The theory of Bernoulli Shifts

by Paul C. Shields
Professor Emeritus of Mathematics
University of Toledo, Toledo, OH 43606

Send corrections to bshifts@earthlink.net

Web Edition 1.01
Reprint of 1973 University of Chicago Press edition
With some error correction

Thanks to Andrew Shields for typing assistance.

PREFACE to this Web Edition 1.01

The University of Chicago Press decided to take this book out of print and kindly returned the copyright to me. I arranged for my son, Andrew, to retype the manuscript as a LaTeX document. The result is uncopyrighted and is available to those who wish to use it without permission. The format has been changed, including page numbering; some errors have been corrected; and there have been some changes in wording in a few places. The theorem and lemma numbering have remained the same. Figures were redone by me to fit into the LaTeX framework.

Please send any corrections and comments to bshifts@earthlink.net.

The first edition was dedicated to my mother and father, this edition is also.

03/10/03. Seattle, WA

PREFACE to 1973 edition.

There are many measure spaces isomorphic to the unit interval with Lebesgue measure, hence there are many ways to describe measure-preserving transformations on such spaces. For example, there are translations and automorphisms of compact metric groups, shifts on sequence spaces (such as those induced by stationary processes), and flows arising from mechanical systems. It is a natural question to ask when two such transformations are isomorphic as measure-preserving transformations. Such concepts as ergodicity and mixing and the study of unitary operators induced by such transformations have provided some rather coarse answers to this isomorphism question.

The first major step forward on the isomorphism quesion was the introduction by Kolmogorov in 1958-59 of the concept of entropy as an invariant for measure-preserving transformation. In 1970, D. S. Ornstein introduced some new approximation concepts which enabled him to establish that entropy was a complete invariant for a class of transformations known as Bernoulli shifts. Subsequent work by Ornstein and others has shown that a large class of transformations of physical and mathematical interest are isomorphic to Bernoulli shifts.

These lecture notes grew out of my attempts to understand and use these new results about Bernoulli shifts. Most of the material in these notes is concerned with the proof that two Bernoulli shifts with the same entropy are isomorphic. This proof makes use of a number of simple ideas about partitions and approximation by periodic transformations. These are carefully presented in Chapters 2-6. The basic results about entropy are sketched in Chapters 7-8. Ornstein's Fundamental Lemma is proved in Chapter 9. This enables one to construct partitions with perfect distribution and entropy close to those which are almost perfect, and is the key to obtaining the isomorphism theorem in Chapter 10. Chapters 11-13 contain extensions of these results, while Chapter 1 contains a summary of the measure theory used in these notes. For a more complete account of recent extensions of these ideas, the reader is referred to D. S. Ornstein's forthcoming notes ([42]).

I am particularly grateful to D. S. Ornstein, who introduced me to most of the ideas in this book. I also wish to thank R. L. Adler, N. A. Friedman, Y. Katznelson, R. McCabe, and B. Weiss for many helpful converstions, and R. Newman, who drew most of the pictures. Thanks are also due to James England, Robert Field, Richard Lacey, Douglas Lind, Stephen Polit, and Michael Steele, who read the original manuscript with great care, correcting numerous errors and giving many suggestions for improvement. The manuscript was typed by Elizabeth Plowman. Special thanks are due her for her patience and care.

This work was supported in part by NSF grants GP 33581X and GJ 776.

CHAPTER 1.
LEBESGUE SPACES

This chapter describes the properties of spaces isomorphic to the unit interval that will be used, frequently without reference, in the sequel.

Our measure spaces (X, Σ, μ) will always be assumed to be finite, complete spaces; that is, $\mu(X)$ is finite and Σ contains all subsets of sets of measure zero. If $\mu(X) = 1$, the space (X, Σ, μ) is called a probability space. All sets and functions are assumed (or must be shown to be) Σ-measurable and our measure spaces are assumed to be probability spaces, unless stated otherwise. Equality is taken to mean "equality mod zero"; for example, two sets A,B are equal if $\mu(A \triangle B) = 0$ where $A \triangle B$ is the symmetric difference $(A - B) \cup (B - A)$.

An isomorphism ϕ of (X^1, Σ^1, μ^1) onto (X^2, Σ^2, μ^2) is a mapping $\phi : X^1 \to X^2$ such that

$$\phi(\Sigma^1) \subseteq \Sigma^2, \quad \phi^{-1}(\Sigma^2) \subseteq \Sigma^1,$$

$$\mu^2(\phi(A_1)) = \mu^1(A_1),\ A_1 \in \Sigma_1;\ \mu^1(\phi^{-1}(A_2)) = \mu^2(A_2),\ A_2 \in \Sigma_2,$$

and ϕ is one-to-one and onto (mod zero), that is, there are sets $\tilde{X}^i \subseteq X^i$ such that $\mu^i(X^i - \tilde{X}^i) = 0$, and ϕ is a one-to-one map of $\tilde{X}^1$ onto $\tilde{X}^2$.

A space isomorphic to the unit interval with Lebesgue sets and Lebesgue measure will be called a *Lebesgue space*. We sketch here Rohlin's characterization of such spaces (see [18]).

A collection $\mathcal{E} \subseteq \Sigma$ *separates* X if there is a set $E \in \Sigma$, $\mu(E) = 0$, such that if $x, y \notin E$, there is a set $A \in \mathcal{E}$ such that $x \in A$, $y \notin A$ or $x \notin A$, $y \in A$. A collection $\mathcal{E}$ *generates* Σ if Σ is the smallest complete σ-algebra containing $\mathcal{E}$. A countable collection $\mathcal{E} = \{A_n\}$ is *complete* in X if each intersection $\cap_{n=1}^{\infty} B_n$ is nonempty where, for each n, B_n is either A_n or A_n^c. If we let A_n be the set of all x in the unit interval such that the n^{th} digit in the binary expansion of x is 0, then it is easy to see that $\{A_n\}$ is a complete, separating, generating collection for the unit interval. The unit interval is also *nonatomic*; that is, any set of positive measure contains subsets of smaller positive measure. The space (X, Σ, μ) is a subspace of $(\overline{X}, \overline{\Sigma}, \overline{\mu})$ if $X \in \overline{\Sigma}$, Σ consists of the sets $A \cap X$, $A \in \bar{\Sigma}$, and $\mu(A) = \bar{\mu}(A)$, $A \in \Sigma$. We then have the result

THEOREM 1.1. A probability space (X, Σ, μ) is a Lebesgue space if and only if it is a subspace of a probability space $(\bar{X}, \bar{\Sigma}, \bar{\mu})$ which has a complete, separating, generating sequence.

This theorem provides us with a large collection of Lebesgue spaces. For example, if X is a compact metric space, μ is a regular nonatomic Borel problility measure, and Σ is the completion of the Borel sets with respect to μ, then (X, Σ, μ) is a Lebesgue space. Also, if (X, Σ, μ) is a Lebesgue space, and if $X^1 \in \Sigma$,

$\mu(X^1) > 0$, then (X^1, Σ^1, μ^1) is a Lebesgue space where $\Sigma^1 = \{A \cap X^1 | A \in \Sigma\}$ and $\mu^1(A) = \mu(A)/\mu(X^1)$. One can also easily show that a countable direct product of Lebesgue spaces is a Lebesgue space.

We list here two important properties of Lebesgue spaces.

THEOREM 1.2. If (X, Σ, μ) is a Lebesgue space, then a collection $\mathcal{E} \subseteq \Sigma$ separates X if and only if it generates Σ.

THEOREM 1.3. Let (X^i, Σ^i, μ^i) be Lebesgue spaces for $i = 1, 2$, and let $\phi : X^1 \to X^2$ be a measurable measure-preserving mapping; that is, $\phi^{-1}(\Sigma^2) \subseteq \Sigma^1$ and $\mu^1(\phi^{-1}(A_2)) = \mu^2(A_2)$, $A_2 \in \Sigma^2$. If ϕ is one-to-one (mod zero) then ϕ is onto (mod sero); in particular, $\phi(X^1)$ must then be Σ^2-measurable.

We will make use of Theorem 1.3 in conjunction with the following more elementary result, which enables us to extend isomorphisms from generating collections to the entire σ-algebra.

THEOREM 1.4. Let $\sigma: X^1 \mapsto X^2$, where (X^1, Σ^1, μ^1) and (X^2, Σ^2, μ^2) are probability spaces. Let $\mathcal{E}$ be a generator for Σ^2, and suppose

(i) $\phi^{-1}(\mathcal{E}) \subseteq \Sigma^1$

(ii) $\mu^1(\phi^{-1}(A)) = \mu^2(A), \quad A \in \mathcal{E}$.

Then $\phi^{-1}(\Sigma^2) \subseteq \Sigma^1$, and (ii) holds for all $A \in \Sigma^2$.

This result is proved by first extending (i) and (ii) to the algebra generated by $\mathcal{E}$, then using the basic uniqueness theorem on extensions of measures (see [8], Theroem A, p. 54).

We now describe a factor space construction that will be useful in Chapter 10. Let (X, Σ, μ) be a probability space, and let Σ_1 be a complete sub-σ-algebra of Σ. We say that $x \sim y$ if x and y cannot be separated by Σ_1, and denote the set of equivalence classes modulo this relation by X^1. For $x \in X$, let $\pi(x)$ denote the equivalence class of x. Define

$$\begin{aligned} \Sigma^1 &= \{A \subset X^1 | \pi^{-1}(A) \in \Sigma_1\} \\ \mu^1(A) &= \mu(\pi^{-1}(A)),\ A \in \Sigma^1. \end{aligned}$$

The space (X^1, Σ^1, μ^1) is called the *factor space* of (X, Σ, μ) by Σ_1. One can now prove:

THEOREM 1.5. If (X, Σ, μ) is a Lebesgue space, and Σ_1 is a complete nonatomic sub-σ-algebra, then the factor space (X^1, Σ^1, μ^1) is a Lebesgue space.

Unless stated otherwise, all spaces in this book are assumed to be Lebesgue spaces, and a *transformation* is an automorphism of such a space; that is, a transformation is an invertible measure-preserving mapping of a space isomorphic to the unit interval.

CHAPTER 2.
SHIFTS AND PARTITIONS

We begin this chapter by giving a precise definition of a Bernoulli shift. Suppose $\pi = (p_1, p_2, \ldots, p_k)$, with $p_i > 0$ and $\Sigma p_i = 1$. Let X be the set of all doubly infinite sequences of the symbols $1, 2, \ldots, k$; that is, the set of all functions from the integers Z into $\{1, 2, \ldots, k\}$. A measure is defined on X as follows: A *cylinder set* is a subset of X determined by a finite number of values, such as

$$C = \{x | x_i = t_i, -m \leq i \leq n\} \tag{2.1}$$

where t_i, $-m \leq i \leq n$, is some fixed finite sequence in $\{1, 2, \ldots, k\}$. Let $\mathcal{E}$ denote the σ-algebra generated by the cylinder sets. There is then a unique measure μ defined on $\mathcal{E}$ such that, if C has the form (2.1), then $\mu(C) = \prod_{i=-m}^{n} p_{t_i}$. The measure space (X, Σ, μ) , where Σ is the completion of $\mathcal{E}$ with respect to μ, will be called the *product space with product measure μ determined by the distribution π.* The transformation T defined by

$$(Tx)_n = x_{n+1}, \; n \in Z,$$

is clearly an invertible μ-measure-preserving transformation. It will be called the *Bernoulli shift with distribution* π and denoted by T_π.

There are many ways of determining the space (X, Σ, μ) ; that is, it has many isomorphisms, so a given Bernoulli shift can be described in many other ways. We give here one simple geometric representation for the case when $\pi = (1/2, 1/2)$. For convenience we shall use the indexing $\{0, 1\}$, rather than $\{1, 2\}$; that is, X will be the set of all doubly infinite sequences of zeros and ones. Given $x \in X$, construct the point $(s(x), t(x))$ in the unit square (using binary digits)

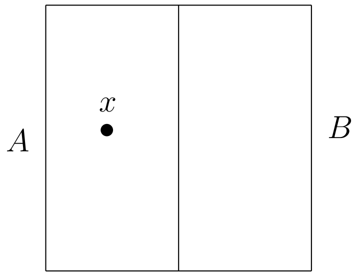
$$\begin{aligned} s(x) &= .x_0 x_1 x_2 \ldots \\ t(x) &= .x_{-1} x_{-2} x_{-3} \ldots \end{aligned}$$

The mapping $x \to (s(x), t(x))$ is easily seen to be one-to-one and onto (after removing a set from X of μ-measure zero). Furthermore, this mapping carries the class Σ onto the class of Lebesgue sets and μ into Lebesgue measure on the unit square. Also,

$$\begin{aligned} s(Tx) &= .x_1 x_2 \ldots \\ t(Tx) &= .x_0 x_{-1} x_{-2} \ldots \end{aligned}$$

so that T is carried onto the Baker's transformation (see Fig. 2.1).

Step 1. Cut unit square into two columns of equal width.



Step 2. Squeeze each column to a rectangle of height 1/2 and base 1.

Step 3. Put A' on top of B' to form a square.

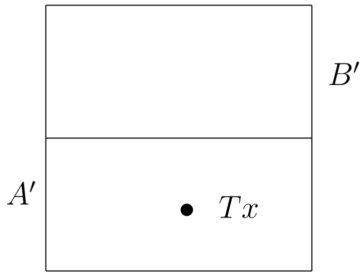
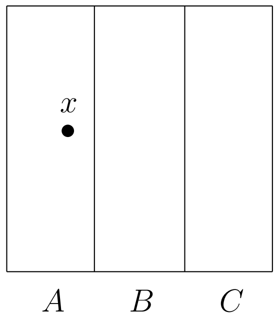


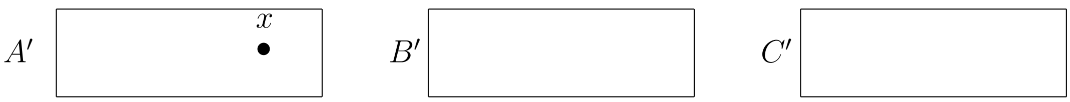
Fig 2.1

Obviously, this construction can be generalized. For example, if $\pi = (1/3, 1/3, 1/3)$, then, using ternary expansions, the shift T with distribution π becomes the Baker's transformation indicated in Figure 2.2.

Step 1. Cut unit square into three columns of equal width.



Step 2. Squeeze each column to a rectangle of height 1/3 and base 1



Step 3. Put B' on top of A' and C' on top of B' to form a square

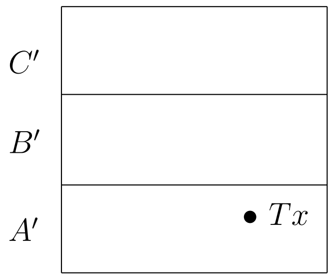


Fig 2.2

Are the transformations of Figures 2.1 and 2.2 isomorphic? That is, can we find an invertible measure-preserving transformation of the square (except for a null set) onto itself which carries one transformation into the other? More generally, if π and $\bar{\pi}$ are given distributions, when will T_π be isomorphic to $T_{\bar{\pi}}$? The answer to this general question is summarized in

THE KOLMOGOROV-ORNSTEIN ISOMORPHISM THEOREM. Two Bernoulli shifts T_π, $T_{\bar{\pi}}$ are isomorphic if and only if

$$\sum_{i=1}^{k} p_i \log p_i = \sum_{i=1}^{\bar{k}} \bar{p}_i \log \bar{p}_i \tag{2.2}$$

where $\pi = (p_1, p_2, \ldots, p_k)$, $\bar{\pi} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_k)$.

The necessity of the condition (2.2) was established by Kolmogorov ([9], [10]), while Ornstein ([13]) established its sufficiency. In this monograph we describe in detail the ideas behind Ornstein's results.

The concepts and terminology associated with partitions will enable us to give a more abstract and useful description of Bernoulli shifts. A *partition* $\mathcal{P}$ of X is an ordered *finite* disjoint collection of (measurable) sets (called the *atoms* of P) whose union is X. If $\mathcal{P}$ and $\mathcal{Q}$ are partitions, then $\mathcal{P}$ *refines* $\mathcal{Q}$ if each atom in $\mathcal{Q}$ is a union of atoms in $\mathcal{P}$. If $\mathcal{P}$ refines $\mathcal{Q}$, we write $\mathcal{P} \supset \mathcal{Q}$ or $\mathcal{Q} \subset \mathcal{P}$. If $\mathcal{P}$ and $\mathcal{Q}$ are partitions, their *join* is

$$\mathcal{P} \vee \mathcal{Q} = \{P_i \cap Q_j | P_i \in \mathcal{P},\ Q_j \in \mathcal{Q}\}$$

with lexicographic ordering. Clearly, $\mathcal{P} \vee \mathcal{Q}$ is the least partition which refines both $\mathcal{P}$ and $\mathcal{Q}$. For sequences of partitions $\mathcal{P}^i$, $-m \leq i \leq n$, we use the notation

$$\bigvee_{i=-m}^{n} \mathcal{P}^i = \mathcal{P}^{-m} \vee \mathcal{P}^{-m+1} \vee \ldots \vee \mathcal{P}^n.$$

A partition $\mathcal{P}$ determines a σ-algebra $\Sigma(\mathcal{P})$, which is just the set of all unions of members of $\mathcal{P}$. Note that

$$\mathcal{P} \supset Q \quad \text{iff} \quad \Sigma(\mathcal{P}) \supset \Sigma(\mathcal{Q}),$$

and that $\Sigma(\mathcal{P} \vee \mathcal{Q})$ is the smallest σ-algebra containing $\Sigma(\mathcal{P})$ and $\Sigma(\mathcal{Q})$.

The *distribution* of a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ is the vector

$$d(\mathcal{P}) = (\mu(P_1), \mu(P_2), \ldots, \mu(P_k)).$$

If T is a transformation and $\mathcal{P}$ is a partition, then $T\mathcal{P} = \{TP_i | P_i \in \mathcal{P}\}$ and, for example,

$$\bigvee_{0}^{n} T^i\mathcal{P} = \mathcal{P} \vee T\mathcal{P} \vee \ldots \vee T^n\mathcal{P}.$$

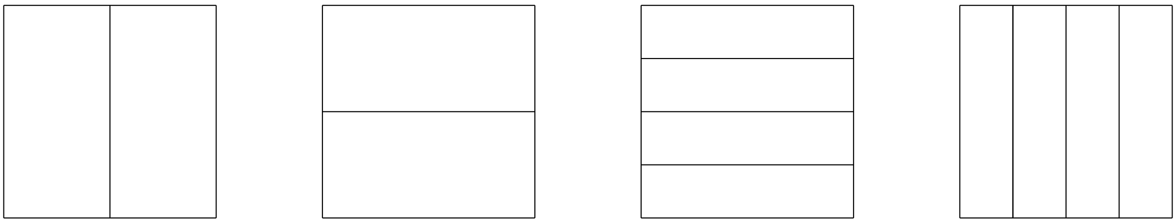
We say that $\mathcal{P}$ is a *generator* for T if Σ is the smallest complete σ-algebra containing $\bigcup_{n=-\infty}^{\infty} T^n\mathcal{P}$.

For example, consider the 2-shift $T = T_\pi$ with $\pi = (1/2, 1/2)$ and the partition $\mathcal{P} = \{P_0, P_1\}$, where

$$P_0 = \{x | x_0 = 0\}, P_1 = \{x | x_0 = 1\}.$$

In this case, $d(\mathcal{P}) = (1/2, 1/2)$, and the sets in $\bigvee_{-m}^{n} T^i\mathcal{P}$ are exactly the cylinder sets $\{x | x_{-i} = t_i, -m \leq i \leq n\}$ as $(t_{-m}, t_{-m+1}, \ldots, t_n)$ ranges over all possible sequences of zeros and ones, $m + n + 1$ units long. The tranformation T is the

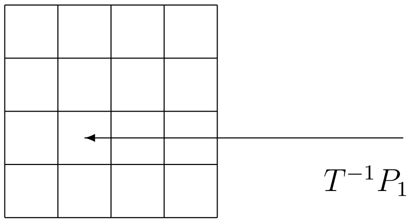
Baker's transformation of Figure 2.1. For this representation, Figure 2.3 illustrates $\mathcal{P}$, $T\mathcal{P}$, $T^2\mathcal{P}$, $T^{-1}\mathcal{P}$, while Figure 2.4 illustrates $T^{-1}\mathcal{P} \vee \mathcal{P} \vee T\mathcal{P} \vee T^2\mathcal{P}$. Note that $\mathcal{P}$ is a generator for T.



$\mathcal{P}$ $\qquad$ $T\mathcal{P}$ $\qquad$ $T^2\mathcal{P}$ $\qquad$ $T^{-1}\mathcal{P}$

Fig 2.3

Fig 2.4



This set is
$T^{-1}P_1 \cap P_0 \cap TP_0 \cap T^2P_1$

We say that the partitions $\mathcal{P}$ and $\mathcal{Q}$ are *independent* if

$$\mu(P_i \cap Q_j) = \mu(P_i)\mu(Q_j), \quad P_i \in \mathcal{P}, \quad Q_j \in \mathrm{Q}.$$

This is just the assertion that $\mathcal{P}$ partitions each set in $\mathcal{Q}$ in exactly the same proportions as it partitions the entire space. We say that the sequence of partitions $\{\mathcal{P}^n\}$, $n \geq 1$, is an *independent* sequence if, for each $n > 1$, $\mathcal{P}^n$and $\bigvee_1^{n-1} \mathcal{P}^i$ are independent. For the 2-shift $T = T_\pi$ of the preceeding paragraph, the sequence $T^{-1}\mathcal{P}$, $\mathcal{P}$, $T\mathcal{P}$, $T^2\mathcal{P}$ is an independent sequence.

A characterization of Bernoulli shifts without reference to product spaces is obtained by generalizing the above construction.

THEOREM 2.1. A transformation T is isomorphic to the Bernoulli shift T_π with distribution $\pi = (p_1, p_2, \ldots, p_k)$ if and only if there is a partition $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ such that

a) $d(\mathcal{P}) = \pi$,

b) $\mathcal{P}$ is a generator for T,

c) $T^n\mathcal{P}$, $n \geq 1$, is an independent sequence.

Proof. Let $T = T^\pi$ be the Bernoulli shift with distribution π, and let X_π be the product space with product measure μ_π determined by π. Put $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$, where

$$P_i = \{x|x_0 = i\}.$$

Clearly, $\mathcal{P}$ is a generator for T (since the cylinder sets are just the atoms of $\bigvee_{-m}^{n} T^i\mathcal{P}$) and $\{T^n\mathcal{P}\}$ is an independent sequence (since the product measure is used). This proves the existence of a $\mathcal{P}$ satisfying (a), (b), (c) for the Bernoulli shift T_π.

The proof of the converse makes use of the ideas sketched in Chapter 1. Assume T and $\mathcal{P}$ satisfy (a), (b), (c), where T is defined on (X, Σ, μ) . We obtain a map ϕ from X into X_π as follows: If $x \in X$, then $\phi(x) = \{x_n\} \in \{1, 2, \ldots, k\}^Z$, where

$$x_n = i \quad \text{iff} \quad T^n x \in P_i.$$

It is obvious that $\phi(Tx) = T^\pi(\phi(x))$. We wish to show that ϕ is an isomorphism; that is, except for a set of measure zero in X and a set of measure zero in X_π, ϕ is one-to-one, onto, measurable and measure-preserving.

The assumption that $\mathcal{P}$ is a generator for T means that the countable collection of sets $\bigcup_{-\infty}^{\infty} T^i\mathcal{P}$ generates Σ; hence it also separates X (see Theorem 1.2). Thus there is a set $E \subset X$ of measure zero such that ϕ is one-to-one on $X - E$.

The proof that ϕ is measurable and measure-preserving is obtained by examining the action of ϕ^{-1} on cylinder sets. Let

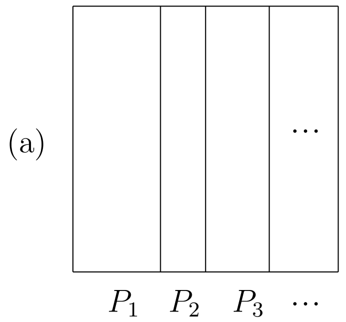
$$A = \bigcap_{i=-m}^{n} T^i P_{t_i}, \ \tilde{A} = \{x \in X_\pi | x_i = t_i, -m \leq i \leq n\}.$$

Then $\phi^{-1}(\tilde{A}) = A$ (in the sense that $\mu(\phi^{-1}(\tilde{A}) \triangle A) = 0$). Also, since the independence condition (c) gives

$$\mu(A) = \prod_{i=-m}^{n} \mu(T^i P_{t_i}) = \prod_{i=-m}^{n} \mu_\pi(\{x \in X_\pi | x_i = t_i\}), \tag{2.3}$$

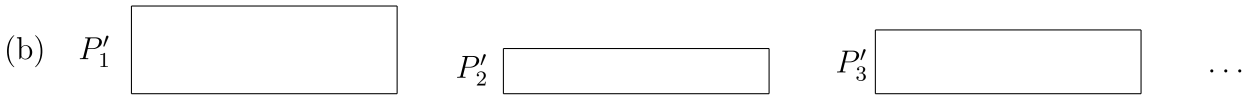
we see that $\mu(\phi^{-1}(\tilde{A})) = \mu_\pi(\tilde{A})$. It follows that ϕ is measurable and measure-preserving (Theorem 1.4), and hence maps onto a measurable set of measure one in X (Theorem 1.3). This proves Theorem 2.1.

