SOME PROBLEMS OF ERGODIC THEORY AND INFORMATION THEORY

By

K. R. Parthasarathy

Research and Training School
Indian Statistical Institute
Calcutta, 1961

PREFACE

This thesis is being submitted to the Indian Statistical Institute in support of the author's application for the degree of Doctor of Philosophy. The thesis embodies research carried out by the author during the period 1959-1961 under the supervision of Prof. C.R. Rao at the Indian Statistical Institute, Calcutta.

The thesis is concerned with the development of a duality between the theory of dynamical systems and information theory and its application to problems of information theory. Some of the results have already appeared in an article by the author [15].

The author is indebted to Messrs. R. Rangarao and S.R.S. Varadhan for their criticisms and suggestions. The author also records his gratefulness to the Research and Training School of the Indian Statistical Institute for providing facilities for research. Thanks are due to Messrs. G.M. Das and J.N. Seth for their efficient typing of the thesis.

Indian Statistical Institute,
Calcutta.
July, 1961.

K.R. Parthasarathy

C O N T E N T S

INTRODUCTION

The subject of ergodic theory which is the mathematical outgrowth of the gas problem in statistical mechanics and the subject of information theory which is the mathematical outgrowth of the problem of information transmission through communication channels have many interesting dual features. In ergodic theory we study collectively and individually the measure preserving transformations of a fixed measure space, whereas in information theory we do the same with the set of all measures invariant under a fixed measurable transformation of a measurable space. The object of this thesis is to study some of the dual relations between the two theories with emphasis on the information-theoretic aspect.

In the first chapter we take up the global problem. Here the problem of ergodic theory is to ensure that there is a large class of ergodic transformations in any non-atomic measure space. If (X, S, μ) is a finite measure space and G the group of all one to one measure preserving transformations, then two interesting topologies can be assigned to G which make it a topological group. In dynamical problems it is of interest to know whether a particular transformation is ergodic or not. Even though this problem has defied solution till now, the existence of a large class of ergodic transformations has been shown by the determination of their category in G. In particular, Halmos [5] proved that the set of weakly mixing transformations is a dense G_δ in G under the weak topology. Similar results were proved earlier by Oxtoby and Ulam [12]. Rokhlin [18] proved that under the same weak topology in G, the set of strongly mixing transformations is a set of first category. Here the main tool is to show the density of what are called periodic transformations.

In problems of information-theoretic interest, we have a fixed measurable space (X, S) and a one-to-one both ways measurable map T of X onto itself. Here, it is of interest to know whether there are a lot of ergodic measures in the space of invariant probability measures. In order to study this problem, we take X to be a topological space, S the Borel σ-field and T a homeomorphism of X onto itself. Taking X to be a complete and separable metric space and assigning the weak topology to the space of invariant probability measures, we show that the set of ergodic measures is a G_δ. Now the question arises as to what are the complete separable metric spaces and what are the homeomorphisms under which the ergodic measures are dense. This classification problem has not been solved even in the case of a compact metric space. But, however, in spaces of information-theoretic importance we have solved this problem. When X is a countable product of complete and separable metric spaces and T is the shift transformation, we show that the ergodic measures form a dense G_δ under the weak topology. In this context it is not without interest to note that the ergodic measures constitute the set of extreme points of the convex set of all invariant probability measures. Examples are given to show that ergodic measures need not be dense in the general case.

We introduce the concept of periodic invariant measures and study their structure. We show that, in the case of shift transformation, the periodic measures are dense in the weak topology and thereby deduce the first category nature of the set of strongly mixing measures. Further, whenever the periodic measures are dense in the weak topology the closure of the set of periodic points is of invariant measure one. The converse problem remains open.

In the second chapter we introduce the notions of entropy of a stationary source and rate of transmission of a stationary channel and study some of their properties. In the problem of classification of measure-preserving transformations the fundamental role of the notion of entropy as a metric invariant has been demonstrated in the recent works of A. N. Kolmogorov [8]. In this connection V. A. Rokhlin [17] points out that there is a large class of measure-preserving transformations with zero entropy and hence in such cases entropy happens to be a trivial invariant. This is done by examining the category of the set of transformations with zero entropy in the space of all measure-preserving transformations under a suitable topology. We have shown that in the case of shift transformation the set of all ergodic probability measures with zero entropy is a dense G_δ in the space of all invariant probability measures.

Then we study the problem of representing the entropy of a stationary information source as an integral of the entropies of its ergodic components. We do the same for the rate of transmission of a stationary input distribution through a stationary communication channel. The continuity properties of different functionals associated with a channel are discussed.

In the last chapter we study the applications of the results obtained in the first and second chapters. We prove that the stationary and ergodic capacities are equal for an arbitrary stationary channel. This has been proved for channels with finite memory by I. P. Tsaragradsky [22], A. Feinstein [4] and L. Breiman[1]. We then study the problem of achievement of capacity for channels of finite memory in the sense of Feinstein. Finally we give a limiting form of the famous Feinstein's fundamental lemma which throws some light on the assumptions under which the lemma is proved.

1. ERGODIC, PERIODIC AND STRONGLY MIXING MEASURES

1.1. Preliminaries

Let $(X, \mathcal{S})$ be any measurable space and T a one to one both ways measurable map of X onto itself. Whenever the space X is a topological space, we take $\mathcal{S}$ to be the Borel σ-field and T a homeomorphism of X onto itself. By a measure or a distribution, we always mean a probability measure. We denote by $\mathcal{M}$, $\mathcal{M}_e$ and $\mathcal{M}_s$ the space of all invariant, ergodic and strongly mixing measures respectively. For these definitions we refer to [5].

A point $x \in X$ will be called periodic if for some integer k, $T^k x = x$. The smallest k for which the equality $T^k x = x$ holds is called the period of x. A measure $\mu \in \mathcal{M}$ is said to be periodic if for some integer k, $\mu(A \cap T^k A) = \mu(A)$ for all sets $A \in \mathcal{S}$. We shall denote by $\mathcal{P}$ and $\mathcal{P}_e$ the class of all invariant periodic measures and the class of all ergodic periodic measures respectively.

When X is a topological space, we assign the weak topology to $\mathcal{M}$ by means of the following convergence: a net $\{\mu_\alpha\}$ in $\mathcal{M}$ converges to μ if and only if $\int f d\mu_\alpha \to \int f d\mu$ for every bounded continuous function defined on X. This topology can actually be defined for the space of all probability measures. Then the space $\mathcal{M}$ can then be viewed upon as a closed subset of the space of all probability measures. In the case when X is a separable metric space the following theorem (which is an immediate consequence of Prohorov's result [17]) concerning the topological nature of $\mathcal{M}$ is of fundamental importance in our study.

Theorem 1.1.1. (Prohorov [17]): When X is a separable metric space the weak topology of $\mathcal{M}$ becomes separable and metric. If further X is complete then $\mathcal{M}$ is also complete.

The following result concerning the weak topology of $\mathcal{M}$ is well known.

Theorem 1.1.2. Sets of the type

$$[\ \mu : \mu(G_i) > \mu_o(G_i) - a_i, \quad i = 1, 2, \ldots, k]$$

$$[\ \mu : \mu(C_i) < \mu_o(C_i) + a_i, \quad i = 1, 2, \ldots, k]$$

where $G_1, \ldots, G_k$ are open sets in X, $C_1, C_2, \ldots, C_k$ are closed sets in X, μ_o is a fixed measure in $\mathcal{M}$ and μ denotes any general invariant measure, form a neighbourhood system at μ_o.

We shall now give a brief description of the way in which ergodic measures are constructed from certain simpler measures and the invariant measures are constructed from ergodic measures. In the case when X is a compact metric space and T is a homeomorphism of X into itself Krylov and Bogolioubov [9] have obtained some important results in this direction and Fomin [13] generalised them to the case of a complete metric space. A detailed account of this is given in Oxtoby [13]. We shall just state the results which will be used in the sequel.

Let X be a complete separable metric space. If $f(x)$ is a real valued function of $x \in X$, let

$$M(f, x, k) = f_k(x) = \frac{1}{k} \sum_{i=1}^{k} f(T^i x) \qquad (k = 1, 2, \ldots)$$

and

$$M(f, x) = f^*(x) = \lim_{k \to \infty} M(f, x, k)$$

in case this exists. A Borel subset E of X is said to have invariant measure one if $\mu(E) = 1$ for every invariant probability measure μ. Let $C(X)$ be the space of all real-valued bounded continuous functions defined on X. We introduce the following definitions:

Definition 1.1.1. A point $x \in X$ is said to be quasi-regular with respect to the space X and the transformation T if

(1) the mean value $M(f, x)$ exists for each $f \in C(X)$

and

(2) for every $\epsilon > 0$ there is a compact set $K \subset X$ ~~such that~~ such that $M(\chi_K, x) > 1 - \epsilon$, χ_K being the characteristic function of K.

With the above definition of a quasi-regular point we have the following theorems.

Theorem 1.1.3. Let X be a complete separable metric space. Then, associated with every quasi-regular point, there is a unique invariant probability measure μ_x defined on the Borel field $\mathcal{S}$ such that

(1.1.1) $$M(f, x) = \int f \, d\mu_x$$

for every $f \in C(X)$.

Theorem 1.1.4. The set of quasi-regular points is Borel-measurable and of invariant measure one.

Definition 1.1.2. A point x is said to be regular if it is quasi-regular and the associated measure μ_x given by (1.1.1) is ergodic.

Let R be the set of all regular points. Then we have

Theorem 1.1.5. The set R of regular points is Borel-measurable and of invariant measure one.

Theorem 1.1.6. For any ergodic measure μ, the set of regular points x such that $\mu_x = \mu$ is of μ-measure one.

Theorem 1.1.7. For any bounded Borel measurable function f on X, $\int f \, d\mu_x$ is a Borel measurable function of x on R, and

$$\int f \, d\mu = \int_R \left[\int f \, d\mu_x \right] d\mu(x)$$

for every invariant Borel measure μ.

Theorem 1.1.8. For any Borel set $E \subset X$, $\mu_x(E)$ is Borel measurable on R, and

$$\mu(E) = \int_R \mu_x(E) \, d\mu(x)$$

for every invariant Borel measure μ.

The above theorems indicate how the invariant measures are built out of the degenerate measures in complete and separable metric spaces. We have to take an arbitrary point x and construct the sequence of measures μ_n where μ_n has mass $\frac{1}{n}$ at the points x, Tx, ... $T^{n-1}x$. If this sequence of measures is compact in the weak topology then there exists a unique limit μ_x which is an invariant probability measure. Under the weak topology we take the closed convex hull generated by all measures of the type μ_x, x being a regular point. This convex hull is precisely the class of all invariant probability measures. It is well known that in the case when X is a complete and separable metric space and T a homeomorphism, the class of all invariant measures is a convex set whose extreme points are precisely the ergodic measures. These points will be made use of in the sequel.

Another important fact which we shall make use of is the following result due to Varadarajan [23] concerning uniformly continuous functions in a separable metric space.

<u>Theorem</u> 1.1.9. If X is a separable metric space, then there exists an equivalent metric d such that the space $U_d(X)$ of functions uniformly continuous with respect to d is separable in the uniform topology.

1.2. Topological nature of ergodic measures in a separable metric space.

In this section we shall prove the following theorem.

Theorem 1.2.1. If X is a separable metric space and T is a homeomorphism of X onto itself, then the set $\mathcal{M}_e$ of all ergodic measures is a G_δ in the space $\mathcal{M}$ of all invariant measures under the weak topology.

