily verified. If $\mathcal{C} \subset \mathcal{B}$ is a σ-algebra we can find (finite) partitions γ_n such that $\gamma_n \uparrow \mathcal{C}$, since we are in a Lebesgue (therefore separable) space. Now using γ_n in place of γ in the basic identities, taking limits, and using the increasing Martingale theorem we obtain,

$$I(\alpha \vee \beta|\mathcal{C}) = I(\alpha|\beta \vee \mathcal{C}) + I(\beta|\mathcal{C}),$$

$$H(\alpha \vee \beta|\mathcal{C}) = H(\alpha|\beta \vee \mathcal{C}) + H(\beta|\mathcal{C})$$

for countable partitions α, β and a sub-σ-algebra $\mathcal{C}$. From these identities we

Cambridge University Press
978-0-521-28794-4 - Classification Problems in Ergodic Theory
William Parry and Selim Tuncel
Excerpt
More information

see that $I(\alpha|\mathcal{C}) \geq I(\beta|\mathcal{C})$, $H(\alpha|\mathcal{C}) \geq H(\beta|\mathcal{C})$ when $\alpha \geq \beta$ (i.e. when $\alpha \supset \beta$ as σ-algebras generated by the partitions). That $H(\alpha|\mathcal{A}) \leq H(\alpha|\mathcal{C})$ when $\mathcal{A} \supset \mathcal{C}$ may be proved with the aid of Jensen's inequality, but the corresponding inequality for information is not generally true.

Suppose $\mathcal{A}$, $\mathcal{C}$ are sub-σ-algebras and let $\alpha_n \uparrow \mathcal{A}$, $\beta_n \uparrow \mathcal{A}$ where α_n, β_n are finite partitions (for $n = 1, 2, \ldots$). Then

$$I(\beta_m|\mathcal{C}) \leq I(\alpha_n \vee \beta_m|\mathcal{C}) = I(\alpha_n|\mathcal{C}) + I(\beta_m|\alpha_n \vee \mathcal{C})$$

and, letting $n \to \infty$,

$$\lim_{n\to\infty} I(\alpha_n|\mathcal{C}) + I(\beta_m|\mathcal{A} \vee \mathcal{C}) \geq I(\beta_m|\mathcal{C})$$

so that, since $\beta_m \subset \mathcal{A}$,

$$\lim_{n\to\infty} I(\alpha_n|\mathcal{C}) \geq I(\beta_m|\mathcal{C}).$$

Letting $m \to \infty$,

$$\lim_{n\to\infty} I(\alpha_n|\mathcal{C}) \geq \lim_{m\to\infty} I(\beta_m|\mathcal{C}) .$$

Clearly the reverse inequality is also true, so that the definition

$$I(\mathcal{A}|\mathcal{C}) = \lim_{n\to\infty} I(\alpha_n|\mathcal{C}) \quad \text{for} \quad \alpha_n \uparrow \mathcal{A}$$

is unambiguous. Note that $\lim_{n\to\infty} I(\alpha_n|\mathcal{C})$ exists in $\mathbb{R} \cup \{\infty\}$ as $I(\alpha_n|\mathcal{C})$ are increasing (for increasing α_n). Again we define $H(\mathcal{A}|\mathcal{C}) = \int I(\mathcal{A}|\mathcal{C})\,dm$. It should be clear that the following sharpened versions of the basic identities are valid:

$$I(\mathcal{A}_1 \vee \mathcal{A}_2|\mathcal{C}) = I(\mathcal{A}_2|\mathcal{A}_1 \vee \mathcal{C}) + I(\mathcal{A}_1|\mathcal{C}),$$

$$H(\mathcal{A}_1 \vee \mathcal{A}_2|\mathcal{C}) = H(\mathcal{A}_2|\mathcal{A}_1 \vee \mathcal{C}) + H(\mathcal{A}_1|\mathcal{C})$$

for sub-σ-algebras $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{C}$. Evidently, $H(\mathcal{A}_1|\mathcal{C}) = H(\mathcal{A}_1 \vee \mathcal{A}_2|\mathcal{C})$ when $\mathcal{A}_2 \subset \mathcal{C}$ and $H(\mathcal{A}|\mathcal{C}) = 0$ iff $\mathcal{A} \subset \mathcal{C}$. When $\mathcal{A} \supset \mathcal{C}_1 \supset \mathcal{C}_2$,

$$I(\mathcal{A}|\mathcal{C}_1) \leq I(\mathcal{A}|\mathcal{C}_2) \quad \text{since}$$

$$I(\mathcal{A}|\mathcal{C}_2) = I(\mathcal{A} \vee \mathcal{C}_1|\mathcal{C}_2) = I(\mathcal{C}_1|\mathcal{C}_2) + I(\mathcal{A}|\mathcal{C}_1) .$$

If T is an endomorphism of $(X, \mathcal{B}, m)$ and α is a countable partition with $H(\alpha) < \infty$, then $h(T, \alpha) = H(\alpha | \bigvee_{i=1}^{\infty} T^{-i}\alpha)$ is called the entropy of T with respect to α. By the Shannon-McMillan-Breiman theorem,

$$\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\alpha) = h(T, \alpha) \qquad (\text{when } H(\alpha) < \infty).$$

The entropy of T is defined by

$$h(T) = \sup \{h(T, \alpha) : \alpha \text{ is a countable partition with } H(\alpha) < \infty\}.$$

Since we are working with Lebesgue spaces,

$$h(T) = \sup \{h(T, \alpha) : \alpha \text{ is a finite partition}\}.$$

If α_n are such that $H(\alpha_n) < \infty$ and $\alpha_n \uparrow \mathcal{B}$ then $h(T) = \lim_{n\to\infty} h(T, \alpha_n)$. If α (with $H(\alpha) < \infty$) is a strong generator or, when T is an automorphism, a generator, then $h(T) = h(T, \alpha)$. The last two results are the main practical tools for the calculation of entropy.