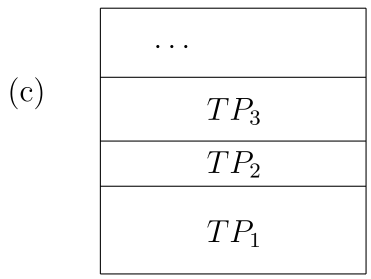
The reader should note that Theorem 2.1 enables one to obtain a general Baker's transformation description for the Bernoulli shift T. The transformation T and partition $\mathcal{P}$ described in Figure 2.5 clearly satisfy (a), (b), (c) of Theorem 2.1.



(a) Cut square into columns $P_1, \ldots, P_k$ with $\mu(P_i) = p_i$

(b) Squeeze each P_i to rectangle of width 1 and height p_i





(c) Put P_2' on top of P_1', P_3' on top of $P_2,'$ …, to form a square.

Fig. 2.5

The proof of Theorem 2.1 includes a useful concept. Suppose $\mathcal{P} = (P_1, P_2, \ldots, P_k)$ is a partition and T is a transformation. The $\mathcal{P}$-*name* of a point x is the bilateral sequence $\{x_n\} \subseteq \{1, 2, \ldots, k\}$, where

$$x_n = i \text{ if } x \in T^{-n}P_i; \text{ that is, } T^n x \in P_i.$$

Theorem 2.1 is just the observation that, if $\mathcal{P}$ is an independent generator for T, then the map

$$x \to \mathcal{P}\text{-name of } X$$

is an isomorphism which carries T into T_π, the Bernoulli shift with distribution $\pi = d(\mathcal{P})$. This result is a special case of the fact that a stochastic process is determined by its joint distributions ([2]). In our case, the process is defined by

$$Z_n(x) = i \text{ if } x \in T^{-n}P_i; \tag{2.4}$$

that is, the $\mathcal{P}$-name of x is the sequence $\{Z_n(x)\}$. This process is stationary, and its joint distributions are given by (2.3). To say that $\{T^n\mathcal{P}\}$ is an independent sequence is just the same as saying that $\{Z_n\}$ is a sequence of independent, identically distributed random variables. These remarks can easily be generalized to yield the following result:

THEOREM 2.2. The transformations T and $\overline{T}$ are isomorphic if and only if there are partitions $\mathcal{P}$ and $\overline{\mathcal{P}}$ which are generators for T and $\overline{T}$, respectively, such that

$$d(\bigvee_0^n T^i\mathcal{P}) = d(\bigvee_0^n \overline{T}^i\overline{\mathcal{P}}), \quad n = 0, 1, 2, \dots .$$

We close this section by stating a version of the law of large numbers, which will be needed in the sequel. Suppose $\mathcal{P}$ is a partition and A is a set in $\bigvee_0^{n-1} T^{-i}\mathcal{P}$, so that A has the form

$$A = P_{i_0} \cap T^{-1}P_{i_1} \cap \dots \cap T^{-n+1}P_{i_{n-1}}.$$

The sequence $(i_0, i_1, \dots, i_{n-1})$ will be called the *P-n-name* of A. Note that, in fact, A consists of all points x such that

$$x_m = i_m, \quad 0 \leq m \leq n-1,$$

where $\{x_m\}$ is the P-name of X. Let $f_A(i,n)$ be the relative frequency of occurrence of i in the P-n-name of A, that is,

$$f_A(i,n) = \frac{|\{t : x_t = i,\ 1 \leq t \leq n\}|}{n}.$$

We then have

THE (WEAK) LAW OF LARGE NUMBERS. If P is a generator for T such that $P, TP, T^2P, \dots$ is an independent sequence and $\varepsilon > 0$, then for all sufficiently large n, there is a collection $\mathcal{E}$ of sets in $\bigvee_0^n T^{-i}P$ of total measure at least $1-\varepsilon$, such that, for all i and all $A \in \mathcal{E}$,

$$|f_A(i,n) - \mu(P_i)| \leq \varepsilon.$$

A proof of this can be found in [2].

CHAPTER 3.
STACKS

The key to an understanding of Ornstein's proof of the isomorphism theorem and to a number of other results is a simple geometric representation of a transformation. The representation is valid for transformations which are aperiodic (that is, for each n, $\mu\{x : T^n x = x\} = 0$), but we shall use it only for ergodic transformations. A transformation is *ergodic* if $TA \subseteq A$ implies $\mu(A) = 0$ or $\mu(A) = 1$.

The simplest way to prove that a Bernoulli shift is ergodic is to establish the much stronger condition

$$\lim_n \mu(T^n A \cap B) = \mu(A)\mu(B), \quad A, B \in \Sigma. \tag{3.1}$$

A transformation satisfying (3.1) is said to be *mixing*. A mixing transformation is obviously ergodic (merely apply (3.1) with $B = A^c$). To verify that a Bernoulli shift is mixing, one first verifies that (3.1) holds for cylinder sets (where it is just the statement that two cylinder sets that depend upon different coordinates are independent). It follows that (3.1) holds for all sets by approximating with cylinder sets.

The following theorem, due to Rohlin (see [7], p. 71), provides us with our desired representation.

ROHLIN'S THEOREM. If T is ergodic, n is a positive integer, and ε is a positive number, then there is a set F such that F, TF, T^2F, ... , $T^{n-1}F$ is a disjoint sequence, and $\mu(\bigcup_{i=0}^{n-1} T^iF) \geq 1 - \varepsilon$.

Proof. We usually picture the theorem as in Figure 3.1 where each T^iF is placed above $T^{i-1}F$, and we think of T as mapping points upwards, the action of T on the roof $T^{n-1}F$, and the set $E = X - \bigcup_0^{n-1} T^iF$ being left unspecified. The picture does not imply that these sets are intervals, although one can obtain intervals by using Theorem 1.1.

Let us sketch a proof for the case $n = 2$. Let B be a set of small positive measure. Since T is ergodic, there are some points $x \in B$ such that $Tx \notin B$. Call the set of these points $\tilde{B}$; that is,

$$\tilde{B} = \{x \in B | Tx \notin B\}.$$

Now picture B and $T\tilde{B}$ as sets with $T\tilde{B}$ above $\tilde{B}$, and think of T as mapping $x \in \tilde{B}$ directly upwards (see Figure 3.2).

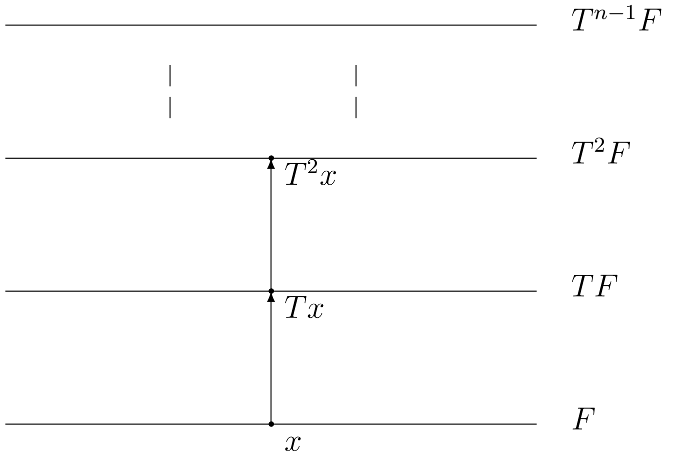


Figure 3.1

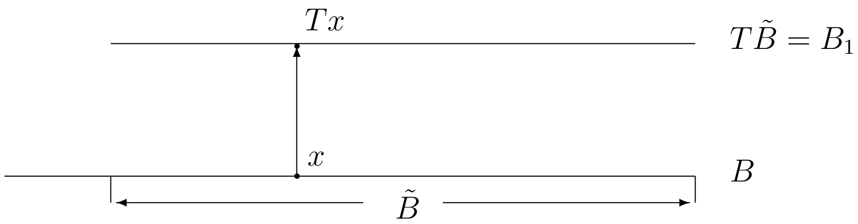


Figure 3.2

Now put $B_1 = T\tilde{B}$, and let $\tilde{B}_1 = \{x \in B_1 | Tx \notin B\}$. Note that, if $x \in \tilde{B}_1$, $Tx \notin B_1$, hence we can picture $T\tilde{B}_1$ as a set above $\tilde{B}_1$. By continuing this process, we obtain Figure 3.3, where T maps x in B_i directly upwards into B_{i+1}, or into B in some unspecified way, depending upon whether or not any part of B_{i+1} lies above x.

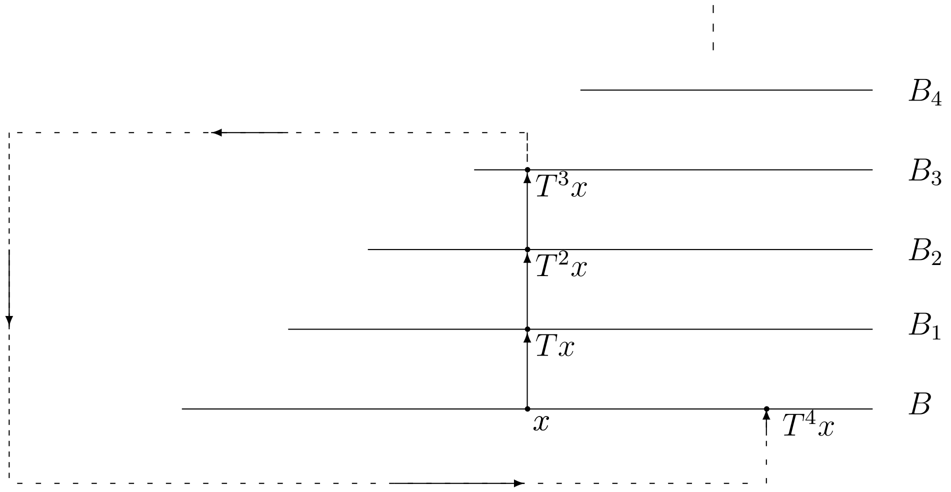
T is ergodic, so the set $B \bigcup_{i=1}^{\infty} B_i = X (\text{mod } 0)$, since this union is invariant and cannot have measure 0. We now let

$$F = T^{-1}B_1 \cup T^{-1}B_3 \cup T^{-1}B_5 \cup \dots \,.$$

Thus F is made up of pieces of B, B_2, B_4, ... , and TF is made up of B_1, B_3, B_5, ... , so, clearly, F and TF are disjoint. Furthermore, the complement of $F \cup TF$ is equal to $\bigcup_{i=0}^{\infty} C_i$, where $C_0 = B - T^{-1}B_1$ and

$$C_i = B_{2i} - T^{-1}B_{2i+1}, \quad i \geq 1$$

(see Figure 3.4).



T^4x returns to B since B_4 doest not lie above T^3x

Figure 3.3

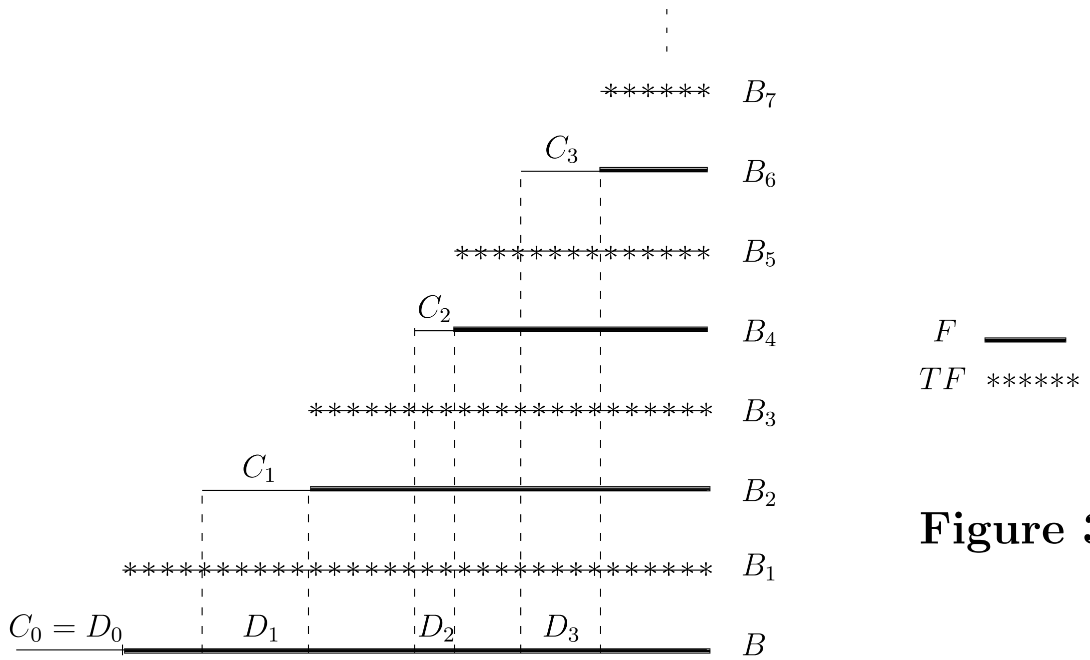


Figure 3.4

C_i = the part of B_{2i} that does <u>not</u> lie below B_{2i+1}

It follows that

$$\mu(\bigcup_{i=0}^{\infty} C_i) = \sum_{i=0}^{\infty} \mu(C_i) = \sum_{i=0}^{\infty} \mu(D_i),$$

where $D_i = T^{-2i} C_i$ (again see Figure 3.4). Since the D_i are disjoint and contained in B, we have

$$\mu(F \cup TF) \geq 1 - \mu(B),$$

so that, if $\mu(B) \leq \varepsilon$, F has the desired properties.

Thus Rohlin's theorem is just the observation that one can obtain a picture like Figure 3.3, then regroup parts of the sets to obtain a picture like Figure 3.1.

Let us introduce some terminology associated with these results. The disjoint sequence $T^i F$, $0 \leq i \leq n-1$, will be called a *stack* of *height* n. In a slight abuse of terminology, we shall use the symbol T for the restriction of T to $\bigcup_{i=0}^{n-2} T^i F$. This restricted T is a map from $\bigcup_{i=0}^{n-2} T^i F$ onto $\bigcup_{i=1}^{n-1} T^i F$ such that $T : T^i F \to T^{i+1} F$.

In order to describe the basic connections between P-names and stacks, we extend some of our previous terminology. If P is a partition and A is a set of positive measure, then the partition *induced* on A by P is

$$P/A = \{P_1 \cap A, P_2 \cap A, \ldots, P_k \cap A\},$$

and the *induced* distribution is the vector

$$d(P/A) = (\mu_A(P_1), \mu_A(P_2), \ldots, \mu_A(P_k)),$$

where μ_A is the conditional measure defined by $\mu_A(B) = \mu(B \cap A)/\mu(A)$. The *$P$-$n$-name* of a point x is the sequence $(i_0, i_1, \ldots, i_{n-1})$ where $T^m x \in P_{i_m}$, $0 \leq m \leq n-1$. Thus $\bigvee_0^{n-1} T^{-i} P/A$ is the partition of A into sets of points with the same P-n-name. To see the relation between this and stacks, suppose $T^i F$, $0 \leq i \leq n-1$, is a stack of height n. P induces a partition $P/T^i F$ on each level $T^i F$ of the stack (see Figure 3.5).

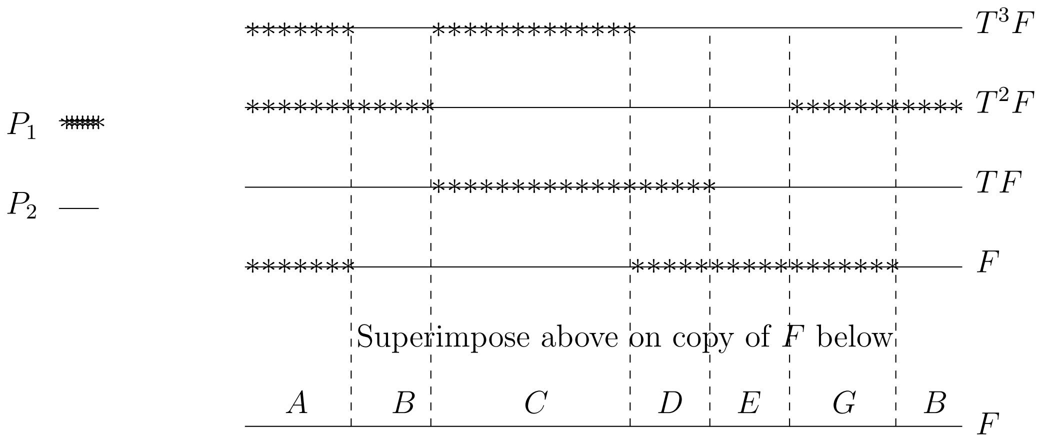


Figure 3.5

$$C = P_2 \cap T^{-1} P_1 \cap T^{-2} P_2 \cap T^{-3} P_1 \cap F$$

The atoms of $\vee_0^3 T^{-i} P/F$ are the sets A, B, C, D, E, G, B

These levels are now brought down to the base and superimposed; that is, we form $T^{-i}(P/T^iF)$, then take the join of these, $\bigvee_0^{n-1} T^{-1}(P/T^iF)$ (see Figure 3.5). It is easy to see that this is the partition $(\bigvee_0^{n-1} T^{-1}P)/F$, which is just the partition of the base into sets of points with the *same* p-n-name. We can use this to partition the stack into substacks, which will be called columns. A *column* is a stack T^iA, $0 \le i \le n-1$, where $A \in (\bigvee_0^{n-1} T^{-1}P)/F$. The set T^iA is called a *column level* (see Figure 3.6).

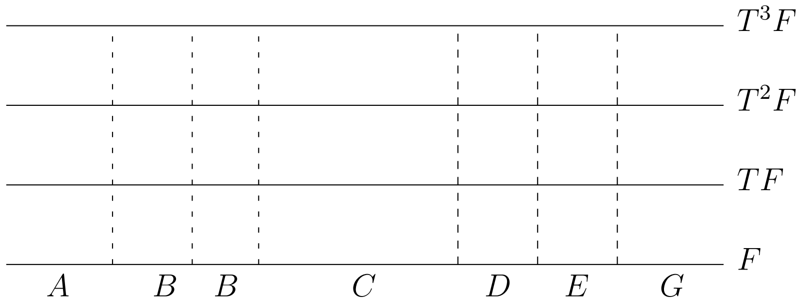


Figure 3.6

The columns of Fig. 3.5 redrawn so B is an interval

Thus the columns partition each level of the stack T^iF, $0 \le i \le n-1$, as follows: $x, y \in T^iF$ belong to the same column level if and only if

$$x_m = y_m, \qquad -i \le m \le n-1-i;$$

that is, *their P-names agree between $-i$ and $n-1-i$.*

Much use of these ideas will be made in subsequent sections. At this point we shall use them to prove the following stronger version of Rohlin's Theorem.

ROHLIN'S THEOREM (STRONG FORM). If T is ergodic, n a positive integer, ε a positive number, and P a partition, then there is a set F such that F, TF, …, $T^{n-1}F$ is a disjoint sequence, $\mu(\bigcup_{i=0}^{n-1} T^iF) \ge 1-\varepsilon$ *and* $d(P/F) = d(P)$.

Proof. We shall sketch the proof for the case $n = 2$. The idea is to take a stack long enough so that top and bottom two levels contribute very little, then split each P column into two subcolumns of the same size, and let F consist of odd levels in the left half and even levels in the right half of each column. To be precise, choose an even number m much larger than n, and use Rohlin's Theorem to select $\bar{F}$ such that $\bar{F}$, $T\bar{F}$, $T^2\bar{F}$, …, $T^{m-1}\bar{F}$ is a disjoint sequence, and $\mu(\bigcup_{i=0}^{m-1} T^i\bar{F}) \ge 1-\varepsilon/2$. Now express each set $A \in \bigvee_0^{m-1} T^{-i}P/\bar{F}$ as $A_1 \cup A_2$, $A_1 \cap A_2 = \phi$, $\mu(A_1) = \mu(A_2)$, and put

$$F_A = \bigcup_{i=0}^{(m/2)-1} T^{2i} A_1 \cup \bigcup_{i=1}^{(m/2)-1} T^{2i-1} A_2$$

(see Figure 3.7).

Clearly, F_A and TF_A are disjoint. We now let F be the union of the sets F_A, $A \in \bigvee_0^{m-1} T^{-i} P/\bar{F}$. It follows that $F \cap TF = \phi$, and that $F \cup TF$ is all of $\bigcup_{i=0}^{m-1} T^i F$, except for at most part of the base and part of the top two levels of the stack $T^i \bar{F}$, $0 \leq i \leq m-1$. Thus if m is large enough, $\mu(F \cup TF)$ will be at least $1 - \varepsilon$. This establishes that

$$d(P/F) = d(P/F \cup TF).$$

One can obviously start with something less than ε, then remove a small piece of F so that the desired stronger equality $d(P/F) = d(P)$ will also hold.

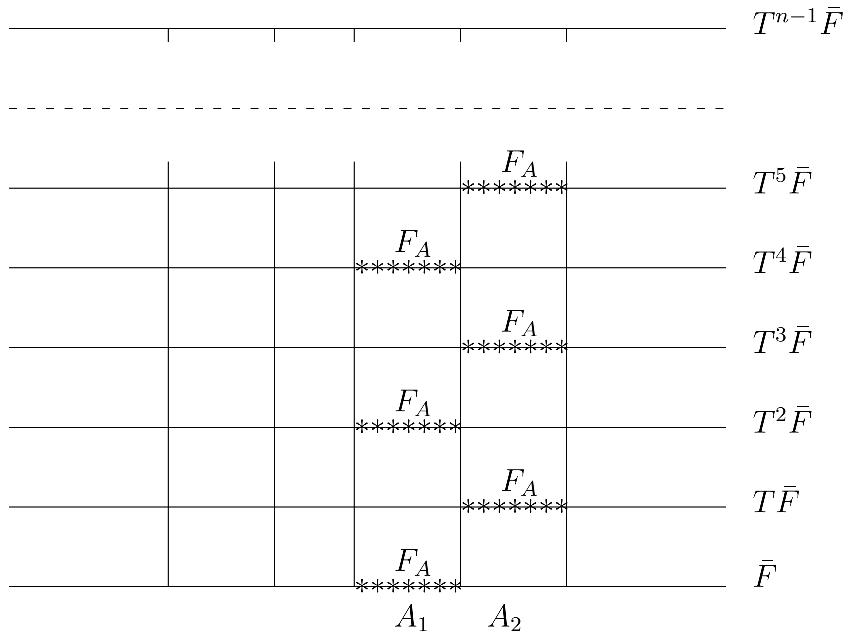


Figure 3.7

Split column into 2 equal subcolumns with bases A_1 and A_2

CHAPTER 4.
GADGETS

We now wish to look more carefully at the column structure induced on an n-stack by a partition P. Labels can be assigned to column levels according to the set in P to which the level belongs. This gives a one-to-one map from columns into P-n-names. It is then shown that any one-to-one map into n-strings of symbols from any alphabet gives rise to a partition Q, which induces the same columns as P. We also show how one can construct isomorphic copies of a given column structure.

To facilitate this discussion, we introduce the terminology used in [13]. A *gadget* is a quadruple (T, F, n, P), where T is a transformation, F a set such that F, TF, T^2F, ... , $T^{n-1}F$ is a disjoint sequence, and P a partition of $\bigcup_{i=0}^{n-1} T^iF$. As was noted in the previous chapter, P partitions each level P/T^iF. If these are brought down to F and superimposed, one obtains

$$\bigvee_0^{n-1} T^{-i}(P/T^iF) = \bigvee_0^{n-1} T^{-i}P/F.$$

The latter is the partition of the base into sets of points with the same P-n-names. The *column* C_A with base $A \in \bigvee_0^{n-1} T^{-i}P/F$ is the stack T^iA, $0 \le i \le n-1$. The points $x, y \in T^iF$ belong to the same column level iff $x_m = y_m$, $\ -i \le m \le (n-1)-i$.

We shall now assign the P-n-name of a column to that column and use this to label the column levels. To be precise, suppose $P = \{P_1, P_2, \dots, P_k\}$. The *$P$-name of a column C_A in the gadget* (T, F, n, P) is the P-n-name of A; that is, the P-name of C_A is $(i_0, i_1, \dots, i_{n-1})$ if and only if the A has the following form.

$$A = F \cap P_{i_0} \cap T^{-1}P_{i_1} \cap \ldots \cap T^{-n+1}P_{i_{n-1}}. \tag{4.1}$$

The mapping from columns into P-names of columns is a *one-to-one* map of columns into sequences of $\{1, 2, \dots, k\}$ of length n. Each column level is now assigned one of the integers in $\{1, 2, \dots, k\}$ according to the corresponding term in its column name; that is, if the P-name of C_A is $(i_0, i_1, \dots, i_{n-1})$, then the *label* of T^mA is i_m. This means (from 4.1) that $A \subseteq T^{-m}P_{i_m}$ (see Figure 4.1).