Proof: It is clear that the class of all Borel sets S with the property $S = TS$ form a σ-field $\mathcal{J}$. Let $\mathcal{C}(X)$ be the space of all real valued bounded continuous functions defined on X. For any fixed measure μ and any $f \in C(X)$, we denote by $E_\mu (f \mid \mathcal{J})$ the conditional expectation of $f(x)$ given the σ-field $\mathcal{J}$. It is easy to see that μ is ergodic if and only if $E_\mu (f \mid \mathcal{J})$ is a constant with probability one for every $f \in C(X)$. This condition can be expressed by the following equation:

$$(1.2.1) \qquad V(f, \mu) = \int [E_\mu (f \mid \mathcal{J})]^2 \, d\mu - (\int f \, d\mu)^2 = 0$$

for every $f \in C(X)$. It is enough if (1.2.1) is satisfied for every bounded uniformly continuous function. This is because of the fact that any bounded continuous function f is a pointwise limit of a uniformly bounded sequence of uniformly continuous functions and the conditional

dominated convergence theorem is applicable (see Doob [ ], pp.
By making use of theorem 1.1.9 we can take the space $U(X)$ of bounded uniformly continuous functions to be separable in the uniform topology. We take a dense sequence $f_1(x)$, $f_2(x)$, ..., in $U(X)$. Thus, in order that an invariant measure μ be ergodic it is necessary and sufficient that

$$(1.2.2) \qquad V(f_k, \mu) = 0, \quad k = 1, 2, \ldots$$

Let

$$(1.2.3) \qquad V_n(f_k, \mu) = \int \left[\frac{f_k(x) + \ldots + f_k(T^{n-1} x)}{n} \right]^2 d\mu - \left(\int f_k d\mu \right)^2 .$$

From the mean ergodic theorem it follows that

$$(1.2.4) \qquad V(f_k, \mu) = \lim_{n \to \infty} V_n(f_k, \mu) = \liminf_{n \to \infty} V_n(f_k, \mu).$$

For each fixed k and n, $V_n(f_k, \mu)$ is a continuous functional in $\mathcal{M}$ under the weak topology. From (1.2.2) and (1.2.4) it follows that

$$(1.2.5) \qquad \mathcal{M}_e = \bigcap_{k=1}^{\infty} \bigcap_{r=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \left[\mu : V_n(f_k, \mu) < \frac{1}{r} \right]$$

The continuity of $V_n(f_k, \mu)$ implies that the set $\left[\mu : V_n(f_k, \mu) < \frac{1}{r} \right]$ is open in the weak topology. This together with (1.2.5) implies that $\mathcal{M}_e$ is a G_δ.

1.3. Measures invariant under the shift transformation in a product space.

Let (M, S) be a separable metric space with its Borel σ-field denoted by S and $(X, \mathfrak{S})$ be the bilateral product of countable number of copies of (M, S). X can be written as

$$X = \prod_{-\infty}^{+\infty} M_i, \quad M_i = M \quad (i = \ldots -1, 0, 1 \ldots \quad)$$

and

$$\mathfrak{S} = \prod_{-\infty}^{+\infty} S_i, \quad S_i = S \quad (i = \ldots -1, \; 0, 1, \ldots \;).$$

Any point $x \in X$ can be represented by

$$x = (\ldots \quad x_{-1}, x_0, \; x_1, \; \ldots \;), \quad x_i \in M_i.$$

We introduce the shift operator T by means of the following definition:

$$Tx = y = (\ldots \; y_{-1}, y_0, y_1, \; \ldots \quad)$$

$$y_i = x_{i-1} \quad (i = \ldots \; -1, 0, 1, \; \ldots).$$

T is obviously a one-to-one both ways measurable map of X onto itself. In the space $\mathfrak{M}$ of measures invariant under T, we introduce a topology $\mathfrak{T}$ by means of the following convergence : a sequence of measures $\mu_n \in \mathfrak{M}$ converges to μ if and only if $\mu_n(A) \to \mu(A)$ as $n \to \infty$ for each finite dimensional measurable subset A.

<u>Theorem</u> 1.3.1. Under the topology $\mathcal{T}$ in $\mathcal{M}$ the set $\mathcal{M}_e$ of ergodic measures is everywhere dense in $\mathcal{M}$

<u>Proof</u> Let μ be any measure in $\mathcal{M}$ and μ_n^r the restriction of μ to the σ-field

$$\mathcal{C}_\tau^{\pi} = \prod_{i=r(2n+1)-n}^{r(2n+1)+n} S_i$$

and ν_π, the product measure given by

$$\nu_\pi = \prod_{r=-\infty}^{+\infty} \mu_n^r$$

which is defined on $\prod_{r=-\infty}^{+\infty} \mathcal{C}_\tau^{\pi} = \mathfrak{S}$. Then ν_π is defined on $\mathfrak{S}$ and is invariant under the transformation T^{2n+1} which is also one-to-one and both ways measurable. It is easy to verify that ν_π is ergodic under T^{2n+1}. Now we write for any set $A \in \mathfrak{S}$

$$(1.3.1) \qquad \mu_n(A) = \frac{\nu_n(T^{-n}A) + \nu_n(T^{-n+1}A) + \dots + \nu_\pi(A) + \dots + \nu_n(T^n A)}{2n+1}$$

From the invariance of ν_n under T^{2n+1}, the invariance of μ_n under T follows immediately. Let now A be any set in $\mathfrak{S}$ which is invariant under T, i.e. $A = TA$. Then $\mu_n(A) = \nu_n(A)$. Since $A = T^{2n+1} A$ and ν_n is ergodic under T^{2n+1} it follows that $\nu_n(A) = 0$ or 1 and hence $\mu_n(A) = 0$ or 1, i.e. $\mu_n(A)$ is ergodic under T and hence belongs to $\mathcal{M}_e$. We shall now prove that μ_n converges to μ under the topology $\mathcal{T}$. Let

$$\mathcal{C}_k = \prod_{r=-k}^{+k} S_i, \qquad k = 1,2, \ldots$$

be the σ-field which is the 2k+1 -fold product of copies of S. $\mathcal{C}_k$ can be considered as a sub σ-field of $\mathcal{S}$. From the construction of ν_n, it is clear that ν_n agrees with μ on $\mathcal{C}_n$. Let now $A \in \mathcal{C}_k$. Then $T^{-n+k} A$, $T^{-n+k+1} A$, ... $T^{n-k} A$ belong to $\mathcal{C}_n$. Thus

$$\nu_n(T^r A) = \mu(A) \quad \text{for } -n+k \leq r \leq n-k \tag{1.3.2}$$

We have from (1.3.1) and (1.3.2),

$$| \mu_n(A) - \mu(A) | = \left| \frac{\nu_n(T^{-n}A) + \ldots + \nu_n(T^n A)}{2n+1} - \mu(A) \right| \leq \frac{4k}{2n+1}. \tag{1.3.3}$$

Thus $\mu_n(A) \to \mu(A)$ as $n \to \infty$ for every $A \in \mathcal{C}_k$. The inequality (1.3.3) shows that not only there is setwise convergence in the σ-field $\mathcal{C}_k$ but there is uniform convergence. Since this is true for each fixed k, $\mu_n \to \mu$ in the topology $\mathcal{T}$. This completes the proof.

The following theorem is almost an immediate corollary of theorems 1.2.1 and 1.3.1.

__Theorem__ 1.3.2. If $X = \prod_{i=-\infty}^{+\infty} M_i$, $M_i = M$ $(i = \ldots -1, 0, 1 \ldots)$ where M is a complete and separable metric space and T is the shift transformation in X, then $\mathfrak{M}_e$ is a dense G_δ in $\mathfrak{M}$ under the weak topology and hence $\mathfrak{M} - \mathfrak{M}_e$ is of first category.

Proof. The first part is an immediate consequence of theorems 1.2.1 and 1.3.1 and the facts that $\mathfrak{M}$ is a complete and separable metric space under the weak topology and convergence under $\mathcal{T}$ implies weak convergence. The second part follows from theorem 1.1.1 and Baire category theorem.

Remarks. A disposition towards the method adopted in proving theorem 1.3.1, may already be found in the works of I.P.Tsaragradsky [22] and A. Feinstein [4] in a different context. We shall have occasion to use this method in later chapters. A result less general than theorem 1.3.1 has been proved recently by M. Nisio [11] by an entirely different procedure. The results we have proved here are contained in a paper by the author [16].

If in theorem 1.3.2 M is a compact metric space, then the space of all totally finite invariant measures becomes a compact convex set with $\mathfrak{M}_e$ as the set of extreme points. From theorem 1.3.2 it follows that $\mathfrak{M}_e$ is a dense G_δ in $\mathfrak{M}$. This is one of the many examples to show that in the infinite dimensional case the structure of the set of extreme points of a compact convex set in a topological vector space is different from that of the finite dimensional situation.

Theorem 1.3.2 states that in the space X with the shift operator, in some sense, the ergodic measures represent the general case.

This can be considered as a dual problem of G. D. Birkhoff's conjecture that, in some sense, ergodic transformations represent the general case. But theorem 1.3.2 is not true in the general case when X is any complete separable metric space and T any homeomorphism of X onto itself. Examples are given at the end of the chapter.

1.4. Periodic measures

We shall now prove the following theorem concerning periodic invariant measures

Theorem 1.4.1. If $X = \prod_{i=-\infty}^{+\infty} M_i$ $(i = \ldots, -1, 0, 1, \ldots)$, $M_i = M$, where M is a complete separable metric space and T is the shift transformation, then the set of periodic measures is dense in the set of all ergodic measures under the weak topology.

Proof. Since the conditions of theorems 1.1.6 are fulfilled, for any ergodic measure μ there exists a point $x \in X$ such that the sequence of measures

$$\mu_n = \frac{m_{T^{-n}x} + m_{T^{-n+1}x} + \cdots + m_{T^n x}}{2n+1}$$

converges weakly to μ, m_x being the degenerate measure with mass one at the point x. We shall now approximate μ_n by means of periodic measures. The point x can be represented by

$$x = (\dots x_{-1}, x_0, x_1, \dots)$$

$$x_i \in M_i, \quad (i = \dots -1, 0, 1, \dots)$$

We write

$$x^n = (\dots . y_{-1}, y_0, y_1 \dots \quad)$$

$$y_{k(2n+1)+r} = x_r \quad \text{for} \quad k = \dots -1, 0, 1, \dots \qquad -n \leq r \leq n$$

Then x^n is a periodic point of period $2n+1$. We consider

$$\nu_n = \frac{m_{T^{-n} x^n} + m_{T^{-n+1} x^n} + \dots + m_{T^n x^n}}{2n + 1}$$

Since $T^{2n+1} x^n = x^n$, ν_n is a periodic measure. Proceeding exactly as in the proof of theorem 1.3.1 it is not difficult to show that for every finite dimensional Borel set A $\mu_n(A) - \nu_n(A) \to 0$. This completes the proof of theorem 1.4.1.

Remarks. Theorem 1.4.1 can be considered as the dual of the following well known result concerning periodic measure-preserving transformations. In the group of all one to one measure preserving transformations of a separable non-atomic measure space the set of periodic transformations is everywhere dense in the uniform topology. We have proved the dual of this result in the special case of shift transformation of a countable product of copies of a complete and separable metric space. The fact that theorem 1.4.1 need not be true in general is clearly brought out by the following

__Theorem__ 1.4.2. If X is a complete separable metric space and the periodic ergodic measures are dense in the set of ergodic measures under the weak topology, then the complement of the closure of periodic points has measure zero for every invariant measure.

In order to prove this theorem we require the following lemmas.

__Lemma__ 1.4.1. If (X, S, μ) is a separable non-atomic measure space and T is a measure preserving transformation of period k at almost all points of X then there exists a measurable set E of measure $\frac{1}{k}$ such that the sets $E, TE, \ldots, T^{k-1} E$ are pairwise disjoint.

