MEASURABLE DYNAMICAL SYSTEM

BY

MISTER SAHLEMARIAM

DEPARTMENT OF MATHEMATICS
ADDIS ABABA UNIVERSITY

SUPERVISOR: **MENGISTU GOA SANGAGO (PhD)**

GRADUATE PROJECT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER SCIENCE IN MATHEMATICS

ADDIS ABABA
JANUARY 2012

Declaration

I declare that this project has been composed by me and that no part of the project has formed the basis for the award of any Degree, Diploma, Associate ship, Fellowship, or any other similar title to me.

Addis Ababa ______________________________

January 2012 **Mister Sahlemariam**

Permission

This is to certify that this project is compiled by ***Ms. Mister Sahlemariam*** *in the Department of Mathematics, Addis Ababa University, under my supervision. I hereby also confirm that the project can be submitted for evaluation by examiners and eventual defense.*

Addis Ababa ________________________________

January 2012 **Mengistu Goa Sangago(Ph.D)**

Abstract

Let $(\Omega, \boldsymbol{\beta}, \mu)$ *be a measure space, and* $\boldsymbol{\varphi}: \Omega \to \Omega$ *such that* $\boldsymbol{\varphi}^{-1}(\boldsymbol{A}) \in \boldsymbol{\beta}$ *for all* $\boldsymbol{A} \in \boldsymbol{\beta}$*. Then* μ *is said to be invariant with respect to* $\boldsymbol{\varphi}$ *if* $\mu(\boldsymbol{A}) = \mu(\boldsymbol{\varphi}^{-1}(\boldsymbol{A}))$ *,* $\boldsymbol{A} \in \boldsymbol{\beta}$*. If* μ *is invariant with respect to* $\boldsymbol{\varphi}$ *and is also a probability measure, then the quadruple* $(\boldsymbol{\Omega}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi})$ *is called a measurable dynamical system. One of the aims of this project is to present a variety of examples of measurable dynamical systems showing their importance and relevance.*

Let $(\boldsymbol{\Omega}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\varphi})$ *be a measurable dynamical system and let* $\boldsymbol{\varphi}^{(\boldsymbol{n})}$ *denotes the* $\boldsymbol{n}^{\boldsymbol{th}}$ *iterate of* $\boldsymbol{\varphi}$*. We also define* $\boldsymbol{\varphi}^{(\boldsymbol{0})}$ *be the identity function on* Ω*. For* $x \in \Omega$*, the sequence* $\boldsymbol{x}$*,* $\boldsymbol{\varphi}(\boldsymbol{x})$*,* $\boldsymbol{\varphi}(\boldsymbol{\varphi}(\boldsymbol{x})), \ldots,$ $\boldsymbol{\varphi}^{(\boldsymbol{n})}(\boldsymbol{x}), \ldots$ *called the orbit of* x*, describes the path of the point* x *as it moves in* Ω *under iteration of the mapping* φ*. The second most important part of this project, called* ***ergodic theory****, studies the properties of this sequence.*

In addition, isomorphism of two measurable dynamical systems is introduced, and the most powerful tool for deciding when two measurable dynamical systems are isomorphic, namely ***entropy****, is presented.*

Acknowledgment

Above all I would like to thank the Almighty God for giving me the patience; next my gratitude goes to my advisor Dr. Mengistu Goa for helpful discussion, comments and providing the necessary materials in the preparation of this paper. Beside I would like to thank my families and my friends who were encouraging me on various section of my work.

Table of contents

Page

Introduction

Dynamical system consists of a set Ω, a σ - algebra β on Ω , a measure μ and a measurable function φ defined on Ω, the iteration of the map is defined by induction

$$\varphi^0 = idenitity, \varphi^n = \varphi \circ \varphi^{n-1} \; for\; all\; n \; \in N.$$

The theory of measurable dynamical system splits in to subfields which differ by the structure which are imposed on Ω and φ.

i. Differential dynamical system:- deals with action by differentiable maps on smooth manifolds.
ii. Topological dynamical system: -deal with actions of continuous maps on topological space usually compact metric space.
iii. Ergodic theory:- deal with measure preserving actions of measurable maps on a measure space.

In this project we discuss the Ergodic theory of dynamical system. The aim of this theory is to describe the behavior of φ^n as $n \to \infty$. The Kolmogrove –Sinai Theorem identify the necessary condition for two measurable dynamical systems to be isomorphic, that is, they have the same entropy.

This project has two parts. Part one: Ergodic theory. In this part we discuss the invariant measure, Ergodicity and prove the point wise Ergodic theorem and presenting several examples. Part two: Entropy. This part mainly focuses the conditions when two measureable dynamical system to be isomorphic, the property and calculation of entropy and entropy of a Bernoulli shift.

PART ONE: ERGODIC THEORY

1.1. Invariant Measure

Suppose that a particle p moves around inside Ω according to the following rule; If p is at $x\ at\ time\ n$,it moves to $\varphi(x)\ at\ time\ n+1$, where $\varphi: \Omega \to \Omega$ is a function that is independent of n. Although, according to this rule, the particle is always moving in Ω, the law governing its movement remains constant for all time. For $A \subset \Omega, let$

$$\mu_n(A) = \begin{cases} 1 & if\ p \in A\ at\ time\ n, \\ 0 & otherwise \end{cases}$$

Then the expression

$$\mu(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu_k(A)$$

represents an average over time of the number of visits of the particle to the set A, that is the number of visits to A per unit time by the particle.

Let β denote the collection of subset of Ω such that the previous limit exist.$\emptyset\ and\ \Omega$ belongs to $\beta\ and\ \mu$ satisfies the following condition

i. $\mu(A) \geq 0 for\ all\ A \in \beta$
ii. $\mu(\emptyset) = 0$
iii. $\mu(\Omega) = 1$
iv. If $A, B \in \beta\ are\ disjoint\ then\ \mu(A \cup B) = \mu(A) + \mu(B)$

$\boldsymbol{Definition\ 1}$: Let Ω be a set. A nonempty collection β of subset of Ω is called a σ-algebra if the following two conditions are satisfied;

i. If $A \in \beta\ then\ A^c \in \beta$
ii. If $\{A_n\} \in \beta\ then\ \bigcup_n A_n \in \beta$

Definition 2: Let Ω be a set and β a σ- algebra of subset of Ω. A measure μ on β is an extended real- valued function satisfying the following condition;

i. $\mu(A) \geq 0 \; for \; all \; A \in \beta$
ii. $\mu(\emptyset) = 0$
iii. If E_1, E_2, E_3 ..., are in β with $E_i \cap E_j = \emptyset \; for \; i \neq j$ then

$$\mu(\cup_n E_n) = \sum_n \mu(E_n)$$

The triple $(\Omega, \boldsymbol{\beta}, \boldsymbol{\mu})$ is called **measure space**.

$If \; \mu(\Omega) = 1 \; in$ the measure space then we says that μ is a probability measure and (Ω, β, μ) is **probability space**.

Definition 3:-Let (Ω, β, μ) be a measure space. Suppose $\varphi: \Omega \to \Omega$ and that $\varphi^{-1}(A) \in \beta \; for \; all \; A \in \beta$.Then μ is said to be **invariant** with respect to φ if

$$\mu(A) = \mu(\varphi^{-1}(A)), A \in \beta.$$

If μ is invariant with respect to φ and is also a probability measure, then the quadruple $(\Omega, \beta, \mu, \varphi)$ is called a **measurable dynamical system.**

Example 1:- Addition modulo one

Let the operations $\dot{+}$ be defined on $[0,1) \; by$

$$x \dot{+} y = (x+y) mod \; 1 = \begin{cases} x+y & if \; x+y < 1 \\ x+y-1 & if \; x+y \geq 1 \end{cases}$$

For fixed $b \in [0,1), let \; \varphi_b(x) = x \dot{+} b \; then \; ([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)}, \varphi_b)$ is a measurable dynamical system.($\mathcal{M}_{[0,1)} \; be \; a \; \sigma -$ algebra $of \; Lebegue \; measureable \; set, \lambda_{[0,1)} \; be \; Lebeg$ue measure)

Proof:

$$\varphi_b(x) = \begin{cases} x+b & if \; x \in [0, 1-b) \\ x+b-1 & if \; x \in [1-b, 1) \end{cases}$$

Claim 1; φ_b is measurable

Let $\alpha \in R, then$

$$\{x: x \in [0,1): \varphi_b(x) \geq \alpha\}$$

$$= \begin{cases} \emptyset & if \alpha \geq 1 \\ [\alpha - b, 1 - b) & if\ b \leq \alpha < 1 \\ [0,1) & if\ \alpha < 0 \\ [0,1 - b) \cup [1 - (b - \alpha), 1) & if\ 0 \leq \alpha < b \end{cases}$$

Therefore, φ_b is measurable.