Figure 4.1

	A	B	C	D	E	G
T^3F	1	2	1	2	2	2
T^2F	1	1	2	2	2	1
TF	2	2	1	1	2	2
F	1	0	2	1	1	1

The labeling of Figure 3.6

The partition P can easily be recovered from this labeling process, for we have

(4.2) $$P_i = \text{union of column levels with label } i.$$

The above labeling process provides a clue to describing other partitions Q for which (T, F, n, Q) has the same columns as (T, F, n, P). Suppose $\mathcal{G} = (T, F, n, P)$ is given, and $C \mapsto \phi(C)$ is any one-to-one function from the columns of $\mathcal{G}$ into the set of n-strings from some alphabet. To be precise, suppose this alphabet is denoted by $\{a_1, a_2, \ldots, a_\ell\}$, and that

$$\phi(C_A) = (a_{i_0}, a_{i_1}, a_{i_2}, \ldots, a_{i_{n-1}}),$$

where A is an atom of $\bigvee_0^{n-1} T^{-1}P/F$. Let us call this n-string the ϕ-name of the column C_A. To each level $T^m A$, one can now assign a *ϕ-label*, which is the symbol a_{i_m}. We then let $Q = \{Q_1, Q_2, \ldots, Q_\ell\}$, where

(4.3) $$Q_i = \text{union of column levels with } \phi\text{-label } a_i.$$

It is then easy to see that (T, F, n, Q) has the same set of columns as (T, F, n, P). Figure 4.2 gives an example of such a partition Q.

Figure 4.2

T^3F	a	a	b	c	c	a
T^2F	c	a	b	b	c	b
TF	b	a	b	a	c	c
F	a	a	b	c	c	a

$Q = \{Q_a, Q_b, Q_c\}$ gives the same columns as $P = \{P_1, P_2\}$

For ease of reference, we summarize this result and its converse in the following lemma.

LEMMA 4.1. Let $\mathcal{G} = (T, F, n, P)$ be a given gadget, and let ϕ be a given one-to-one function from columns of $\mathcal{G}$ into n-strings from some finite alphabet. If Q is the partition formed by (4.3), then (T, F, n, Q) has the same set of columns as (T, F, n, P). Conversely, if $Q = \{Q_1, Q_2, \dots, Q_\ell\}$ is any partition such that (T, F, n, Q) has the same set of columns as (T, F, n, P), then the mapping

$$C \mapsto Q\text{-name of } C$$

is a one-to-one mapping from the columns into the set of Q-n-names.

The following lemma gives a further connection between the partition P and the partition Q, when (T, F, n, P) and (T, F, n, Q) have the same set of columns.

LEMMA 4.2. Suppose (T, F, n, P) and (T, F, n, Q) have the same set of columns. Let $\mathcal{H}$ denote the two-set partition $\{F, F^c\}$ of the set $G = \bigcup_{i=0}^{n-1} T^i F$. Then

$$P/G \subset \bigvee_{-n+1}^{n-1} T^i(Q \vee \mathcal{H})/G.$$

Proof. The lemma is a consequence of the following simple fact.

$$\text{(4.4)} \qquad \begin{array}{l} \text{If } B \text{ and } \bar{B} \text{ are distinct column levels with } B \subseteq T^i F,\ \bar{B} \subseteq \\ T^j F, \text{ then either } i \neq j \text{ or } i = j, \text{ and, for some } m,\ -i \leq m \leq \\ n-1-i, \text{ the two sets } T^m B \text{ and } T^m \bar{B} \text{ have different } Q\text{-labels.} \end{array}$$

To complete the proof, let $\mathcal{E}$ be the partition into column levels

$$\mathcal{E} = \{C_A \cap T^i F | A \in \bigvee_0^{n-1} T^{-j} P/F, 0 \leq i \leq n-1\}.$$

The ordering of $\mathcal{E}$ is not important. We can then rephrase (4.4) as the statement

$$\mathcal{E} \subset \bigvee_{-n+1}^{n-1} T^i(Q \vee \mathcal{H}).$$

Since P/G is refined by $\mathcal{E}$, the lemma follows.

Now we turn to the question of isomorphism of gadgets. We shall say that (T, F, n, P) is *isomorphic* to $(\bar{T}, \bar{F}, n, \bar{P})$ if

$$d(\bigvee_0^{n-1} T^{-1} P/F) = d(\bigvee_0^{n-1} \bar{T}^{-}1 \bar{P}/\bar{F});$$

that is, P-n-names partition F in the *same* proportions as corresponding $\bar{P}$-n-names partition $\bar{F}$. It is implicit in this definition that the two gadgets have the same height, and that P and $\bar{P}$ have the same number of sets. It is easy to see that (T, F, n, P) is isomorphic to $(\bar{T}, \bar{F}, n, \bar{P})$ if and only if there is an invertible mapping $S : \bar{F} \mapsto F$ such that, for all measurable $\bar{A} \subset \bar{F}$ and $A \subset F$, we have

$$\bar{\mu}(\bar{A})/\bar{\mu}(\bar{F}) = \mu(S\bar{A})/\mu(F),$$
$$\bar{\mu}(S^{-1}A)/\bar{\mu}(\bar{F}) = \mu(A)/\mu(F),$$

and, for $x \in \bar{F}$, the $\bar{P}$-n-name of x and the P-n-name of Sx are the same. In other words, except for a possible change of scale, two gadgets are isomorphic if one cannot distinguish between them by examining their column structures.

The statement that two gadgets are isomorphic says very little about their respective transformations, for Rohlin's Theorem and a simple construction combine to give the following result.

LEMMA 4.3. If (T, F, n, P) is any gadget and $\bar{T}$ is any ergodic transformation, then, for any $\epsilon > 0$, there is a set $\bar{F}$ and a partition $\bar{P}$ such that $(\bar{T}, \bar{F}, n, \bar{P})$ is a gadget isomorphic to (T, F, n, P) and $\mu(\bigcup_{i=0}^{n-1} \bar{T}^i \bar{F}) \geq 1 - \epsilon$.

Proof. The proof makes use of the fact that, given any partition P of X and any nonatomic space Y, one can partition Y in the same proportions as P partitions X; that is, there is a partition Q of Y such that $d(P) = d(Q)$. With this in mind, use Rohlin's Theorem to find $\bar{F}$ such that $\bar{F}$, $\bar{T}\bar{F}$, ..., $\bar{T}^{n-1}\bar{F}$ is a disjoint sequence and $\mu(\bigcup_{i=0}^{n-1} \bar{T}^i \bar{F}) \geq 1 - \epsilon$. Then let $\bar{Q}$ be a partition of $\bar{F}$ such that

$$d(\bar{Q}) = d(\bigvee_{0}^{n-1} T^{-1}P/F). \tag{4.5}$$

A $\bar{Q}$-column will then be a stack $\bar{T}^i\bar{A}$, $\bar{A} \in \bar{Q}$. The correspondence between P-n-names and sets of $\bar{Q}$ given implicitly by (4.5) then gives a one-to-one map ϕ of $\bar{Q}$-columns into P-n-names. Just as before, this means that each $\bar{Q}$-column is assigned a P-n-name, which means that each $\bar{Q}$-level is assigned a P-label. This gives a partition $\bar{P}$ of $\bigcup_{i=0}^{n-1} \bar{T}^i \bar{F}$ into sets with the same label (as in (4.3)). Clearly, (T, F, n, P) will then be isomorphic to $(\bar{T}, \bar{F}, n, \bar{P})$.

We will make use of an extension of Lemma 4.3, which is easily established by similar arguments:

LEMMA 4.4. Suppose (T, F, n, P) is isomorphic to $(\bar{T}, \bar{F}, n, \bar{P})$, and Q is a partition of $\bigcup_0^{n-1} T^i F$. Then there is a partition $\bar{Q}$ of $\bigcup_0^{n-1} \bar{T}^i \bar{F}$ such that $(T, F, n, P \vee Q)$ is isomorphic to $(\bar{T}, \bar{F}, n, \bar{P} \vee \bar{Q})$.

Of course, Lemmas 4.3 and 4.4 tell us nothing about the action of T and $\bar{T}$ on the top and complement of their gadgets, where almost anything can happen. We shall later see how entropy can be used to control the relationsip between T and $\bar{T}$.

CHAPTER 5.
METRICS ON PARTITIONS

Our subsequent discussion will make use of a number of approximation ideas. In this chapter, we introduce metrics that measure the distance between partitions, gadgets, and processes. The first of these is the *distribution distance* given by

$$|d(P) - d(Q)| = \sum_{i=1}^{k} |\mu(P_i) - \mu(Q_i)|. \tag{5.1}$$

Here we assume that P and Q each have k sets, but it is not required that they partition the same space. Note that, if $|d(P) - d(Q)| = 0$, then P and Q have the same distribution.

A stronger form of closeness is the *partition distance*

$$|P - Q| = \sum_{i=1}^{k} \mu(P_i \triangle Q_i), \tag{5.2}$$

where "$\triangle$" denotes the symmetric difference (that is, $A \triangle B = (A-B) \cup (B-A)$), and it is assumed that P and Q each have k sets and partition the same space. In this case, $|P-Q| = 0$ means that $\mu(P_i \triangle Q_i) = 0$, $1 \leq i \leq k$; that is, P and Q agree except on a set of measure zero. This is, of course, just the precise meaning of the statement $P = Q$.

Let us note that

$$0 \leq |d(P) - d(Q)| \leq 2 \text{ and } 0 \leq |P - Q| \leq 2,$$

and, if P and Q partition the same space, then

$$|d(P) - d(Q)| \leq |P - Q|.$$

We also note that the set of all sequences $\{p_i\}$, with $p_i \geq 0$, $\sum p_i = 1$, is a closed subset of ℓ_1 (the space of absolutely summable sequences). If we also require that $p_i = 0$, $i > k$, then this set is compact. Thus the collection of all distributions of partitions is complete, and the set of all distributions of partitions with no more than k sets is compact in the distribution metric (5.1). It is also easy to show that the set of partitions of a given space X is complete in the partition metric (5.2). This set is *not* compact, even if we restrict the number of sets in a partition. (For example, if $\{P^n\}$ is an independent sequence of two-set partitions, all with the same distribution $(1/2, 1/2)$, then $|P^i - P^j| = 1$ for all $i \neq j$.)

The metric of interest to us for gadgets measures how well one gadget can be superimposed upon another of the same height so that the levels fit well on the average. More precisely, suppose $\mathcal{G} = (T, F, n, P)$ and $\bar{\mathcal{G}} = (\bar{T}, \bar{F}, n, \bar{P})$. Let $\mathcal{H}$

be the set of all partitions Q of $\bigcup_{i=0}^{n-1} T^iF$ such that (T, F, n, Q) is isomorphic to $(\bar{T}, \bar{F}, n, \bar{P})$. In other words, if $Q \in \mathcal{H}$, then (T, F, n, Q) is a copy of $\bar{\mathcal{G}}$ on the stack T^iF, $0 \le i \le n-1$. The *gadget distance* is then

$$\bar{d}_n(\mathcal{G}, \bar{\mathcal{G}}) = \inf_{Q \in \mathcal{H}} \frac{1}{n} \sum_{i=0}^{n-1} |P/T^iF - Q/T^iF|. \tag{5.3}$$

Since a gadget isomorphism can be implemented by a transformation, this definition can also be formulated as follows: Let $\mathcal{E}$ be the set of all invertible mappings $S : \bar{F} \mapsto F$, such that

$$\bar{\mu}(\bar{A})/\bar{\mu}(\bar{F}) = \mu(S\bar{A})/\mu(F), \quad \bar{A} \subset \bar{F},$$

$$\bar{\mu}(S^{-1}A)/\bar{\mu}(\bar{F}) = \mu(A)/\mu(F), \quad A \subset F.$$

Extend S to a mapping of $\bigcup_{i=0}^{n-1} \bar{T}^i\bar{F}$ onto $\bigcup_{i=0}^{n-1} T^iF$ by defining $S\bar{T}^ix = T^iSx$, $x \in \bar{F}$. Then

$$\bar{d}_n(\mathcal{G}, \bar{\mathcal{G}}) = \inf_{S \in \mathcal{E}} \frac{1}{n} \sum_{i=0}^{n-1} |P/T^iF - S\bar{P}/T^iF|. \tag{5.3a}$$

It is quite easy to show that $\bar{d}_n(\mathcal{G}, \bar{\mathcal{G}}) = 0$ if and only if $\mathcal{G}$ and $\bar{\mathcal{G}}$ are isomorphic. We also note that

$$\bar{d}_n(\mathcal{G}, \bar{\mathcal{G}}) \le |d(\bigvee_0^{n-1} T^{-i}P/F) - d(\bigvee_0^{n-1} \bar{T}^{-1}\bar{P}/\bar{F})|,$$

so that, if the P-n-names and the $\bar{P}$-n-names of points in the respective bases are close in distribution, the gadgets are close. The converse of this is not true. There is, however, a sense in which the P-n-names and $\bar{P}$-n-names are close. This surprising and useful result is a consequence of the following lemma.

LEMMA 5.1. Suppose (T, F, n, P) and (T, F, n, Q) are gadgets satisfying

$$\frac{1}{n} \sum_{i=0}^{n-1} |P/T^iF - Q/T^iF| < \epsilon^2. \tag{5.4}$$

Let E be the set of points $x \in F$ such that the P-n-name and the Q-n-name of x differ in more than ϵn places. Then $\mu(E) \le \epsilon\mu(F)$.

Proof. Let E_j be the set of points $x \in F$ such that the P-n-name and Q-n-name of X differ in the j^{th} place; that is,

$$E_j = T^{-j} \bigcup_{i=1}^{k} [(P_i \cap T^jF) \triangle (Q_i \cap T^jF)].$$

The condition (5.4) then gives

$$\sum_{j=0}^{n-1} \mu(E_j) \leq n\epsilon^2 \mu(F),$$

while the definition of E tells us that

$$\epsilon n X_E \leq \sum_{j=0}^{n-1} X_{E_j},$$

where X_A denotes the characteristic function of A. Now integrate to obtain

$$\epsilon n \mu(E) \leq \sum_{j=0}^{n-1} \mu(E_j) \leq n\varepsilon^2 \mu(F),$$

which gives the desired conclusion. This proves Lemma 5.1.

If the names of most points agree in most places, then the levels must be close on the average. This converse to Lemma 5.1 is stated as

LEMMA 5.2. Let E be the set of points $x \in F$ such that the P-n-name and Q-n-name of x differ in more than ϵn places, and suppose $\mu(E) \leq \epsilon \mu(F)$. Then

$$\frac{1}{n} \sum_{i=0}^{n-1} |P/T^i F - Q/T^i F| \leq 3\epsilon. \tag{5.5}$$

In particular, $|P/G - Q/G| = 1/n \sum_{i=0}^{n-1} |P/T^i F - Q/T^i F|$, so

$$|P/G - Q/G| \leq 3\epsilon, \tag{5.6}$$

where $G = \bigcup_{i=0}^{n-1} T^i F$.

Proof. This is proved by using the column structure of the gadget. Let $\mathcal{E}$ be the class of sets of the form

$$C = A \cap B \cap E^c,$$

$$A \in \bigvee_{0}^{n-1} T^{-i} P/F, \quad B \in \bigvee_{0}^{n-1} T^{-i} Q/F.$$

The column $T^j C$, $0 \leq j \leq n-1$, is then the intersection of the two columns $\{T^j(A \cap E^c)\}$ and $\{T^j(B \cap E^c)\}$.

Since $C \cap E = \phi$, all except at most ϵn of the levels $T^j C$ have identical P- and Q-labels. In particular, we have

$$\frac{1}{n} \sum_{j=0}^{n-1} |P/T^j C - Q/T^j C| \leq 2\epsilon. \tag{5.7}$$

For each j, we have

$$\begin{aligned}|P/T^jF - Q/T^jF| = \quad &= \sum_i \frac{\mu(T^{-j}(P_i \triangle Q_i) \cap F)}{\mu(F)} \\ &= \sum_{C \in \mathcal{E}} |P/T^jC - Q/T^jC| \cdot \frac{\mu(C)}{\mu(F)} \\ &+ \sum_i \frac{\mu(T^{-j}(P_i \triangle Q_i) \cap E)}{\mu(F)}.\end{aligned}$$

The hypothesis that $\mu(E) \leq \epsilon\mu(F)$ and the result (5.7) then yield the desired result (5.5). This proves Lemma 5.2.

A number of our later results will be most easily stated in terms of an extension of this gadget metric to processes. Let T and $\bar{T}$ be transformations defined X, $\bar{X}$, respectively, and with respective partitions P and $\bar{P}$. The *process distance* is defined by

$$\bar{d}((T,P),\ (\bar{T},\bar{P})) = \sup_n \inf_{S \in \mathcal{E}} \frac{1}{n+1} \sum_{i=0}^{n} |T^iP - S\bar{T}^i\bar{P}|, \tag{5.8}$$

where $\mathcal{E}$ is the class of all isomorphisms of X onto $\bar{X}$. The full significance of this metric is still somewhat unclear (see some of the discussion in Chapter 10 below and [16]). It will primarily be used in this paper to simplify the statements of a number of results.

We mention here some of the properties of the process metric (5.8). First, we note that the supremum used in (5.8) is actually a limit. This follows from the fact that, if $\inf_{S\in\mathcal{E}} 1/n \sum_{i=0}^{n-1} |T^iP - S\bar{T}^i\bar{P}| = \alpha$, then, for all r, $\inf_{S\in\mathcal{E}} 1/nr \sum_{i=0}^{nr-1} |T^iP - S\bar{T}^i\bar{P}| \geq \alpha$. We also observe that

(5.9) If P and $\bar{P}$ are generators for T and $\bar{T}$, and and if $\bar{d}((T,P),(\bar{T},\bar{P})) = 0$, then T is isomorphic to $\bar{T}$.

The proof of (5.9) is as follows: The condition that $\bar{d}((T,P),(\bar{T},\bar{P})) = 0$ implies that, for each n,

$$\inf_{S\in\mathcal{E}} \frac{1}{n} \sum_{i=0}^{n-1} |T^iP - S\bar{T}^i\bar{P}| = 0,$$

and hence that $d(\bigvee_0^{n-1} T^iP) = d(\bigvee_0^{n-1} \bar{T}^i\bar{P})$, $n = 1, 2, \ldots$. Theorem 2.2 then implies that (5.9) is true.

The following lemma is established in much the same way as Lemmas 5.1 and 5.2.

LEMMA 5.3. If $\bar{d}((T,P),(\bar{T},\bar{P})) < \epsilon^2$, then, for each n, there is an isomorphism S_n from $\bar{X}$ to X such that the set of points $x \in X$ for which the P-n-name of x

and the $\bar{P}$-n-name of $S_n^{-1}x$ differ in more than ϵn places has measure less than ϵ. Conversely, the existence of such an S_n implies that $\bar{d}((T,P),\,(\bar{T},\bar{P})) < 3\epsilon$.

An equivalent form of the definition (5.8) can be obtained by superimposing T, P and $\bar{T}, \bar{P}$ on a third space. In fact, let Y be a fixed nonatomic probability space, and let $\mathcal{E}$ denote the class of all isomorphisms of the T-space onto Y, and let $\bar{\mathcal{E}}$ denote the class of all isomorphisms of the $\bar{T}$-space onto Y. We then have

$$\bar{d}((T,P),\ (\bar{T},\bar{P})) = \sup_n \inf_{\substack{\bar{S}\in\bar{\mathcal{E}} \\ S\in\mathcal{E}}} \frac{1}{n}\sum_{i=0}^{n-1} |ST^iP - \bar{S}\bar{T}^i\bar{P}|. \tag{5.10}$$

This enables one to establish easily (for ergodic transformations) that closeness in the process metric is equivalent to closeness in the gadget metric for arbitrarily long gadgets.

LEMMA 5.4. If T and $\bar{T}$ are ergodic, then $\bar{d}((T,P),\,(\bar{T},\bar{P})) < \epsilon$ if and only if, for each n and each $\delta > 0$, there are gadgets $\mathcal{G} = (T, F, n, P)$ and $\bar{\mathcal{G}} = (\bar{T}, \bar{F}, n, \bar{P})$ such that $\mu(\bigcup_{i=0}^{n-1} T^iF) \ge 1-\delta$, $\bar{\mu}(\bigcup_{i=0}^{n-1} \bar{T}^i\bar{F}) \ge 1-\delta$, and $\bar{d}(\mathcal{G},\bar{\mathcal{G}}) < \epsilon$.

Proof. If $\bar{d}((T,P),\,(\bar{T},\bar{P})) < \epsilon$, one can use the strong form of Rohlin's Theorem to find the desired $\mathcal{G}$ and $\bar{\mathcal{G}}$ so that

$$d(\bigvee_0^{n-1} T^{-1}P/F) = d(\bigvee_0^{n-1} T^{-1}P),\ d(\bigvee_0^{n-1} \bar{T}^{-1}\bar{P}/\bar{F}) = d(\bigvee_0^{n-1} \bar{T}^{-1}\bar{P}),$$

and then use (5.10) to conclude that $\bar{d}_n(\mathcal{G},\bar{\mathcal{G}}) < \epsilon$. The converse of this is trivial (and does not even require that T and $\bar{T}$ be ergodic).

CHAPTER 6.
INDEPENDENCE AND ϵ-INDEPENDENCE

The original proof of the isomorphism theorem made use of a concept of approximate independence known as ϵ-independence ([13]). The concept appears to be useful in many settings, and will be introduced here. After discussing some of the alternative ways to define ϵ-independence, we shall establish the main result of this section, namely that an ϵ-independent sequence can be modified to give an independent sequence.

Recall that two partitions P and Q are *independent* if

$$\mu(P_i \cap Q_j) = \mu(P_i)\mu(Q_j),\ \ P_i \in P,\ \ Q_j \in P.$$

This is the same as the statement $d(P/Q_j) = d(P)$, $Q_j \in Q$, which is shorthand for the idea that P partitions each set in Q in the same proportions that P partitions X (see Figure 2.3).

The definition of approximate independence we shall use merely asserts (roughly) that P partitions most sets in Q in almost the same way that P partitions X. To be precise, we say that P is ϵ*-independent* of Q if there is a class $\mathcal{E}$ of sets in Q such that

$$\text{(6.1)} \qquad \begin{aligned} &\text{a)} \quad \mu(\cup\mathcal{E}) \geq 1-\epsilon, \\ &\text{b)} \quad |d(P/Q_j) - d(P)| \leq \epsilon, \quad Q_j \in \mathcal{E}. \end{aligned}$$

This definition is not quite symmetric in P and Q. However, one can easily show that if P is ϵ-independent of Q then Q is $\sqrt{3\epsilon}$-independent of P (see Lemma 6.2 below). It is easy to see that P and Q are independent if and only if P is ϵ-independent of Q for each $\epsilon > 0$.

One can establish ϵ-independence by obtaining independence on a large portion of the space. This is stated precisely in the following lemma.

LEMMA 6.1. If $\mu(E) \geq 1-\epsilon^2$, where P/E and Q/E are independent, then P is 3ϵ-independent of Q.

Proof. Suppose $\mu(E) \geq 1-\epsilon^2$, and P/E and Q/E are independent. Let

$$\mathcal{E} = \{Q_j \in Q\colon \mu(Q_j \cap E) \geq (1-\epsilon)\mu(Q_j)\}.$$

Certainly we have $\mu(\cup\mathcal{E}) \geq 1-\epsilon$. Furthermore, if $Q_j \in \mathcal{E}$, then

$$\sum_i \left|\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} - \mu(P_i)\right| \leq$$
$$\leq \sum_i \left|\frac{\mu(P_i \cap Q_j \cap E)}{\mu(Q_j)} - \mu(P_i \cap E)\right| + \sum_i \left|\frac{\mu(P_i \cap Q_j \cap E^c)}{\mu(Q_j)} - \mu(P_i \cap E^c)\right|.$$

Since P/E and Q/E are independent, the first sum on the right is equal to

$$\sum_i |\frac{\mu(P_i \cap E)\mu(Q_j \cap E)}{\mu(Q_j)\mu(E)} - \mu(P_i \cap E)| = \mu(E)|\frac{Q_j \cap E)}{\mu(Q_j)\mu(E)} - 1|,$$

and it is easy to see that the latter quantity cannot exceed ϵ. Also,

$$\sum_i |\frac{\mu(P_i \cap Q_j \cap E^c)}{\mu(Q_j)} - \mu(P_i \cap E^c)| \leq$$
$$\leq \sum_i \frac{\mu(P_i \cap Q_j \cap E^c)}{\mu(Q_j)} + \sum_i \mu(P_i \cap E^c) = \frac{\mu(Q_j \cap E^c)}{\mu(Q_j)} + \mu(E^c).$$

If $Q_j \in \mathcal{E}$, then $\mu(Q_j \cap E^c) < \epsilon\mu(Q_j)$, so $|d(P/Q_j) - d(P)| \leq \epsilon + \epsilon + \mu(E^c) \leq 3\epsilon$. This proves the lemma.

The following lemma shows how one can give a more symmetric definition of approximate independence. We prefer to use the more geometric definition of ϵ-independence given by (6.1).

LEMMA 6.2. If P is ϵ-independent of Q, then

$$\sum_i \sum_j |\mu(P_i \cap Q_j) - \mu(P_i)\mu(Q_j)| \leq 3\epsilon.$$

Conversely, if this inequality holds, then P is $\sqrt{3\epsilon}$-independent of Q.

Proof. Left to the reader.