__Lemma__ 1.4.2. If T is an antiperiodic measure preserving transformation of a separable non-atomic measure space (X, S, μ), then for every positive integer n and for every positive number ε there exists a measurable set E such that the sets $E, TE, \ldots, T^{n-1} E$ are pairwise disjoint and such that $\mu(E \cup TE \quad \ldots \cup T^{n-1} E) > 1 - \varepsilon$.

__Lemma__ 1.4.3. If X is a complete separable metric space and μ is an ergodic measure with period k, then there exists a point $x_0 \in X$ such that $T^k x_0 = x_0$ and $\mu(\{x_0\}) = \mu(\{Tx_0\}) = \ldots$ $= \mu(\{T^{k-1} x_0\}) = \frac{1}{k}$.

For the proofs of lemmas 1.4.1 and 1.4.2, we refer to P. R. Halmos ([5], see pp. 70-71). We pass on to the proof of lemma 1.4.3. By theorem 1.1.6 and the remarks made in section 1.1 it is clear that, if μ is an ergodic measure, there exists a point $x \in X$ such that μ is the weak limit of a sequence of measures μ_n where μ_n has mass $\frac{1}{n}$ at each of the points $x, Tx, \ldots$ This shows that an ergodic measure is either purely atomic or purely non-atomic. In the atomic case the lemma is obvious. In the purely non-atomic case we can apply lemmas 1.4.1 and 1.4.2. An ergodic transformation is either periodic or antiperiodic. Let us suppose that T is antiperiodic (almost everywhere with respect to μ). Then by lemma 1.4.2 there exists a set E such that E and $T^k E$ are disjoint but $\mu(E) \neq 0$. Since by the definition of periodic measure $\mu(E \cap T^k E) = \mu(E)$ we arrive at a contradiction. Thus T can only be periodic. Hence by lemma 1.4.1. there exists a set E of measure $1/k$ such that $E, TE, \ldots, T^{k-1} E$ are disjoint. Let $F \subset E$ be any Borel set. Then $F, TF, \ldots, T^{k-1} F$ are disjoint and $F \cup TF \cup \ldots \cup T^{k-1} F$ is an invariant set. Since μ is ergodic the measure of $F \cup TF \cup \ldots \cup T^{k-1} F$ is either 0 or 1. Thus the measure of F is either zero or $1/k$. Since every Borel subset of E has this property the mass of the measure μ in the set E is concentrated at a point. This completes the proof of the lemma.

<u>Proof of theorem 1.4.2.</u> Let P be the set of all periodic points, $\bar{P}$ its closure and $G = X - \bar{P}$. Then G is an open subset of X. We shall now show that, for every ergodic measure μ, $\mu(G) = 0$. Then an

application of theorem 1.1.8 and Tonelli-Fubini theorem will complete the proof.

Let, if possible, $\mu(G) > 0$ for some ergodic measure μ. Since by hypothesis $\mathcal{P}_e$ is dense in $\mathcal{M}_e$ there exists a sequence $\mu_n \in \mathcal{P}_e$ such that μ_n converges weakly to μ. Since G is open, $\liminf_{n \to \infty} \mu_n(G) \geq \mu(G) > 0$. Thus there exists an n such that $\mu_n(G) > 0$. By lemma 1.4.3 there exists a point x_o such that $T^{a_n} x_o = x_o$ (a_n being the period of μ_n) and $\mu_n(x_o) = 1/a_n$. From the fact that $\mu_n(G) > 0$, it immediately follows that

$$\frac{\chi_G(x_o) + \chi_G(Tx_o) + \dots + \chi_G(T^{a_n - 1} x_o)}{a_n} > 0$$

where χ_G is the characteristic function of G. Thus for some r, $T^r x_o \in G$. Since $T^r x_o$ is a periodic point we arrive at a contradiction. This completes the proof.

Remark The converse of theorem 1.4.2 is still an open problem. It is true, for example, in the case when the system (X, T) is L-stable (see [13]).

1.5. Strongly mixing measures

The dual of Rokhlin's first category theorem for strongly mixing transformations [18] is contained in the following

Theorem 1.5.1. When X and T are the same as in theorem 1.4.2 the set $\mathcal{M}_s$ of strongly mixing measures is of first category in $\mathcal{M}$ under the weak topology.

Proof. Let $0 < \varepsilon < \frac{1}{2}$, $\eta < \frac{2}{3}\varepsilon^2$, δ any rational number $0 < \delta < \eta/2$, r any rational number in $0 \leq r \leq 1$, F_1 and F_2 two disjoint closed sets and G any open set such that $G \supset F_1$. We write

$$(1.5.1) \qquad \xi\,(F_1, F_2, G, \varepsilon, r, \delta, n) =$$

$$= \bigcap_{k=n}^{\infty} \{\mu : \mu(F_1) \geq \varepsilon,\ \mu(F_2) \geq \varepsilon,\ \mu(G \cap T^k G) \leq r + \delta,\ r \leq \mu^2(F_1) + \eta\},$$

where μ denotes any general invariant probability measure. Since F_1 and F_2 are closed and G is open, by theorem 1.1.2, the set (1.5.1) is closed under the weak topology. Let

$$(1.5.2) \qquad \xi\,(F_1, F_2, G, \varepsilon) = \bigcup_{0 \leq r \leq 1} \ \bigcup_{0 < \delta < \eta/2} \ \bigcup_{n=1}^{\infty} (F_1, F_2, G, \varepsilon, r, \delta, n)$$

It is not difficult to verify that

$$(1.5.3) \quad \xi\,(F_1, F_2, G, \varepsilon) = \bigcup_{0 \leq r \leq 1} \ \bigcup_{0 < \delta < \eta/2} \{\mu : \mu(F_1) \geq \varepsilon,\ \mu(F_2) \geq \varepsilon,$$

$$\limsup_{k \to \infty} \mu(G \cap T^k G) \leq r + \delta,\ r \leq \mu^2(F_1) + \eta\}$$

Let G_n be a sequence of open sets descending to F_1. Since, for a strongly mixing measure, $\lim_{k \to \infty} \mu(G \cap T^k G) = \mu^2(G)$ it is clear that all strongly mixing measures with the property $\mu(F_1) \geq \varepsilon$, $\mu(F_2) \geq \varepsilon$ belong to the set

(1.5.4) $$\bigcup_{G_n \downarrow F_1} \mathcal{E}(F_1, F_2, G_n, \varepsilon)$$

We shall now show that the set (1.5.4) is of the first category. From (1.5.2), (1.5.3) and (1.5.4), it is clear that the set (1.5.4) is a countable union of the closed sets $\mathcal{E}(F_1, F_2, G, \varepsilon, r, \delta, n)$. It is enough to show that these closed sets are nowhere dense or their complements are everywhere dense.

Let P_k be the set of all periodic measures of period k and $P_n^* = \bigcup_{k>n} P_k$. Since by theorem 1.4.1 periodic ergodic measures are dense in $\mathcal{M}_e$ it follows that the set of periodic invariant measures $\mathcal{P}$ is dense in $\mathcal{M}$. Thus P_n^* is everywhere dense in $\mathcal{M}$. We shall complete the proof by showing that

(1.5.5) $$P_n^* \subset \mathcal{M} - \mathcal{E}(F_1, F_2, G, \varepsilon, r, \delta, n).$$

The inclusion relation (1.5.5) is satisfied if

$$P_k \subset \mathcal{M} - [\mu: \mu(F_1) \geq \varepsilon, \mu(F_2) \geq \varepsilon, \mu(G \cap T^k G) \leq r+\delta, r \leq \mu^2(F_1)+\gamma]$$

Let now μ_o be any periodic measure of period k. If any one of the inequalities $\mu(F_1) \geq \varepsilon$, $\mu(F_2) \geq \varepsilon$ is violated, then we are through. Otherwise, since F_1 and F_2 are disjoint we have

$$\varepsilon \leq \mu_o(F_1) \leq 1 - \varepsilon, \quad \mu_o(G \cap T^k G) = \mu_o(G).$$

Since $0 < \delta < \eta/2$, it is enough to prove that

$$\mu_o(G) \geq \mu_o^2 (F_1) + \frac{3\eta}{2} \; . \tag{1.5.6}$$

Since $0 < \varepsilon < \frac{1}{2}$, $\varepsilon \leq \mu_o(F_1) \leq 1 - \varepsilon$, $G \supset F_1$ and the function $x - x^2 \geq \varepsilon(1 - \varepsilon)$ in $\varepsilon \leq x \leq 1 - \varepsilon$, $0 < \varepsilon < \frac{1}{2}$, we have

$$\mu_o(G) - \mu_o^2 (F_1) \geq \mu_o(F_1) - \mu_o^2(F_1) \geq \varepsilon(1 - \varepsilon) > \varepsilon^2 = \frac{3\eta}{2} .$$

Thus we have proved (1.5.5).

Let now $S(F_1, F_2, \varepsilon)$ denote the class of all strongly mixing measures with the property

$$\mu(F_1) \geq \varepsilon, \quad \mu(F_2) \geq \varepsilon.$$

We have proved that $S(F_1, F_2, \varepsilon)$ is of first category. Now we take a dense sequence of points and consider all closed spheres of rational radii with centers at these points. We denote this class of sets by A. Then A is a countable class. It is clear that the set of all non-degenerate strongly mixing measures is the same as

$$\bigcup_{\substack{F_1, F_2 \in A \\ F_1 \cap F_2 = \phi}} \ \bigcup_{\substack{0 < \varepsilon < \frac{1}{2} \\ \varepsilon \text{ rational}}} S(F_1, F_2, \varepsilon)$$

Since the set of degenerate strongly mixing measures is of first category and any other strongly mixing measure is non-atomic, we have completed the proof.

Remarks and Examples

We shall now give some examples to show that theorem 1.3.2 need not be true in general.

1) Let X_o be a compact group with at least one periodic element and the transformation T_o be the translation of X by a periodic element. Then the ergodic probability measures form a closed set under the weak topology.

2) Let (X_o, T_o) be as obove and (X_1, T_1) be the product space with the shift transformation. Let $X = X_o \times X_1$ and $T = T_o \times T_1$ be defined in the obvious manner. If X_1 is a complete separable metric space, then the set of ergodic measures is neither closed nor dense.

It was originally conjectured by the author that the density theorem (theorem 1.3.2) should be true whenever there exists a dense orbit. However, in the example discussed by Oxtoby [14], there exists a dense orbit and nevertheless the ergodic measures form a closed set. Thus it would be very interesting to get a characterisation of all those homeomorphisms of a complete separable metric space for which the density theorem is true. Nothing is known in this direction even in the case of a compact metric space.

From the proof of the first category theorem 1.5.1. it is clear that it holds good as soon as $\mathcal{P}_e$ is dense in $\mathcal{M}_e$ in the weak topology. Thus arises the problem of obtaining necessary and sufficient conditions on the homeomorphism T so that the periodic measures may be dense. This is true, for example, in the case when the system (X, T) is L-stable [13]. A necessary condition is given by theorem 1.4.2. It is conjectured that the converse of this result is true.

2. ENTROPY OF A SOURCE AND RATE OF TRANSMISSION THROUGH A CHANNEL

In this chapter we shall introduce the notions of entropy and rate of transmission and study some of the properties of invariant measures in this context.

2.1. Entropy of finite schemes and sources

Let A be a finite alphabet consisting of a symbols $\theta_1, \theta_2, \ldots, \theta_a$ and let a probability distribution be defined over A such that the probability of θ_i is p_i. Then $p_i \geq 0$ for $i = 1, 2, \ldots, a$ and $\sum_{i=1}^{a} p_i = 1$. Then we write

$$H(A) = - \sum_{i=1}^{a} p_i \log p_i .$$

(all logarithms will be with respect to the base 2). The quantity $H(A)$ is called the entropy of the finite scheme under the distribution $p_1, p_2, \ldots, p_a$.