Claim 2: $\lambda_{[0,1)}$ is invariant with respect to φ_b ($i, e\ \lambda_{[0,1)}(E) = \lambda_{[0,1)}(\varphi^{-1}(E))$)

Let $E \subseteq [0,1)$ be measurable. Since φ_b is measurable, $\varphi^{-1}{}_b(E)$ is measurable

$$\varphi_b\ (x) = \begin{cases} x + b & if\ x \in [0, 1 - b) \\ x + b - 1 & if\ x \in\ [1 - b, 1) \end{cases}$$

1. $\varphi_b([0,1 - b)) = [b, 1) := B_1$

2. $\varphi_b([1 - b, 1)) = [0, b) := B_2$

3. $\varphi_b{}^{-1}([b, 1)$ =$[0,1 - b)$

4. $\varphi_b{}^{-1}([0, b)$ =$[1 - b, 1)$

Consider that $E = (E \cap B_1) \cup (E \cap B_2)$

$\varphi_b{}^{-1}(E) = \varphi_b{}^{-1}(E \cap B_1)\ \cup\ \varphi_b{}^{-1}(E \cap B_2)$, But

$$\varphi_b{}^{-1}(E \cap B_1)\ = \{x \in [0,1): \varphi_b(x) \in E \cap B_1\}$$

$$= \{x \in [0,1): x + b \in E \cap B_1\}$$

$$= \{x \in: x \in (E \cap B_1) - b\}$$

Hence, $\varphi_b{}^{-1}(E \cap B_1) = (E \cap B_1) - b$. Similarly

$$\varphi_b{}^{-1}(E \cap B_2) = \{x: \varphi_b(x) \in E \cap B_2\}$$

$$= \{x: x + b - 1 \in E \cap B_2\}$$

$$= \{x: x \in (E \cap B_1) - (b - 1)\}$$

Hence, $\varphi_b{}^{-1}(E \cap B_2) = (E \cap B_2) - (b - 1)$

Thus, $\varphi_b{}^{-1}(E) = \varphi_b{}^{-1}(E \cap B_1) \ \cup \ \varphi_b{}^{-1}(E \cap B_2)$,

$$= \big((E \cap B_1) - b\big) \cup ((E \cap B_2) - (b - 1))$$

$\lambda_{[0,1)}\big(\varphi_b{}^{-1}(E)\big) = \lambda_{[0,1)}(\varphi_b{}^{-1}(E \cap B_1) + \lambda_{[0,1)}(\varphi_b{}^{-1}(E \cap B_2))$

$$= \lambda_{[0,1)}\Big(\big((E \cap B_1) - b\big)\Big) + \lambda_{[0,1)}\big((E \cap B_2) - (b - 1)\big)$$

$= \lambda_{[0,1)}(E \cap B_1) + \lambda_{[0,1)}(E \cap B_2)$ (Since, Lebesgue measure is translation invariant)

$$= \lambda_{[0,1)}(E)$$

Therefore, $\lambda_{[0,1)}\big(\varphi_b{}^{-1}(E)\big) = \lambda_{[0,1)}(E)$ .Hence, $\lambda_{[0,1)}$ *is invariant with respect to* φ.

Claim 3:- $\lambda_{[0,1)}$ is probability measure

$\lambda_{[0,1)}$ *is a lebesgue measure, so* $\mathcal{M}[a,b) = l[I]$ so, $\boldsymbol{\mathcal{M}}[0,1) = 1$

Hence $\lambda_{[0,1)}$ is a probability measure. Therefore, $([0,1), \mathcal{M}_{[0,1)}\ \lambda_{[0,1)}, \varphi_b)$ is a measurable dynamical system.

1.2. Pointwise Ergodic Theorem

Let $(\Omega, \beta, \mu, \varphi)$ be a measurable dynamical system and let $\varphi^{(n)}$ denotes the n^{th} iterate of φ. We also define $\varphi^{(0)}$ be the identity function on Ω.

For $x \in \Omega$, *the sequence*

$x, \varphi(x), \varphi\big(\varphi(x)\big), \dots \varphi^{(n)}(x), \dots$ Called the orbit of x, describes the path of the point x as it moves in Ω under iteration of mapping φ.

Theorem1: point wise Ergodic Theorem.

For each $f \in L^1(\mu)$ the limit

$f^* = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^{(k)}$ (1.1)

exist μ ae .Further more , $f^* \in L^{'}(\mu)$ and satisfies $\int f^*\, d\mu = \int f\, d\mu$ (1.2)

Proof:- Let $f = x_B$

Let $S_n(x)$ *be the number of visits to the first* n^{th} terms of orbit of x,

$$S_n(x) = \sum_{k=0}^{n-1} x_B \circ \varphi^{(k)}(x)$$

Let $A_n(x)$ *be* the average number of visit,

$$A_n(x) = \frac{S_n(x)}{n}.$$

Suppose $\bar{A} = \lim_{n\to\infty} sup\, A_n(x)$ and $\underline{A} = \lim_{n\to\infty} inf\, A_n(x)$

Claim 1:- $\int \overline{A} d\mu \le \mu(B)$ *and* $\int \underline{A} d\mu \ge \mu(B)$

To verify the claim, will make use of the following properties of the functions $\overline{A}$ *and* $\underline{A}$:

$0 \le \underline{A} \le \overline{A} \le 1$ (1.3)

And

$\underline{A} \circ \varphi = \underline{A}$ and $\overline{A} \circ \varphi = \overline{A}$. (1.4)

This means,

$$\underline{A} \circ \varphi(x) = \underline{A}\big(\varphi(x)\big) = \lim_{n\to\infty} inf\, A_n(\varphi(x))$$

$$= \lim_{n\to\infty} inf \frac{S_n(\varphi(x))}{n}$$

$$= \lim_{n\to\infty} inf \frac{1}{n}\sum_{k=0}^{n-1} x_B \circ \varphi^{(k+1)}(x)$$

$$= \lim_{n\to\infty} inf \frac{1}{n}\big(x_B \circ \varphi + x_B \circ \varphi^{(2)} + x_B \circ \varphi^{(3)} + x_B \circ \varphi^{(4)} + \cdots + x_B \circ \varphi^{(n)}\big)(x)$$

$$= \lim_{n\to\infty} inf \frac{1}{n}\left(\sum_{k=0}^{n-1} x_B \circ \varphi^k - x_B + x_B \circ \varphi^{(n)}\right)(x)$$

$$= \lim_{n\to\infty} inf \frac{1}{n}(\sum_{k=0}^{n-1} x_B \circ \varphi^k)(x) = \underline{A}$$ (since, as $n \to \infty$, $the\ term$

$(-x_B + x_B \circ \varphi^{(n)}) \to 0$)

Therefore, $\underline{A} \circ \varphi = \underline{A}$.

Similarly,

$$\overline{A} \circ \varphi(x) = \overline{A}\big(\varphi(x)\big) = \lim_{n\to\infty} sup\, A_n\big(\varphi(x)\big)$$

$$= \lim_{n\to\infty} sup \frac{S_n(\varphi(x))}{n}$$

$$= \lim_{n\to\infty} sup \frac{1}{n}\sum_{k=0}^{n-1} x_B \circ \varphi^{(k)}\,(\varphi(x))$$

$$= \lim_{n\to\infty} sup \frac{1}{n}\left(x_B \circ \varphi + x_B \circ \varphi^{(2)} + x_B \circ \varphi^{(3)} + x_B \circ \varphi^{(4)} + \cdots + x_B \circ \varphi^{(n)}\right)(x)$$

$$= \lim_{n\to\infty} sup \frac{1}{n}\left(\sum_{k=0}^{n-1} x_B \circ \varphi^k - x_B + x_B \circ \varphi^{(n)}\right)(x)$$ (Since, as $n \to \infty$, $the\ term$

$(-x_B + x_B \circ \varphi^{(n)}) \to 0$)

$$= \lim_{n\to\infty} sup \frac{1}{n}(\sum_{k=0}^{n-1} x_B \circ \varphi^k)(x) = \overline{A}$$

Therefore, $\overline{A} \circ \varphi = \overline{A}$. It is also true that $\overline{A} \circ \varphi^{(k)} = \overline{A}\ and\ \underline{A} \circ \varphi^{(k)} = \underline{A}$

Let $\varepsilon > 0$ and $\tau_\varepsilon(x)$ denote the 1st time that average number of visits exceeds $\overline{A}(x) - \varepsilon$.

$$\tau_\varepsilon(x) = min\{n \in N: A_n(x) > \overline{A}(x) - \varepsilon\}$$

We observe that by (1.3), $\tau_\varepsilon(x)$ is always a positive integer. From

$$\{x: \tau_\varepsilon(x) > c\} = \bigcap_{n<c}\{x: A_n(x) \le \overline{A}(x) - \varepsilon\}$$

It follows that τ_ε is β-measurable.