We say that a sequence $\{P^i\}$ is an ϵ-*independent* sequence if, for each $n > 1$, P^n is ϵ-independent of $\bigvee_0^{n-1} P^i$. The main result of this section is that an ϵ-independent sequence $\{T^i P\}$ can be modified slightly so as to obtain an independent sequence. To facilitate the statement and proof of this result, we extend the $\bar{d}$-metric (defined in Chapter 5 for processes) to arbitrary sequences of partitions. Let P^i, $1 \leq i \leq n$, be a sequence of k set partitions of X, and let $\bar{P}^i$, $1 \leq i \leq n$, be a sequence of k set partitions of $\bar{X}$. Let Y be a fixed Lebesgue space. Then

$$\bar{d}_n(\{P^i\}_{i=1}^n, \{\bar{P}^i\}_{i=1}^n) = \inf \frac{1}{n} \sum_{i=1}^{n-1} |Q^i - \bar{Q}^i|,$$

where this infimum is taken over all sequences Q^i, $1 \leq i \leq n$, and $\bar{Q}^i$, $1 \leq i \leq n$, of k set partitions of Y such that

$$d(\bigvee_1^n Q^i) = d(\bigvee_1^n P^i), \quad d(\bigvee_1^n \bar{Q}^i) = d(\bigvee_1^n \bar{P}^i).$$

Note that $\bar{d}_n$ does not depend on Y. We shall prove

LEMMA 6.3. Let P^i, $1 \le i \le n$, and $\bar{P}^i$, $1 \le i \le n$, be sequences of k set partitions with the following properties.

a) $\{\bar{P}^i\}$ is independent.

b) $\{P^i\}$ is ϵ-independent.

c) $|d(P^i) - d(\bar{P}^i)| < \epsilon$, $1 \le i \le n$.

Then

d) $\bar{d}_n(\{P^i\}_{i=1}^n \{\bar{P}^i\}_{i=1}^n) < 4\epsilon$.

Proof. The proof makes use of the fact that a Lebesgue space can be partitioned according to any given distribution. We leave to the reader the proof of the following sharper form of this fact.

(6.2) If P and $\bar{P}$ are partitions of any two probability spaces, and Y is any Lebesgue space, then there are partitions $Q, \bar{Q}$ of Y so that $|Q - \bar{Q}| = |d(P) - d(\bar{P})|$, $d(Q) = d(P)$, and $d(\bar{Q}) = d(\bar{P})$.

This result and condition (c) of the lemma immediately imply that (d) holds for $n = 1$, so let us assume the lemma is true for n. Suppose $\{P^i\}$ and $\{\bar{P}^i\}$ satisfy the hypothesis for $n+1$. We can apply the induction hypothesis to choose partitions Q^i, $1 \le i \le n$, and $\bar{Q}^i$, $1 \le i \le n$, of a given nonatomic Y so that

$$
\begin{aligned}
&\text{i)} && d\left(\textstyle\bigvee_1^n Q^i\right) = d\left(\textstyle\bigvee_1^n P^i\right) \\
&\text{ii)} && d\left(\textstyle\bigvee_1^n \bar{Q}^i\right) = d\left(\textstyle\bigvee_1^n \bar{P}^i\right) \\
&\text{iii)} && \tfrac{1}{n}\textstyle\sum_{i=1}^n |Q^i - \bar{Q}^i| < 4\epsilon.
\end{aligned}
\tag{6.3}
$$

We will show how to construct Q^{n+1} and $\bar{Q}^{n+1}$ so that (6.3) holds for $n+1$ in place of n. We shall do this by defining Q^{n+1} and $\bar{Q}^{n+1}$ on the sets $A \cap \bar{A}$, where $A \in \bigvee_1^n Q^i$, $\bar{A} \in \bigvee_1^n \bar{Q}^i$, so that (6.3i) and (6.3ii) will hold for $n+1$. We shall use the hypotheses (a), (b), and (c) to show that the partitioning result (6.2) can be applied so that (6.3iii) will hold for $n+1$.

First, use (b) to chooses $\mathcal{E}_n \subseteq \bigvee_1^n P^i$ such that $\mu(\cup \mathcal{E}_n) \ge 1 - \epsilon$ and

$$|d(P^{n+1}/B) - d(P^{n+1})| < \epsilon, \quad B \in \mathcal{E}_n.$$

The hypotheses (a) and (c) then imply that for $B \in \mathcal{E}_n$ and $\bar{B} \in \bigvee_1^n \bar{P}^i$ the following holds.

$$|d(P^{n+1}/B) - d(\bar{P}^{n+1}/\bar{B})| < 2\epsilon. \tag{6.4}$$

Let $\mathcal{E}_n^*$ denote the sets in $\bigvee_1^n Q^i$ that correspond to those in $\mathcal{E}_n$. Now apply (6.2) to the space $Y = A \cap \bar{A}$, $A \in \mathcal{E}_n^*$, $\bar{A} \in \bigvee_1^n \bar{Q}^i$, and the two partitions P^{n+1}/B, $\bar{P}^{n+1}/\bar{B}$, where B corresponds to A, and $\bar{B}$ to $\bar{A}$. We thus obtain partitions $Q^{n+1}/A \cap \bar{A}$, and $\bar{Q}^{n+1}/A \cap \bar{A}$ so that

$$(6.5) \qquad \begin{array}{rl} \text{i)} & d(Q^{n+1}/A \cap \bar{A}) = d(P^{n+1}/B), \\ \text{ii)} & d(\bar{Q}^{n+1}/A \cap \bar{A}) = d(\bar{P}^{n+1}/\bar{B}), \\ \text{iii)} & |Q^{n+1}/A \cap \bar{A} - \bar{Q}^{n+1}/A \cap \bar{A}| < 2\epsilon. \end{array}$$

If $A \notin \mathcal{E}_n^*$, we just define $Q^{n+1}/A \cap \bar{A}$ and $\bar{Q}^{n+1}/A \cap \bar{A}$ so that (6.5i) and (6.5ii) hold. The fact that $\mu(\cup \mathcal{E}_n^*) \geq 1 - \epsilon$ then tells us that

$$(6.6) \qquad |Q^{n+1} - \bar{Q}^{n+1}| < 4\epsilon,$$

while (6.5i,ii) and (6.3i,ii) imply that

$$(6.7) \qquad d(\bigvee_1^{n+1} Q^i) = d(\bigvee_1^{n+1} P^i), \quad d(\bigvee_1^{n+1} \bar{Q}^i) = d(\bigvee_1^{n+1} \bar{P}^i).$$

This proves the lemma, for (6.6), (6.3ii), and (6.7) combine to show that (d) holds. In fact, we have established a stronger result; namely, that the Q^i and $\bar{Q}^i$ can be chosen so that, for all i, $|Q^i - \bar{Q}^i| < 4\epsilon$. It is enough for later applications that the average given by (d) holds.

The following result is a restatement of these results for gadgets.

LEMMA 6.4. Suppose $\mathcal{G} = (T, F, n, P)$ and $\bar{\mathcal{G}} = (\bar{T}, \bar{F}, n, \bar{P})$ are gadgets satisfying the following conditions:

a) $d(\bigvee_0^{n-1} T^{-i}P/F) = d(\bigvee_0^{n-1} T^{-i}P)$, $d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P}/\bar{F}) = d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P})$.

b) The seqeunce $\{T^{-i}P/F\}$ is ϵ-independent.

c) The sequence $\{\bar{T}^{-i}\bar{P}/\bar{F}\}$ is independent.

d) $|d(P) - d(\bar{P})| < \epsilon$.

Then

e) $\bar{d}_n(\mathcal{G}, \bar{\mathcal{G}}) < 4\epsilon$.

The above lemma can be restated in terms of the process distance as follows:

LEMMA 6.5. If $\{T^iP\}$ is an ϵ-independent sequence and $\{\bar{T}^i\bar{P}\}$ is an independent sequence where $|d(P) - d(\bar{P})| < \epsilon$, then $\bar{d}((T, P), (\bar{T}, \bar{P})) \leq 4\epsilon$.

This can be proved by using the strong form of Rohlin's Theorem to build gadgets for T and $\bar{T}$ of height n which nearly fill the space, then applying Lemma 6.4. A direct proof that models the proof given in Lemma 6.3 can be found in Chapter 12 (see the proof of Theorem 12.3).

CHAPTER 7. ENTROPY

We now introduce the concept of entropy. The entropy of a transformation T relative to a partition P will be a number H with the property that, for n large enough, one can use binary strings of length $n(H+\epsilon)$ to code unambiguously P-n-names, except for a collection of P-n-names with total probability less than ϵ. The development of these ideas is due to C. Shannon in his fundamental paper on information theory ([19]). The existence of such an H was later rigorously established for ergodic transformations by McMillan ([11]). The entropy H provides the necessary control over the size of atoms in $\bigvee_0^{n-1} T^i P$. In this chapter, we shall define the entropy of T relative to P, then calculate it and discuss one form of McMillan's theorem in the case when $\{T^i\}$ is an independent sequence. We shall then describe some of the general properties of entropy, and establish a theorem relating ϵ-independence and entropy. In the next chapter, it will be shown that, if T is fixed, its relative entropy is largest when P is a generator. This will provide an invariant for transformations, and solve part of the isomorphism problem. Our treatment in both sections will omit many proofs. The reader is referred to Billingsley's excellent book ([3]) for detailed proofs.

Let us begin with the definitions. If $P=\{P_1,P_2,\ldots,P_k\}$ is a partition, then the *entropy* of P is

$$H(P) = -\sum_i \mu(P_i)\log\mu(P_i). \tag{7.1}$$

Any logarithmic base can be used here. We shall always use base 2. The *entropy of T relative to P* is

$$H(T,P) = \lim_n \frac{1}{n} H\Big(\bigvee_1^n T^i P\Big). \tag{7.2}$$

It will be shown later that this limit exists.

These definitions are most easily understood in the independent case. We first prove

LEMMA 7.1. If $\{T^iP\}$ is an independent sequence, then $H(T,P)=H(P)$.

Proof. To prove the lemma, note that $H(P)$ depends on the distribution of P so that $H(TP)=H(P)$ if T is a measure-preserving transformation. The lemma is a consequence of this fact and the following result.

(7.3) If P and Q are independent then $H(P\vee Q)=H(P)+H(Q)$.

To prove this, suppose P and Q are independent. Then $\mu(P_i\cap Q_j)=\mu(P_i)\mu(Q_j)$, so that

$$\log\mu(P_i\cap Q_j) = \log\mu(P_i)+\log\mu(Q_j).$$

Therefore,

$$\begin{aligned} H(P \vee Q) &= -\sum_{i,j} \mu(P_i \cap Q_j) \log \mu(P_i \cap Q_j) \\ &= -\sum_j \mu(Q_j) \sum_i \mu(P_i) \log \mu(P_i) \\ &\quad -\sum_i \mu(P_i) \sum_j \mu(Q_j) \log \mu(Q_j) \\ &= H(P) + H(Q), \end{aligned}$$

since $\sum_j \mu(Q_j) = 1 = \sum_i \mu(P_i)$.

The result (7.3) immediately implies that $H(\bigvee_1^n T^i P) = \sum_{i=1}^n H(T^i P) = nH(P)$, if $\{T^i P\}$ is an independent sequence, hence the lemma is established.

We shall now state a strong form of McMillan's theorem. We give a proof only for the case when $\{T^i P\}$ is an independent sequence. The general proof can be found in Billingsley ([3], pp. 129ff.).

THE SHANNON-McMILLAN-BREIMAN THEOREM. If T is ergodic, P a finite partition, and $\epsilon > 0$, there is an N such that, for $n \geq N$, there is a collection $\mathcal{E}_n$ of atoms in $\bigvee_0^{n-1} T^i P$ such that $\mu(\cup \mathcal{E}_n) \geq 1 - \epsilon$, and

i) $2^{-(h(T,P)+\epsilon)n} \leq \mu(A) \leq 2^{-(h(T,P)-\epsilon)n}$, for $A \in \mathcal{E}_n$.

ii) $\mathcal{E}_n$ contains at *least* $(1-\epsilon)^{-1} 2^{(h(T,P)-\epsilon)n}$ and at most $2^{(h(T,P)+\epsilon)n}$ atoms.

Proof. Note that (ii) is an immediate consequence of (i). To simplify our discussion, we assume that $\{T^i P\}$ is an independent sequence. Suppose A is an atom of $\bigvee_0^{n-1} T^i P$ so that A can be uniquely expressed in the form

$$A = \bigcap_{j=0}^{n-1} T^j P_{i_j}.$$

Let $n_i = n_i(A)$ be the number of occurrences of P_i in this expression for A; that is, $n_i(A)$ is the number of indices j, $0 \leq j \leq n-1$, such that $T^{-j} A \subseteq P_i$. The assumption that $\{T^i P\}$ is an independent sequence tells us that

$$\mu(A) = \mu(P_1)^{n_1} \mu(P_2)^{n_2} \ldots \mu(P_k)^{n_k}.$$

This can be rewritten in the form

$$\log \mu(A) = \sum_{i=1}^k n_i \log \mu(P_i). \tag{7.4}$$

The law of large numbers tells us that, for large n, there is a collection $\mathcal{E}_n$ of atoms of $\bigvee_0^{n-1} T^i P$ such that $\mu(\cup \mathcal{E}_n) \geq (1-\epsilon)$ and if $A \in \mathcal{E}_n$ then

$$\left|\frac{n_i(A)}{n} - \mu(P_i)\right| \leq \delta,\ i = 1, 2, \ldots, k.$$

If δ is small enough, we combine this with (7.4) to obtain our desired conclusion (i).

To facilitate our further discussion of entropy, we introduce the concept of the *conditional entropy* of P given Q. This is defined as

$$H(P|Q) = H(P \vee Q) - H(Q). \tag{7.5}$$

An easy calculation establishes the formula

$$H(P|Q) = -\sum_j \mu(Q_j) \sum_i \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \log \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}. \tag{7.6}$$

This shows that $H(P|Q)$ is the average over the atoms of Q of the entropies of the induced partitions P/Q_j.

It was noted above (see (7.3)) that, if P and Q are independent, then $H(P \vee Q) = H(P) + H(Q)$; that is, $H(P|Q) = H(P)$. It is also obvious from (7.5) that, if $P \subset Q$, then $H(P|Q) = 0$. In fact, the formula (7.6) and the strict convexity properties of $x \log x$ imply the converse of these results. In summary,

$$\begin{gathered} 0 \leq H(P|Q) \leq H(P), \text{ with } H(P|Q) = 0 \text{ if and only if } P \subset Q, \\ \text{and } H(P|Q) = H(P) \text{ if and only if } P \text{ and } Q \text{ are independent.} \end{gathered} \tag{7.7}$$

The function $H(P|Q)$ is decreasing in Q and increasing in P; that is,

$$\begin{aligned} &\text{a)}\quad H(P|Q) \leq H(\bar{P}|Q) \text{ if } P \subset \bar{P}, \\ &\text{b)}\quad H(P|Q) \geq H(P|\bar{Q}) \text{ if } Q \subset \bar{Q}. \end{aligned} \tag{7.8}$$

The definition (7.5) and the results (7.8) enable us to show that $\lim_n n^{-1} H(\bigvee_1^n T^i P)$ indeed exists. First note that $H(\bigvee_1^{n+1} T^i P) = H(\bigvee_0^n T^i P)$ since T is measure-preserving. Formula (7.5) then gives

$$H(\bigvee_0^n T^i P) - H(\bigvee_1^n T^i P) = H(P| \bigvee_1^n T^i P).$$

The sequence $H(P| \bigvee_1^n T^i P)$ is decreasing from (7.8)). It is an elementary exercise to show that, if $\{a_n\}$ is an increasing sequence, $a_n \geq 0$, and $\{a_{n+1} - a_n\}$ is a decreasing sequence, then $\{n^{-1} a_n\}$ converges, and $\lim n^{-1} a_n = \lim(a_{n+1} - a_n)$. This then establishes that the limit used in (7.2) actually exists, and furthermore that

$$H(T, P) = \lim_n H(P| \bigvee_1^n T^i P). \tag{7.9}$$

This result is sometimes summarized by saying that the entropy of T relative to P is the conditional entropy of the present P relative to the entire past. (Remember that the i^{th} coordinate of the P-name of x is the index of the set in $T^{-i}P$ to which x belongs.)

We shall later use the following alternative version of (7.9):

$$H(T,P) = \lim_n H(P|\bigvee_1^n T^{-i}P). \tag{7.9a}$$

This follows from the fact that $H(\bigvee_1^n T^iP) = H(\bigvee_1^n T^{-i}P)$ since T^{n+1} is measure-preserving, and hence $H(T,P) = H(T^{-1},P)$. This with (7.9) establishes (7.9a).

Lemma 7.1 asserted that if $\{T^iP\}$ is an independent sequence then $H(T,P) = H(P)$. The converse of this is also true. In summary,

(7.10) $H(T,P) \le H(P)$, with equality if and only if $\{T^iP\}$ is an independent sequence.

Let us show that $\{T^iP\}$ is indeed an independent sequence if $H(T,P) = H(P)$. The hypothesis $H(T,P) = H(P)$ combines with (7.9) to tell us that $H(P|\bigvee_1^n T^iP) = H(P)$ for $n \ge 1$, and hence (7.9) implies that P is independent of $\bigvee_1^n T^iP$ for $n \ge 1$. This proves (7.10). It is important for our later results that this result has an approximate form.

LEMMA 7.2. Given k and $\epsilon > 0$, there is a $\delta = \delta(\epsilon, k) > 0$ such that if P has k sets and $H(T,P) \ge H(P) - \delta$ then $\{T^iP\}$ is an ϵ-independent sequence; that is, for each n, T^nP is ϵ-independent of $\bigvee_0^{n-1} T^iP$.

Proof. Smorodinsky ([22]) showed that δ can be chosen to be independent of k also. We shall give here the simpler proof of Ornstein for the case when δ is allowed to depend upon k. We first note that (7.9a) gives

$$H(T^nP) - H(T^nP|\bigvee_0^{n-1} T^iP) = H(P) - H(P|\bigvee_1^n T^{-i}P) \le H(P) - H(T,P),$$

so it is enough to prove the following lemma.

LEMMA 7.3. Given k and $\epsilon > 0$, there is a $\delta > 0$ such that, if P has k sets and $H(P) - H(P/Q) \le \delta$, then P is ϵ-independent of Q.

Proof of Lemma 7.3. We would like to show that $H(P) - H(P|Q)$ is bounded away from zero on the set of all pairs (P,Q) such that P has k sets and P is *not* ϵ-independent of Q. We first show how Q can be replaced by a two set partition. Suppose P has k sets and is *not* ϵ-independent of Q, so that the collection $\mathcal{E}$ of atoms A of Q for which

$$|d(P/A) - d(P)| \ge \epsilon,$$

has total measure greater than ϵ. Thus

$$\sum_i \sum_{A\in\mathcal{E}} |\mu(P_i\cap A) - \mu(P_i)\mu(A)| \geq \epsilon^2,$$

so that there is a P_i and a subcollection $\mathcal{E}'\subseteq\mathcal{E}$ such that

$$|\sum_{A\in\mathcal{E}'} \mu(P_i\cap A) - \mu(P_i)\mu(A)| \geq \epsilon^2/2k.$$

Let S be the union of the sets in $\mathcal{E}'$, and note that

$$\mu(S) \geq \epsilon^2/2k \text{ and } |d(P/S) - d(P)| \geq \epsilon^2/2k. \tag{7.11}$$

Since $\mathcal{S} = \{S, S^c\}$ is refined by Q, we have $H(P|\mathcal{S}) \geq H(P|Q)$. Furthermore, $\mathcal{S}$ is not independent of P, so that $H(P) > H(P|\mathcal{S})$. We therefore have

$$0 < H(P) - H(P|\mathcal{S}) \leq H(P) - H(P|Q). \tag{7.12}$$

Let K be the set of all $(3k+1)$-tuples

$$d(P), d(P/S), d(P/S^c), \mu(S),$$

where P is a k set partition and (7.11) holds. The set K is compact, and $H(P) - H(P|\mathcal{S})$ is a continuous non-vanishing function on K, hence it must be bounded away from 0. This, along with (7.12), shows that Lemma 7.3, and hence Lemma 7.2, is true.

If Lemma 7.2 is combined with Lemma 6.4, we obtain the following fundamental result.

ORNSTEIN'S COPYING THEOREM. Suppose $\bar{T}$ is a Bernoulli shift with independent generator $\bar{P}$, where $\bar{P}$ has k sets. Given $\epsilon > 0$, there is a $\delta > 0$ such that, if T is any ergodic transformation and P any k set partition such that

$$\text{(i) } |d(P) - d(\bar{P})| < \delta \text{and (ii) } |H(\bar{P}) - H(T, P)| < \delta,$$

then

$$\text{(iii) } \bar{d}((\bar{T},\ \bar{P}),\ (T, P)) \leq \epsilon.$$

Proof. Given $\delta' > 0$, if δ is small enough, we have $|H(P) - H(\bar{P})| < \delta'$, so that (i) and (ii) will imply that $|H(P) - H(T, P)| < \delta + \delta'$. Thus Lemma 7.2 tells us that, if $\delta + \delta'$ is small enough, $\{T^iP\}$ will then be an ϵ-independent sequence, so that, if $\delta < \epsilon$, Lemma 6.4 will imply that $\bar{d}((\bar{T}, \bar{P}), (T, P)) \leq 4\epsilon$. Replace ϵ by $\epsilon/4$, and (iii) is then established.

The conclusion (iii) of this theorem means that there is a measure-preserving transformation S from the T-space to the $\bar{T}$-space such that, except for a set of measure less than $\sqrt{\epsilon}$ (which depends on n), the P-n-name of x and the $\bar{P}$-n-name of Sx disagree in less than $\sqrt{\epsilon}n$ place. All that is required for this to hold is that the pair T, P be close enough to the pair $\bar{T}, \bar{P}$ in distribution and entropy. This result will be very useful in our proof of the isomorphism theorem. Later it will also be shown that, if $\bar{P}$ is *any* generator for a Bernoulli shift $\bar{T}$, a slightly weaker version of the copying theorem will hold. This result will, in fact, be the characterization of Bernoulli shifts which enables one to show that many transformations are Bernoulli even when one cannot explicitly construct an independent generator.

CHAPTER 8.
THE ENTROPY OF A TRANSFORMATION

The entropy $H(T,P)$ of a transformation T relative to a partition P was defined in the previous section. The number $H(T,P)$ depends upon the partition P. To obtain an invariant for T, we define the *entropy* of T as

$$H(T) = \sup\{H(T,P)\colon\ P \text{ is a finite partition}\}. \tag{8.1}$$

This is clearly an invariant for T; that is, if S is isomorphic to T, then $H(S) = H(T)$. At first glance, one might think that $H(T)$ would always be infinite. The following result of Komogorov and Sinai ([9], [10], [20]) gives us a means for calculating $H(T)$, and establishes that $H(T)$ is finite for a large class of transformations.

KOLMOGOROV-SINAI THEOREM. If P is a generator for T, and Q is any partition, then $H(T,P) \geq H(T,Q)$. In particular, $H(T) = H(T,P)$ for any generator P.

Proof. The proof of this result depends upon two lemmas.

LEMMA 8.1. If $Q = \bigvee_{-k}^{k} T^iP$, then $H(T,Q) = H(T,P)$.

Proof. We have

$$H(T,Q) = \lim_n H(Q|\bigvee_1^n T^iQ) = \lim_n H(\bigvee_{-k}^{k} T^iP| \bigvee_{-k+1}^{n+k} T^iP).$$

Note that the definition of conditional entropy gives

$$H(\bigvee_{-k}^{k} T^iP| \bigvee_{-k+1}^{n+k} T^iP) = H(\bigvee_{-k}^{n+k} T^iP) - H(\bigvee_{-k+1}^{n+k} T^iP),$$

which is equal to $H(T^{-k}P| \bigvee_{-k+1}^{n+k} T^iP)$. Replace P by T^kP to obtain

$$H(T,Q) = \lim_n H(P| \bigvee_1^{n+2k} T^iP) = H(T,P).$$

This proves Lemma 8.1.

LEMMA 8.2. Given k and $\epsilon > 0$, there is a $\delta > 0$ such that if P and Q have k sets and $|P - Q| \leq \delta$ then $|H(T,P) - H(T,Q)| \leq \epsilon$.

Proof. Fix $\delta < 1/2$, and suppose $|P - Q| < \delta$. Let

$$R_0 = \bigcup_{i=1}^{k} P_i \cap Q_i,\ R_i = P_i - R_0,\ 1 \leq i \leq k,$$

and let R denote the partition $\{R_0, R_1, R_2, \ldots, R_k\}$. The strict convexity of $-x \log x$, $0 \leq x \leq 1$, then shows that, for R_0 fixed, the largest value of $H(R)$ is obtained when $\mu(R_1) = \mu(R_2) = \ldots = \mu(R_k)$. Thus

$$H(R) \leq -\delta \log \delta - (1-\delta)\log(1-\delta) + \delta \log k,$$

so we can choose δ so small that $H(R)$ cannot exceed ϵ. Furthermore, $P \vee Q = Q \vee R$, so that

$$H(T, P) \leq H(T, P \vee Q) = H(T, Q \vee R),$$

and the latter does not exceed $H(T, Q) + H(R)$, so we must have $H(T, P) \leq H(T, Q) + \epsilon$. A similar argument shows that, if δ is small enough, then $H(T, Q) \leq H(T, P) + \epsilon$, and this completes the proof of Lemma 8.2.

Proof of the Kolmogorov-Sinai Theorem. Suppose P is a generator for T, and Q is some arbitrary finite partition. Given $\delta > 0$, we can find a large k and a partition $\bar{Q}$ with the same number of sets as Q, such that $\bar{Q} \subseteq \bigvee_{-k}^{k} T^i P$ and $|\bar{Q} - Q| \leq \delta$. This follows from the hypothesis that P is a generator and the assumption that Q has finitely many sets, for, given any atom A of Q, the set A can be approximated by sets in the σ-algebra generated by $\bigvee_{-k}^{k} T^i P$ for large enough k.