If $A = (\theta_1, \theta_2, \ldots, \theta_a)$ and $B = (\varphi_1, \varphi_2, \ldots, \varphi_b)$ are two alphabets and a joint distribution $P(\theta_i, \varphi_j)$ is defined over the product alphabet AB such that

(1) $P(\theta_i) = p_i$

(2) $P(\varphi_j \mid \theta_i) = p_{ij}$

then the conditional entropy $H_A(B)$ of the scheme B given the scheme A is defined by

$$H_A(B) = - \sum_i p_i \sum_j p_{ij} \log p_{ij}.$$

Let

$$A^I = \prod_{i=-\infty}^{+\infty} A_i, \quad A_i = A, \ (i = \ldots -1, 0, 1, \ldots)$$

be the product of the alphabet taken over all integers. Let T denote the shift transformation introduced in the last chapter. Assigning the discrete topology to A and the corresponding product topology to A^I we make A^I a compact metric space. Let F_A be the Borel σ-field in the space A^I and $\mathfrak{M}_A$ the space of all probability measures defined on F_A and invariant under T. Assigning the weak topology to $\mathfrak{M}_A$ makes it a compact metric space.

<u>Definition 2.1.1.</u> If $\mu \in \mathcal{M}_A$ is an invariant probability measure, then the pair $[A^I, \mu]$ is called an information source.

Let $[A^I, \mu]$ be an information source. We shall denote by $[x_{i_1}, x_{i_2}, \ldots, x_{i_k}]$ the cylinder set of all points x in A^I for which the $i_1^{th}, \ldots, i_k^{th}$ co-ordinates are $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ respectively. Sometimes we refer to $i_1, i_2, \ldots$ as time points. The class of all cylinder sets $[x_1, x_2, \ldots, x_n]$ of length n constitute a finite scheme if we restrict the distribution to the n-dimensional cylinder sets. Let $H_n(\mu, A)$ denote the entropy of this scheme. Then $\frac{H_n(\mu, A)}{n}$ can be called the rate at which information is emitted by the source during the time period 1 to n. If the limit of $\frac{H_n(\mu, A)}{n}$ exists as $n \to \infty$ it may reasonably be called as the rate at which information is emitted by the source or the entropy of the source. One of the fundamental results of information theory asserts the following:

<u>Theorem 2.1.1.</u> The entropy of a source is always well defined i.e. the limit $\lim_{n \to \infty} \frac{H_n(\mu, A)}{n}$ exists.

Hereafter we shall denote by $H(\mu, A)$ the entropy of the source $[A^I, \mu]$.

2.2. <u>Sources with zero entropy.</u>

The space A^I together with the shift transformation becomes a compact dynamical system when A^I is considered as a compact metric space and T as a homeomorphism of A^I onto itself. One of the important problems in the theory of dynamical systems is to find the complete set of invariants for a dynamical system with an invariant measure. Let us consider the class of measures defined on F_A and invariant under T. Let μ_1 and μ_2 be two such measures. We say that $[A^I, \mu_1, T]$ and $[A^I, \mu_2, T]$ are isomorphic if there exists a one-to-one measurable transformation U of A^I onto itself such that $UT = TU$ and the induced measure $\mu_1 U^{-1}$ is the same as μ_2. From the results of A. N. Kolmogorov [8] it follows that if $[A^I, \mu_1, T]$ and $[A^I, \mu_2, T]$ are isomorphic then $H(\mu_1, A) = H(\mu_2, A)$. This means that entropy is a metric invariant. We shall now show that there is a large class of invariant measures with zero entropy. In such cases entropy is a trivial invariant. This is a dual to the result of Rokhlin [19] concerning measure preserving transformations in a Lebesgue space.

<u>Theorem</u> 2.2.1. In the space A^I with the shift transformation the set of ergodic measures with zero entropy is a dense G_δ in $\mathcal{M}_A$ under the weak topology.

<u>Proof</u>: Let us first prove that the set of distributions with zero entropy is a G_δ. Let μ be any invariant measure and $H(\mu, A)$ its entropy. By definition 2.1.1.

$$(2.2.1) \qquad H(\mu, A) = \lim_{n \to \infty} \frac{H_n(\mu, A)}{n} = \liminf_{n \to \infty} \frac{H_n(\mu, A)}{n}$$

where $H_n(\mu, A)$ is the entropy of the finite scheme obtained by restricting μ to the n-dimensional cylinder sets (starting from the time point 1 and ending with time point n). $H_n(\mu, A)$ is a continuous functional on the space $\mathcal{M}_A$ under the weak topology. By making use of (2.2.1), the set of measures with zero entropy can be easily verified to be the same as

$$\bigcap_{r=1}^{\infty} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \left[\mu : \frac{H_n(\mu, A)}{n} < \frac{1}{r}\right]$$

Since $H_n(\mu, A)$ is a continuous functional $[\mu: \frac{H_n(\mu, A)}{n} < \frac{1}{r}]$ is open in $\mathcal{M}_A$. Thus the set of measures with zero entropy is a G_δ. The fact that the set of all ergodic measures is a G_δ implies that the set of ergodic measures with zero entropy is a G_δ.

In order to prove the density of measures with zero entropy, we note that the set of periodic ergodic measures is everywhere dense (see theorem 1.4.1) and prove that every periodic ergodic measure has zero entropy. By lemma 1.4.3 we see that, for any periodic ergodic measure μ of period k, there exists a point x_0 such that μ has mass $1/k$ at the points $x_0, T x_0, \ldots, T^{k-1} x_0$. The complement of the set of points $x_0, T x_0, \ldots, T^{k-1} x_0$ has measure zero. For sufficiently large n, $H_n(\mu, A) = \log k$ and hence

$$\lim_{n \to \infty} \frac{H_n(\mu, A)}{n} = 0.$$

This completes the proof of theorem 2.2.1.

2.3. Properties of the entropy functional

We shall now study the continuity properties of the entropy functional. The following result was observed by the author and Breiman [ ] independently.

Theorem 2.3.1. If $\{\mu_n\}$ is a sequence of measures in $\mathfrak{M}_A$ and μ_n converges weakly to μ, then

$$\limsup_{n \to \infty} H(\mu_n, A) \leq H(\mu, A)$$

i.e. the entropy is an upper semi continuous functional in the weak topology.

Proof: To prove the upper semi continuity property of a functional it is enough to show that it is a limit of a monotonically decreasing sequence of functionals. To this purpose we use the following well known inequality (see [7]): for any $\mu \in \mathfrak{M}_A$.

(2.3.1) $$H_{n+m}(\mu, A) \leq H_n(\mu, A) + H_m(\mu, A).$$

From (2.3.1) we have

$$\frac{H_{2^k}(\mu)}{2^k} \leq \frac{H_{2^{k-1}}(\mu)}{2^{k-1}}, \quad k = 2, 3, \ldots$$

The entropy $H(\mu, A)$ of μ is the limit of the sequence $\frac{H_{2^k}(\mu, A)}{2^k}$ and $H_{2^k}(\mu, A)$ is a continuous functional on $\mathcal{M}_A$. This completes the proof.

__Remark__ That entropy is not necessarily a continuous functional follows from the fact that the set of measures with zero entropy is everywhere dense

The following result was first noted by Breiman [1]. Our proof is slightly different.

__Theorem__ 2.3.2. (Breiman [1]). If $\mu = a\mu_1 + (1 - a)\mu_2$, where μ, μ_1 and μ_2 are invariant measures on A^I and $0 \leq a \leq 1$, then

$$H(\mu, A) = a\, H(\mu_1, A) + (1 - a)H(\mu_2, A).$$

__Proof.__ Let $x = (\ldots x_{-1}, x_0, x_1, \ldots)$ be any point in A^I and let $[x_1, x_2, \ldots, x_n]$ denote the cylinder set of points whose first n co-ordinates are $x_1, x_2, \ldots x_n$ respectively. Let

$$\text{(2.3.2)} \qquad f_n(x) = -\frac{1}{n} \log \mu[x_1, x_2, \ldots, x_n].$$

Then by Mcmillan's theorem [7], the sequence of functions f_n converges almost everywhere to a function $f(x)$ whose expectation with respect to μ is precisely the entropy of μ. Thus

$$(2.3.3) \qquad H(\mu, A) = E \lim_{n \to \infty} -\frac{1}{n} \log \mu[x_1, \ldots, x_n]$$

If $a = 0$ or 1 the theorem is obvious. Let therefore $0 < a < 1$. Then

$$-\frac{1}{n} \log \mu[x_1, \ldots, x_n] = -\frac{1}{n} \log (a\mu_1 + (1-a)\mu_2)[x_1, \ldots, x_n] =$$

$$= -\frac{1}{n} \log \mu_1[x_1, \ldots, x_n] - \frac{1}{n} \log \left(a + (1-a) \frac{\mu_2[x_1, \ldots, x_n]}{\mu_1[x_1, \ldots, x_n]} \right)$$

Further the existence of the limit

$$\lim_{n \to \infty} \frac{\mu_2[x_1, \ldots, x_n]}{\mu_1[x_1, \ldots, x_n]} \qquad \text{a.e.} \quad (\mu_1)$$

as a function integrable with respect to μ_1 is a consequence of Doob's martingale theorem. Since $a > 0$, we have

$$(2.3.4) \quad \lim_{n \to \infty} -\frac{1}{n} \log \mu[x_1, \ldots, x_n] = \lim_{n \to \infty} -\frac{1}{n} \log \mu_1[x_1, \ldots, x_n] \quad \text{a.e} \ (\mu_1)$$

Similarly, since $(1-a) > 0$, we have

$$(2.3.5) \quad \lim_{n \to \infty} -\frac{1}{n} \log \mu[x_1, \ldots, x_n] = \lim_{n \to \infty} -\frac{1}{n} \log \mu_2[x_1, \ldots, x_n] \quad \text{a.e.} (\mu_2)$$

Thus (2.3.3), (2.3.4) and (2.3.5) imply that

$$H(\mu, A) = aH(\mu_1, A) + (1-a) H(\mu_2, A).$$

Thus we have proved that the entropy is an upper semi-continuous linear functional in the space $\mathcal{M}_A$. We shall now prove something more in the sense that this linear functional is actually defined through an integral. To this end we shall introduce the following notations.

Let F_A^- be the Borel σ-field generated by cylinder sets $[x_{i_1}, \ldots, x_{i_k}]$ where $i_1, \ldots, i_k$ vary over negative integers only. Let Z_α denote the cylinder set of points with zeroth coordinate equal to α. Corresponding to any finite measure μ we consider the following conditional probability function $g_\mu(x, \alpha)$ given by

$$\mu(E \cap Z_\alpha) = \int_E g_\mu(x, \alpha)\, d\mu(x) \tag{2.3.6}$$

for any Borel set E in F_A^-. We shall now prove the following theorem concerning $g_\mu(x, \alpha)$.