If ϕ is a homomorphism from $(X_1, \mathcal{B}_1, m_1)$ to $(X_2, \mathcal{B}_2, m_2)$ then $I(\alpha|\mathcal{C}) \circ \phi = I(\phi^{-1}\alpha|\phi^{-1}\mathcal{C})$ whenever α is a partition of X_2 and $\mathcal{C} \subset \mathcal{B}_2$ is a σ-algebra. It follows that $H(\alpha|\mathcal{C}) = H(\phi^{-1}\alpha|\phi^{-1}\mathcal{C})$ and that $h(T_1) \geq h(T_2)$ when T_2 is a factor of T_1. Thus, $h(T_1) = h(T_2)$ when T_1 and T_2 are isomorphic.

Having established entropy as an isomorphism invariant, we conclude the section by listing some results as exercises.

Exercise. Let $(X, \mathcal{B}, m)$ be a Lebesgue space and let $\mathcal{A} \subset \mathcal{B}$ be a σ-algebra. Show that there exists a sequence of finite partitions α_n with $\alpha_n \uparrow \mathcal{A}$.

Exercise. Let T be an endomorphism of $(X, \mathcal{B}, m)$. Show that $h(T^n) = n\,h(T)$ for $n = 0, 1, 2, \ldots$ and that $h(T^{-1}) = h(T)$ when T is invertible.

Exercise. For $i = 1, 2$, let T_i be an endomorphism of $(X_i, \mathcal{B}_i, m_i)$. Show that $h(T_1 \times T_2) = h(T_1) + h(T_2)$.

Exercise. If T is the Bernoulli automorphism given by the probability vector $(p(1), \ldots, p(n))$, show that $h(T) = -\sum_{i=1}^{n} p(i) \log p(i)$.

Exercise. If T is the Markov automorphism given by the irreducible stochastic matrix P with invariant probability vector p ($pP = p$), show that $h(T) = -\sum_{i,j} p(i)P(i, j)\log P(i, j)$.

Exercise. Let α, β be finite partitions of $(X, \mathcal{B}, m)$. Show that $H(\alpha \vee \beta) = H(\alpha) + H(\beta)$ iff α and β are independent (i.e. iff $m(A \cap B) = m(A) \cdot m(B)$ for all $A \in \alpha$, $B \in \beta$).

Exercise. Let α, β, γ be finite partitions of $(X, \mathcal{B}, m)$. For $B \in \beta$ denote by (B, m_B) the normalization of the restriction of m to B with $m_B(E) = \frac{m(E \cap B)}{m(B)}$. By using the last exercise on each of (B, m_B), show that $H(\alpha|\beta \vee \gamma) = H(\alpha|\beta)$ iff $\frac{m(A \cap B \cap C)}{m(B \cap C)} = \frac{m(A \cap B)}{m(B)}$ for all $A \in \alpha$, $B \in \beta$, $C \in \gamma$.

7. TYPES OF CLASSIFICATION

As we have already indicated, these notes are mainly concerned with various classifications of processes (in particular, Markov chains) and of topological Markov chains. This section is intended as a preview of the basic definitions of these classifications.

Let $(X_i, \mathcal{B}_i, \mathcal{A}_i, m_i, T_i)$, where $T_i^{-1}\mathcal{A}_i \subset \mathcal{A}_i \subset \mathcal{B}_i$ and $T_i^n\mathcal{A}_i \uparrow \mathcal{B}_i$, be processes ($i = 1, 2$). An isomorphism $T_1 \overset{\phi}{\to} T_2$ is said to be regular if $\phi^{-1}\mathcal{A}_2 \subset T_1^p\mathcal{A}_1$ and $\phi\mathcal{A}_1 \subset T_2^p\mathcal{A}_2$ for some integer $p \geq 0$. The idea is that the code ϕ (and its inverse) should depend, perhaps on the entire past but, only on a bounded amount of the future. The main result for regular isomorphisms is that the information functions $I(\mathcal{A}_1|T_1^{-1}\mathcal{A}_1)$ and $I(\mathcal{A}_2|T_2^{-1}\mathcal{A}_2) \circ \phi$ are related by an equation. We exploit this equation in various ways to obtain invariants. The same equation holds, in a slightly weaker form, for quasi-regular isomorphisms. Quasi-regular isomorphisms are defined by insisting that the pasts of the processes should not be too distant from each other in a sense we shall make precise later.

If $(X, \mathcal{B}, m, T)$ and $(X', \mathcal{B}', m', T')$, $X = \prod_{-\infty}^{\infty} X_0$ and $X' = \prod_{-\infty}^{\infty} X_0'$, are countable state processes, a homomorphism $T \overset{\phi}{\to} T'$ is completely determined by a function $\phi_0 : X \to X_0'$ ($\phi(x) = \{\phi_0(T^n x)\}$). For a point $x \in X$ the present, $\phi_0(x)$, may depend on the entire past and future of x. If we require for each $x \in X$ that $\phi_0(x)$ is determined by a finite section of x, we have a