Since $\bar{Q} \subseteq \bigvee_{-k}^{k} T^i P$, Lemma 8.1 implies that $H(T, \bar{Q}) \leq H(T, P)$, while Lemma 8.2 guarantees that, if δ is small enough, $H(T, \bar{Q})$ will be close to $H(T, Q)$. Thus we must have $H(T, Q) \leq H(T, P)$, which completes the proof of the theorem.

If T is a Bernoulli shift, then the Kolmogorov-Sinai Theorem and Lemma 7.1 enable one to compute the entropy of T. Suppose $P = \{P_1, P_2, \ldots, P_k\}$ is a generator for T such that $\{T^i P\}$ is an independent sequence. We then have

$$H(T) = -\sum \mu(P_i) \log \mu(P_i). \tag{8.2}$$

To prove this, use the Kolmogorov-Sinai Theorem to obtain $H(T) = H(T, P)$, and then use Lemma 7.l to obtain $H(T, P) = H(P)$.

At this point, note that these results solve part of the isomorphism problem; namely, two Bernoulli shifts cannot possibly be isomorphic unless they have the same entropy. For example, the Bernoulli shifts T and $\bar{T}$, with respective distributions $\pi = (1/2, 1/2)$ and $\bar{\pi} = (1/3, 1/3, 1/3)$, can not be isomorphic for $H(T) = \log 2$ and $H(\bar{T}) = \log 3$.

For any given $\pi = (p_1, p_2, \ldots, p_k)$, there are an uncountable number of distributions $\bar{\pi} = (\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_{\bar{k}})$ such that

$$\sum p_i \log p_i = \sum \bar{p}_i \log \bar{p}_i, \tag{8.3}$$

and the Kolmogorov-Sinai Theorem gives us no positive information about whether the two Bernoulli shifts T_π and $T_{\bar{\pi}}$ are isomorphic. Meshalkin ([12]) and later Blum and Hanson ([4]) developed special coding techniques for establishing isomorphisms when, in addition to (8.3), the p_i and $\bar{p}_j$ satisfy special algebraic relations. For example, Meshalkin showed that, if

$$\pi = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) \text{ and } \bar{\pi} = (\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}),$$

T_π and $T_{\bar{\pi}}$ are isomorphic.

It will be shown in the next two sections that (8.3) is sufficient for isomorphism of the Bernoulli shifts T_π and $T_{\bar{\pi}}$; that is, entropy is a complete invariant for Bernoulli shifts. We sketch here some of the background of the proof of this theorem.

Suppose T and $\bar{T}$ are Bernoulli shifts with the same entropy. We can therefore find generators P and $\bar{P}$, respectively, such that $\{T^i P\}$ and $\{\bar{T}^i \bar{P}\}$ are each independent sequences. Furthermore, $H(T) = H(P)$ and $H(\bar{T}) = H(\bar{P})$, so we must have $H(P) = H(\bar{P})$. To prove the isomorphism of T and $\bar{T}$, it is enough (from Theorem 2.1) to find a partition Q such that

(8.4)
a) Q is a generator for T,
b) $\{T^i Q\}$ is an independent sequence,
c) $d(Q) = d(\bar{P})$.

At this point, we mention that Sinai ([20]) showed how to find Q such that (b) and (c) hold. His construction is so difficult that it is not easy to show how one might choose Q so that (a) will also hold. Ornstein established a much stronger version of Sinai's result, showing how one can choose a Q satisfying (b) and (c), which is not too far away from a $\bar{Q}$ that almost satisfies (b) and (c). Precise statements of these results are stated here.

SINAI'S THEOREM. If $\bar{T}$ is a Bernoulli shift with independent generator $\bar{P}$, and if T is any ergodic transformation such that $H(T) \geq H(\bar{T})$, then there is a partition Q satisfying

a) $\{T^i Q\}$ is an independent sequence,

b) $d(Q) = d(\bar{P})$.

ORNSTEIN'S FUNDAMENTAL LEMMA. If $\bar{T}$ is a Bernoulli shift with independent generator $\bar{P}$, and $\epsilon > 0$, there is a $\delta > 0$ such that, if T is any ergodic transformation with $H(T) \geq H(\bar{T})$ and P is any partition with the same number of sets as $\bar{P}$ such that

i) $|d(P) - d(\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

then there is a partition Q satisfying the following three conditions.

iii) $\{T^i Q\}$ is an independent sequence,

iv) $d(Q) = d(\bar{P})$.

v) $|Q - P| \leq \epsilon$.

Note that the fundamental lemma asserts that, once we have found a P satisfying (i) and (ii), then close to P (in the partition metric) is a partition satisying conditions (a) and (b) of Sinai's Theorem. Gadget constructions can be used to obtain partitions satisfying (i) and (ii), hence we can control the location of partitions satisfying (iii) and (iv). This will enable us then to modify Q so that it will satisfy (iii) and (iv) and "almost" generate, and condition (v) will guarantee that our sequence of modifications will converge in the partition metric.

The next chapter will contain a proof of the fundamental lemma. The number δ will come from the copying theorem of Chapter 7; that is, δ will be chosen so that (i) and (ii) guarantee that $\bar{d}((T, P), (\bar{T}, \bar{P}))$ is small. This will enable us to copy gadgets involving $\bar{T}, \bar{P}$ onto those involving T, P and will be the key to controlling the location of Q. Underlying these constructions will be the following principle.

(8.5) In order to be certain that $H(T, Q)$ is close to $H(T)$, choose Q so that, for some n, $\bigvee_{-n}^{n} T^i Q \supseteq \bar{Q}$, where $\bar{Q}$ is close to a generator P for T.

Thus, if $|P - \bar{Q}| \leq \delta$ and δ is small enough, Lemma 8.2 and the Kolmogorov-Sinai Theorem imply that $H(T, \bar{Q})$ will be close to $H(T, P) = H(T)$. Lemma 8.1 then implies that $H(T, Q) \geq H(T, \bar{Q})$, so $H(T, Q)$ will indeed be close to $H(T)$.

CHAPTER 9.
THE FUNDAMENTAL LEMMA

In this section, we shall prove the fundamental lemma. For ease of reference, we restate this result here.

ORNSTEIN'S FUNDAMENTAL LEMMA. If $\bar{T}$ is a Bernoulli shift with independent generator $\bar{P}$, and $\epsilon > 0$, there is a $\delta > 0$ such that, if T is any ergodic transformation with $H(T) \geq H(\bar{T})$ and P is any partition with the same number of sets as $\bar{P}$ such that

i) $|d(P) - d(\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

there is a Q satisfying the following three conditions

iii) $\{T^i Q\}$ is an independent sequence,

iv) $d(Q) = d(\bar{P})$,

and

v) $|Q - P| \leq \epsilon$.

In order to facilitate the understanding of the proof of this lemma, we first prove a much simpler result, which shows that, for any $\delta > 0$, there is a P satisfying (i) and (ii).

LEMMA 9.1. If $\bar{T}$ is a Bernoulli shift with independent generator $\bar{P}$, and T is any ergodic transformation with $H(T) \geq H(\bar{T})$, then, for any given $\delta > 0$, there is a partition Q with the same number of sets as $\bar{P}$ such that the following two conditions hold.

i) $|d(Q) - d(\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, Q) \leq \delta$.

Proof. The key to the proof is the use of gadgets to select new partitions, Lemma 4.2 and the Shannon-McMillan-Breiman Theorem to control entropy, and the law of large numbers to control the distribution. We first choose a partition R with good entropy, that is, so that

$$0 < H(\bar{T}, \bar{P}) - H(T, R) < \alpha, \tag{9.1}$$

where α is a small positive number to be specified later.

The existence of R follows from the fact that $H(T,R)$ is a continuous function of R in the partition metric, relative to which the set of all partitions is connected. Of course, neither the number of sets nor the distribution of R have any relation to $\bar{P}$.

Our goal is to construct a gadget (T,F,n,R) and relabel its columns with $\bar{P}$-n-names in such a way that the resulting partition Q satisfies (i) and (ii). To do this, we need to control the number of columns in the gadget. The Shannon-McMillan-Breiman (SMB) Theorem provides us with this control. Let us write $\mu(A) \sim 2^{-n(H\pm\epsilon)}$ if

$$2^{-n(H+\epsilon)n} \le \mu(A) \le 2^{-n(H-\epsilon)}.$$

Let β be a small positive number to be specified later, and use the SMB theorem to choose n so large that the there is a collection $\mathcal{E} \subset \bigvee_0^{n-1} R$ and a collection $\bar{\mathcal{E}} \subset \bigvee_0^{n-1} \bar{T}^{-1}\bar{P}$, each of total measure at least $1-\beta$, such that

$$\text{(9.2)} \qquad \begin{aligned} &\text{(a)}\ \ \mu(A) \sim 2^{-n(H(T,R)\pm\beta)},\ \ A \in \mathcal{E}. \\ &\text{(b)}\ \ \mu(\bar{A}) \sim 2^{-n(H(\bar{T},\bar{P})\pm\beta)},\ \ A \in \bar{\mathcal{E}}. \end{aligned}$$

By choosing β small enough and n large enough, we can assume that

$$\text{(9.3)} \qquad 2^{-(H(\bar{T},\bar{P})-\beta)n} \le 2^{-(H(T,R)+\beta)n}.$$

This uses the assumption that $H(T,R) < H(\bar{T},\bar{P})$. The inequality (9.3) has the consequence that, if β is small enough,

(9.4) There are more sets in $\bar{\mathcal{E}}$ than in $\mathcal{E}$.

We can sharpen this result even further by using the law of large numbers. Let $f_{\bar{A}}(i,n)$ be the relative frequency of i in the $\bar{P}$-n-name of the atom $\bar{A} \in \bigvee_0^{n-1} \bar{T}^{-i}\bar{P}$. We can assume that n and $\bar{\mathcal{E}}$ satisfy

$$\text{(9.5)} \qquad \sum_{\bar{P}_i \in \bar{P}} |f_{\bar{A}}(i,n) - \mu(\bar{P}_i)| \le \beta, \quad \bar{A} \in \bar{\mathcal{E}}.$$

The strong form of Rohlin's Theorem implies that there is a gadget $(T,\tilde{F},n,R)$ such that

$$\text{(9.6)} \qquad d(\bigvee_0^{n-1} T^{-1}R/\tilde{F}) = d(\bigvee_0^{n-1} T^{-i}R) \text{ and } \mu(\bigcup_{i=0}^{n-1} T^i\tilde{F}) \ge 1-\beta.$$

We can cut down the number of columns in this gadget by replacing $\tilde{F}$ with $F = (\cup\mathcal{E}) \cap \tilde{F}$. Put $\mathcal{E}_F = \{A \cap F | A \in \mathcal{E}\}$. The conditions (9.6), (9.4), and (9.2a) imply that $\mu(\cup_{i=0}^{n-1} T^i F) > 1 - 2\beta$, and that there are more sets in $\bar{\mathcal{E}}$ than in $\mathcal{E}_F$. Thus

(9.7) There is a one-to-one function ϕ from the columns of (T, F, n, R) into the set of $\bar{P}$-n-names of the atoms in $\bar{\mathcal{E}}$.

As in Chapter 4, for $A \in \mathcal{E}_F$, we label $T^m A$ with i_m, if $\phi(A) = (i_0, i_1, \ldots, i_{n-1})$, and let Q_i be the union of the sets labelled i. This defines the partition Q on $\bigcup_{i=0}^{n-1} T^i F$, and one can then define Q on the complement of $\bigcup_{i=0}^{n-1} T^i F$ in some arbitrary fashion. Let us show that if α and β are small enough and n is large enough the following hold.

$$\text{(i) } |d(Q) - d(\bar{P})| \leq \delta \text{ and (ii) } 0 \leq H(\bar{T}, \bar{P}) - H(T, Q) \leq \delta.$$

The fact that (i) will hold follows from (9.5), for, if C is any column of (T, F, n, R), (9.5) implies that

$$\sum_i |\mu(Q_i \cap C) - \mu(\bar{P}_i)\mu(C)| \leq \beta\mu(C).$$

Thus (i) will indeed hold if β is small enough. The proof of (ii) is as follows: Lemma 4.2 implies that, on the set $\cup_{i=0}^{n-1} T^i F$, we have

$$R \subset \bigvee_{-n}^{n} T^i(Q \vee \mathcal{F}), \; \mathcal{F} = \{F, F^c\}.$$

If β is small enough, there is an

$$R' \subset \bigvee_{-n}^{n} T^i(Q \vee \mathcal{F})$$

such that R' is close to R, and hence we can make $H(T, R')$ close to $H(T, R)$. Since $R' \subset \bigvee_{-n}^{n} T^i(Q \vee \mathcal{F})$, we will also have $H(T, R') \leq H(T, Q \vee \mathcal{F})$, and the latter will be close to $H(T, Q)$ since $H(\mathcal{F})$ is very small. Thus if n is sufficiently large and α and β are sufficiently small, both (i) and (ii) will hold. This proves Lemma 9.1.

The choice of ϕ in (9.7) is completely arbitrary. All that is required is that it be a one-to-one function from the columns of (T, F, n, R) into the set of $\bar{P}$-n-names of atoms in $\bar{\mathcal{E}}$. Once ϕ is selected, it determines a Q. If we would like to Q to lie close to a given P, we need to select ϕ with some care. The next lemma shows that, if P has good enough distribution and entropy, then ϕ can be chosen so that P-n-names and Q-n-names of most points in the base of the gadget agree in most places. This will imply that P and Q are close.

LEMMA 9.2. If $\bar{T}$ is a Bernoulli shift with independent k-set generator $\bar{P}$, and $\epsilon > 0$, there is a $\delta > 0$ with the following properties: If T is any ergodic transformation with $H(T) \geq H(\bar{T})$, and P is any k-set partition such that

$$\text{(i) } |d(P) - d(\bar{P})| \leq \delta, \text{ and (ii) } 0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta,$$

then, for any $\bar{\delta} \leq 0$, there is a k-set partition Q such that

$$\text{(iii) } |d(Q) - d(\bar{P})| \leq \bar{\delta}, \quad \text{(iv) } 0 \leq H(\bar{T}, \bar{P}) - H(T, Q) \leq \bar{\delta},$$

and

$$\text{(v) } |P - Q| \leq \epsilon.$$

Proof. We are going to proceed much as we did in the proof of Lemma 9.1, then make use of the gadget metric to copy $\bar{T}, \bar{P}$ close to T, P and a marriage lemma to show that ϕ can be chosen so that Q will be close to P. The number δ comes from the Ornstein Copying Theorem in Chapter 7. We use that theorem to find $\delta > 0$ such that, if P satisfies (i) and (ii), then

$$\bar{d}((\bar{T}, \bar{P}), (T, P)) < \bar{\epsilon}, \tag{9.8}$$

where $\bar{\epsilon}$ will be specified later.

Let T, P be given satisfying (i) and (ii) and hence (9.8). Without loss of generality, we can also assume that $\delta < \epsilon$. Furthermore, it can be supposed that

$$0 < H(\bar{T}, \bar{P}) - H(T, P) \leq \delta. \tag{9.9}$$

If this were not so, we could modify P by a small amount in the σ-algebra generated by $\bigcup_{-\infty}^{\infty}\{T^i P\}$. Either there will be a modification satisfying (9.9), or all modifications satisfy $H(\bar{T}, \bar{P}) = H(T, Q)$. In the latter case, we merely choose Q so that (v) holds, and $d(Q) = d(P)$. This Q would then satisfy (iii) and (iv) for all $\bar{\delta} > 0$, and we would be finished. Thus we can assume that (9.9) holds.

Now choose $R \supset P$ such that

$$0 < H(\bar{T}, \bar{P}) - H(T, R) < \alpha, \tag{9.1$'$}$$

where α is to be specified later. To do this, select R satisfying (9.1), and replace R by $R \vee P$. Continuity of entropy and connectedness of the space of partitions which refine P then imply that (9.1$'$) can be achieved for some $R \supset P$.

Choose β so small and n so large that (9.2), (9.3), (9.4), and (9.5) all hold, then choose $(T, \tilde{F}, n, R)$ such that (9.6) holds. Since R refines P, (9.6) implies

$$d\Big(\bigvee_0^{n-1} T^{-i} P\Big) = d\Big(\bigvee_0^{n-1} T^{-i} P / \tilde{F}\Big). \tag{9.6$'$}$$

Furthermore, each atom $A \in \bigvee_0^{n-1} T^{-i}R/\tilde{F})$ is contained in a unique atom of $\bigvee_0^{n-1} T^{-i}P$; hence such atoms have P-n-names.

Let $F = \cup\mathcal{E} \cap \tilde{F}$, $\mathcal{F} = \{A \cap F | A \in \mathcal{E}\}$. As before, $\mu(\bigcup_0^{n-1} T^iF) > 1 - 2\beta$, and there are more atoms in $\bar{\mathcal{E}}$ than in $\mathcal{E}_F$. We shall prove that there is a function ϕ satisfying (9.7) *and* the following condition.

(9.10) There is a collection $\mathcal{A} \subseteq \mathcal{E}_F$, with $\mu(\cup\mathcal{A}) \geq (1-\epsilon/6)\mu(F)$ such that the P-n-name of $A \in \mathcal{A}$ and the P-n-name of $\phi(A) \in \bar{\mathcal{E}}$ disagree in less than $\frac{\epsilon}{6}n$ places.

If (9.10) is true, on each column C_A, $A \in \mathcal{A}$, there will be at most $\frac{\epsilon}{6}n$ indices i for which $Q_i \cap C_A \neq P_i \cap C_A$, and hence

$$\sum_{i=1}^{k} \mu[(Q_i \cap C_A) \triangle (P_i \cap C_A)] \leq \frac{2\epsilon}{6}\mu(C_A).$$

It follows that if $G = \bigcup_{i=0}^{n-1} T^iF$ then

$$|Q/G - P/G| \leq \frac{2\epsilon}{3},$$

and hence $|Q - P|$ will not exceed ϵ, if $2\beta < \epsilon/6$. Thus once (9.10) is proved, Lemma 9.2 will be established.

To establish (9.10), we make use of the copying condition (9.8). First, choose a gadget $(\bar{T}, \bar{F}, n, \bar{P})$ so that

$$d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P}/\bar{F}) = d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P}),$$

and then (using Lemma 5.4 along with (9.8)) choose P^* so that $(T, \tilde{F}, n, P^*)$ is isomorphic to $(\bar{T}, \bar{F}, n, \bar{P})$ and

$$\frac{1}{n}\sum_{i=0}^{n-1} |P/T^i\tilde{F} - P^*/T^I\tilde{F}| < \bar{\epsilon}.$$

Now replace $\tilde{F}$ by F. If β is small enough, we will then have

(9.11)

$$\text{(a)} \quad |d(\bigvee_0^{n-1} \bar{T}^{-i}\bar{P}/\bar{F}) - d(\bigvee_0^{n-1} T^{-i}P^*/F)| < \beta,$$

$$\text{(b)} \quad \frac{1}{n}\sum_{i=0}^{n-1} |P/T^iF - P^*/F^iF| < 2\bar{\epsilon}$$

Let $\mathcal{E}_F^*$ denote the atoms in $\bigvee_0^{n-1} T^{-i}P^*/F$ corresponding to those in $\bar{\mathcal{E}}$. If β is sufficiently small, then (9.4) implies that there are more sets in $\mathcal{E}_F^*$ than in $\mathcal{E}_F$. Our final arguments will be simplified if we assume that β is so small that

$$\mu(A) \geq 4\mu(A^*),\ A \in \mathcal{E}_F,\ A^* \in \mathcal{E}_F^*, \tag{9.12}$$

so that at least four sets in $\mathcal{E}_F^*$ are needed to cover a set in $\mathcal{E}_F$.

Let $\mathcal{A}$ be the class of all sets $A \in \mathcal{E}_F$ such that more than half of A is covered by sets $A^* \in \mathcal{E}_F^*$ such that the P-n-name of A and the P^*-n-name of A^* differ in no more than $n\sqrt{2\epsilon}$ place. First, we shall show that $\cup\mathcal{A}$ fills up most of F. Towards this end, let E be the set of points $x \in F$ such that the P-n-name and P^*-n-name of x disagree in more than $n\sqrt{2\epsilon}$ places. Lemma 5.1 implies that $\mu(E) \leq \sqrt{2\epsilon}\mu(F)$. Furthermore, if $B = \cup\mathcal{E}_F^*$, it follows easily that

$$\sum_{\substack{A\in\mathcal{E}_F \\ A\notin\mathcal{A}}} \mu(A) \leq 2[\mu(F-B) + \mu(E)],$$

and hence, if $\bar{\epsilon}$ is small enough and β is small enough, we shall indeed have

$$\mu(\cup\mathcal{A}) \geq (1-\epsilon/6)\mu(F).$$

To complete the proof of (9.10), we make use of Hall's Matching Lemma (See [6], p. 45). Clearly, the definition of $\mathcal{A}$ and property (9.12) imply that any t elements A_1, A_2, …, A_t in $\mathcal{A}$ intersect at least t elements in $\mathcal{E}_F^*$ whose P^*-n-name differs from the P-n-name of at least one of the A_i in no more than $n\sqrt{2\epsilon}$ places. The marriage lemma thus implies that there is a one-to-one function ϕ from $\mathcal{A}$ to $\mathcal{E}_F^*$ such that the P-n-name of $A \in \mathcal{A}$ and the P^*-n-name of $\phi(A)$ differ in no more than $n\sqrt{2\epsilon}$ places. Thus defining ϕ in any arbitrary one-to-one manner on $\mathcal{E}_F - \mathcal{A}$, property (9.10) will hold for small $\bar{\epsilon}$. The resulting Q will then satisfy (v), for the distribution of $\bar{P}$-n-names is nearly the same as the distribution of P^*-n-names (from (9.11)). The proof of Lemma 9.1 shows that, if α and β are small enough and n is large enough, Q will also saitsfy (iii) and (iv). This completes the proof of Lemma 9.2.

Proof of Ornstein's Fundamental Lemma. Let δ be the number given by Lemma 9.2 for $\epsilon/2$, suppose T is an ergodic transformation with $H(T) \geq H(\bar{T})$, and P satisfies (i) and (ii) of Lemma 9.2. Let δ_n be the number given by Lemma 9.2 for $\epsilon/2^n$, and by induction choose k-set partitions $Q^{(n)}$ such that $Q^{(0)} = P$ and

a) $|d(Q^{(n)}) - d(\bar{P})| \leq \delta_{n+1}$,

b) $0 \leq H(\bar{T},\bar{P}) - H(T,Q^{(n)}) \leq \delta_{n+1}$,

c) $|Q^{(n)} - Q^{(n-1)}| \leq \epsilon/2^n$.

The limiting partition $Q = \lim_n Q^{(n)}$ then satisfies the three conditions

d) $d(Q) = d(\bar{P})$,

e) $H(T, Q) = H(\bar{T}, \bar{P})$,

f) $|Q - P| \leq \epsilon$.

The conditions (d) and (e) imply (see (7.10)) that $\{T^i Q\}$ is an independent sequence. This completes the proof of the fundamental lemma.

CHAPTER 10.
THE ISOMORPHISM THEOREM

In this chapter, we shall complete the proof that two Bernoulli shifts with the same entropy are isomorphic. First, we introduce some convenient shorthand.

If P and Q are partitions, then we write $Q \subset_\epsilon P$ if there is a partition $\bar{Q} \subset P$ with the same number of sets as Q such that $|\bar{Q} - Q| \le \epsilon$.

LEMMA 10.1. *P is a generator for T if and only if, for each partition Q and each $\epsilon > 0$, there is an n such that $Q \subset_\epsilon \bigvee_{-n}^{n} T^i P$.*

Proof. To say that P is a generator is to say that, for each set A of positive measure and each $\epsilon > 0$, there is an n and a set B in the σ-algebra generated by $\bigvee_{-n}^{n} T^i P$ such that $\mu(A \triangle B) \le \epsilon$. This is easily seen to be equivalent to the lemma.

We shall say that T, P is a *copy* of $\bar{T}, \bar{P}$, and write $(T, P) \sim (\bar{T}, \bar{P})$ if

$$d(\bigvee_0^n T^i P) = d(\bigvee_0^n \bar{T}^i \bar{P}), \quad n = 0, 1, 2, \dots .$$

This means that, for each n, the distribution of P-n-names is the same as the distribution of $\bar{P}$-n-names. Note for example that, if $\{\bar{T}^i \bar{P}\}$ is an independent sequence and $(T, P) \sim (\bar{T}, \bar{P})$, then $d(P) = d(\bar{P})$, and $\{T^i P\}$ is an independent sequence.

Throughout the remainder of this section, *T and $\bar{T}$ will denote Bernoulli shifts with the same entropy, with respective independent generators P and $\bar{P}$.* We shall prove

THEOREM 10.1. *There is a partition Q such that*

i) $(T, Q) \sim (\bar{T}, \bar{P})$,

ii) Q is a generator for T.

Theorem 10.1 and the Kolmogorov-Sinai Theorem then imply the Kolmogorov-Ornstein Theorem.

There is no problem constructing Q so that $(T, Q) \sim (\bar{T}, \bar{P})$. In fact, one can use Lemma 9.1 to construct Q', with $d(Q')$ very close to $d(\bar{P})$ and $H(T, Q')$ very close to $H(\bar{T}, \bar{P})$. Then apply Ornstein's Fundamental Lemma of Chapter 9 to construct Q close to Q' so that $(T, Q) \sim (\bar{T}, \bar{P})$. Our goal is to show that such a Q can be modified so that (i) and (ii) of Theorem 10.1 will both hold. We first introduce some further notation.