<u>Theorem</u> 2.3.3. If μ, μ_1 and μ_2 are invariant measures in A^I, $\mu = a\mu_1 + (1-a)\mu_2$ $(0 \leq a \leq 1)$, and μ_1 and μ_2 are orthogonal, then

$$g_\mu (x, \alpha) = g_{\mu_1} (x, \alpha) \qquad \text{a.e. } x\ (\mu_1)$$

Proof: Since μ_1 and μ_2 are invariant and orthogonal, the critical sets in which their masses are concentrated can be taken to be invariant and hence in F_A. It is then immediate from the definition of conditional probabilities that

$$g_\mu(x, \alpha) = g_{\mu_1}(x, \alpha) \qquad \text{a.e. } x\ (\mu_1)$$

Theorem 2.3.4. If μ is an invariant probability measure, then

$$g_\mu(x, \alpha) = g_{\mu_p}(x, \alpha) \qquad \text{a.e. } x(\mu_p)$$

for almost all $p(\mu)$, where μ_p is the ergodic measure associated with the regular point p in theorem 1.1.3

Proof: For any invariant measure μ, we have from (2.3.6) and theorem 1.1.1,

(2.3.7)
$$\mu(E \cap Z_\alpha) = \int_E g_\mu(x, \alpha)\, d\mu(x) = \int_R \left[\int_E g_\mu(x, \alpha) d\mu_p(x) \right] d\mu(p)$$

where R is the set of regular points (see definition 1.1.2) in A^I. From theorem 1.1.8 and (2.3.6) we have

(2.3.8)
$$\mu(E \cap Z_\alpha) = \int \mu_p(E \cap Z_\alpha)\, d\mu(p) = \int_R \left[\int_E g_{\mu_p}(x, \alpha)\, d\mu_p(x) \right] d\mu(p)$$

where E is any set in F_A^-.

For any invariant set A for which $\mu(A)$ is neither zero nor one we can write

$$\mu = a\mu_1 + (1 - a)\mu_2$$

where $a = \mu(A)$, $\mu_1(E) = \mu(E \cap A)/\mu(A)$, and $\mu_2(E) = \mu(E \cap A')/\mu(A')$ for any Borel set E. Then μ_1 and μ_2 are invariant and orthogonal. Hence by theorem 2.3.3,

$$g_\mu(x, \alpha) = g_{\mu_1}(x, \alpha) \qquad \text{a.e. } x(\mu_1)$$

Substituting μ_1 for μ in (2.3.7) and (2.3.8), equating the two expressions and making use of theorem 1.1.7, we obtain

$$(2.3.9) \quad \int_{A \cap R} \left[\int_E g_\mu(x, \alpha)\, d\mu_p(x) \right] d\mu(p) = \int_{A \cap R} \left[\int_E g_{\mu_p}(x, \alpha) d\mu_p(x) \right] d\mu(p)$$

for any invariant set A and any set E in $\bar{F_A}$. Since the functions of p within the square brackets in (2.3.9) are invariant and thus measurable with respect to the σ-field of invariant Borel sets, we have, for all cylinder sets $E \in F_A$ and almost all $p(\mu)$

$$\int_E g_\mu(x, \alpha) d\mu_p(x) = \int_E g_{\mu_p}(x, \alpha)\, d\mu_p(x).$$

The required result is now an immediate consequence of the fact that the Radon-Nykodym ~~is~~ derivative is unique.

<u>Theorem</u> 2.3.5 : There exists a function $h(p)$ defined over the set R of regular points in A^I such that for every invariant probability measure μ,

$$H(\mu, A) = \int_R h(p)d\mu(p).$$

<u>Proof</u>: For any point $x = (\dots x_{-1}, x_0, x_1, \dots)$ in A^I, let

$$h_\mu(x) = g_\mu(x, x_0)$$

where g_μ is defined by (2.3.6). Then by Mcmillan's theorem [7], $-\log h_\mu(x)$ is integrable with respect to μ, and

$$H(\mu, A) = -\int \log h_\mu(x)d\mu.$$

Define

$$h(p) = -\int \log g_{\mu_p}(x)d\mu_p(x)$$

where μ_p is the ergodic measure associated with the regular point p by means of theorem 1.1.3. By theorem 2.3.4

$$h_\mu(x) = h_{\mu_p}(x) \qquad \text{a.e. } x(\mu_p)$$

for almost all $p(\mu)$. Since $-\int \log h_{\mu_p} d\mu_p$ is finite for almost all $p(\mu)$, by theorem 1.1.7 and Fubini's theorem we have

$$H(\mu, A) = -\int \log h_\mu(x)d\mu = -\int_R [\int \log h_\mu(x)d\mu_p(x)]d\mu$$
$$= -\int_R [\int \log h_{\mu_p}(x)d\mu_p(x)]d\mu = \int_R h(p)d\mu(p).$$

This completes the proof.

Remarks : Theorem 2.3.5 was proved by the author [15]. Theorem 2.3.5 can be considered as a dual to that of Rokhlin [19] who expresses the entropy of any measure preserving transformation of a Lebesque space as an integral of the entropies of its factor automorphisms. That theorem 2.3.5 is true for any upper-semi continuous linear functional was recently communicated to the author by K. Jacobs [6]. This has been further generalised by K. Jacobs for general compact subsets in locally convex topological vector spaces. Since our interest is confined to information theory we do not go into these details any further.

2.4 Rate of transmission through a channel.

We shall now briefly discuss the concept of a channel and study the properties of some functionals associated with a channel.

Definition 2.4.1. A channel $[A, \nu_x, B]$ consists of two finite alphabets A and B, called the input and output alphabets respectively, and a collection of probability distributions ν_x, where ν_x is a measure on F_B associated with the point x in A^I and possessing the stationarity property, viz., $\nu_x(F) = \nu_{Tx}(TF)$ for any Borel set F in B^I, T being the usual shift operator. For each fixed Borel set F B^I the function $\nu_x(F)$ is assumed to be measurable.

Any invariant measure on the space A^I will be called an input measure. For any input measure μ, we define

(2.4.1) $$\omega(E \times F) = \int_E \nu_x(F) d\mu(x)$$

(2.4.2) $$\eta(F) = \int_{A^I} \nu_x(F) d\mu(x)$$

where E is any set in F_A and F is any set in F_B. ω is called the joint input-output distribution and η the output distribution corresponding to the input μ. The space $A^I \times B^I$ can be considered, in an obvious manner, as a countable product of the product alphabet AB and ω as a stationary distribution. Thus we now have the following three sources corresponding to any input measure μ :

$$[A^I, \mu], \quad [A^I \times B^I, \omega], \quad [B^I, \eta].$$

Let their entropies be $H(\mu, A)$, $H(\omega, AB)$ and $H(\eta, B)$ respectively. Then we have the following

Definition 2.4.2. The functional $R(\mu) = H(\mu, A) + H(\eta, B) - H(\omega, AB)$ is called the rate of transmission through the channel $[A, \nu_x, B]$ for the input measure μ.

Now we shall prove the following representation theorem.

Theorem 2.4.1. For any channel $[A, \nu_x, B]$ and any input measure μ,

$$R(\mu) = \int_R R(\mu_p) d\mu(p)$$

where μ_p is the ergodic measure associated with any regular point[1) p in A^I and R is the set of all regular points in A^I.

Proof : By theorem 1.1.8, for any Borel set E,

(2.4.3) $$\mu(E) = \int \mu_p(E) d\mu(p).$$

1) For the definition of a regular point and the associated ergodic measure see definition 1.1.2 and theorems 1.1.1 and 1.1.5.

Let ω_p and η_p be the joint input-output and output measures corresponding to the input measure μ_p. If ω and η are as in (2.4.1) and (2.4.2), then (2.4.3) gives

$$\omega[E \times F] = \int \omega_p(E \times F)d\mu(p), \tag{2.4.4}$$

$$\eta(F) = \int \eta_p(F)d\mu(p). \tag{2.4.5}$$

An application of theorem 2.3.5, (2.4.4) and (2.4.5) give

$$R(\mu) = \int R(\mu_p)d\mu(p).$$

This completes the proof.

Let

$$H_n(\mu, B|A) = -\frac{1}{n} \int \log \nu_x[y_1 \dots y_n]d\,\omega(x,y) \tag{2.4.6}$$

where is the joint input-output measure corresponding to the in ut μ.

Then we have the following

<u>Theorem</u> 2.4.2. The limit

$$\lim_{n \to \infty} H_n(\mu, B|A)$$

exists and if it is denoted by $H(\mu, B|A)$, then

$$H(\mu, B|A) = \int_{A^I} H(x)d\mu(x)$$

where

$$H(x) = -\int \log g(x,y)d\,\nu_x(y) \tag{2.4.7}$$

and

$$(2.4.8) \qquad g(x,y) = \lim_{n \to \infty} \frac{\nu_x[y_{-(n-1)}, \ldots, y_0]}{\nu_x[y_{-(n-1)} \ldots y_{-1}]} \qquad \text{a.e. } y(\nu_x)$$

for every x.

In order to prove this theorem we require several lemmas.

<u>Lemma</u> 2.4.1. Let μ_o be any probability measure (not necessarily invariant) on A^I. Let

$$E_{n,r} = [x : r \leq -\log g_n(x) < r+1]$$

where

$$(2.4.9) \qquad g_n(x) = \frac{\mu_o[x_{-(n-1)} \ldots x_o]}{\mu_o[x_{-(n-1)} \ldots x_{-1}]} \quad .$$

Then

$$-\int_{E_{n,r}} \log g_n(x) d\mu_o(x) \leq a(r+1)2^{-r}$$

<u>Lemma</u> 2.4.2. Given L > 0, let $A_{n,L}$ be the set of all points $x \varepsilon A^I$ for which $-\log g_n(x) > L$ where g_n is given by (2.4.9). Then, given $\varepsilon > 0$, an $L_o = L_o(\varepsilon)$ can be found such that for $L > L_o$,

$$-\int_{A_{n,L}} \log g_n(x) \, d\mu_o(x) < \varepsilon.$$

<u>Lemma</u> 2.4.3. Given $\varepsilon > 0$, a $\delta > 0$ can be found such that, for $E \varepsilon F_A$ and $\mu_o(E) < \delta$

$$-\int_E \log g_n(x) d\mu_o(x) < \varepsilon, \qquad n = (2, 3 \ldots).$$

<u>Lemma</u> 2.4.4. The limit function

$$g(x) = \lim_{n \to \infty} g_n(x) \qquad \text{a.e. } x(\mu_o)$$

is well defined.

Lemma 2.4.5. The function $-\log g(x)$ is integrable with respect to μ_0 where $g(x)$ is as in lemma 2.4.4.

Lemma 2.4.6. $\lim_{n \to \infty} \int_{A^I} |\log g_n(x) - \log g(x)| d\mu_0 = 0$ where $g(x)$ is as in lemma 2.4.4.

The proofs of lemmas 2.4.1, 2.4.2 and 2.4.3 are identical with the proofs of lemmas 7.3, 7.4 and 7.5 on pages 67-68 of Khinchin's book [7]. Lemma 2.4.4 is a consequence of the well known martingale theorem. Lemmas 2.4.5 and 2.4.6 are immediate consequences of lemmas 2.4.1 to 2.4.4. We remark that the assumption of stationarity is nowhere made use of in the proof of lemmas 7.3 to 7.7 on pages 67-70 of [7].

Replacing the alphabet A by the alphabet B, fixing the point x and writing $\mu_0 = \nu_x$ in the above lemmas we have the following lemma

Lemma 2.4.7. Let

$$g_n(x,y) = \frac{\nu_x[y_{-(n-1)}, \dots y_0]}{\nu_x[y_{-(n-1)}, \dots y_{-1}]}$$

Then

$$g(x,y) = \lim_{n \to \infty} g_n(x,y) \qquad \text{a.e. } y(\nu_x)$$

is well-defined for each fixed x and

$$\lim_{n \to \infty} \int |\log g_n(x,y) - \log g(x,y)| d\nu_x(y) = 0$$

for all x. In particular $-\log g(x,y)$ is integrable with respect to the measure ν_x for every x.