Case1. $\tau_\varepsilon \epsilon L^\infty(\mu)$. Then we can choose an integer M such that

$$\mu\left({\tau_\varepsilon}^{-1}(M,\infty)\right) = 0 \tag{1.5}$$

For each $x \epsilon \Omega$ we consider the sequence of integers

$$\tau_1(x) = \tau_\varepsilon(x), \tau_2(x) = \tau_\varepsilon\left(\varphi^{\tau_1(x)}(x)\right), \tau_3(x) = \left(\varphi^{\tau_1(x)+\tau_2(x)}(x)\right), \ldots.$$

It follows from (1.5) and the invariant of μ with respect to φ that, for μ almost all x, we have

$$\tau_j(x) \le M, j \epsilon N \tag{1.6}$$

Suppose that x satisfy (1.6).Let n be a positive integer than M and q be such that

$$\sigma_q \le n < \sigma_{q_{+1}} \tag{1.7}$$

Where we are using the notation $\sigma_q = \tau_1 + \tau_2 + \tau_3 + \cdots \tau_{\sigma_q}$ then $s_n(x) \ge s_{\sigma_q}(x)$

$$\begin{aligned} s_{\sigma_q}(x) &= \sum_{k=0}^{\sigma_q - 1} x_{B\ 0}\varphi^k \\ &= \sum_{k=0}^{\sigma_1 - 1} x_{B\ 0}\varphi^{(k)} + \sum_{k=\sigma_1}^{\sigma_2 - 1} x_{B\ 0}\varphi^{(k)} + \sum_{k=\sigma_2}^{\sigma_3 - 1} x_{B\ 0}\varphi^{(k)} + \cdots + \sum_{k=\sigma_{q-1}}^{\sigma_q - 1} x_{B\ 0}\varphi^{(k)} \\ &= s_{\tau_1}(x) + s_{\tau_2}\left(\varphi^{(\sigma_1)}(x)\right) + s_{\tau_3}\left(\varphi^{(\sigma_2)}(x)\right) + \cdots + s_{\tau_q}(\varphi^{(\sigma_q - 1)}(x) \end{aligned}$$

It follows from (1.4) and the definition of τ_ε that

$$s_{\tau_1}(x) \ge \tau_1(\overline{A}(x) - \varepsilon)$$

$$s_{\tau_2}\left(\varphi^{\sigma_1}(x)\right) \ge \tau_2(\overline{A}\left(\varphi^{\sigma_1}(x)\right) - \varepsilon) = \tau_2\left(\overline{A}(x) - \varepsilon\right)$$

$$s_{\tau_3}(\varphi^{\sigma_2}(x) \ge \tau_3(\overline{A}\left(\varphi^{\sigma_2}(x)\right) - \varepsilon) = \tau_3\left(\overline{A}(x) - \varepsilon\right)$$

.

.

.

$s_{\tau_q}(\varphi^{\sigma_{q-1}}(x)) \geq \tau_q(\overline{A}(\varphi^{\sigma_{q-1}}(x)) - \varepsilon) = \tau_q(\overline{A}(x) - \varepsilon)$

$$s_{\sigma_q}(x) = s_{\tau_1}(x) + s_{\tau_2}\left(\varphi^{(\sigma_1)}(x)\right) + s_{\tau_3}\left(\varphi^{(\sigma_2)}(x)\right) + \cdots + s_{\tau_q}(\varphi^{(\sigma_q-1)}(x)$$

$$\geq \tau_1(\overline{A}(x) - \varepsilon) + \tau_2(\overline{A}(x) - \varepsilon) + \cdots + \tau_q(\overline{A}(x) - \varepsilon)$$

$$= (\tau_1 + \tau_2 + \tau_3 + \cdots + \tau_q)(\overline{A}(x) - \varepsilon)$$

$$= \sigma_q(\overline{A}(x) - \varepsilon)$$

From the definition of q, we have $n \leq \sigma_{q+1} \Rightarrow n \leq \tau_1 + \tau_2 + \cdots + \tau_q + \tau_{q+1}$

$$\Rightarrow n \leq \sigma_q + \tau_{q+1}$$

$$\Rightarrow n - \tau_{q+1} \leq \sigma_q$$

Hence,$s_n(x) \geq \sigma_q(\overline{A}(x) - \varepsilon) \geq (n - \tau_{q+1})(\overline{A}(x) - \varepsilon)$

Applying (1.6) we conclude that, for μ-almost all x, we have

$$s_n(x) \geq (n - M)(\overline{A}(x) - \varepsilon) \qquad (1.8)$$

holds for all $x \, \varepsilon \, \Omega$. And,

$$x_B \circ \varphi(x) = \begin{cases} 1 \;\; if \; \varphi(x) \in B \\ 0 \; if \; \varphi(x) \notin B \end{cases}$$

$$= \begin{cases} 1 \;\; if \; x \in \varphi^{-1}(B) \\ 0 \; if \; x \notin \varphi^{-1}(B) \end{cases}$$

$$= x_{\varphi^{-1}(B)}(x)$$

Integrating both sides of (1.8) and using the invariant of μ with respect to φ, we obtain

$$\int s_n(x)d\mu = \int \sum_{k=0}^{n-1} x_B \circ \varphi^{(k)}\, d\mu(x)$$

$$= \sum_{k=0}^{n-1} \int x_B \circ \varphi^{(k)}\, d\mu$$

$$= \sum_{k=0}^{n-1} \int x_{\varphi^{-k}(B)} d\mu$$

$$= \sum_{k=0}^{n-1} \mu\,(\varphi^{-k}(B)) = \sum_{k=0}^{n-1} \mu\,(B) = n\mu(B) \qquad (*)$$

Since μ is invariant with respect to φ.

$n\mu(B) = n(\mu(\varphi^{-1}(B)))$ *and using* the fact that μ is also invariant with respect to φ^k.

$$n\mu(B) = \sum_{k=0}^{n-1} \mu\,(\varphi^{-k}(B))$$

Applying (1.8) and (*),we get

$$\int s_n(x)d\mu(x) \geq \int (n-M)\big(\overline{A}(x) - \varepsilon\big)\, d\mu(x)$$

$$n\mu(B) \geq \int (n-M)(\overline{A}(x) - \varepsilon)d\mu(x)$$

$$\mu(B) \geq \frac{1}{n}\int (n-M)(\overline{A}(x) - \varepsilon)d\mu(x)$$

Dividing by n and letting $n \to \infty$, we get

$$\mu(B) \geq \int \big(\overline{A}(x) - \varepsilon\big)d\mu(x)$$

$$\geq \int \overline{A}(x)d\mu(x) - \varepsilon$$

Since ε was arbitrary, $\mu(B) \geq \int \overline{A}(x)d\mu(x)$

Similarly, we define,

$$\tau_\varepsilon(x) = min\big\{n\ \varepsilon\ N: A_n(x) < \underline{A}(x) + \varepsilon\big\}\text{, for } \varepsilon > 0$$

From (1.7) *and the definition of* τ_ε, *we have,*

$s_n(x) \leq s_{\sigma_{q+1}}(x)$ But $\sigma_{q+1} = \tau_1 + \tau_2 + \tau_3 + \cdots + \tau_q + \tau_{q+1}$

$$= \sigma_q + \tau_{q+1} \leq n + \tau_{q+1}$$

From (1.5) and the definition of τ_ε , that

$$s_{\sigma_{q+1}}(x) = \sum_{k=0}^{\sigma_{q+1}-1} x_{B\ 0}\varphi^{(k)}(x)$$

$$= \sum_{k=0}^{\sigma_1-1} x_{B\ 0}\varphi^{(k)}\,(x) + \sum_{k=\sigma_1}^{\sigma_2-1} x_{B\ 0}\varphi^{(k)}(x) + \sum_{k=\sigma_2}^{\sigma_3-1} x_{B\ 0}\varphi^{(x)}(x) + \cdots + \sum_{k=\sigma_q}^{\sigma_{q+1}-1} x_B \circ \varphi^k(x)$$

$$= s_{\tau_1}(x) + s_{\tau_2}\left(\varphi^{(\sigma_1)}(x)\right) + s_{\tau_3}\left(\varphi^{(\sigma_2)}(x)\right) + \cdots + s_{\tau_{q+1}}(\varphi^{(\sigma_{q+1}-1)}(x)$$

$$\leq \tau_1\big(\underline{A}(x) + \varepsilon\big) + \tau_2\big(\underline{A}(x) + \varepsilon\big) + \cdots + \tau_{q+1}(\underline{A}(x) + \varepsilon)$$

$$= (\tau_1 + \tau_2 + \tau_3 + \cdots + \tau_{q+1})(\underline{A}(x) + \varepsilon)$$

$$= \sigma_{q+1}(\underline{A}(x) + \varepsilon)$$

Hence, $s_n(x) \leq \sigma_{q+1}\big(\underline{A}(x) + \varepsilon\big) \leq (n + \tau_{q+1})(\underline{A}(x) + \varepsilon)$