The transformation T is defined on (X, Σ, μ) where Σ is generated by the sets in $\bigcup_{-\infty}^{\infty}\{T^n P\}$. If Q is another partition of X, we let Σ_Q denote the complete σ-algebra generated by $\bigcup_{-\infty}^{\infty}\{T^n Q\}$, and let T_Q denote the restriction of T to the measure space (X, Σ_Q, μ). If $(T, Q) \sim (\bar{T}, \bar{P})$, then, of course, T_Q is a Bernoulli shift with independent generator Q. (This is a slight abuse of terminology. The σ-algebra Σ_Q will not, if $\Sigma_Q \neq \Sigma$, separate the points of X, so (X, Σ_Q, μ) is not a Lebesgue space. Throughout this discussion, we are tacitly making the identification between sets in Σ_Q and their projections onto the factor space (X_Q, Σ_Q, μ), which is a Legesgue space, defined by Σ_Q (see Chapter 1).)

We also need some notation to clarify the relationship between partitions refined by $\bigvee_{-n}^{n} T^i P$ and P-names. Suppose P has k sets. The correspondence $A \leftrightarrow \{t_i(A)\}$ where

$$A = \bigcap_{-n}^{n} T^i P_{t_i}$$

is a one-to-one correspondence between atoms $A \in \bigvee_{-n}^{n} T^i P$ and sequences $\{t_i\}$ of the symbols $\{1, 2, \ldots, k\}$ of length $2n+1$. (To connect to our prior terminolgy note that, if $s_{n-i} = t_i$, then $\{s_i\}$ is the P-$(2n+1)$-name of $T^{-n}A$.) The correspondence $A \leftrightarrow \{t_i(A)\}$ induces a correspondence between partitions of $(2n+1)$-sequences and partitions refined by $\bigvee_{-n}^{n} T^i P$. Thus, if $\Pi = \{\Pi_1, \Pi_2, \ldots, \Pi_m\}$ is a partition of $(2n+1)$-sequences, then $\Pi(P) = \{\Pi_1(P), \Pi_2(P), \ldots, \Pi_m(P)\}$ is a partition refined by $\bigvee_{-n}^{n} T^i P$, where $\Pi_j(P)$ is the union of those atoms $A \in \bigvee_{-n}^{n} T^i P$ such that $\{t_i(A)\} \in \Pi_j$. Furthermore, all partitions refined by $\bigvee_{-n}^{n} T^i P$ are of this form.

We now prove

LEMMA 10.2. Suppose $(T, Q) \sim (\bar{T}, \bar{P})$, and let $\epsilon > 0$. Assume P has k sets. Choose N so that $Q \subset_\epsilon \bigvee_{-N}^{N} T^i P$. Then there is a partition $\tilde{P}$ such that

i) $(T, \tilde{P}) \sim (T, P)$,

ii) $\tilde{P}$ consists of sets in Σ_Q,

iii) $Q \subset_{2\epsilon} \bigvee_{-N}^{N} T^i \tilde{P}$.

In fact, if Π is a partition of $(2N+1)$-sequences of $\{1, 2, \ldots, k\}$ such that $|\Pi(P) - Q| \leq \epsilon$, then $|\Pi(\tilde{P}) - Q| \leq 2\epsilon$.

Proof. First, use the fact that P is a generator to choose $N_1 > N$ such that

$$Q \subset_\alpha \bigvee_{-N_1}^{N_1} T^i P, \tag{10.1}$$

where α will be specified later. Given n, also to be determined later, apply Rohlin's Theorem to the transformation T_Q to choose a set $F \in \Sigma_Q$ such that $T^i F$, $0 \leq 1 \leq n-1$, is a disjoint sequence and $\mu(\bigcup_{i=0}^{n-1} T_Q^i F) \geq 1 - \alpha$.

The set F and partition Q give us two isomorphic gadgets. First, we have the gadget (T_Q, F, n, Q) which lies in the space (X, Σ_Q, μ). Second,since $\Sigma_Q \subseteq \Sigma$, we have the gadget (T, F, n, Q) which lies in the space (X, Σ, μ) . These are clearly isomorphic, hence we can apply Lemma 4.4 to select a partition P^*, consisting of sets in Σ_Q such that

$$(T, F, n, Q \vee P) \sim (T_Q, F, n, Q \vee P^*), \tag{10.2}$$

where $\sim$ indicates gadget isomorphism.

This gadget isomorphism forces a number of relations between P and P^* to hold. First, note that, if $G = \bigcup_{i=0}^{n-1} T^i F$, then $d(P/G) = d(P^*/G)$ so that, if α is small, $d(P)$ will be close to $d(P^*)$. Furthermore, suppose $\bar{\Pi}$ is a partition of $(2N+1)$-sequences such that $|\bar{\Pi}(P) - Q| \leq \alpha$. If n is so large that the top and bottom N_1-levels of the stack plus the complement of the stack have small measure, then $\bar{\Pi}(P^*)$ will be close to Q. Thus, if α is sufficiently small and n is sufficiently large, we will have

$$|\bar{\Pi}(P^*) - Q| \leq 2\alpha. \tag{10.3}$$

The condition (10.3), of course, means that

$$Q \subset_{2\alpha} \bigvee_{-N_1}^{N_1} T^i P^*. \tag{10.4}$$

In particular, if α is small enough, $H(T, P^*)$ will be very close to $H(T, Q) = H(Q) = H(P) = H(T, P)$. In summary, if δ is any given positive number, then α and n can be chosen so that

$$|d(P^*) - d(P)| \leq \delta \text{ and } 0 \leq H(T, P^*) - H(P) \leq \delta \tag{10.5}$$

Now apply Ornstein's Fundamental Lemma to T_Q to choose δ and then α and n so that there is a partition $\tilde{P}$, consisting of sets in Σ_Q, such that

$$\text{(a) } (T, \tilde{P}) \sim (T, P), \text{and (b) } |\tilde{P} - P^*| \leq \beta, \tag{10.6}$$

where β is a given positive number. If β is small, then (10.6b) will guarantee that $\Pi(\tilde{P})$ is close to $\Pi(P^*)$, hence β can be chosen so that (10.6b) implies that

$$|\Pi(\tilde{P}) - Q| \leq 2\epsilon. \tag{10.7}$$

This completes the proof of Lemma 10.2.

Lemma 10.2 provides us with a partition which we can copy to show that close to Q is a copy of $(\bar{T}, \bar{P})$ that almost generates.

LEMMA 10.3. Suppose $(T, Q) \sim (\bar{T}, \bar{P})$. Given $\epsilon > 0$, there is a partition $\tilde{Q}$ and a number K such that

i) $(T, \tilde{Q}) \sim (\bar{T}, \bar{P})$,

ii) $P \subset_\epsilon \bigvee_{-K}^{K} T^i \tilde{Q}$,

iii) $|Q - \tilde{Q}| \leq \epsilon$.

Proof. The idea of the proof is to choose $\tilde{P}$, consisting of sets in Σ_Q, so that $(T, \tilde{P}) \sim (T, P)$, and Q is close to a partition that is refined by $\bigvee_{-K}^{K} T^i \tilde{P}$. If K is large enough then $\tilde{P}$ will be close to a partition that is refined by $\bigvee_{-K}^{K} T^i Q$. We then construct a long gadget (T, F, n, P), $F \in \Sigma$, and an isomorphic gadget $(T_Q, E, n, \tilde{P})$, $E \in \Sigma_Q$, and use these to construct Q^* so that $(T, F, n, P \vee Q^*)$ is isomorphic to $(T_Q, E, n, \tilde{P} \vee Q)$. If this is done carefully, Q^* will be so close to Q that the fundamental lemma can be applied to modify it so that the desired $\tilde{Q}$ will exist.

To carry out the above plan, let α be a positive number to be specified later. Choose N_1 so that $Q \subset_\alpha \bigvee_{-N_1}^{N_1} T^i P$, then apply Lemma 10.2 to construct a partition Π_1 of $(2N{+}1)$-sequences of $\{1, 2, \ldots, k\}$, where P has k sets, so that $|\Pi(P) - Q| \leq \alpha$, and a $\tilde{P}$ such that

$$(10.8) \qquad \begin{aligned} &\textbf{a)}\ (T, \tilde{P}) \sim (T, P), \\ &\textbf{b)}\ \tilde{P} \text{ consists of sets in } \Sigma_Q, \\ &\textbf{c)}\ |\Pi_1(\tilde{P}) - Q| \leq 2\alpha. \end{aligned}$$

Since (10.8b) holds, we can choose $N_2 > N_1$ so that $\tilde{P} \subset_\alpha \bigvee_{-N_2}^{N_2} T^i Q$. Thus there is a partition Π_2 of $(2n_2 + 1)$-sequences of $\{1, 2, \ldots, m\}$, where Q has m sets, so that

$$(10.9) \qquad |P - \Pi_2(Q)| \leq \alpha.$$

Now choose $N_3 > N_2$ so that $\tilde{P} \subset_\beta \bigvee_{-N_3}^{N_3} T^i Q$, where β will be specified later as a function of α. Hence there is a partition Π_3 of $(2N_3{+}1)$-sequences of $\{1, 2, \ldots, m\}$ such that

$$(10.10) \qquad |\tilde{P} - \Pi_3(Q)| \leq \beta.$$

Let n be a large number, and use the strong form of Rohlin's Theorem to choose $F \in \Sigma$ such that $T^i F$, $0 \leq i \leq n-1$, is a disjoint sequence, and

$$(10.11) \qquad d\Big(\bigvee_0^{n-1} T^{-i} P / F\Big) = d\Big(\bigvee_0^{n-1} T^{-i} P\Big).$$

In a similar manner, choose $E \in \Sigma_Q$ such that $T^i E$, $0 \leq i \leq n-1$, is a disjoint sequence and

$$d(\textstyle\bigvee_0^{n-1} T^{-i}\tilde{P}/E) = d(\textstyle\bigvee_0^{n-1} T^{-i}\tilde{P}). \tag{10.11$'$}$$

The condition (10.8a), along with (10.11) and (10.11′), implies that the two gadgets (T, F, n, P) and $(T_Q, E, n, \tilde{P})$ are isomorphic. Hence (from Lemma 4.4) there is a Q^* such that

$$(T, F, n, P \vee Q^*) \text{ is isomorphic to } (T_Q, E, n, \tilde{P} \vee Q). \tag{10.12}$$

We shall show that, if β is small enough and n is large enough, then Q^* can be modified to obtain the desired $\tilde{Q}$. First, we can assume that n is so large that the top and bottom N_3-levels of each stack $\{T^i F\}$ and $\{T^i E\}$ plus their complements have such small measure that (10.10) and (10.12) imply

$$|\Pi_3(Q^*) - P| \leq 2\beta \text{ and } |d(Q^*) - d(Q)| \leq 2\beta, \tag{10.13}$$

and that (10.12), (10.9), and (10.8c) imply

$$|\Pi_2(Q^*) - P| \leq 2\alpha \tag{10.14}$$

and

$$|\Pi_1(P) - Q^*| \leq 3\alpha. \tag{10.15}$$

If β is small enough, the conditions (10.13) imply that $H(T, Q^*)$ will be very close to $H(T, Q)$. Thus we can assume that β is so small that there is a $\tilde{Q}$ satisfying

$$\text{(a) } (T, \tilde{Q}) \sim (T, Q), \quad \text{(b) } |\tilde{Q} - Q^*| \leq \gamma \tag{10.16}$$

Furthermore, if N_2 (which does not depend on β) is fixed, we can assume that γ is so small that (10.16b) implies that $\bigvee_{-N_2}^{N_2} T^i\tilde{Q}$ is very close to $\bigvee_{-N_2}^{N_2} T^i Q^*$. Thus, if γ is small enough, this and (10.14) imply that

$$|\Pi_2(\tilde{Q}) - P| \leq 3\alpha. \tag{10.17}$$

Note also that (10.15) and the relation (10.16b) imply that $|\Pi_1(P) - \tilde{Q}| \leq 3\alpha + \gamma$. Since we started with the hypothesis that $|\Pi_1(P) - Q| \leq \alpha$, we therefore can choose β so small that

$$|Q - \tilde{Q}| \leq 5\alpha. \tag{10.18}$$

The results (10.16a), (10.17), and (10.18) are the desired results, for we merely need to begin with $\alpha = \epsilon/5$. This proves Lemma 10.3.

Proof of Theorem 10.1. First, use Lemma 9.1 combined with Ornstein's Fundamental Lemma to select $Q^{(0)}$ so that $(T, Q^{(0)}) \sim (\bar{T}, \bar{P})$. Let ϵ_n go to zero rapidly, and apply Lemma 10.3 to select $Q^{(n)}$ and K_n such that

$$\text{(10.19)} \qquad \begin{array}{ll} \text{a)} & (T, Q^{(n)}) \sim (\bar{T}, \bar{P}), \\ \text{b)} & P \subset_{\epsilon_n} \bigvee_{-K_n}^{K_n} T^i Q^{(n)}, \\ \text{c)} & |Q^{(n)} - Q^{(n-1)}| \le \epsilon_n. \end{array}$$

One can clearly assume that the K_n increase. Furthermore, if ϵ_n goes to zero fast enough, we can use (10.19c) to assume that

$$\text{(10.20)} \qquad P \subset_{\epsilon_j + \ldots + \epsilon_n} \bigvee_{-K_j}^{K_j} T^i Q^{(n+1)},\ 1 \le j \le n.$$

If we assume that $\sum \epsilon_n < \infty$, then (10.19c) implies that $Q = \lim_n Q^{(n)}$ exists. Then (10.19a) implies that $(T, Q) \sim (\bar{T}, \bar{P})$, while (10.20) will imply that P consists of sets which are measurable Σ_Q. Thus Q will also be a generator. This completes the proof of Theorem 10.1.

CHAPTER 11.
FINITELY DETERMINED PARTITIONS

The key result used to prove the isomorphism theorem of Chapter 10 was the lemma that enabled us to modify partitions with good entropy and distribution to obtain arbitrarily better entropy and distribution (Lemma 9.2). There is no reason why one cannot lengthen the gadgets used in this proof so as to obtain good joint distribution as well. This simple observation is the basis for showing that many transformations which arise in other parts of mathematics are isomorphic to Bernoulli shifts (see Chapters 12 and 13). In this section, we shall show how to modify Lemma 9.2 so as to obtain a characterization of generators for Bernoulli shifts.

Suppose $\bar{T}$ is a Bernoulli shift with independent k-set generator P. The property of the pair $\bar{T}, \bar{P}$ used in the proof of Lemma 9.2 was the fact that, if T is any ergodic transformation with $H(T) \geq H(\bar{T}, \bar{P})$ and P is any k-set partition such that (a) $|d(P) - d(\bar{P})| \leq \delta$, and (b) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$, then, if δ is small enough, we have (c) $\bar{d}((T, P), (\bar{T}, \bar{P})) < \epsilon$. This is the Ornstein Copying Theorem of Chapter 7. This property of an independent generator was used in the proof of Lemma 9.2 to insure that the constructed Q was close to the given P. This property of independent generators is contained (in a more general form) in the following definition:

Let $\bar{T}$ be an ergodic transformation and $\bar{P}$ a k-set partition. We say that $\bar{P}$ is *finitely determined* (relative to $\bar{T}$) if, given $\epsilon > 0$, there is an $n > 0$ and a $\delta > 0$ such that, if T is any ergodic transformation with $H(T) \geq H(\bar{T}, \bar{P})$ and P is any k-set partition such that

i) $|d(\bigvee_0^n T^i P) - d(\bigvee_0^n \bar{T}^i \bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

then

iii) $\bar{d}((T, P), (\bar{T}, \bar{P})) < \epsilon$.

The Ornstein Copying Theorem asserts that, if $\{\bar{T}^i \bar{P}\}$ is an independent sequence, then $\bar{P}$ is finitely determined relative to $\bar{T}$. As we shall see below, the finitely determined partitions are precisely those which generate transformations isomorphic to Bernoulli shifts.

To obtain a stronger version of Lemma 9.2, we need a stronger form of the law of large numbers which we now describe. Let $(t_0, t_1, \ldots, t_{m-1})$ be the P-m-name of an atom of $B \in \bigvee_0^{m-1} T^{-i} P$; that is, $B = \cap_{i=0}^{m-1} T^{-i} P_{t_i}$. Fix $n \geq m$, and let $(s_0, s_1, \ldots, s_{n-1})$ be the P-n-name of an atom $A \in \bigvee_0^{n-1} T^{-i} P$. Let us use the symbol $f_A(B, n)$ to denote the relative frequency of occurrence of the P-m-name of

B in consecutive positions in the P-n-name of A; that is, $nf_A(B,n)$ is the number of indices j, $0 \leq j \leq n-1$, such that $T^jA \subseteq B$.

THEOREM 11.1. If T is ergodic, and m and $\epsilon > 0$ are given, then for all sufficiently large n, there is a collection $\mathcal{E}_n$ of sets in $\bigvee_0^{n-1} T^{-i}P$ of total measure at least $1-\epsilon$ such that, for all $B \in \bigvee_0^{m-1} T^{-i}P$ and all $A \in \mathcal{E}_n$, we have $|f_A(B,n) - \mu(B)| \leq \epsilon$.

This is, of course, a simple consequence of the so-called individual ergodic theorem which asserts that for any integrable g and ergodic T,

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} g(T^i x) = \int g d\mu \ \text{ a.e.}$$

A proof of this ergodic theorem can be found in [3], pp. 20f. We then obtain Theorem 11.1 by letting g be the characteristic function of B, and using the fact that a.e. convergence implies almost uniform convergence (see Egoroff's Theorem in [8], p. 88).

The following is a general version of Lemma 9.2.

LEMMA 11.1. If $\bar{P}$ has k sets and is finitely determined relative to $\bar{T}$ then, given $\epsilon > 0$, there is an n_1 and a $\delta > 0$ such that, if T is any ergodic transformation with $H(T) \geq H(\bar{T}, \bar{P})$, and P any k-set partition such that

i) $|d(\bigvee_0^{n_1} T^i P) - d(\bigvee_0^{n_1} \bar{T}^i \bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

then, for any $\bar{\delta} > 0$ and any $\bar{n}$, there is a Q such that

iii) $|d(\bigvee_0^{\bar{n}} T^i Q) - d(\bigvee_0^{\bar{n}} \bar{T}^i \bar{P})| \leq \bar{\delta}$,

iv) $0 \leq H(\bar{T}, \bar{P}) - H(T, Q) \leq \bar{\delta}$,

and

v) $|P - Q| \leq \epsilon$.

The proof of this lemma is identical to the proof of Lemma 9.2 with three exceptions. First, we use the definition of finitely determined to choose n_1 and $\delta > 0$ such that (i) and (ii) imply that $\bar{d}((T,P), \bar{T}, \bar{P})) < \bar{\epsilon}$. Second, we must make sure that (ii) can be replaced by

ii') $0 < H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$, so that we can find $R \subset P$ for which

$$0 < H(\bar{T}, \bar{P}) - H(T, R) < \alpha$$

(see statements (9.2) and (9.1') in the proof of Lemma 9.2). This is proved in Lemma 11.2 below. Our third modification in the proof is that we use Theorem 11.1 above in place of the law of large numbers to be sure that the class $\bar{\mathcal{E}}$ (see (9.2) ff.) satisfies the property

$$\sum_{B \in \bigvee_0^{\bar{n}-1} \bar{T}^i \bar{P}} |f_A(B, n) - \mu(B)| \leq \beta, \ A \in \bar{\mathcal{E}}.$$

The constructed Q will then, for suitably small $\bar{\epsilon}$ and β, satisfy (iii), (iv), and (v).

We shall show that we can assume (ii') holds by showing that, if it cannot be made to hold, a Q satisfying (iii), (iv), and (v) can be constructed directly from the definition of finitely determined partition. Let $\bar{T}, \bar{P}$ be finitely determined, let ϵ be a positive number, and choose n_1 and δ so that, if (i) and (ii) hold, then $\bar{d}((T, P), (\bar{T}, \bar{P})) < \epsilon$. Let T, P be given, satisfying (i) and (ii).

LEMMA 11.2. Suppose that $H(T, Q) \geq H(\bar{T}, \bar{P})$ for any k-set partition Q for which $|Q - P| < 2\epsilon$. Then given $\bar{n}$ and $\bar{\delta} > 0$, there is a Q satisfying (iii) and (iv) of Lemma 11.1, and $|Q - P| < 2\epsilon$.

Proof. Given $\alpha > 0$, use Lemma 5.4 to choose gadgets $\mathcal{G} = (T, F, \bar{n}, P)$ and $\bar{\mathcal{G}} = (\bar{T}, \bar{F}, \bar{n}, \bar{P})$ so that $\mu(\bigcup_{i=0}^{\bar{n}-1} T^i F) \geq 1 - \alpha$, $\bar{\mu}(\bigcup_{i=0}^{\bar{n}-1} \bar{T}^i \bar{F}) \geq 1 - \alpha$, and $\bar{d}_{\bar{n}}(\mathcal{G}, \bar{\mathcal{G}}) < \epsilon$. Thus we can choose Q so that

$$\bar{\mathcal{G}} \text{ is isomorphic to } (T, F, \bar{n}, Q) \tag{11.1}$$

and

$$\frac{1}{\bar{n}} \sum_{i=0}^{\bar{n}-1} |P/T^i F - Q/T^i F| < \epsilon. \tag{11.2}$$

If $\alpha \leq \epsilon$, condition (11.2) implies $|Q - P| < 2\epsilon$, so that our hypothesis implies that $H(T, Q) \geq H(\bar{T}, \bar{P})$. We can choose Q in the σ-algebra generated by P under T, and so obtain $H(T, Q) \leq H(T, P) \leq H(\bar{T}, \bar{P})$ and therefore be sure that (iv) holds. If $\alpha \leq \delta/2$, we also obtain (iii) from condition (11.1). This completes the proof of Lemma 11.2, and thus Lemma 11.1 is established.

Ornstein's Fundamental Lemma in the following general form follows easily from Lemma 11.1.

ORNSTEIN'S FUNDAMENTAL LEMMA (GENERAL FORM). If $\bar{P}$ and $\bar{T}$ are as in Lemma 11.1, then, given $\epsilon > 0$, there is an n_1 and a $\delta > 0$ such that, if T is any ergodic transformation satisfying $H(T) \geq H(\bar{T}, \bar{P})$, and P is any k-set partition satisfying

i) $|d(\bigvee_0^{n_1} T^iP) - d(\bigvee_0^{n_1} \bar{T}^i\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

then there is a Q such that

iii) $d(\bigvee_0^n T^iQ) = D(\bigvee_0^n \bar{T}^i\bar{P})$, $n = 0, 1, 2, \ldots$,

iv) $|P - Q| \leq \epsilon$.

The Ornstein isomorphism theorem (Theorem 10.1) now extends to the following result.

THEOREM 11.2. Any two transformations T and $\bar{T}$ with finitely determined generators P and $\bar{P}$ are isomorphic if they have the same entropy; that is, there is a Q such that

i) $(T, Q) \sim (\bar{T}, \bar{P})$,

ii) Q is a generator for T.

This is proved in much the same way as the isomorphism theorem (Theorem 10.1). The only change which is needed in that proof is to choose gadgets long enough so that one gets good joint distributions for large n_1, so that one can apply the general form of the Fundamental Lemma above. We leave the details to the reader.

We complete this chapter by proving

THEOREM 11.3. If $\bar{P}$ is a generator for a Bernoulli shift $\bar{T}$, then $\bar{P}$ is finitely determined.

Proof. Once can actually prove that *any* partition $\bar{P}$ for a Bernoulli shift is finitely determined ([14]). The proof of this is quite complicated. We give here the simpler proof for the case when $\bar{P}$ is a generator (in fact, all that is used about $\bar{P}$ is that $H(\bar{T}, \bar{P}) = H(\bar{T})$.

Let $\bar{B}$ be an independent generator for $\bar{T}$, and assume $\bar{P}$ is a k-set generator. Fix δ and n_1 (to be specified later), and suppose T is an ergodic transformation with $H(T) \geq H(\bar{T})$, and that P is a k-set partition such that

i) $|d(\bigvee_0^{n_1} T^iP) - d(\bigvee_0^{n_1} \bar{T}^i\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$.

We wish to show that, if δ is small enough and n_1 is large enough, then

iii) $\bar{d}((T,P),(\bar{T},\bar{P}))<\epsilon$.