Proof of theorem 2.4.2. We write

$$a_n = -\int \log \nu_x[y_1, \dots y_n] \, d\omega(x,y)$$

$$b_n = n$$

and to the sequence $H_n(\mu, B|A)$ (see (2.4.6)) we apply the following well known result : if a_n and b_n are two sequences such that b_n is monotonic and $b_n \to \infty$ as $n \to \infty$, then

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{a_n - a_{n-1}}{b_n - b_{n-1}}$$

whenever the second limit exists. Then we have

$$(2.4.10) \qquad \lim_{n\to\infty} H_n(\mu_1, B|A) = \lim_{n\to\infty} -\int \log \frac{\nu_x[y_1 \dots y_n]}{\nu_x[y_1 \dots y_{n-1}]} \, d\omega(x,y)$$

provided the limit on the right side of (2.4.10) exists. Using the stationarity property of the channel and the measure ω, we can write

$$(2.4.11) \quad \lim_{n\to\infty} H_n(\mu, B|A) = \lim_{n\to\infty} -\int \log \frac{\nu_x[y_{-(n-1)} \dots y_0]}{\nu_x[y_{-(n-1)} \dots y_{-1}]} \, d\omega(x,y)$$

$$= \lim_{n\to\infty} -\int \log g_n(x,y) \, d\omega(x,y).$$

Now applying lemma 2.4.7., we have

$$(2.4.12) \qquad H(\mu, B|A) = \lim_{n\to\infty} H_n(\mu, B|A)$$

$$= -\int \log g(x,y) \, d\nu_x(y) \, d\mu(x)$$

$$= \int H(x) \, d\mu(x)$$

where $H(x)$ is given by (2.4.7).

<u>Remarks</u>. Another method of defining the rate of transmission through a channel for any input measure μ is by means of the following functional: $R^*(\mu) = H(\eta, B) - H(\mu, B|A)$ where η is the corresponding output measure and

(2.4.13) $$H(\mu, B|A) = \lim_{n \to \infty} H_n(\mu, B|A).$$

The existence of (2.4.13) is assured by theorem 2.4.2. It is not known whether $R(\mu) = R^*(\mu)$ for any arbitrary channel. It is not difficult to prove that $R^*(\mu) \geq R(\mu)$ for all μ. However, for what are known as channels with finite memory in the sense of Khinchin (which we shall introduce in the next chapter) the two definitions of rate of transmission become identical.

The following theorem is an immediate corollary of theorems 2.4.2, 2.3.5 and 1.1.7.

<u>Theorem</u> 2.4.3. $R^*(\mu) = \int_R R^*(\mu_p) d\mu(p)$

where μ_p is the ergodic measure associated with any regular point p and R is the set of all regular points. In particular,

$$H(\mu, B|A) = \int_R H(\mu_p, B|A) d\mu(p).$$

We shall now study some of the continuity properties of the functional $H(\mu, B|A)$ associated with a channel $[A, \nu_x, B]$. To this end we require the following lemmas.

<u>Lemma</u> 2.4.8. Let A, B, C be three finite schemes, BC the joint scheme containing B and C and $H_B(A)$ and $H_{BC}(A)$ the conditional entropies (see section 2.1) of A given the schemes B and BC respectively. Then $H_B(A) \geq H_{BC}(A)$.

For proof we refer to Khinchin [7].

Lemma 2.4.9. The sequence of functions

$$H_n(x) = -\sum_{[\mathcal{Y}_{(n-1)} \cdots y_0]} \nu_x[\mathcal{Y}_{(n-1)} \cdots y_0] \log \frac{\nu_x[\mathcal{Y}_{(n-1)} \cdots y_0]}{\nu_x[\mathcal{Y}_{(n-1)} \cdots \mathcal{Y}_1]} \tag{2.4.14}$$

$(n = 2, 3, \ldots)$

is monotonically decreasing for each fixed x.

Proof. Keeping x fixed and hence the probability distribution ν_x fixed, we see that $[y_0]$, $[\mathcal{Y}_{(n-1)} \cdots \mathcal{Y}_1]$ and $[\mathcal{Y}_n]$ are three finite schemes. If we denote them by A, B and C respectively, we see that

$$H_n(x) = H_B(A); \; H_{n+1}(x) = H_{BC}(A)$$

and hence an applic tion of lemma 2.4.8 completes the proof.

Theorem 2.4.4. The functional $H(\mu, B|A)$ is upper semi continu us in the weak topology.

Proof. To prove this theorem it is enough to show that $H(\mu, B|A)$ is the limit of a monotonically decreasing sequence of continuous functionals. From (2.4.11), (2.4.12) and (2.4.14) we have

$$H(\mu, B|A) = \lim_{n \to \infty} \int H_n(x) d\mu(x).$$

Since $\int H_n(x)\, d\mu$ is continuous in μ (in the weak topology) an application of lemma 2.4.9 completes the proof.

Theorem 2.4.5. The sequence $H_n(\mu, B|A)$[1] is monotonically decreasing.

We define $H_n(x)$ as in (2.4.14) for n = 2, 3, ... and $H_1(x)$ as $-\sum_{[y_0]} \nu_x[y_0] \log \nu_x[y_0]$. We write

---

1) For the definition of $H_n(\mu, B|A)$ see (2.4.6).

$$H_n'(x) = - \sum_{[y_1 \cdots y_n]} \nu_x[y_1 \cdots y_n] \log \frac{\nu_x[y_1 \cdots y_n]}{\nu_x[y_1 \cdots y_{n-1}]}$$

$$H_1'(x) = H_1(x)$$

Then

$$\int H_n'(x) d\mu(x) = \int H_n(x) d\mu(x), \quad n = 1, 2, \ldots$$

further

$$H_n(\mu, B|A) = \frac{\sum_{r=1}^{n} \int H_r'(x) d\mu(x)}{n} = \frac{\sum_{r=1}^{n} \int H_r(x) d\mu(x)}{n}$$

Since the sequence $\int H_n(x) d\mu(x)$ is monotonically decreasing it follows that $H_n(\mu, B|A)$ is monotonically decreasing.

Remarks. It is not known that under what minimum conditions on the channel the functional $H(\mu, B|A)$ is continuous in the weak topology. In the next chapter when we discuss the properties of channels with finite memory in the sense of Feinstein we shall prove the continuity of $H(\mu, B|A)$ in this special case. In this context we recall that the entropy is not a continuous functional eventhough it is upper semi-continuous.

3. APPLICATIONS

In this chapter we shall deal with some of the applications of the results of the earlier chapters to problems of Information theory. We introduce different definitions of capacity of a channel and prove their equivalence in the case of channels with finite memory. We study the problem of achievement of capacity, the continuity properties of the functional $H(\mu, B|A)$ associated with a channel with finite memory and finally state a problem about additive noise channels.

3.1 Capacity of a channel:

Definition 3.1.1. The ergodic capacity C_e of a channel $[A, \nu_x, B]$ is defined by

$$C_e = \sup_{\mu\text{-ergodic}} R(\mu).$$

Definition 3.1.2. The stationary capacity C_s of a channel $[A, \nu_x, B]$ is defined by

$$C_s = \sup_{\mu\text{-stationary}} R(\mu).$$

Another possible definition of ergodic and stationary capacities is obtained by replacing $R(\mu)$ by $R^*(\mu)$ in the above . Let the corresponding capacities be C_e^* and C_s^* . We shall refer to these as star capacities.

In problems of information theory the quantities C_e and C_s play a fundamental role. It has been clearly pointed out by Khinchin that in proving the converse of Shannon's fundamental theorems for channels with finite memory (see definitions 3.2.1 and 3.2.2), earlier authors have been careless in defining the capacity of a channel. He pointed out that, in the case of channels with finite memory, the fundamental lemma of Feinstein is valid only when we take C_e as the capacity. The converse of Shannon's theorems became clear when I.P. Tsaragradsky [22] and A. Feinstein [4] proved that $C_e = C_s$ in this special case. By making use of the integral representation theorem 2.3.5 which we have proved in the last chapter we shall give a very simple proof of the fact that the ergodic and stationary capacities of an arbitrary channel are equal.

<u>Theorem</u> 3.1.1. For any channel $C_e = C_s$; $C_e^* = C_s^*$.

<u>Proof</u>. The proof is identical in both the cases. We shall prove that $C_e = C_s$. For any stationary measure μ on A^I, we have by theorem 2.4.1

$$(3.1.1) \qquad R(\mu) = \int_R R(\mu_p)\, d\mu(p)$$

where R is the set of regular points in A^I and μ_p is the ergodic measure associated with p by means of theorem 1.1.3. Thus

$$R(\mu) \leq \sup_{p \in R} R(\mu_p) \leq C_e.$$

Taking supremum of the left side over all $\mu \in \mathcal{M}_A$, we have

$$C_s \leq C_e.$$

That $C_e \leq C_s$ is obvious. Thus $C_e = C_s$. Replacing $R(\mu)$ by $R^*(\mu)$ everywhere we get the second part.

<u>Definition</u> 3.1.3. The common value $C_e = C_s = C$ will be called the capacity of the channel. The corresponding value $C^* = C_e^* = C_s^*$ will be called the star capacity of the channel.

Another consequence of the representation (3.1.1) is the following

<u>Theorem</u> 3.1.2. The set of stationary measures at which the capacity may be attained is a convex set with ergodic extreme points. In particular, if the capacity is attained at some measure μ then it is attained at an ergodic measure.

Remarks. It is not yet known whether the capacity of a channel is always attained. We shall later give sufficient conditions for the capacity to be attained. This includes all channels which are of finite memory in the sense of Feinstein (see definition 3.2.2).

3.2 Channels with finite memory.

Definition 3.2.1. A channel $[A, \nu_x, B]$ is said to be of finite memory m in the sense of Khinchin if, for any cylinder set $[y_1 \ldots y_n]$ in B^I,

$$\nu_x[y_1 \ldots y_n] = \nu_{x'}[y_1 \ldots y_n]$$

as soon as the co-ordinates of x and x′ agree at the time points $-(m-1)$, $-(m-2)$, ... -1, 0, 1, ... n.

Definition 3.2.2. A channel $[A, \nu_x, B]$ is said to be of finite memory m in the sense of Feinstein if it is of finite memory m in the sense of Khinchin and

$$\nu_x[y_1, \ldots, y_k, \quad y_{k+m+1}, \quad y_{k+m+2}, \ldots y_n]$$

$$= \nu_x[y_1 \ldots y_k] \cdot \nu_x[y_{k+m+1}, y_{k+m+2}, \ldots y_n]$$

for all cylinders of the type $[y_1, \ldots, y_k, y_{k+m+1}, \ldots, y_n]$.

At the end of this chapter we shall describe a class of channels with finite memory in the sense of Khinchin but not in the sense of Feinstein, for which the Feinsteins fundamental lemma and hence Shannon's theorems are valid. We now prove the equivalence of the two definitions of rate of transmission through channels with finite memory in the sense of Khinchin.

Theorem 3.2.1. For channels with finite memory in the sense of Khinchin $R(\mu) = R^*(\mu)$ for every input measure μ and hence $C = C^*$.

<u>Proof.</u> For the input measure μ, let ω and η be the corresponding joint input-output and output measures respectively. Let $x = (\ldots x_{-1}, x_0, x_1, \ldots)$ and $y = (\ldots y_{-1}, y_0, y_1 \ldots)$ be points in A^I and B^I. Then we have

$$\omega\,[x_{-(m-1)} \ldots x_n ; y_{-(m-1)} \ldots y_n] \leq \omega\,[x_{-(m-1)}, \ldots x_n ; y_1 \ldots y_n]$$

$$\leq \omega\,[x_1 \ldots x_n, y_1 \ldots y_n].$$

From this inequality it is easily seen that the entropy $H(\omega, AB)$ is given by

(3.2.1) $$H(\omega, AB) = \lim_{n \to \infty} -\frac{1}{n} E \log \omega\,[x_{-(m-1)}, \ldots x_n ; y_1 \ldots y_n]$$

where the expectation is taken with respect to the distribution .