Applying (1.6) we conclude that for μ-almost all x, we have the inequality

$$s_n(x) \leq (n + M)(\underline{A}(x) + \varepsilon) \qquad (1.9)$$

Integrating both sides of (1.9) and using the invariant of μ with respect to φ we obtain,

$$n\mu(B) = \textstyle\sum_{k=0}^{n-1} \mu\,(B) = \sum_{k=0}^{n-1} \mu\,(\varphi^{-k}(B)$$

$$\int s_n(x)d\mu\,(x) \leq \int (n + M)(\underline{A}(x) + \varepsilon)$$

Dividing by n and letting $n \to \infty$, we get

$$\mu\,(B) \leq \int \underline{A}d\mu(x) + \varepsilon$$

As $\varepsilon > 0$ was arbitrary, we obtain the claim, thus the proof of the theorem is complete.

Case2: $\tau_\varepsilon \notin L^\infty(\mu)$. that is , $\tau_\varepsilon\ is$ not essentially bounded. Because τ_ε is a finite –valued, we can choose a positive integer M such that $\mu(\tau^{-1}(M,\infty)) < \varepsilon$

Set $B^\varepsilon = B \cup \left(\tau^{-1}{}_\varepsilon(M,\infty)\right), {s_n}^\varepsilon(x) = \sum_{k=0}^{n-1} x_{B^\varepsilon} \circ \varphi^k$ and ${A_n}^\varepsilon = \frac{{s_n}^\varepsilon(x)}{n}$

Let $\tau^\varepsilon = min\{n \in N: {A_n}^\varepsilon > \overline{A}(x) - \varepsilon\}$ hence τ^ε is A-measurable and $\tau^\varepsilon \leq \tau_\varepsilon$

Claim: $\tau^\varepsilon(x) \leq M, x \in \ \Omega$

For,if $\tau^\varepsilon(x) > M\ then\ \tau_\varepsilon \geq \tau^\varepsilon > M$.Hence,

$${A_1}^\varepsilon(x) = x_{B^\varepsilon} = x_{B\cup\tau^{-1}(M,\infty)} \geq 1 > \ \overline{A}(x) - \varepsilon$$

It implies, $\tau^\varepsilon(x) = 1 < M$ it contradict our assumption. So, it must be the case $\tau^\varepsilon(x) \leq M$.

Now let n be a positive integer than M and let q be such that $\sigma_q \leq n \leq \sigma_{q+1}$

$${s_n}^\varepsilon(x) \geq {s_{\sigma_q}}^\varepsilon(x) = \sum_{k=0}^{\sigma_1 - 1} x_{B^\varepsilon} \circ \varphi^{(k)} + \sum_{k=\sigma_1}^{\sigma_2 - 1} x_{B^\varepsilon} \circ \varphi^{(k)} + \cdots + \sum_{k=\sigma_{q-1}}^{\sigma_q - 1} x_{B^\varepsilon} \circ \varphi^{(k)}$$

$$= {s_{\tau_1}}^\varepsilon(x) + {s_{\tau_2}}^\varepsilon\left(\varphi^{(\sigma_1)}(x)\right) + \cdots + {s_{\tau_q}}^\varepsilon(\varphi^{(\sigma_q - 1)}(x))$$

$$\geq \tau^\varepsilon{}_1\left(\overline{A}(x) - \varepsilon\right) + \tau^\varepsilon{}_2\left(\overline{A}(x) - \varepsilon\right) + \cdots + \tau^\varepsilon{}_q(\overline{A}(x) - \varepsilon)$$

$$= \sigma_q\left(\overline{A}(x) - \varepsilon\right) \geq (n - M)(\overline{A}(x) - \varepsilon)$$

$${s_n}^\varepsilon(x) \geq (n - M)(\overline{A}(x) - \varepsilon)$$

Integrate both sides, $\int {s_n}^\varepsilon\,(x) d\mu(x) \geq \int (n - M)(\overline{A}(x) - \varepsilon)\, d\mu(x)$

As $n \to \infty$, we have,$\mu(B^\varepsilon) \geq \int \overline{A}(x)\, d\mu(x) - \varepsilon$

Therefore $\mu(B^\varepsilon) \geq \int \overline{A}(x)\, d\mu(x) - \varepsilon$

$$\mu(B) + \varepsilon > \mu(B) + \mu\left(\tau^{-1}(M,\infty)\right), since\ \mu\left(\tau^{-1}(M,\infty)\right) < \varepsilon$$

$$= \mu(B^{\varepsilon}) \geq \int \overline{A}(x)d\mu - \varepsilon$$

Thus , $\mu(B) + \varepsilon \geq \int \overline{A}(x)d\mu - \varepsilon$

Since ε was arbitrary, then

$$\mu(B) \geq \int \overline{A}(x)d\mu \qquad (*)$$

By similar argument, take $\tau^{\varepsilon} = min\{n \in N: A_n{}^{\varepsilon} < \underline{A}(x) + \varepsilon\}$.

Define q such that $\sigma_q \leq n \leq \sigma_{q+1} now$ by definition of τ^{ε} and $A_n{}^{\varepsilon}$,

$s_{\sigma_q}(x) \leq s_n(x) < n(\underline{A}(x) + \varepsilon)$ and $\sigma_{q+1} = \sigma_q + \tau_{q+1} \leq n + \tau_{q+1}$

$s_n(x) \leq (n + M)(\underline{A}(x) + \varepsilon)$

$\mu(B^{\varepsilon}) \leq \int \underline{A}(x) + \varepsilon$

$\mu(B) - \varepsilon < \mu(B) - \mu(\tau^{-1}(M, \infty)), since\ -\varepsilon < -\mu(\tau^{-1}(M, \infty))$

$$\leq \mu(B) + \mu(\tau^{-1}(M, \infty))$$

$$= \mu(B^{\varepsilon}) \leq \int \underline{A}(x) + \varepsilon$$

$\mu(B) - \varepsilon) \leq \int \underline{A}(x) + \varepsilon$

Since ε was arbitrary then

$$\mu(B) \leq \int \underline{A}(x) \qquad (**)$$

From (*) and (**) we get

$$\mu(B) \leq \int \underline{A}(x)d\,\mu(x)\ and\ \int \overline{A}\,(x)d\,\mu(x) \leq \mu(B)$$

Hence our claim holds. ■

Definition 4 :.(Ergodicity)

A measurable dynamical system $(\Omega, \beta, \mu, \varphi)$ is called ergodic if

$$E \in \beta \,\&\, E = \varphi^{-1}(E) \Rightarrow \mu(E) = 0 \; or \; \mu(E) = 1$$

Remark:

Ergodicity is equivalent to,

$$E \in \beta \,\&\, E \subset \varphi^{-1}(E) \Rightarrow \mu(E) = 0 \; or \; \mu(E) = 1$$

Theorem 2:

Let $(\Omega, \beta, \mu, \varphi)$ be a measurable dynamical system. Then the following are equivalent:

a) $(\Omega, \beta, \mu, \varphi)$ is ergodic
b) For each $f \in L^1(\mu)$, the average
$f^* = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ \varphi^k$ is constant $\mu \; a.e$

C) If $f \in L^1(\mu)$ and $f \circ \varphi = f \; \mu \; a.e$ then f is constant $\mu \; a.e$

Proof: $(a \Rightarrow c)$ suppose $(\Omega, \beta, \mu, \varphi)$ is ergodic. If $f \in L^1(\mu) and f \circ \varphi = f \; \mu \; a.e$

Claim: f is constant

Let f is real valued. Let $D = \{x \in \Omega: f(x) \neq f \circ \varphi(x)\}$.then

$\mu(D) = 0. letting \; \varphi^{-k} = \varphi^{k^{-1}}$ and using the fact that μ is invariant with respect to $\varphi^{(k)}$.
$\mu(D) = 0 = \mu(\varphi^{k^{-1}}(D))$ for all k.