This will be accomplished by choosing a copy T, B of $\bar{T}, \bar{B}$ so that P-names are related to B-names in (almost) the same way that $\bar{P}$-names are related to $\bar{B}$-names. Let α and $\alpha_1 \le \alpha$ be positive numbers to be specified later, and choose N and $N_1 \ge N$ so that

$$\bigvee_{-N}^{N} \bar{T}^i\bar{B} \subset_\alpha \bar{P} \text{ and } \bigvee_{-N_1}^{N_1} \bar{T}^i\bar{B} \subset_{\alpha_1} \bar{P}. \tag{11.3}$$

Then construct gadgets $(\bar{T},\bar{F},n_1,\bar{P})$ and (T,F,n_1,P) so that

$$d(\bigvee_{0}^{n_1-1} T^{-i}P) = d(\bigvee_{0}^{n_1-1} T^{-i}P/F), \tag{11.4}$$

$$d(\bigvee_{0}^{n_1-1} \bar{T}^{-i}\bar{P}) = d(\bigvee_{0}^{n_1-1} \bar{T}^{-i}\bar{P}/\bar{F}),$$

and, for each gadget, the complement and top and bottom N_1-levels contain less than α_1 of the space. If δ is small enough, equation (11.4) and condition (i) imply that we can choose $F^* \subset F$, $\bar{F}^* \subset \bar{F}$ so that

$$(T,F^*,n_1,P) \sim (\bar{T},\bar{F}^*,n_1,\bar{P}), \tag{11.5}$$

and $\mu(F-F^*)$, $\bar{\mu}(\bar{F}-\bar{F}^*)$ are both very small, and $\sim$ indicates gadget isomorphism. Now Lemma 4.4 is used to select B^* so that

$$(\bar{T},\bar{F}^*,n_1,\bar{P}\vee\bar{B}) \sim (T,F^*,n_1,P\vee B^*). \tag{11.6}$$

If n_1 is large enough and δ is small enough, the conditions (11.6) and (11.3) guarantee

$$\bigvee_{-N_1}^{N_1} T^iB^* \subset_{3\alpha_1} P \text{ and } \bigvee_{-N}^{N} T^iB^* \subset_{3\alpha} P, \tag{11.7}$$

since we have assumed that $\alpha_1 \le \alpha$. Thus in particular, we can choose α_1 so that $d(B^*)$ is close enough to $d(\bar{B})$, and $H(T,B^*)$ is close enough to $H(\bar{T},\bar{B})$ so we can apply Ornstein's Fundamental Lemma to obtain B such that

$$(T,B) \sim (\bar{T},\bar{B}) \text{ and } |B-B^*|<\beta, \tag{11.8}$$

where β is any quantity specified in advance. If β is small enough, we can be sure that

$$|\bigvee_{-N}^{N} T^iB - \bigvee_{-N}^{N} T^iB^*| < \alpha, \tag{11.9}$$

and hence (11.7) gives

$$\bigvee_{-N}^{N} T^i B \subset_{4\alpha} P. \tag{11.10}$$

Let S be the measure-preserving transformation which carries T, B onto $\bar{T}, \bar{B}$. The conditions (11.3) and (11.10) then give $|SP - P| \leq 5\alpha$, so, if $\alpha < \epsilon/5$, we obtain the desired inequality

$$\bar{d}((T,P),(T,P)) < \epsilon.$$

This completes the proof of Theorem 11.3.

The reader is referred to [16] for a further discussion of the $\bar{d}$-metric and the concept of finitely determined partition.

CHAPTER 12.
WEAK AND VERY WEAK BERNOULLI PARTITIONS

The results of Chapter 11 can be used to show that many transformations of physical and mathematical interest are isomorphic to Bernoulli shifts. This is usually done by showing that there is a generating partition which satisfies some condition that implies that it is finitely determined. Two such conditions are discussed in this section, each presented in the context of a specific example.

A class of transformations of interest are the Markov shifts. These are the shifts associated with stationary Markov chains, and are defined as follows: Let Π be a $k \times k$-matrix with nonnegative entries, each row of which has sum 1, and let π be a k-tuple of nonnegative numbers which sum to 1. The space X is the set of doubly infinite sequences of the symbols $1, 2, \ldots, k$, and the measure μ is the unique complete extension of the measure defined on cylinder sets by the formula

$$\mu\{x | x_i = t_i,\ -m \leq i \leq n\} = \pi_{t_{-m}} \Pi_{t_{-m}t_{-m+1}} \ldots \Pi_{t_{n-1}t_n}. \tag{12.1}$$

We also assume that $\pi\Pi = \Pi$, from which it follows that the shift T defined by $(Tx)_n = x_{n+1}$, $n = 0, \pm1, \pm2, \ldots$, is μ-invariant. T is called the *Markov shift* with initial distribution π and transition matrix Π.

The reader is referred to Billingsley's book [3] for a discussion of the properties of Markov shifts. For our purposes, we are interested in the following. Let E be the $k \times k$-matrix each row of which is π. Then we have $\Pi E = E\Pi = E$. Furthermore,

$$T \text{ is mixing iff } \lim_n (\Pi^n - E) = 0. \tag{12.2}$$

The following result, due to Friedman and Ornstein ([5]), was the first generalization of the original isomorphism paper ([13]).

THEOREM 12.1. A mixing Markov shift is isomorphic to a Bernoulli shift.

Except in a few cases, it is very difficult to construct a generator for a Markov shift whose iterates are independent. It is fairly easy, however, to find a generator that satisfies an asymptotic condition which implies it is finitely determined. This generator is the partition $P = \{P_1, P_2, \ldots, P_k\}$, where

$$P_i = \{x | x_0 = i\}. \tag{12.3}$$

It is obvious that P is a generator for T. We shall prove that P is weak Bernoulli, a concept defined as follows: a partition P is called *weak Bernoulli* for an ergodic transformation T if, given $\epsilon > 0$, there is an N such that, for all $m \geq 1$,

$$\bigvee_{-m}^{0} T^i P \text{ is } \epsilon\text{-independent of } \bigvee_{N}^{N+m} T^i P.$$

For the process T, P, this says that the future and distant past are approximately independent, and is a strong form of the well known Kolmogorov $0-1$ law, which states that any set measurable with respect to the arbitrarily distant past has measure 0 or 1. If P is weak Bernoulli, then the process T, P certainly satisfies the $0-1$ law, but the converse of this is false ([17]).

Later in this chapter, it will be shown that a weak Bernoulli partition is finitely determined. At this point, we shall show that the partition P of (12.3) is weak Bernoulli for a mixing Markov shift. Towards this end, fix m and N, and let $A \in \bigvee_{-m}^{0} T^i P$, $B \in \bigvee_{N}^{N+m} T^i P$ so that A and B have the form

$$A = P_{t_0} \cap T^{-1} P_{t_{-1}} \cap \ldots \cap T^{-m} P_{t_{-m}}, \quad B = T^N P_{t_N} \cap \ldots \cap T^{N+m} P_{t_{N+m}}.$$

The formula (12.1) then gives

$$\mu(A \cap B) = \sum_{t_1^{N-1}} \pi_{t_{-m}} \Pi_{t_{-m} t_{-m+1}} \ldots \Pi_{t_{N+m-1} t_{N+m}},$$

where the sum is taken over all possible $(N-1)$-tuples $t_1^{N-1} = t_1, \ldots, t_{N-1}$ of the symbols $1, 2, \ldots, k$. It is easy to see that

$$\sum_{t_1, \ldots, t_{N-1}} \Pi_{t_0 t_1} \ldots \Pi_{t_{N-1} t_N} = \Pi_{t_0 t_N}^{(N)},$$

where $(\Pi_{ij}^{(N)})$ is the N^{th} power of Π. We also have

$$\mu(A)\mu(B) = \pi_{t_{-m}} \Pi_{t_{-m} t_{-m+1}} .. \Pi_{t_{-1} t_0} \pi_{t_N} \Pi_{t_N t_{N+1}} .. \Pi_{t_{N+m-1} t_{N+m}}$$

so that

$$\mu(A \cap B) - \mu(A)\mu(B) = \mu(A)[\Pi_{t_0 t_N}^{(N)} - \pi_{t_N}] \Pi_{t_N t_{N+1}} \ldots \Pi_{t_{N+m-1} t_{N+m}},$$

from which it easily follows that

$$\sum_{A,B} |\mu(A \cap B) - \mu(A)\mu(B)| \leq \sum_{i,j=1}^{k} |\Pi_{i,j}^{(N)} - \pi_j|,$$

where the left-hand sum is taken over all $A \in \bigvee_{-m}^{0} T^i P$ and all $B \in \bigvee_{N}^{N+m} T^i P$. Mixing then implies (from 12.2) that, uniformly in m,

$$\lim_{N \to \infty} \sum_{A,B} |\mu(A \cap B) - \mu(A)\mu(B)| = 0,$$

and this implies (see Lemma 6.2) that P is indeed weak Bernoulli. Thus Theorem 12.1 is established once we have proved the following.

THEOREM 12.2. A weak Bernoulli partition is finitely determined.

This can be proved by generalizing the proof that a partition with independent iterates is finitely determined (Ornstein's Copying Theorem, Chapter 7). We shall describe instead a weaker condition (called "very weak Bernoulli"), and give a proof later that very weak Bernoulli implies finitely determined.

The definition of very weak Bernoulli is given in terms of the extension of the $\bar{d}$-metric to arbitrary sequences of partitions, as discussed in Chapter 7. Let $\{P^i\}$ and $\{\bar{P}^i\}$ be two finite sequences of k set partitions. Then

$$\bar{d}_n(\{P^i\}_{i=0}^{n-1}, \{\bar{P}^i\}_{i=0}^{n-1}) = \inf \frac{1}{n} \sum_{i=0}^{n-1} |Q^i - \bar{Q}^i|,$$

where this infinum is taken over all sequences $\{Q^i\}_{i=0}^{n-1}$ and $\{\bar{Q}^i\}_{i=0}^{n-1}$ of k-set partitions of a given Lebesgue space X that satisfy

$$d(\bigvee_0^{n-1} P^i) = d(\bigvee_0^{n-1} Q^i),\ d(\bigvee_0^{n-1} \bar{P}^i) = d(\bigvee_0^{n-1} \bar{Q}^i).$$

A partition P is called *very weak Bernoulli for T* if, given $\epsilon > 0$, there is an N such that, for all $m > 0$,

$$\bar{d}_N(\{T^i P/A\}_{i=0}^{N-1},\ \{T^i P\}_{i=0}^{N-1}) < \epsilon$$

for all atoms A in a collection $\mathcal{A}_m$ of atoms in $\bigvee_{-m}^{-1} T^i P$ of total measure at least $1-\epsilon$. This says (roughly) that, for each $m > 0$, one can, for most atoms A (in the past) find a measure-preserving transformation S from A into the whole space such that, for most points $x \in A$, the P-N-name of x and the P-N-name of Sx agree in most places.

It is easy to see that weak Bernoulli implies very weak Bernoulli. Later we show that very weak Bernoulli implies finitely determined (see Theorem 12.3 below). The concept of very weak Bernoulli was introduced by Ornstein ([15]) as an aid in showing that a Bernoulli shift can be embedded in a flow. We shall at this point sketch this proof as it illustrates the kind of argument which is used in the proofs that many other transformations are Bernoulli.

Let S denote the Bernoulli shift on the unit interval X, with a 2-set independent generator $Q = \{Q_0, Q_1\}$ satisfying $\mu(Q_0) = \mu(Q_1) = 1/2$. Let f be defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in Q_0 \\ \alpha & \text{if } x \in Q_1 \end{cases},$$

where α is a fixed irrational larger than 1. $\bar{X}$ will denote the space $\{(x,y)|x \in X,\ 0 \le y \le f(x)\}$, with Lebesgue measure (normalized so that $\mu(\bar{X}) = 1$). A *flow*

$\{T_t\}$ is defined on $\bar{X}$ as follows: A point on the fiber with base x moves upwards with uniform speed until it hits the graph of f. It then moves instantaneously to the fiber with base Sx and continues moving at uniform speed (see Figure 12.1).

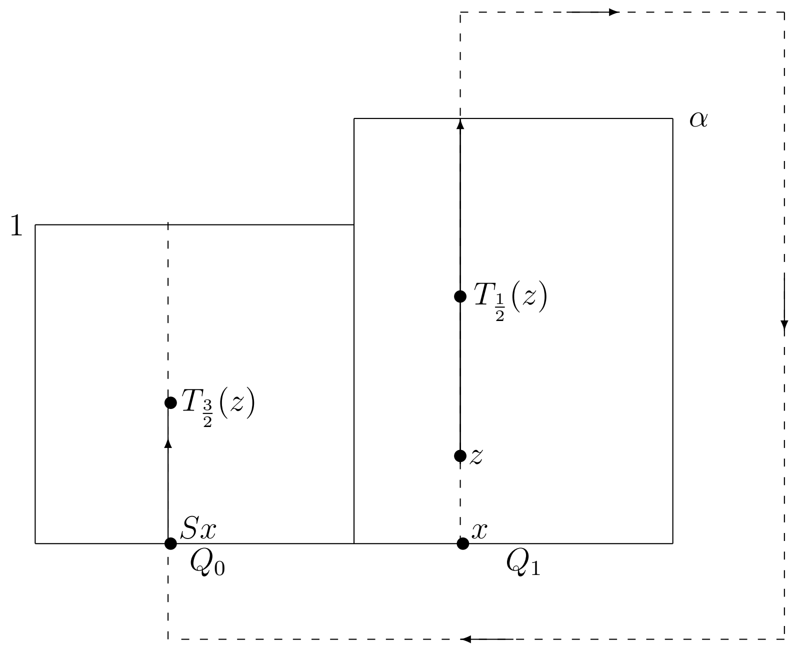


Figure 12.1

The flow $\{T_t\}$ is called the flow with *base S built under the function f*. Clearly, $\{T_t\}$ is a one-parameter group of measure-preserving transformations on the Lebesgue space $\bar{X}$. More general flows can be constructed by using a more general base transformation and function f. In fact, Ambrose ([11]) showed that every ergodic flow is isomorphic to a flow built under some function. (See Totoki ([23]) for a discussion of properties of such flows, including the flow of Figure 12.1.)

Let P denote the two-set partition of $\bar{X}$ given by

(12.4) $$P_0 = \{(x, y)|x \in Q_0\},\ P_1 = \{(x, y)|x \in Q_1\}.$$

We shall prove that each member T_t, $t \neq 0$, is isomorphic to a Bernoulli shift. This will be done in three stages. First, it will be shown that P is a generator for T_γ if $0 < \gamma < 1/4$. Second, it will be shown that P is very weak Bernoulli for any T_γ, $\gamma \neq 0$. Third will come the proof that very weak Bernoulli implies finitely determined. This will prove that each T_γ, $0 < \gamma < 1/4$, and hence that each T_t, $t \neq 0$, is isomorphic to a Bernoulli shift.

An extension of our terminology will be helpful in these proofs. If $z = (x, y)$, then the *continuous P-name* of z is the partition of the real line into the two sets

$$\mathcal{P}_0(z) = \{t|\ T_t z \in P_0\},\ \mathcal{P}_1(z) = \{t|\ T_t z \in P_1\}.$$

Note that $\mathcal{P}_0(z)$ is a countable disjoint union of intervals of integer lengths, and $\mathcal{P}_1(z)$ is a countable disjoint union of intervals whose lengths are integer multiples of α. Also note that $\mathcal{P}_i(T_t z) = -t + \mathcal{P}_i(z)$. The ergodic theorem (Theorem 11.1) applied to the base transformation S implies that, for almost all points z, all possible finite sequences of intervals occur in the continuous name of z. To make this precise, let n_i, $0 \leq i \leq k$, and m_i, $0 \leq i \leq k$, be sequences of positive integers such that $n_0 = m_0$ and define the two sequences

$$\begin{aligned} u_\ell &= \sum_{i=0}^{\ell} n_i + \alpha \sum_{i=0}^{\ell} m_i, \ 0 \leq \ell \leq k, \\ v_\ell &= n_{\ell+1} + u_\ell, \quad 0 \leq \ell \leq k-1. \end{aligned}$$

We then have the following.

(12.5) There is a null set $\bar{E} \subset \bar{X}$ such that if $z \notin \bar{E}$, then there is a t such that for any k and $\{n_i\}, \{m_i\}$ the intervals $(u_\ell, v_{\ell+1})$ belong to $\mathcal{P}_0(T_t z)$ and the intervals $(v_\ell, u_{\ell+1})$ belong to $\mathcal{P}_1(T_t z)$.

Clearly, two distinct points z and $\bar{z}$ will have distinct continuous P-names. We would like to show that, if γ is fixed, $0 < \gamma < 1/4$, and $T = T_\gamma$, then distinct points will have distinct discrete T-P-names. This is easy to show (since $\gamma \leq 1/4$) if $z = (x, y)$, $\bar{z} = (\bar{x}, \bar{y})$, and $x \neq x$, for then x and $\bar{x}$ have sistinct S-Q-names. Let us suppose that $z = (x, y)$, $\bar{z} = (x, \bar{y})$, and that $\bar{y} - y = \delta > 0$. We also suppose $z \notin \bar{E}$, where $\bar{E}$ is the set of measure zero of (12.5). We now make use of the following elementary result (whose proof is left to the reader; see also [3], pp. 15-16).

(12.6) If α_1/α_2 is irrational, and $(\cdot)$ denotes fractional part, the numbers $(n\alpha_1 + m\alpha_2)$, n, m positive integers,are dense in the unit interval.

Choose a positive integer k such that $\gamma/k < \delta$, then use (12.6) to choose sequences n_ℓ, m_ℓ, $1 \leq \ell \leq k$, such that

$$\frac{\ell - 1}{k} \leq \left(\sum_{i=1}^{\ell} n_i/\gamma + \sum_{i=1}^{\updownarrow} m_i \alpha/\gamma \right) \leq \frac{\ell}{k}. \tag{12.7}$$

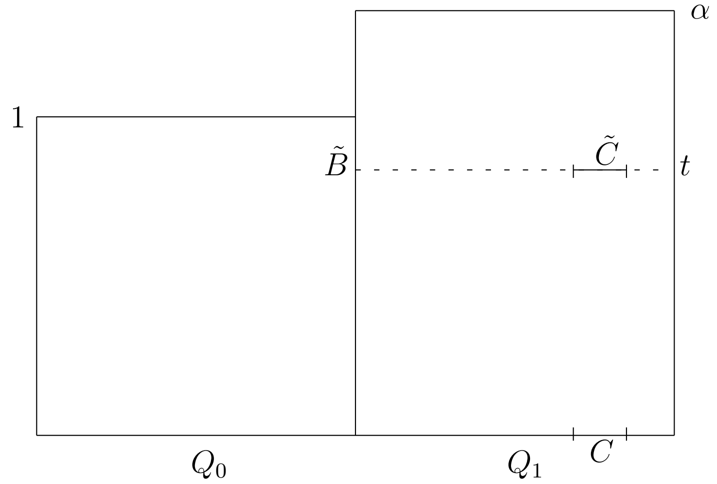


Figure 12.2

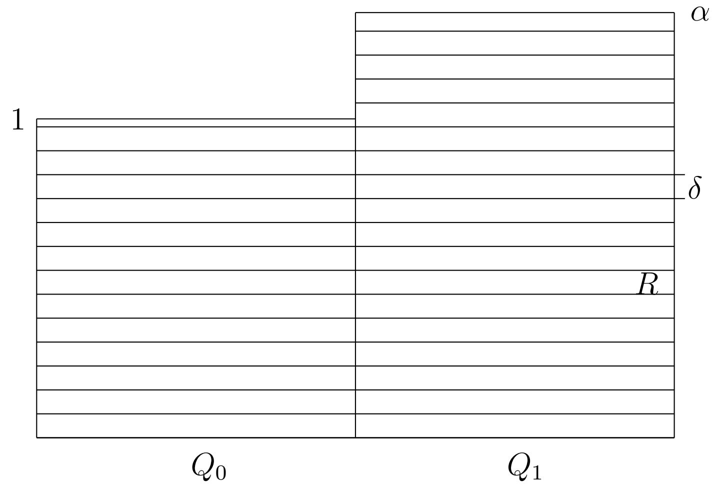


Figure 12.3

Approximately the same proportion of $T_\gamma^N \tilde{B}$ lies in each δ rectangle

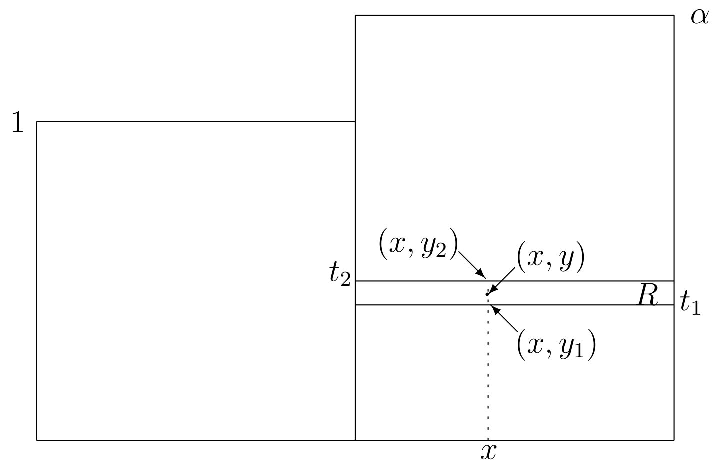


Figure 12.4

Choose t for z from the result (12.5), and choose n such that $n\gamma - t = \gamma'$, where $0 < \gamma' \leq \gamma$. From (12.7), we have that there is an ℓ, $1 \leq \ell \leq k$, such that $|\gamma' - w_\ell \gamma| < \delta$, where $w_\ell = (\sum n_i/\gamma + \sum m_i \alpha/\gamma)$, and this tells us that there is an integer K such that $K\gamma + \delta \in \mathcal{P}_0(z)$ while $K\gamma \in \mathcal{P}_1(z)$. Since $\mathcal{P}_0(z) = \delta + \mathcal{P}_0(\bar{z})$, we therefore have $T^k z \in P_1$ and $T^k \bar{z} \in P_0$. This completes the proof that P is a generator for T.

Now for a sketch of the proof that P is very weak Bernoulli for each T_γ, $\gamma \neq 0$. Fix $\gamma \neq 0$, and put $T = T_\gamma$. Let $C \in \bigvee_{-k}^{-1} S^i Q$, and suppose $C \subset P_1$. Fix t, $0 < t < \alpha$, and put $\tilde{C} = T_t C$, $\tilde{B} = T_t Q_1$ (see Figure 12.2). We assert that for all $n \geq 0$

$$d(\bigvee_0^n T^i P/\tilde{C}) = d(\bigvee_0^n T^i P/\tilde{B}). \tag{12.8}$$

This follows from the fact that each time we go through the top in Figure 12.2, a past set in the base is split into two parts of equal size, one of which goes to the left half, the other to the right half.

We next make the observation that, for some large N, the set $T^N \tilde{B}$ is *nearly* uniformly distributed throughout the space; that is, given $\delta > 0$ and a subdivision into rectangles of height δ (or very close to δ) as shown in Figure 12.3, there is an N such that $T^N \tilde{B}$ appears in nearly the same (linear) proportion in each rectangle. This follows from the renewal theorem ([2], p. 219) applied to the random walks with jump ahead of 1 or α, each with probability 1/2. It will be assumed that such an N has been selected.

Consider one of the rectangles R of Figure 12.3, say one on the right-hand side. The intersection $T^N \tilde{B} \cap R$ is made up of translates upwards of past (relative to N) atoms from the base; that is, $T^N \tilde{B} \cap R$ is a union of sets $T_t A$, where A is an atom of some $\bigvee_k^n S^i Q$, $n \leq N$. Fix such an A, and put $\tilde{B}_t = T_t Q_1$. Choose t_1, t_2 so that $T_{t_1} Q_1$ and $T_{t_2} Q_1$ are the bottom and top, respectively, of R. We then have $t_1 < t < t_2$, $|t_1 - t_2| \leq \delta$. Let $(x, y) \in \tilde{B}_t$, $(x, y_1) \in T_{t_1} Q_1$, $(x, y_2) \in T_{t_2} Q_1$ (see Figure 12.4). If δ is small, the T, P-names of (x, y_1) and (x, y_2) can differ only in a few places; that is, one can choose δ so small that for $m \geq 1$

$$\bar{d}_m(\{T^i P/R\}_{i=N+1}^{i=N+m},\ (\{T^i P/\tilde{B}_t\}_{i=N+1}^{i=N+m}) \leq \epsilon. \tag{12.9}$$

The uniform distribution of $T^n \tilde{B}$ and the result (12.8) then imply that, for all n large enough,

$$\bar{d}_n(\{T^i P\}_{i=0}^{n-1},\ \{T^i P/\tilde{C}\}_{i=0}^{n-1}) \leq \epsilon$$

for all atoms $C \in \bigvee_{-k}^{-1} S^i Q$ and for all $k \geq 1$. Simple integration establishes that P is indeed very weak Bernoulli for T. The existence of a Bernoulli flow is therefore a consequence of the following theorem.

THEOREM 12.3. A very weak Bernoulli partition is finitely determined.

Proof. The proof of this is obtained by extending as much as we can the ideas used in the proof that an independent generator is finitely determined (see the Ornstein Copying Theorem, Chapter 7). Our first task is to sharpen Lemma 7.2. Let $\bar{P}$ be very weak Bernoulli for $\bar{T}$. Given $\epsilon > 0$, choose N so that, for each $m > 0$, there is a collection $\bar{\mathcal{E}}_m$ of atoms of $\bigvee_{-m}^{-1} \bar{T}^i\bar{P}$ of total measure at least $1-\epsilon$ such that, for $\bar{A} \in \bar{\mathcal{E}}_m$,

$$\bar{d}_N(\{\bar{T}^i\bar{P}/\bar{A}\}_{i=0}^{N-1},\ \{\bar{T}^i\bar{P}\}_{i=0}^{N-1}) < \epsilon. \tag{12.10}$$

We would like now to show that, if T, P is close enough to $\bar{T}, \bar{P}$ in joint distribution and entropy, then a result analogous to (12.10) will hold. This is

LEMMA 12.1. If $\bar{T}, \bar{P}$ and N are as above, there is a $\delta > 0$ and an n^* such that if T is any ergodic transformation with $H(T) \geq H(\bar{T}\bar{P})$, and P is any partition (with the same number of sets as $\bar{P}$) such that

i) $|d(\bigvee_0^{n^*} T^iP) - d(\bigvee_0^{n^*} \bar{T}^i\bar{P})| \leq \delta$,

ii) $0 \leq H(\bar{T}, \bar{P}) - H(T, P) \leq \delta$,

then, for each $m > 0$, there is a class $\mathcal{E}_m \subset \bigvee_{-m}^{-1} T^iP$ of measure at least $1 - 3\epsilon$ such that, for $A \in \mathcal{E}_m$,

iii) $\bar{d}_N(\{T^iP/A\}_{i=0}^{N-1},\ \{T^iP\}_{i=0}^{N-1}) \leq 3\epsilon$.