From (2.4.6) and definition 3.2.1 we have

$$H_n(\mu, B | A) = -\frac{1}{n} E \log \nu_x[y_1 \ldots y_n]$$

$$= -\frac{1}{n} E \log \omega\,[x_{-(m-1)} \ldots x_n ; y_1 \ldots y_n] +$$

$$+ \frac{1}{n} E \log \mu\,[x_{-(m-1)} \ldots x_n]$$

and hence

(3.2.2) $$H(\mu, B | A) = H(\omega, AB) - H(\mu, A).$$

The definition of $R^*(\mu)$ and (3.2.2) give

$$R(\mu) = R^*(\mu).$$

This completes the proof. Thus we have the following

<u>Corollary</u> 3.2.1. For a channel with finite memory in the sense of Khinchin

$$C_e = C_s = C_e^* = C_s^* = C.$$

Earlier we have proved that the entropy functional $H(\mu, A)$ and the functional $H(\mu, B|A)$ associated with a channel are upper semi continuous. However, we noted that entropy is not continuous in the weak topology. We shall now prove the continuity of $H(\mu, B|A)$ in the special case of channels with finite memory in the sense of Feinstein.

<u>Theorem</u> 3.2.2. For channels with finite memory in the sense of Feintein the functional $H(\mu, B|A)$ is continuous in the weak topology. Furthermore

$$\lim_{n \to \infty} \sup_{\mu \in \quad A} | H_n(\mu, B|A) - H(\mu, B|A)| = 0.$$

<u>Proof</u>. We have proved the upper semi continuity of $H(\mu, B|A)$ in theorem 2.4.4. For channels with finite memory, we have by (3.2.2)

$$H(\mu, B|A) = H(\omega, AB) - H(\mu, A)$$

where ω is the joint input-output distribution corresponding to the input distribution μ. We shall now prove its lower semi continuity by showing that $H(\mu, A) - H(\omega, AB)$ is the limit of a monotonically decreasing sequence of continuous functionals. We now follow Breiman [ ]. Let the memory of the channel be m and let $n = k(\ell+m)$, $k > 1$. Let x, y be points in A^I and B^I respectively and u_1, v_1, w_1 be the parts of the x and y sequences as in these diagrams

$$\overbrace{x_1, \cdots x_{m+\ell}}^{w_1},\ \overbrace{x_{m+\ell+1} \cdots x_{2(m+\ell)}}^{w_2},\ \cdots,\ \overbrace{x_{(k-1)(m+\ell)+1}, \cdots x_{k(m+\ell)}}^{w_k}$$

$$\overbrace{y_{m+1} \cdots y_{m+\ell}}^{u_1},\ \overbrace{y_{m+\ell+1} \cdots y_{2m+\ell}}^{v_1},\ \overbrace{\cdots y_{2(m+\ell)}}^{u_2},\ \overbrace{\cdots y_{2m+3\ell}}^{v_2} \cdots,$$

$$\overbrace{y_{(k-1)(m+\ell)+1}, \dots y_{km+(k-1)\ell}}^{v_{k-1}}, \overbrace{y_{km+(k-1)\ell+1} \dots y_{k(m+\ell)}}^{u_k}.$$

w sequences are of length $\ell+m$, u sequences of length ℓ and v – sequence of length m. From the stationarity of the channel and (2.4.6) we have

$$- \frac{(n-m)}{n} H_{n-m}(\mu, B|A) =$$

$$= \frac{1}{n} \sum_{u,v,w} P(u_1, v_1, \dots v_{k-1}, u_k | w_1 \dots w_k) \mu[w_1 \dots w_k] .$$

$$\log P(u_1, v_1, \dots u_k | w_1 \dots w_k)$$

where $P(u_1, v_1, \dots v_{k-1}, u_k | w_1 \dots w_k)$ stands for $\nu_x[y_{m+1} \dots y_n]$. (Since the channel is of finite memory m this depends only on $w_1 \dots w_k$). Further from definition 3.1.2 we have

$$\log P(u_1, v_1 \dots v_{k-1}, u_k | w_1 \dots w_k) \leq \log P(u_1, u_2, \dots u_k | w_1 \dots w_k) =$$

$$= \sum \log P(u_i | w_i).$$

Therefore

$$(3.2.3) \qquad - \frac{(n-m)}{n} H_{n-m}(\mu, B|A) \leq \frac{1}{k(\ell+m)} \sum_{i=1}^{k} \sum_{u_i, w_i} P(u_i|w_i)\mu(w_i) \log P(u_i|w_i)$$

$$= \frac{1}{\ell+m} \sum_{u_1, w_1} P(u_1|w_1)\mu(w_1) \log P(u_1|w_1) =$$

$$= - \frac{\ell}{\ell+m} H_{(\ell+m)}(\mu, B|A)$$

Let now

$$G_m(\mu) = - \frac{(n-m)}{n} H_{n-m}(\mu, B|A)$$

The inequality (3.2.3) shows that the sequence $\{G_n(\mu)\}$ has a monotonic decreasing subsequence (since l is arbitrary). Further

$$\lim_{n \to \infty} G_n(\mu) = H(\mu, A) - H(\omega, AB).$$

This shows that $H(\mu, A) - H(\omega, AB)$ is upper semi continuous and hence completes the proof of the first part.

As remarked earlier the space $\mathcal{M}_A$ is a compact metric space and the continuous function $H(\mu, B|A)$ is the limit of a monotonically decreasing sequence $H_n(\mu, B|A)$ (see theorem 2.4.5). Hence by a well known theorem of Dini it follows that $H_n(\mu, B|A)$ converges to $H(\mu, B|A)$ uniformly.

Utilising the continuity property of $H_n(\mu, B|A)$ we shall now prove the following theorem due to Breiman.

<u>Theorem</u> 3.2.3. The capacity of a channel with finite memory in the sense of Feinstein is attained at some ergodic measure.

<u>Proof</u>. From theorems 2.3.1 and 3.2.2 it is clear that the rate of transmission $R(\mu)$ is an upper semi continuous functional on the space $\mathcal{M}_A$ under the weak topology. Let C be the capacity of the channel. By definition there exists a sequence of measures $\mu_n \in \mathcal{M}_A$ such that

$$C = \lim_{n \to \infty} R(\mu_n).$$

By the compactness of $\mathcal{M}_A$, it follows that there exists a subsequence μ_{n_k} such that $\lim_{n_k \to \infty} \mu_{n_k} = \mu$ exists. By the upper semi continuity of $R(\mu)$ we have

$$C = \lim_{n_k \to \infty} R(\mu_{n_k}) = \limsup_{n_k \to \infty} R(\mu_{n_k}) \leq R(\mu) \leq C.$$

Thus $C = R(\mu)$. Since the capacity is attained at a measure μ, an application of theorem 3.1.2 shows that there exists an ergodic measure μ_o such that $C = R(\mu_o)$. This completes the proof.

Remarks. From the proof of the above theorem it is clear that the capacity C is attained for all those channels for which the functional $H(\mu, B|A)$ is continuous in the weak topology. There is as yet no precise characterisation of these channels.

The above theorem shows that the capacity is attained at an ergodic measure. Is it true that the capacity is attained at an m-dependent stationary measure ?(By an m-dependent stationary measure we mean that measure for which the corresponding stochastic process $(\ldots x_{-1}, x_o, x_1 \ldots)$ is stationary and the collections of random variables $(\ldots x_{-1}, x_o)$ and $(x_{m+1}, x_{m+2}, \ldots)$ are independent). It is well known that the capacity of channels with zero memory in the sense of Feinstein is attained at a product measure. There is as yet no method of computing the capacity and the measure at which it is attained.

3.3 A limiting form of Feinstein's fundamental lemma.

Let us denote by u an arbitrary cylinder $[x_1, x_2, \ldots x_n]$ of length n and call it an u-sequence. Let v denote cylinders of the type $[y_{m+1}, y_{m+2}, \ldots y_n]$ obtained by taking n-m length sequences in B^I. Then the well known Feinstein's fundamental lemma asserts the following :

Theorem 3.3.1. If C is the capacity of a channel $[A, \nu_x, B]$ with finite memory in the sense of Feinstein, then, for any $\varepsilon > 0$ and sufficiently large n, there exist u-sequences $u_1, u_2, \ldots u_N$ and sets $B_1, \ldots, B_N$ of v-sequences such that

(1) $B_i \cap B_j = \phi$, $i \neq j$

(2) $\nu_{u_i}(B_i) > 1 - \varepsilon$

(3) $N > 2^{n(C-\varepsilon)}$

The following result is a limiting form of the above theorem when the length of the sequences becomes infinite. Here the global structure of the family of all ergodic measures brings to light the content of the conditions under which Feinstein's lemma is proved. The crucial property of the family of ergodic measures that comes into play here is the fact that the whole space A^I can be partitioned into ergodic sets and a set of invariant measure zero.(See theorems 1.1.5 and 1.1.6)

Theorem 3.3.2. Let $[A, \nu_x, B]$ be a channel with finite memory m in the sense of Feinstein and non zero capacity. Then there exists an uncountable number of points $\{x_\alpha\}$ in A^I and an uncountable number of mutually disjoint Borel subsets $\{B_\alpha\}$ in B^I such that $\nu_{x_\alpha}(B_\alpha) = 1$.

Proof. From theorem 1.1.6 we see that, associated with every ergodic measure μ on A^I, there exists an invariant set E_μ such that $\mu(E_\mu) = 1$ and all the sets E_μ are mutually disjoint. E_μ can be taken to be the set of all regular points p in A^I such that the associated ergodic measure $\mu_p = \mu$. The set E_μ is called the ergodic set of μ. A similar result holds for ergodic measures ν on B^I. Let F_ν be the ergodic set of ν. For channels with finite memory in the sense of Feinstein it is well known that, for any ergodic input measure μ, the corresponding joint input-output and output measures are ergodic (see [3]). The output measure ν is given by

$$(3.3.1) \qquad \nu(F) = \int_{A^I} f_x(F)\,d\mu(x)$$

for any Borel set F in B^I. Hence we have

$$(3.3.2) \qquad 1 = \nu(F_\nu) = \int_{A^I} f_x(F_\nu)\,d\mu(x),$$

F_ν being the ergodic set of ν. Since $0 \leq f_x(F_\nu) \leq 1$, we have from (3.3.2)

$$f_x(F_\nu) = 1 \quad \text{a.e. } x(\mu).$$

Thus there exists at least one point $x_\mu \in \mathcal{E}_\mu$ such that

$$(3.3.3) \qquad f_{x_\mu}(F_\nu) = 1$$

i.e. if ν is the output measure corresponding to some input measure μ then there exists a point x_μ such that (3.3.3) holds. Since the ergodic sets of two distinct measures ν are disjoint it is enough to show that there is an uncountable number of ergodic measures ν which are outputs of input measures on the space A^I. This problem can be looked upon as follows: the channel probability $f_x(F)$ associates with every input measure μ an output measure ν by means of equation (3.3.1). Thus it is a linear transformation of the space $\mathcal{F}_A$ into the space $\mathcal{F}_B$. It transforms ergodic measures into ergodic measures. The question is whether there is an uncountable number of ergodic measures in the range of the transformation induced by the channel probability. We shall now prove that this is actually so.