Hence $\mu(\bigcup_{k=0}^{\infty} \varphi^{-k}(D)) \leq \sum_{k=0}^{\infty} \mu\left(\varphi^{k^{-1}}(D)\right) = 0$

Let $b \in R \; and \; E = f^{-1}\big((-\infty, b)\big) \setminus \bigcup_{k=0}^{\infty} \varphi^{k^{-1}}(D)$ .then we have
$\mu(E) = \mu\left(f^{-1}\big((-\infty, b)\big)\right) equals \; 0 \; or \; 1$

It show that f is constant $\mu \; a.e$

$(c \Rightarrow a)$ suppose $f \in L^1(\mu)$ and $f \circ \varphi = f$ μ $a.e$ then f is constant

Let $E \in \beta$ $such$ $that$ $E = \varphi^{-1}(E)$ then $x_E \circ \varphi = x_{\varphi^{-1}(E)} = x_E$

By (c), x_E is constant μ $a.e.$ it follows $\mu(E) = 0$ or $\mu(E) = 1$.

$(b \Rightarrow c)$: Suppose $f^* = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^k$ for each $f \in L^1(\mu)$, f^*,is constant.

Let $f \circ \varphi = f$ μ $a.e$ $then$ $f \circ \varphi^{(k)} = f$ .we went to show that f is $constant$.

By(b), $f^* = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^k$

$= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f = f$. But f^* is constantμ $a.e$, so f is constant μ $a.e$.

$(c \Rightarrow b)$: suppose (c) is hold,

$f^* = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^k = f$, so by(c) f is $constant$, $hence$ f^* is constant μ $a.e$.

Corollary 2.1: If $(\Omega, \beta, \mu, \varphi)$ is ergodic. For each $f \in L^1(\mu)$

$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ \varphi^{(k)}(x) = \int f\, d\mu$ for almost all $x \in \Omega$.

1.3. Examples of measurable dynamical system and ergodic

Example2: Rotation through an angle

Let E be the map from [0,1) on to the unit circle T in the complex plane define by $E(x) = e^{2\pi i x}$ and let $\beta = \{A \subset T : E^{-1}(A) \in \mathcal{M}\}$ is σ- algebra..Define the measure μ on β by

$\mu(A) = \lambda(E^{-1}(A))$, so $that$ μ is normalize arc length measure on T. also, for fixed $b \in [0,1)$, define $\varphi_b : T \to T by$ $\varphi_b(z) = e^{2\pi i b} z.$ $then$ $(T, \beta, \mu, \varphi_b)$ is a measurable dynamical system and is ergodic if and only if b is irrational.

Proof: (I)

Let $E: [0,1) \to T$ by $E(x) = e^{2\pi i x}$

$\varphi_b: T \to T$ by $\varphi_b(z) = e^{2\pi ib} z$ and $\mu(A) = \lambda(E^{-1}(A))$

$\beta = \{A \subset T: E^{-1}(A) \in \mathcal{M}\}$

Claim1: φ_b is measurable

$\varphi_b: T \to T,\ \varphi_b(z) = e^{2\pi ib} z, z \in T$

For each $w \in T$ there is $z \in T$ such that , $\varphi_b(z) = w$

Let $w = e^{i\theta} \in T. Put\ z = e^{i(\theta - 2\pi b)}, then$

$$\varphi_b(z) = e^{2\pi ib} z = e^{2\pi ib}\ e^{i(\theta-2\pi b)} = e^{i\theta} = w.$$

Hence φ_b is surjective.

Let $w_1 \in T$ and $w_2 \in T, suppose$ $\varphi_b(w_1) = \varphi_b(w_2)$ then $e^{2\pi ib} w_1 = e^{2\pi ib} w_2$ it implies $w_1 = w_2$

Hence φ_b is injective. Therefore φ_b is bijective. For each $z \in T, {\varphi^{-1}}_b(z) \in T$

Hence φ_b is measurable.

Claim2: μ is invariant with respect to φ_b.

$${\varphi^{-1}}_b(E) = \{z \in T: \varphi_b(z) \in E\}$$

$$= \left\{z \in T: e^{2\pi ib} z \in E\right\}$$

$$= \left\{z \in T: z \in E. e^{-2\pi ib}\right\} = \left\{z \in T \cap E. e^{-2\pi ib}\right\} = \left\{z = E. e^{-2\pi ib}\right\}$$

$${\varphi^{-1}}_b(E) = E. e^{-2\pi ib}$$

$$\mu\left({\varphi^{-1}}_b(E)\right) = \mu\left(E. e^{-2\pi ib}\right) = \left|e^{-2\pi ib}\right| \mu(E) = \mu(E)$$

Therefore, μ is invariant with respect to φ.

ii) $(T, \beta, \mu, \varphi_b)$ is ergodic if and only if b is irrational.

Suppose $f \in L^1(\mu)$ in such that $f \circ \varphi_b = f$. Then the Fourier coefficient of the function $g(x) = f(e^{ix})$ must satisfy $\hat{g}(n) = e^{2\pi inb}\, \hat{g}(n)$.If b is irrational, then $\hat{g}(n) = 0$ for all non zero integer n, since $g \in L^1(\mu)$ and $\hat{g}(n) = 0$ then $g = 0\ a.e$,for all non zero integer n.Thus,f is constant. Therefore $(T, \beta, \mu, \varphi_b)$ is ergodic.

On the other hand, if b is rational, say,$b = \frac{p}{q}$ where p and q are integers, then the function $f(z) = z^q$ is non constant and satisfies $f \circ \varphi_b = f$. Hence $(T, \beta, \mu, \varphi_b)$ is not ergodic.

Example 3: Multiplication by $2 mod\ one$

Let the mapping φ_b in example 1 be replaced by

$$\varphi(x) = 2x mod\ 1 = x \dotplus x$$

Then, $([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)}, \varphi)$ is measurable dynamical system.

Proof: i) $([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)}, \varphi)$ is measurable dynamical system

Claim 1: φ is measurable

The map $\varphi: [0,1) \to [0,1)$ $is\ define\ by$

$$\varphi(x) = x \dotplus x = \begin{cases} 2x & if\ x \in [0, \frac{1}{2}) \\ 2x - 1 & if\ x \in [\frac{1}{2}, 1) \end{cases}$$

Let $\alpha \in R$. Then

$$\{x \in [0,1) : \varphi(x) \geq \alpha\} = \begin{cases} \emptyset & if \alpha \geq 1 \\ \left[\frac{\alpha}{2}, \frac{1}{2}\right) \cup \left[\frac{\alpha+1}{2}, 1\right) & if\ 0 \leq \alpha < 1 \\ [0,1) & if \alpha < 0 \end{cases}$$

It show that φ is measurable.

Claim2. $\lambda_{[0,1)}$ is invariant with respect to φ.

$$\varphi(x) = \begin{cases} 2x & if\ x \in [0, \frac{1}{2}) \\ 2x - 1 & if\ x \in [\frac{1}{2}, 1) \end{cases}$$

Let $C_1 = \left[0, \frac{1}{2}\right)$ and $C_2 = \left[\frac{1}{2}, 1\right)$

$for\ a\ measurable\ set\ \ A \subset [0,1)$

$$\varphi^{-1}(A) = (\varphi^{-1}(A) \cap C_1) \cup (\varphi^{-1}(A) \cap C_2)$$

But $\varphi^{-1}(A) \cap C_1 = \{x : x \in \varphi^{-1}(A) \cap C_1\}$

$$= \{x : x \in \varphi^{-1}(A)\ and\ x\ \in C_1\}$$

$$= \{x : \varphi(x) \in A\ and\ x\ \in C_1\}$$

$$= \{x : 2x \in A\ \ and\ x\ \in C_1\} = \left\{x : x \in \frac{1}{2}A\ and\ x\ \in C_1\right\}$$

$$= \left\{x : x \in \frac{1}{2}A\ \cap C_1\right\} = \frac{1}{2}A$$

Similarly $\varphi^{-1}(A) \cap C_2 = \{x : x \in \varphi^{-1}(A) \cap C_2\}$

$$= \{x : x \in \varphi^{-1}(A)\ and\ x\ \in C_2\}$$

$$= \{x : \varphi(x) \in A\ and\ x\ \in C_2\}$$

$$= \{x : 2x - 1 \in A\ \ and\ x\ \in C_2\} = \left\{x : x \in \frac{1}{2}A + \frac{1}{2} and\ x\ \in C_2\right\}$$

$$= \left\{x : x \in (\frac{1}{2}A + \frac{1}{2}) \cap C_2\right\} = \frac{1}{2}A + \frac{1}{2}$$

Since Lebesgue measure is translate in variant, so we get

$$\lambda_{[0,1)}(\varphi^{-1}(A)) = \lambda_{[0,1)}(\varphi^{-1}(A) \cap C_1) + \lambda_{[0,1)}(\varphi^{-1}(A) \cap C_2)$$

$$= \lambda_{[0,1)}\left(\frac{1}{2}A\ \cap C_1\right) + \lambda_{[0,1)}\left(\left(\frac{1}{2}A + \frac{1}{2}\right) \cap C_2\right) = \lambda_{[0,1)}\left(\frac{1}{2}A\right) + \lambda_{[0,1)}\left(\frac{1}{2}A\right)$$

$$= \frac{1}{2}\lambda_{[0,1)}(A) + \frac{1}{2}\lambda_{[0,1)}(A)\ = \lambda_{[0,1)}(A)$$

Hence, $\lambda_{[0,1)}$ is invariant with respect to φ.