Proof. One can easily generalize Lemma 7.3 to see that there is a $\bar{\delta} = \bar{\delta}(\epsilon, k) > 0$ so that, if P is any k-set partition, and Q and R are any partitions satisfying

$$H(P|Q) - H(P|Q \vee R) \leq \bar{\delta}, \tag{12.11}$$

then there is a class $\mathcal{E} \subset Q$ of measure at least $1 - \epsilon$ such that, for $A \in \mathcal{E}$, the partitions P/A and R/A are ϵ-independent. We then select $\bar{n}$ so large that

$$H(\bigvee_0^{N-1} \bar{T}^i\bar{P}|\bigvee_{-\bar{n}}^{-1} \bar{T}^i\bar{P}) < NH(\bar{T}, \bar{P}) + \epsilon\bar{\delta}, \tag{12.12}$$

where $\bar{\delta} = \bar{\delta}(\epsilon, k)$, and k is chosen to be $\tilde{k}^N$; $\tilde{k}$ = number of sets in P. Now choose δ so that (i), with $n^* = \bar{n}$, implies that there is a class $\mathcal{E}_{\bar{n}} \subset \bigvee_{-\bar{n}}^{-1} T^iP$ of measure at least $1 - \epsilon$ such that, for $A \in \mathcal{E}_{\bar{n}}$, we have

$$\bar{d}_N(\{T^iP/A\}_{i=0}^{N-1},\ \{T^iP\}_{i=0}^{N-1}) < \epsilon. \tag{12.13}$$

Certainly (iii) now holds for $m \leq \bar{n}$. If $m > \bar{n}$, the condition (ii) is used. If δ is small enough, (12.12) and conditions (i) and (ii) imply that

$$H(\bigvee_0^{N-1} T^iP|\bigvee_{-\bar{n}}^{-1} T^iP) < NH(T, P) + \epsilon\bar{\delta}.$$

This and condition (ii) now guarantee that, for all $m \geq \bar{n}$,

$$H(\bigvee_{0}^{N-1} T^iP | \bigvee_{-\bar{n}}^{-1} T^iP) - H(\bigvee_{0}^{N-1} T^iP | \bigvee_{-m}^{-1} T^iP) \leq \bar{\delta} \tag{12.14}$$

since, for any T, P, m, we always have $NH(T,P) \leq H(\bigvee_0^{N-1} T^iP | \bigvee_{-m}^{-1} T^iP)$.

Fix $m > \bar{n}$, and use (12.11) with P replaced by $\bigvee_0^{N-1} T^iP$, Q replaced by $\bigvee_{-\bar{n}}^{-1} T^iP$, and R replaced by $\bigvee_{-m}^{-\bar{n}-1} T^iP$. Thus (12.11) and (12.13) imply that there is a class $\mathcal{E}_{\bar{n}} \subset Q$ of measure at least $1 - 2\epsilon$ such that (12.13) holds for $A \in \mathcal{E}_{\bar{n}}$, *and* the partitions $\bigvee_0^{N-1} T^iP/A$ and $\bigvee_{-m}^{-\bar{n}-1} T^iP/A$ are ϵ-independent. It is easy to see that this implies that (iii) holds, so the proof of Lemma 12.1 is completed where $n^* = \bar{n} + N$.

Proof of Theorem 12.3 (continued). The remainder of the proof of Theorem 12.3 is very similar to the proof of Lemma 7.3. Suppose $\bar{P}$ is very weak Bernoulli for T, and $\epsilon > 0$. Choose N so that (12.10) holds; then choose $\delta > 0$ and n^* from Lemma 12.1. Let T be an ergodic transformation such that $H(T) \geq H(\bar{T}, \bar{P})$, and let P be a partition so that (i) and (ii) of Lemma 12.1 hold. We shall show how to construct partitions $\{Q^i\}$, $\{\bar{Q}^i\}$ of a Lebesgue space X so that for all $n \geq 0$

$$\text{a)}\quad d(\bigvee_{i=0}^{n-1} Q^i) = d(\bigvee_{0}^{n-1} T^iP),\ d(\bigvee_{0}^{n-1} \bar{Q}^i) = d(\bigvee_{0}^{n-1} \bar{T}^i\bar{P}), \tag{12.15}$$

$$\text{b)}\quad \frac{1}{n}\sum_{i=0}^{n-1} |Q^i - \bar{Q}^i| \leq 13\epsilon.$$

This will establish the desired conclusion that conditions (i) and (ii) of Lemma 12.1 imply that $d((T,P), \bar{T}, \bar{P})) < 13\epsilon$.

In the proof of Lemma 7.3, we constructed the desired Q's one step at a time. Here we do it N steps at a time. One can certainly choose Q^0 and $\bar{Q}^0$ by assuming that $\delta < 13\epsilon$ and $n^* > 0$. Suppose we have found sequences Q^i and $\bar{Q}^i$, $0 \leq i \leq m-1$, so that (12.15) holds for $n = m$. The hypothesis (12.10) and the result (iii) of Lemma 12.1 then imply that there is a class $\mathcal{E}_m \subset \bigvee_0^{m-1} T^iP$ and a class $\bar{\mathcal{E}}_m \subset \bigvee_0^{m-1} \bar{T}^i\bar{P}$ so that $\mu(\cup\mathcal{E}_m) \geq 1 - 3\epsilon$, $\bar{\mu}(\cup\bar{\mathcal{E}}_m) \geq 1 - \epsilon$, and

$$\bar{d}_N(\{T^{m+i}P/B\}_{i=0}^{N-1},\ \{\bar{t}^{m+1}\bar{p}/\bar{B}\}_{i=0}^{N-1}) < 5\epsilon, \tag{12.16}$$

for all $B \in \mathcal{E}_m$, $\bar{B} \in \bar{\mathcal{E}}_m$; provided we choose $\delta < \epsilon$.

Let A be the set in $\bigvee_0^{n-1} Q^i$ corresponding to $B \in \bigvee_0^{n-1} T^iP$, and let $\bar{A}$ be the set in $\bigvee_0^{m-1} \bar{Q}^i$ corresponding to $\bar{B} \in \bigvee_0^{m-1} \bar{T}^i\bar{P}$ under the correspondence given by (12.15a). The result (12.16) implies that, on the Lebesgue space $A \cap \bar{A}$ where $B \in \mathcal{E}$,

$\bar{B} \in \mathcal{E}_m$, we can construct partitions $Q^i/A \cap \bar{A}$ and $\bar{Q}^i/A \cap \bar{A}$, $m \le i \le m+N-1$, so that

$$
\begin{aligned}
d(\bigvee_{m}^{m+N-1} Q^i/A \cap \bar{A}) &= d(\bigvee_{m}^{m+N-1} T^i P/B), \\
d(\bigvee_{m}^{m+N-1} \bar{Q}^i/A \cap \bar{A}) &= d(\bigvee_{m}^{m+N-1} \bar{T}^i \bar{P}/\bar{B})
\end{aligned}
\tag{12.17}
$$

and

$$
\frac{1}{N} \sum_{i=m}^{m+N-1} |Q^i/A \cap \bar{A} - \bar{Q}^i/A \cap \bar{A}| < 5\epsilon. \tag{12.18}
$$

For those sets $A \cap \bar{A}$, for which the corresponding B is not in $\mathcal{E}_m$ or the corresponding $\bar{B}$ is not in $\bar{\mathcal{E}}_m$, we merely define $Q^i/A \cap \bar{A}$ and $\bar{Q}^i/A \cap \bar{A}$ so that (12.17) holds. This will guarantee that (12.15a) holds for $n = m + N$. The fact that (12.18) holds and that $\mu(\cup \mathcal{E}_m) \ge 1 - 3\epsilon$, $\bar{\mu}(\cup \bar{\mathcal{E}}_m) \ge 1 - 3\epsilon$ then imply that

$$
\frac{1}{N} \sum_{i=m}^{m+N-1} |Q^i - \bar{Q}^i| < 13\epsilon,
$$

and this certainly implies that (12.15b) holds for $n = m + N$. This completes the proof of Theorem 12.3.

CHAPTER 13.
FURTHER RESULTS AND QUERIES

We list here a number of recent results which make use of the ideas described in this work. We also list some open questions.

1. K-AUTOMORPHISMS. A natural generalization of the class of Bernoulli shifts is the class of Kolmogorov or K-automorphisms. T is a K-*automorphism* if it has a generator P such that every set in the *tail field* $\cap_n \Sigma_n$ has measure 0 or 1, where Σ_n is the σ-algebra generated by $\cup_{i \geq n} T^i P$.

An alternative definition of K-automorphisms is due to Rohlin and Sinai ([45]). T is a K-automorphism if $H(T, P) > 0$ for *every* nontrivial partition P. Thus K-automorphisms are *completely nondeterministic* in the sense that no nontrivial partition is measurable with respect to its past.

The significance of K-automorphisms is due primarily to the work of a number of Russian authors in the 1960's, who established that many transformations of physical and mathematical interest are K-automorphisms. These include the class of ergodic (group) automorphisms of separable compact groups, the so-called Anosov diffeomorphisms, geodesic flow on manifolds of constant negative curvature, and the flow associated with the hard sphere gas (see the references in [48], [36]). Much effort has gone into the problem of showing that these transformations are actually Bernoulli shifts, particularly after it was shown that *not* every K-automorphism is Bernoulli ([17]). We describe here some of the results which have been obtained.

Adler and Weiss in [26] showed that an ergodic automorphism of the torus (i.e., the product of two circles) is a mixing Markov shift. The results of Chapter 12 therefore imply that these are Bernoulli. Katznelson ([31]) used Fourier approximation methods to show that, if T is an ergodic automorphism of the n-torus (the product of n circles), then T is Bernoulli. This is done by taking a partition P into rectangles, then approximating the characteristic function of a set A in P by a Fejer polynomial. In this way, one can show that given $\epsilon > 0$ and M, there is an N such that, for all $m > 0$, $\bigvee_{-m}^{-1} T^i P$ is ϵ-independent of $\bigvee_{N}^{N+N^2} T^i P$.

This condition is slightly weaker than weak Bernoulli, but it is easy to see that it implies very weak Bernoulli. The proof of Katznelson's theorem is completed by using the following result of Ornstein (see [14]).

THEOREM 13.1. If $\{\Sigma_n\}$ is an increasing sequence of σ-algebras which are invariant under T and T^{-i}, and the restriction of T to each Σ_n is Bernoulli, then T is Bernoulli on the join of the Σ_n.

The above results have been modified slightly in an unpublished work of Katznelson and Weiss to show that an ergodic automorphism of a solenoidal group (i.e., a group whose dual is a subgroup of the rationals) is Bernoulli. It remains an open question whether these results can be extended to arbitrary compact groups.

An Anosov or C-diffeomorphism is a diffeomorphism of a manifold with a smooth invariant measure which has foliations into expanding and contracting directions (see [27], pp. 53f). The horizontal and vertical foliations of a Baker's transformation can be regarded, respectively, as expanding and contracting foliations. The existence of such foliations has been exploited to establish that such transformations are Bernoulli. The recent paper of Ornstein and Weiss ([44]) has an extensive discussion of this foliation property, using it to prove that geodesic flow on a manifold of negative curvature is Bernoulli.

Transformations associated with other physical models such as the Ising model are of interest. Gallavotti ([30]) has shown that the shift in the one-dimensional model with finite first moment is Bernoulli, and has utilized the recent results on commuting families ([32]) to show that certain two-dimensional shifts are Bernoulli. These results are clearly only a beginning.

The problem of classifying K-automorphisms is completely open. It is known that there is an uncountable family of nonisomorphic K-automorphisms with the same entropy ([17]), and that there exists a K-automorphism with no square root ([39]), and a K-automorphism with no roots at all (unpublished work of J. Clark). It is possible that the rigid block structure of the family in [17] will form the basis for further invariants.

If Σ' is a sub-σ-algebra invariant under T and T^{-1}, the transformation T on the space (X, Σ', μ) is called a *factor* of T, and Σ' is called a *factor algebra* of T. If T is a K-automorphism, then T has Bernoulli factors of full entropy (this follows from Sinai's Theorem, Chapter 9), and hence Theorem 13.1 above implies that T has maximal Bernoulli factors of full entropy. It is unlikely that such maximal factors are unique; in fact, one can show that

THEOREM 13.2. If T is a K-automorphism which is not Bernoulli, then some power T^N has more than one maximal Bernoulli factor.

A sketch of the proof of this result follows: Let P be a generator for T. The condition that T be K implies that, for some N, the sequence $\{S^iP\}$ is ϵ-independent where $S = T^N$. It follows that, given $\epsilon > 0$, there is an N and a partition Q such that $|Q - P| \le \epsilon$ and $\{S^iQ\}$ is independent, where $S = T^N$. Clearly, each of the sequences $\{S^iQ\}$, $\{S^i(TQ)\}$, $\{S^i(T^{N-1}Q)\}$ will then be independent. If Theorem 13.2 were false, then the partition $\bigvee_0^{N-1} T^iQ$ would be finitely determined for S, hence Q would be finitely determined for T. Hence, if Theorem 13.2 were false, there would exist a sequence $Q^{(n)}$ such that $|P - Q^{(n)}| \le 1/n$, and $Q^{(n)}$ is finitely determined for T. This implies that T must be Bernoulli ([36]). This contradiction establishes the theorem.

Related to this is the recent result of the author who showed that if S has the same factor algebras as a Bernoulli shift T, then S is isomorphic to T ([47]). This may, in fact, characterize Bernoulli shifts.

2. BERNOULLI FLOWS. Ornstein has extended the isomorphism theorem to Bernoulli flows ([38]). He has shown that if S_t is a flow such that $H(T_1) = \log 2$ and S_1 is Bernoulli, then there is a measure-preserving U such that $U^{-1}S_tU$ is the flow discussed in Chapter 12. The proof is obtained by suitably adapting the Fundamental Lemma to flows. Note that it implies in particular that if one member of a flow S_1 is Bernoulli, then each S_t, $t \neq 0$, is Bernoulli. This is easy to prove for rational t, but appears to need the isomorphism theorem of [38] for irrational t. For flows, one can prove that if T_1 is a K-automorphism, then so is T_t for each $t \neq 0$. It is not known whether two flows $\{T_t\}$ and $\{S_t\}$, for which T_1 and S_1 are isomorphic K-automorphisms, are isomorphic.

Many other open questions about flows remain. For example, if S is the base of a flow T_t built under f, one would like to have necessary and sufficient conditions on S and f for which the flow is Bernoulli, or at least strong enough sufficient conditions to cover all cases of interest. Recently, Bunimovitsch ([28]) has shown that if S satisfies a condition like weak Bernoulli relative to Q, and if f satisfies a Lipschitz condition relative to Q, then $\{T_t\}$ is a Bernoulli flow.

It would also be useful for various physical models to have an isomorphism theory for commuting flows. Towards this end, D. Lind ([33]) has obtained a Rohlin-type theorem for such flows.

3. DIRECT AND SKEW PRODUCTS. If T_1 and T_2 are given transformations on X_1 and X_2, the *direct product* $T = T_1 \times T_2$ is defined on the product space $X_1 \times X_2$ (with the product measure) by

$$T(x_1, x_2) = (T_1x_1, T_2x_2).$$

A more general concept is that of the skew product. Here we are given T_1 on X_1 and a family $\{\phi_{x_1} | x_1 \in X_1\}$ such that $(x_1, x_2) \to \phi_{x_1}(x_2)$ is measurable. The *skew product* T of T_1 with $\{\phi_{x_1}\}$ is defined by

$$T(x_1, x_2) = (T_1x_1, \phi_{x_1}(x_2)).$$

There are a number of results about direct and skew products and many open questions. For example, the Pinsker conjecture asserted that any ergodic transformation is a direct product of a K-automorphism and a transformation of entropy zero. Ornstein constructed a K-automorphism with no square root ([39]), which disproves the Pinsker conjecture for ergodic transformations, and then showed that it is also not true for mixing transformations ([40]). It is possible, however, that the Pinsker conjecture holds for classes of transformations of physical interest, for example, for factors of direct products of rotations with Bernoulli shifts.

Skew products are of interest for the following reason: If Σ_1 is a subalgebra which is invariant under T and T^{-1}, let xRy be the relation that x and y cannot be separated by sets in Σ_1. This gives a new space X_1 of equivalence classes, a representation of T as a skew product of T_1 (the action of T on X_1), and a family $\{\phi_x\}$, where ϕ_x is the action of T on the equivalence class of x. The transformation T_1 is called a factor of T. Sinai's Theorem, Chapter 9, implies that any ergodic transformation of positive entropy has Bernoulli factors of that entropy, and hence has a skew product representation where T_1 is Bernoulli.

The author and R. Adler have investigated some of the possiblities when T_1 is Bernoulli and each ϕ_x is a rotation ([24]). It was shown that if ϕ_x is constant on the sets of an independent generator T_1, and T is mixing, then T is Bernoulli. This can be generalized to Markow generators, but little else is known.

4. CUTTING AND STACKING. Transformations (and partitions) can be defined in stages by cutting and stacking gadgets made up of intervals. For example, the author has shown how to construct Bernoulli shifts in this way ([46]). Modifications of the method were used to construct non-Bernoulli K-automorphisms ([17]) and also yield a mixing transformation (of entropy 0), which has *no* nontrivial invariant subalgebras and commutes only with its powers ([41]). One can also use such methods to show that any transformation of finite entropy has a finite generator, a result originally due to Krieger (see [34], [22]). There are many open questions about constructions of this type. Rohlin's theorem says (roughly) that any ergodic transformation is obtained by cutting and stacking of intervals. The kinds of transformations and properties like mixing, Bernoulli, etc., which one can *effectively* construct in this way need much further study.

5. INFINITE ENTROPY. Let X be the countable direct product of copies of a given separable probability space Y, and let T be the shift on X. T is called a *(generalized) Bernoulli shift*. This monograph has been concerned with the case when Y is a finite set, but many of the results extend to the case when Y is countable, or when Y has a non-atomic part. If $Y = \{y_i\}$ is countable with mass p_i at y_i, then

$$H(T) = -\sum_i p_i \log p_i,$$

and this may or may not be finite. If it is finite, Smorodinsky showed that T has a finite independent generator ([22]). If Y has a nonatomic part, then $H(T) = \infty$. Ornstein showed that two Bernoulli shifts of infinite entropy are isomorphic ([37]), and has extended other results to the infinite entropy case ([14], [38]).

The author and R. McCabe showed that the shift associated with a continuous state discrete time stationary ergodic Markov process is Bernoulli if the spectrum of the associated Markov operator is contained (except for the eigenvalue 1) in a circle interior to the disc ([35]). The author and Ornstein extended this to the case of Markov operators of kernel type, which showed that Brownian motion in a reflecting rectangle region is a Bernoulli flow ([43]). J. Feldman and M. Smorodinsky have shown that a continuous finite state Markov process is Bernoulli ([29]).

Bibliography

The following are the books and articles referred to in Chapters 1-12.

1. Ambrose, W. Representation of ergodic flows. *Ann. of Math.* 42 (1941):723-739.

2. Breiman, L. *Probability.* Reading, Mass.: Addison-Wesley, 1968.

3. Billingsley, P. *Ergodic theory and information.* New York: Wiley, 1965.

4. Blum, J., and Hanson, D. On the isomorphism problem for Bernoulli schemes. *Bull. Amer. Math. Soc.* 69 (1963): 221-223.

5. Friedman, N., and Ornstein, D. On isomorphism of weak Bernoulli transformations. *Advances in Math.* 5 (1970): 365-394.

6. Hall, M. *Combinatorial theory.* Waltham, Mass.: Blaisdell, 1967.

7. Halmos, P. *Lectures on ergodic theory.* Tokyo: Math. Soc. of Japan, 1956.

8. Halmos, P. *Measure theory.* Princeton, N.J.: van Nostrand, 1950.

9. Kolmogorov, A. N. A new invariant for transitive dynamical systems. *Dokl. Akad. Nauk SSSR* 119 (1958): 861-864.

10. Kolmogorov, A. N. Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk SSSR* 124 (1959): 754-755.

11. McMillan, B. The basis theorems of information theory. *Ann. Math. Statist.* 24 (1958): 196-219.

12. Meshalkin, L. D. A case of isomorphism of Bernoulli schemes. *Dokl. Akad. Nauk SSSR* 128 (1959): 41-44.

13. Ornstein, D. S. Bernoulli shifts with the same entropy are isomorphic. *Advances in Math.* 4 (1970): 337-352.

14. Ornstein, D. S. Factors of Bernoulli shifts are Bernoulli shifts. *Advances in Math.* 5 (1970): 349-364.

15. Ornstein, D. S. Imbedding Bernoulli shifts in flows. *Springer Lecture Notes* 160 (1970): 178-218.

16. Ornstein, D. S. An application of ergodic theory to probability. *Ann. of Probability* 1 (1973): 43-58.

17. Ornstein, D. S., and Shields, P. An uncountable family of K-automorphisms. *Advances in Math.* 10 (1973): 63-88.

18. Rohlin, V. A. On the fundamental ideas of measure theory. *Mat. Sb.* 25 (1949): 107-150; *Amer. Math. Soc. Translations*, ser. 1, vol. 10 (1962): 1-54.

19. Shannon, C. A mathematical theory of communication. *Bell System Tech. J.* 27 (1948): 379-423, 623-656.

20. Sinai, J.G. On the notion of entropy of a dynamical system. *Dokl. Akad. Nauk SSSR* 124 (1959): 768-771.

21. Sinai, J. G. A weak isomorphism of transformations with invariant measure. *Mat. Sb.* 63 (1964): 23-42; *Amer. Math. Soc. Translations*, ser. 2, vol. 57 (1966): 123-143.

22. Smorodinsky, M. Ergodic theory, entropy. *Springer Lecture Notes* 214 (1970).

23. Totoki, H. Ergodic theory. *Aarhus Lecture Notes* 14 (1970).

Additional References:

24. Adler, R., and Shields, P. Skew products of Bernoulli shifts with rotations. *Israel J. Math.* 12 (1972): 215-220.

25. Adler, R., Shields, P., and Smorodinsky, M. Irreducible Markov shifts. *Ann. Math. Statist.* 43 (1972): 1027-1029.

26. Adler, R., and Weiss, B. Similarity of automorphisms of the torus. *Mem. Amer. Math. Soc.* 98.

27. Arnold, V. I., and Avez, A. *Ergodic problems of classical mechanics.* New York: Benjamin, 1968.

28. Bunimovitsch, L. A. Private communication.

29. Feldman, J., and Smorodinsky, M. Bernoulli flows with infinite entropy. *Ann. Math. Statist.* 42 (1971): 381-382.

30. Gallavotti, G. Ising model and Bernoulli schemes in one dimension. Preprint.

31. Katznelson, Y. Ergodic automorphisms of T^n are Bernoulli shifts. *Israel J. Math.* 10 (1971): 186-195.

32. Katznelson, Y., and Weiss, B. Commuting measure preserving transformations. *Israel J. Math.* 12 (1972): 161-173.

33. Lind, D. Rohlin's Theorem for n-dimensional flows. To appear.

34. Krieger, W. On entropy and generators of measure preserving transformations. *Trans. Amer. Math. Soc.* 199 (1970): 453-464.

35. McCabe, R., and Shields, P. A class of Markov shifts which are Bernoulli shifts. *Advances in Math.* 6 (1971): 323-328.

36. Ornstein, D. S. Some new results in the Kolmogorov-Sinai theory of entropy and ergodic theory. *Bull. Amer. Math. Soc.* 77 (1971): 878-890.

37. Ornstein, D. S. Two Bernoulli shifts wtih infinite entropy are isomorphic. *Advances in Math.* 5 (1970): 339-348.

38. Ornstein, D. S. The isomorphism theory for Bernoulli flows. *Advances in Math.* 10 (1973): 124-142.

39. Ornstein, D. S. A K-automorphism with no square root and Pinsker's conjecture. *Advances in Math.* 10 (1973): 89-102.

40. Ornstein, D. S. A mixing transformation for which Pinsker's conjecture fails. *Advances in Math.* 10 (1973): 103-123.

41. Ornstein, D. S. A mixing transformation that commutes only with its powers. *Proc. 6th Berkeley Symp. Math. Stat. Prob.* II, Univ. of Calif. Press (1967): 335-360.

42. Ornstein, D. S. *Ergodic theory, randomness and dynamical systems.* New Haven: Yale Univ. Press, 1973.

43. Ornstein, D. S., and Shields, P. Mixing Markov shifts of kernel type are Bernoulli. *Advances in Math.* 10 (1973): 143-146.

44. Ornstein, D. S., and Weiss, B. Geodesic flows are Bernoullian. *Israel J. Math.*, to appear.

45. Rohlin, V. A., and Sinai, J. G. Construction and properties of invariant measure partitions. *Dokl. Akad. Nauk SSSR* 141 (1961): 1038-1041.

46. Shields, P. Cutting and stacking of intervals. *Math. Systems Theory* 7 (1973): 1-4.

47. Shields, P. A Bernoulli shift is determined by its factor algebras. *Proc. Amer. Math. Soc.*, to appear.

48. Sinai, J. G. Dynamical systems with elastic reflections. *Uspehi Mat. Nauk.* 27 (1972): 137-189.

Index