We assume that $\nu_x(F)$ is not independent of x. This is legitimate since the capacity is non zero. Let the memory be m. Then, for at least two u-sequences u_1 and u_2 of length n and one v-sequence v_o length n-m, we have

$$\nu_{u_1}(v_o) \neq \nu_{u_2}(v_o).$$

Let $\nu_{u_1}(v_o) = C_1$, $\nu_{u_2}(v_o) = C_2$ and $C_1 < C_2$ without loss of generality. Let p(u) be a distribution defined on the space of u-sequences. Then for any v-sequence we write

$$(3.3.4) \qquad \sum_u \nu_u(v)\ p(v) = q(v)$$

where the summation is over all u sequences. If $p(u_1) = 1$ and the rest are zero, we get $q(v_o) = C_1$. If $p(u_2) = 1$ and the rest are zero, then $q(v_o) = C_2$. Thus as p(u) *varies* over all distributions on the space of u-sequences $q(v_o)$ takes at least two different values C_1 and C_2. $q(v_o)$ being a continuous linear function of the probability distribution p(u) takes every value between C_1 and C_2 as p(u) varies over all probability distributions on the space of u-sequences. Out of every distribution p(u) we build the distribution

$$p^* = \prod_{r=-\infty}^{+\infty} p_r$$

by taking the product of distributions p_r which are identical copies of p. The product is taken over all integers only and p* is defined on the space A^I. p* is invariant and ergodic under the transformation T^n. Substituting p* instead of μ in equation (3.3.1) we get a measure q* which is also invariant and ergodic under T^n in the space R^I. (Ergodicity follows from the fact that the channel is of finite memory in the sense of Feinstein and hence transforms measures ergodic under T^n into measures ergodic under T^n). As the distribution $p(w)$ over the u-sequences varies we get an uncountable number of measures q^*. This is because of the fact that the restriction of q^* to v-sequences gives an uncountable number of distributions of the type $q(v)$ (see (3.3.4)). We now construct two measures

$$\mu(E) = \frac{p^*(E) + p^*(TE) + \dots + p^*(T^{n-1} E)}{n}$$

$$\eta(F) = \frac{q^*(F) + q^*(TF) + \dots + q^*(T^{n-1} F)}{n}$$

It is obvious that η is the output measure corresponding to the input μ. μ is invariant and ergodic under T and hence η is ergodic and invariant under T. Thus measures η which are constructed in the above manner are in the range of the transformation induced by the channel probabilities. We have only to show that there is an uncountable number of such measures.

If we take two distributions $p_1(w)$ and $p_2(w)$ over u-sequences we finally arrive at two input measures μ_1 and μ_2 and two corresponding output measures η_1 and η_2 respectively. Let

$$\eta_1(F) = \frac{q_1^*(F) + q_1^*(TF) + \dots + q_1^*(T^{n-1} F)}{n}$$

$$\eta_2(F) = \frac{q_2^*(F) + q_2^*(TF) + \dots + q_2^*(T^{n-1}F)}{n}$$

for any Borel set $F \subset B^I$. Any two of the measures $q_1^*(T^r F)$ $r = 0, 1, \dots n-1$ and $q_2^*(T^r F)$, $r = 0, 1, \dots n-1$, being ergodic under T, are either identical or orthogonal. Hence it follows that $\eta_1 = \eta_2$ if and only if $q_1^*(F) = q_2^*(T^r F)$ for all Borel sets F and some fixed integer r lying in the interval $-n \leq r \leq n$. We now go back to the space of distributions p(u) over u-sequences and say that two distinct distributions q_1^* and q_2^* on B^I (corresponding to some $p_1(u)$ and $p_2(u)$) are equivalent if and only if the corresponding η_1 and η_2 are identical i.e. if and only if $q_1^*(F) = q_2^*(T^r F)$ for all Borel sets F and some integer r in $-n \leq r \leq n$. Thus the equivalence class corresponding to a q^* consists of $q^*(T^r F)$, $r = -n, -(n-1), \dots 0, 1, \dots n$. Thus each equivalence class contains utmost 2n+1 distributions. Since there is an uncountable number of distributions of the type q^* and different η's correspond to different equivalence classes it is clear that it is possible to construct an uncountable number of distinct distributions of the type η. This completes the proof.

3.4 <u>Channels with additive noise</u>

We have discussed in some detail the properties of a channel with finite memory in the sense of Khinchin. The most general channel for which the Feinstein's fundamental lemma and Shannon's fundamental theorems have been proved is the one with finite memory in the sense of Feinstein. As a justification for introducing channels with finite memory in the sense of

Khinchin we shall discuss here some of the properties of a particular class of channels which have finite memory in the sense of Khinchin but not in the sense of Feinstein. We now give a description of those channels.

Let the alphabet A be some finite group. For simplicity we shall take A to be an abelian group. Let the alphabet B be the same as A. In a natural way the space A^I becomes an abelian group. We denote by + and − the addition and inverse operation in the group A^I. For any set E we write

$$E - x = [z : z \in A^I, z + x \in E].$$

Let μ_0 be an invariant measure defined on A^I. Then the probability distributions

(3.4.1) $$\nu_x(F) = \mu_0(F-x), \quad F \in F_A$$

(associated with points $x \in A^I$) define a stationary channel. The output measure ν corresponding to any input measure μ is obtained by convoluting μ with μ_0. It is easy to verify that this channel is of zero memory in the sense of Khinchin and not necessarily in the sense of Feinstein. The channel defined by (3.4.1) is called a channel with additive noise.

Now let us suppose that the defining measure μ_0 is ergodic. Even then the channel need not be of zero memory in the sense of Feinstein. It need not even possess the property that the joint input-output measure corresponding to an ergodic input measure μ is ergodic. This is because of the fact that the convolution of two ergodic measures need not be ergodic. As an example one may consider the convolution of two different periodic ergodic measures with same period k. However the fundamental lemma of Feinstein and hence Shannon's theorems are valid for channels with additive noise as soon as the defining measure μ_o is ergodic.

Theorem 3.4.1. The capacity C of an additive noise channel $[A, \nu_x, A]$ defined by a finite group A of order a and an invariant measure μ_o is equal to $\log a - H(\mu_o, A)$ where $H(\mu_o, A)$ is the entropy of μ_o. It is attained at the Haar measure on the space A^I.

Proof. Since the channel is of zero-memory the two definitions of rate of transmission coincide. Thus for an input measure μ with output measure η and joint input-output measure ω we have

$$R(\mu) = R^*(\mu) = H(\eta) - H(\mu, B|A)$$

where B coincides with A. It is obvious from the definition of $H(\mu, B|A)$ that

$$H(\mu, B|A) = \lim_{n \to \infty} - \frac{1}{n} \iint \log \nu_x[y_1 \dots y_n] \, d\nu_x(y) d\mu(x)$$

$$= \lim_{n \to \infty} \frac{1}{n} \int H_n(\mu_o, A) d\mu(x) =$$

$$= \lim_{n \to \infty} \frac{H_n(\mu_o, A)}{n} = H(\mu_o, A)$$

where $H_n(\mu_o, A)$ is as in theorem 2.1.1. Thus

$$R(\mu) = H(\eta) - H(\mu_o, A).$$

Since $H(\eta,B) \leq \log a$ and equal to $\log a$ when and only when η is the Haar measure, we have

$$R(\mu) \leq \log a - H(\mu_0, A).$$

If μ is the Haar measure then η is also the Haar measure and hence

$$C = \log a - H(\mu_0, A).$$

is attained at the Haar measure.

<u>Theorem</u> 3.4.2. For any additive noise channel $[A, \nu_x, A]$ with channel probability $\nu_x(F) = \mu_0(F - x)$ where μ_0 is an ergodic measure the Feinstein's fundamental lemma is valid.

Let us denote the Haar measure by λ. The output measure corresponding to the input measure λ is the same as λ. Let the joint input-output measure corresponding to λ be ω. Then it is easy to verify that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\nu_x[y_1 \ldots y_n]}{\lambda[y_1 \ldots y_n]} =$$

$$= \log a + \lim_{n\to\infty} \frac{1}{n} \log \nu_x[y_1 \cdots y_n] =$$

$$= \log a - H(\mu_0, A) = C \quad \text{a.e. } (x,y)\ (\omega).$$

Then a verbatim repetition of the argument which Takano [21] uses to prove the Feinstein's fundamental lemma completes the proof.

Finally we state a problem concerning additive noise channels. We say that a system of u-sequences $u_1, u_2, \ldots u_N$ of length n and sets $B_1, \ldots, B_N$ of v-sequences of length n-m $[A, \nu_x, B]$ of memory m in the sense of Khinchin if

(1) $B_i \cap B_j = \emptyset$, $i \neq j$

(2) $\nu_{u_i}(B_i) > 1 - \varepsilon$

In the case of additive noise channels we say that the code is a translatory code if in addition to (1) and (2) there exists a set F of v sequences of length n ($m = 0$) such that $B_i = F + u_i$, $i = 1, 2, \ldots N$. The following problem remains open. Does there exist a translatory code for any additive noise channel defined by an ergodic measure with $N > 2^{n(C-\varepsilon)}$ for sufficiently large n? Is it possible to choose $u_1, u_2, \ldots u_N$ as a subgroup of A^n where A^n is the Cartesian product of A taken n times?

BIBLIOGRAPHY

1. L. Breiman, On achieving channel capacity in finite memory channels, Illinois. J. Math., Vol.4 (1960), pp 246-252.

2. J.L. Doob, Stochastic processes, New York (1953).

3. A. Feinstein, Foundations of Information Theory, New York (1958).

4. A. Feinstein, On the coding theorem and its converse for finite memory channels, Inform. and Control., Vol.2 (1959) pp 25-44.

5. P.R. Halmos, Lectures on Ergodic Theory, Mathematical Society of Japan (1956).

6. K. Jacobs, Personal communication.

7. A.I. Khinchin, Mathematical foundations of information theory, New York (1957).

8. A.N. Kolmogorov, On the entropy per unit time as a metric invariant of automorphisms, Dok. Akad. Nauk SSSR, Vol. 124 (1959) pp 754-755.

9. N. Kryloff et N. Bogoliouboff, La théorie générale de la mesure dans son application à l'étude des systèms dynamiques de la mécanique non linéaire, Ann. of Math. (2), Vol.38 (1937), pp 65-113.

10. B. Mcmillan, The basic theorems of information theory, Ann. Math. Statistics, Vol.24 (1953), pp 196-219.

11. M. Nisio, On polynomial approximation for strictly stationary processes, Journal of the Math. Soc. Japan, Vol. 12 (1960), pp 207-226.

12. J.C. Oxtoby and S. Ulam, Measure preserving homeomorphisms and metrical transitivity, Ann. of Math., Vol.42 (1941), pp 847-920.

13. J.C. Oxtoby, Ergodic sets, Bull. Amer. Math. Soc., Vol.58 (1952), pp 116-136.

14. J.C. Oxtoby, Stepanoff flows on the torus., Proc. Amer. Math. Soc. Vol.4 (1953), pp 982-987.

15. K.R. Parthasarathy, On the integral representation of the rate of transmission of a stationary channel, Illinois. J. Math., Vol.5 (1961) pp 299-305.

16. K.R. Parthasarathy, On the category of ergodic measures. (To appear in Illinois. J. Math.)

17. Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, Teor. Veroyatnost. i Primenen, Vol. 1 (1956) pp 177-238 (in Russian)

18. V.A. Rokhlin, In general a measure preserving transformation is not mixing, Doklady. Akad. Nauk, Vol.60 (1948), pp 349-351.

19. V.A. Rokhlin, New progress in the theory of transformations with invariant measure. Uspekhi. Mat. Nauk, 4(94), Vol.15 (1960).

20. C.E. Shannon, The mathematical theory of communication, Bell Syst. Techn. Journ. Vol.27 (1948), pp 379-423; 623-656.

21. K. Takano, On the basic theorems of information theory, Ann. Inst. Stat. Math. (Japan), Vol.9 (1957-58), pp 53-77.

22. I.P. Tsaragradsky, On the capacity of a stationary channel with finite memory, Teor. Veroyatnost. i Primenen, Vol.3 (1958), pp 84-96 (in Russian)

23. V.S. Varadarajan, Weak convergence of measures on separable metric spaces, Sankhya, Vol.19 (1958), pp 15-22.