Therefore, $([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)}, \varphi)$ is measurable dynamical system.

Example 4: Bernoulli schemes

Let $S = \{1,2,\dots,N\}$ $where\ N \geq 2\ and\ let\ p = (p_1, p_2, \dots, p_N), where\ p_j \geq 0$ for each $j \in S\ and\ \sum_{j=1}^{N} p_j = 1$ then vector p defines a probability measure $\mu_0\ on\ S\ via\ \mu_0(\{j\}) = p_j$

Let Ω be the Cartesian product ($\Omega = S^z$) consists of all functions on the integers having values in S or alternatively all doubly infinite sequence of elements of S.

From μ_0 we will construct a probability measure $\mu\ on\ \Omega$. That is, let $(\Omega_k, \beta_k, \mu_k), 1 \leq k \leq n$ be σ-finite measure space and

Let $X_{k=1}^{n}\Omega_k$=$\{(x_1, x_2, \dots, x_n): x_k \in \Omega_k\}$ be the Cartesian product of $\Omega_1, \Omega_2, \dots, \Omega_n$.

Let $X_{k=1}^{n}\beta_k$ be the σ-algebra generated by the n-dimensional measure.

Then there exist a measure $X_{k=1}^{n}\mu_k\ on\ X_{k=1}^{n}\beta_k$ such that

$$(X_{k=1}^{n}\mu_k)(X_{k=1}^{n}E_k) = \prod_{k=1}^{n}(\mu_n)(E_k)$$

Let F be a finite set of integers and **a** be a function from F in to S. Then we define

$C_{F,\boldsymbol{a}} = \{f \in \Omega: f(j) = \boldsymbol{a}(j)\ for\ j \in F\}$

Define by $\mathcal{C}$, the collection of subset of Ω consisting of $\emptyset, \Omega\ and\ sets\ of\ the\ form\ C_{F,\boldsymbol{a}}$.

Claim: $\mathcal{C}$ is a semi algebra of subset of Ω.

Given that $\mathcal{C}$ be the collection of sub set of Ω consisting of $\emptyset, \Omega, and\ all\ sets\ of\ the\ form$ $C_{F,\boldsymbol{a}}$, it implies that $\mathcal{C}$ satisfy the first condition of a semi algebra.

Let $A, B \in \mathcal{C}$ by definition of $\mathcal{C}$, $A\ and\ B\ of\ the\ form\ C_{F,\boldsymbol{a}}$,

For a set A, we have, $C_{F,a} = \{f \in \Omega: f(j) = \boldsymbol{a}(j) for\ j \in F\}$ similarly for a set B we have $C_{F,a} = \{g \in \Omega: g(j) = \boldsymbol{a}(j)\ for\ j \in F\}$.since,$f , g \in \Omega(the\ cartisian\ product)$, now $f\ andg$ can be written as

$$f = (x_1, x_2, \dots, x_n), x_i \in \Omega, for\ each\ i$$

$$g = (y_1, y_2, \dots, y_n), y_i \in \Omega, for\ each\ i$$

Then$f \ \cap g = (x_1, x_2, \dots, x_n) \cap \ (y_1, y_2, \dots, y_n), x_i, y_i \in \Omega, for\ each\ i$

$= (x_1 \cap y_1, x_2 \cap y_2, \dots, x_n \cap y_n), x_i \cap y_i \in \Omega, for\ each\ i$

$= (\prod_{i=1}^{n} f_i) \cap (\prod_{i=1}^{n} g_i)$

$= \prod_{i=1}^{n} (f_i \cap \ g_i)$

Put $h = f \ \cap g, then\ h(j) = (f \ \cap g)(j) = f(j) \cap g(j)$

$= \boldsymbol{a}(j) \cap \boldsymbol{a}(j) = \boldsymbol{a}(j)$

Hence,$\{h \in \Omega: h(j) = \boldsymbol{a}(j)\}$. Therefore $h \in \mathcal{C}$.

$A \in \mathcal{C}\ , then, \mathrm{C}_{F,\boldsymbol{a}} = \{f \in \Omega: f(j) = \boldsymbol{a}(j)\ for\ j \in F\}$

$=> A^c = \{f \in \Omega: f(j) \neq \boldsymbol{a}(j)\ for\ j \in F\} = \emptyset$

Hence $\mathcal{C}$ is semi algebra.

Next we define a set function $\mathcal{J}$ on $\mathcal{C}$ by letting

$$\mathcal{J}(\emptyset) = 0,\ \mathcal{J}\ (\Omega) = 1$$

$$\mathcal{J}(\mathrm{C}_{F,a}) = \prod_{j \in F} \mu_0(\{a(j)\})$$

Claim: $\mathcal{J}$ satisfy

i. If $\{C_k\}_{k=1}^{n}$ is a finite sequence of pair wise disjoint member of $\mathcal{C}$, whose union is in $\mathcal{C}$ then $\mathcal{J}(\cup_{k=1}^{n} C_k) = \sum_{k=1}^{n} \mathcal{J}(C_k)$

Proof:

If $\{C_k\}_{k=1}^{n}$ is a finite sequence of pair wise disjoint member of $\mathcal{C}$, whose union is in $\mathcal{C}$, by definition of $\mathcal{C}$, $C_{F,a} = \{\bigcup_{i=1}^{n} f_i : \bigcup_{i=1}^{n} f_i(j) = \boldsymbol{a}(j)\}$

$\mathcal{J}\,(\bigcup_{k=1}^{n} C_k) = \prod_{j \in F} \mu_0(\{\boldsymbol{a}(j)\}) = \prod_{j \in F} \mu_0(\{\bigcup_{i=1}^{n} f_i(j)\})$

$= \prod_{j \in F} \bigcup_{i=1}^{n} \mu_0\,(f_i(j) = \bigcup_{i=1}^{n} \prod_{j \in F} \mu_0(\{f_i(j)\})$

$= \sum_{i=1}^{n} \prod_{j \in F} \mu_0(\{f_i(j)\}) = \sum_{k=1}^{n} \iota(C_k$

ii. $ii)$ if $c, c_1, c_2, \ldots \in C$ and $C \subset \bigcup_n c_n$ then $\mathcal{J}(c) \le \sum_n \mathcal{J}(C_k)$

$\boldsymbol{proof}$: $c, c_1, c_2, \ldots \in C$ and $C \subset \bigcup_n c_n$

$\mathcal{J}(c) \le \mathcal{J}(\bigcup_n c_n) = \sum_n \mathcal{J}(C_k)$ and $\mathcal{J}\,(c) < \infty$

Consequently, the above conditions are hold then $\mathcal{J}$ extended to be a probably measure μ on a σ $algebra$ β $generated$ by C.

We have a probability space (Ω, β, μ).

Next we define the function

$$\varphi: \Omega \to \Omega \; by \; \varphi(f)(j) = f(j+1)$$

If we consider the elements of Ω doubly infinite sequences then the effect of φ is to move each term of a sequence f one place to the left. For this reason, the mapping φ is often called a Bernoulli shift.

now, for $F^* = \{j+1 : j \in F\}$ and $\boldsymbol{a}^*(j+1) = \boldsymbol{a}(j)$,

$$\varphi^{-1}(C_{\boldsymbol{F},\boldsymbol{a}}) = \{f \in \Omega : \varphi(f)(j) \in C_{F,\boldsymbol{a}}\}$$

$$= \{f \in \Omega : f(j+1) \in C_{F,\boldsymbol{a}}\}$$

$$= \{f \in \Omega : f(j+1) = \boldsymbol{a}(j), j \in F\}$$

$$= \{f \in \Omega : f(j+1) = \boldsymbol{a}^*(j+1), j+1 \in F^*\} = C_{F^*,a^*}$$

Therefore, $\varphi^{-1}(C_{F,\boldsymbol{a}}) = C_{F^*,\boldsymbol{a}^*}$. It follows that the σ-algebra

$\{A \subset \Omega: \varphi^{-1}(A) \in \beta\}$ contains C and hence $\varphi^{-1}(A) \in \beta for\ each\ A \in \beta$

Claim: $\mu\ is\ invariant\ with\ respect\ to\ \varphi$

Let the measure v be defined on$\beta\ by\ v(A) = \mu(\varphi^{-1}(A))$

$$v(\mathrm{C}_{F,a}) = \mu\left(\varphi^{-1}(\mathrm{C}_{F,a})\right) = \mu(\mathrm{C}_{F^*,a^*})$$

$$= \prod_{j \in F^*} \mu_0\ (a^*(j))$$

$$= \mu(\mathrm{C}_{F,a})$$

Thus,$v\ and\ \mu\ agree\ on\ c$.Hence $v = \mu$. This means that μ is invariant with respect to φ. there for $(\Omega, \beta, \mu, \varphi)$ is a measurable dynamical system.

This system is known in the literature as a Bernoulli scheme and is often de$noted\ by\ \ B(p_1, p_2, \dots, p_N)$.

PART TWO: ENTROPY

2.1. Isomorphism of Measurable Dynamical System

Definition 5: Two measurable dynamical systems $(\Omega, \beta, \mu, \varphi)$ and (Λ, S, v, ψ) are said to be isomorphic if there are mapping $J: \Omega \to \Lambda and\ K: \Lambda \to \Omega$ such that:

i. $J^{-1}(B) \in \beta\ for\ each\ B \in S,$
ii. $K^{-1}(A) \in S\ for\ each\ A \in \beta,$
iii. $\mu\big(J^{-1}(B)\big) = v(B) for\ each\ B \in S,$
iv. $v\big(K^{-1}(A)\big) = \mu(A) for\ each\ A \in \beta,$
v. $J \circ \varphi = \psi \circ J\ \mu\ ae,$
vi. $K \circ \psi = \varphi \circ K\ v\ ae,$
vii. $K \circ J(x) = x\ \mu\ ae,$
viii. $J \circ K(y) = y\ \ v\ ae,$

$each\ of\ the\ mapping\ J\ and\ K\ is\ called\ an\ isomorphism.$

***Example* 5:** example (1) and (2) **,**in the above, are isomorphic.

Given two measurable dynamical systems, determine whether they are isomorphic. An invariant of a measurable dynamical system$(\Omega, \beta, \mu, \varphi)$ is a number or property, $I(\Omega, \beta, \mu, \varphi)$,such that if $(\Omega, \beta, \mu, \varphi)$ and(Λ, S, v, ψ) are isomorphic ,then $I(\Omega, \beta, \mu, \varphi) and\ I(\Lambda, S, v, \psi)$ are identical.

2.2. Entropy

Let (Ω, β, μ) be probability space. Suppose that the distribution of the location of a partition p, in Ω is given by the probability measure μ. that is, for each $A \in \beta$, the probability that p is in A equals $\mu(A)$. The object of the experiment is to locate the position of p as closely as possible.

Definition 6: Let (Ω, β, μ) be a probability space. A measurable partition $\mathfrak{B}$ of Ω is a finite collection of sets $\{A_1, A_2, \dots, A_n\}$ such that

i) $A_k \in \beta \; for \; all \; k$.

ii) $A_k \cap A_j = \emptyset \; for \; i \neq k$

iii) $\bigcup_{k=1}^{n} A_k = \Omega$.

Let $\mathfrak{B}$ be a measurable partition of (Ω, β). Some partitions tell us more than others about the location of p. for example, for the probability space $([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)})$, we expect more information from $\mathfrak{B} = \left\{\left[0, \frac{1}{2}\right), [\frac{1}{2}, 1)\right\} \; than \; \mathcal{D} = \left\{\left[0, \frac{1}{100}\right), [\frac{1}{100}, 1)\right\}$. This is because we are guaranteed that $\mathfrak{B}$ will reduce by 50% the measure of the set where we have to look for p, whereas $\mathcal{D}$ will reduce it by only 1% . The number of information gained from a measurable partition is called the entropy of the measurable partition.

Definition 7: let (Ω, β, μ) be a probability space and $\mathfrak{B}$ a measurable partition of (Ω, β)

a) A numerical value $I(A)$, the information contained in the event p is in A is

$$I(A) = -\log \mu(A)$$

b) The information function of a measurable partition is

$$I(\mathfrak{B}) = \sum_{A \in \mathfrak{B}} I(A) x_A = -\sum_{A \in \mathfrak{B}} \log \mu(A) x_A$$

$c)$ $the \; entr$opy of $\mathfrak{B}$ is the average of the information contents of element, denoted by $H(\mathfrak{B})$,

$$H(\mathfrak{B}) = -\sum_{A \in \mathfrak{B}} \mu(A) \log \mu(A)$$

$$where \; we \; use \; the \; convention \; that \; 0 \log 0 = 0$$

The entropy of a two element partition in the probability space $([0,1), \mathcal{M}_{[0,1)}, \lambda_{[0,1)})$

$$H(\{A, A^C\}) = -\lambda(A) \log \lambda(A) - (1 - \lambda(A)) \log(1 - \lambda(A))$$

is maximized when $\lambda(A) = \lambda(A)^C) = \frac{1}{2}$.

Definition 8: Let $\mathfrak{B}$ and $\mathcal{D}$ be two measurable partitions of (Ω, β) in the probability space (Ω, β, μ) then the conditional entropy $H(\mathfrak{B} \backslash \mathcal{D})$ of $\mathfrak{B}$ is given by

$$H(\mathfrak{B} \backslash \mathcal{D}) = -\sum_{B \in \mathcal{D}} \mu(B) \sum_{A \in \mathfrak{B}} \frac{\mu(A \cap B)}{\mu(B)} \log \frac{\mu(A \cap B)}{\mu(B)} \quad where\ \mu(A \backslash B) = \frac{\mu(A \cap B)}{\mu(B)}.$$

We say that a measurable partition $\mathcal{D}$ is a **refinement** of the measurable partition $\mathfrak{B}$ and write $\mathfrak{B} \ll \mathcal{D}$ if every element of $\mathfrak{B}$ is a union of elements of $\mathcal{D}$**.** For any two measurable partition $\mathfrak{B}$ and $\mathfrak{N}$, there is a smallest **common refinement** given by

$$\mathfrak{B} \vee \mathfrak{N} = \{A \cap B : A \in \mathfrak{B}, \mathrm{B} \in \mathfrak{N}\}.$$

Proposition1:

Let (Ω, β, μ) be a probability space and $\mathfrak{B}$, $\mathcal{D}$ and $\mathfrak{N}$ be measurable partitions of (Ω, β).then the following hold:

a) If $\mathfrak{B} \ll \boldsymbol{\mathcal{D}}$ ***then*** $\boldsymbol{H}(\mathfrak{B}) \leq \boldsymbol{H}(\mathcal{D})$
b) $H(\mathfrak{B} \vee \mathfrak{N}) \leq H(\mathfrak{B}) + H(\mathfrak{N})$

Proof: a) Suppose $\mathfrak{B} \ll \mathcal{D}$**,** by definition of refinement each $A \in \mathfrak{B}$ is a disjoint union of members of $\mathcal{D}$. Thus for $\mu(A) > 0$, we have

$$-\mu(A) \log \mu(A) = -\sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(A)$$

$$= -\sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(A) + \sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B) - \sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B)$$

$$= -\sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B) + \sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \frac{\mu(B)}{\mu(A)}$$

$$\leq -\sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B) + \sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \frac{\mu(A)}{\mu(A)}$$

$$\leq -\sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B)$$

Summing over $A \in \mathfrak{B}$ we obtain

$$H(\mathfrak{B}) = -\sum_{A \in \mathfrak{B}} \mu(A) \log \mu(A) \leq -\sum_{A \in \mathfrak{B}} \sum_{\substack{B \subset A \\ B \in \mathcal{D}}} \mu(B) \log \mu(B) = H(\mathcal{D})$$

b) Define the function $g(t) = -t\log t$ and is concave on [0,1]; that is, g satisfies

$$\sum_{j=1}^{n} c_j\, g(t_j) \leq g(\sum_{j=1}^{n} c_j\, t_j) \tag{1}$$

for all convex combinations of elements of [0,1]. Without loss of generality we can assume that $\mu(C) > 0 for\ all\ C \in \mathfrak{N}$. Thus, we can write

$$\mu(A) = \sum_{C\in\mathfrak{N}} \left(\frac{\mu(A\cap C)}{\mu(C)}\right)\mu(C) \tag{2}$$

For each $A \in \mathfrak{B}$. it follows from (1) and (2) that

$$g\left(\sum_{j=1}^{n} c_j\, t_j\right) = g(\sum_{C\in\mathfrak{N}} \mu(C)(\frac{\mu(A\cap C)}{\mu(C)}))$$

$$= -\sum_{C\in\mathfrak{N}} \mu(C)(\frac{\mu(A\cap C)}{\mu(C)}) \log \sum_{C\in\mathfrak{N}} \mu(C)(\frac{\mu(A\cap C)}{\mu(C)})$$

$$= -\mu(A)\log\mu(A) \geq \sum_{C\in\mathfrak{N}} \mu(C) g\left(\frac{\mu(A\cap C)}{\mu(C)}\right) = -\sum_{C\in\mathfrak{N}} \mu(C)\frac{\mu(A\cap C)}{\mu(C)}\log\frac{\mu(A\cap C)}{\mu(C)}$$

$$-\mu(A)\log\mu(A) \geq -\sum_{C\in\mathfrak{N}} \mu(A\cap C)\log(\frac{\mu(A\cap C)}{\mu(C)})$$

$$= -\sum_{C\in\mathfrak{N}} \mu(A\cap C)\log\mu(A\cap C) + \sum_{C\in\mathfrak{N}} \mu(A\cap C)\log\mu(C)$$

Summing over $A \in \mathfrak{B}$ we get

$$H(\mathfrak{B}) \geq -\sum_{A\in\mathfrak{B}}\sum_{C\in\mathfrak{N}} \mu(A\cap C)\log\mu(A\cap C) + \sum_{C\in\mathfrak{N}}\sum_{A\in\mathfrak{B}} \mu(A\cap C)\log\mu(C)$$

$$= -\sum_{A\in\mathfrak{B}}\sum_{C\in\mathfrak{N}} \mu(A\cap C)\log\mu(A\cap C) + \sum_{C\in\mathfrak{N}} \mu(C)\log\mu(C)$$

$$= H(\mathfrak{B}\vee\mathfrak{N}) - H(\mathfrak{N})$$

Thus, $H(\mathfrak{B}\vee\mathfrak{N}) \leq H(\mathfrak{B}) + H(\mathfrak{N})$.

2.3. Entropy and measurable dynamical systems

Suppose the particle p to move according to the following rule: if p is at x at time 0, then its position at time 1 is $\varphi(x)$, its position at time 2 is $\varphi(\varphi(x))$,etc

If we use a measurable partition $\mathfrak{B}$ to obtain information about the location of p at time 0, then the measurable partition

$$\varphi^{-n}\mathfrak{B} = \left\{\left(\varphi^{(n)}\right)^{-1}(A) : A \in \mathfrak{B}\right\}$$

yields corresponding information about the particle's location at time n, and the measurable partition

$$\mathfrak{B}^{(n)} = \mathfrak{B} \vee \varphi^{-1}\mathfrak{B} \vee \ldots \vee \varphi^{-(n-1)}\mathfrak{B}$$

yields corresponding information about the path of successive positions of p at times 0 through n-1 as it moves in Ω under the action of φ.

Proposition 2:

Let $(\Omega, \beta, \mu, \varphi)$ be a measurable dynamical system and $\mathfrak{B}$ a measurable partition of (Ω, β). Then the following hold:

a) $H(\varphi^{-k}\mathfrak{B}) = H(\mathfrak{B})$
b) $H\big((\varphi^{-k}\mathfrak{B})^{(n)}\big) = H(\mathfrak{B}^{(n)})$
c) $H\big(\mathfrak{B}^{(n+m)}\big) \leq H\big(\mathfrak{B}^{(n)}\big) + H(\mathfrak{B}^{(m)})$

Proof: a) Using the definition of entropy of a measurable partition and the invariance ofμ $with$h respect to φ.

$$H(\mathfrak{B}) = -\sum_{A\in\mathfrak{B}} \mu(A) \log \mu(A)$$

$$= -\sum_{A\in\mathfrak{B}} \mu(\varphi^{-k}A) \log \mu(\varphi^{-k}A)$$

$$= H(\varphi^{-k}\mathfrak{B})$$

Therefore, $H(\mathfrak{B}) = H(\varphi^{-k}\mathfrak{B})$

b)$H\big(\mathfrak{B}^{(n)}\big) = -\sum_{A\in\mathfrak{B}^{(n)}} \mu(A) \log \mu(A)$ but $\mathfrak{B}^{(n)} = \mathfrak{B} \vee \varphi^{-1}\mathfrak{B} \vee \ldots \vee \varphi^{-(n-1)}\mathfrak{B}$

$$= -\sum_{A_i\in\mathfrak{B}} \mu(A_1 \cap \varphi^{-1}A_2 \cap \ldots \cap \varphi^{-(n-1)}A_n) \log \mu(A_1 \cap \varphi^{-1}A_2 \cap \ldots \cap \varphi^{-(n-1)}A_n)$$

$$= -\sum_{A_i\in\mathfrak{B}} \mu(\textstyle\bigcap_{i=0}^{n-1} \varphi^{-i}A_{i+1}) \log \mu(\textstyle\bigcap_{i=0}^{n-1} \varphi^{-i}A_{i+1})$$

$$= -\sum_{A_i\in\mathfrak{B}} \mu(\varphi^{-k}(\textstyle\bigcap_{i=0}^{n-1} \varphi^{-i}A_{i+1})) \log \mu(\varphi^{-k}(\textstyle\bigcap_{i=0}^{n-1} \varphi^{-i}A_{i+1})) \quad \text{(by a)}$$

$$= -\sum_{A_i \in \mathfrak{B}} \mu(\bigcap_{i=0}^{n-1} \varphi^{-i}(\varphi^{-k} A_{i+1})) \log \mu(\bigcap_{i=0}^{n-1} \varphi^{-i}(\varphi^{-k} A_{i+1}))$$

$$= -\sum_{B_i \in \varphi^{-k}\mathfrak{B}} \mu(\bigcap_{i=0}^{n-1} \varphi^{-i} B_{i+1}) \log \mu(\bigcap_{i=0}^{n-1} \varphi^{-i} B_{i+1})$$

$$= -\sum_{B_i \in (\varphi^{-k}\mathfrak{B})^{(n)}} \mu(\varphi^{-i} B_{i+1}) \log \mu(\varphi^{-i} B_{i+1})$$

$$= -\sum_{B_i \in (\varphi^{-k}\mathfrak{B})^{(n)}} \mu(B_{i+1}) \log \mu(B_{i+1})$$

$$= H((\varphi^{-k}\mathfrak{B})^{(n)})$$

Therefore, $H(\mathfrak{B}^{(n)}) = H((\varphi^{-k}\mathfrak{B})^{(n)})$

c) $H(\mathfrak{B}^{(n+m)}) = \mathfrak{B} \vee \varphi^{-1}\mathfrak{B} \vee \dots \vee \varphi^{-(n-1)}\mathfrak{B} \vee \varphi^{-n}\mathfrak{B} \vee \dots \vee \varphi^{-(n+m-1)}\mathfrak{B}$

$$= \mathfrak{B}^{(n)} \vee (\varphi^{-n}\mathfrak{B} \vee \varphi^{-(n+1)}\mathfrak{B} \vee \dots \vee \varphi^{-(n+m-1)}\mathfrak{B})$$

$$= \mathfrak{B}^{(n)} \vee (\varphi^{-n}\mathfrak{B} \vee \varphi^{-1}(\varphi^{-n}\mathfrak{B}) \vee \dots \vee \varphi^{-(m-1)}(\varphi^{-n}\mathfrak{B}))$$

$$= \mathfrak{B}^{(n)} \vee (\varphi^{-n}\mathfrak{B})^{(m)}$$

Thus, $H(\mathfrak{B}^{(n+m)}) = \mathfrak{B}^{(n)} \vee (\varphi^{-n}\mathfrak{B})^{(m)}$

It follows from proposition 1 that

$$H(\mathfrak{B}^{(n+m)}) \le H(\mathfrak{B}^{(n)}) + H((\varphi^{-n}\mathfrak{B})^{(m)}) = H(\mathfrak{B}^{(n)}) + H((\mathfrak{B})^{(m)})$$

Lemma 1: Suppose that $\{a_n\}_{n=1}^{\infty}$ is a sequences of real numbers satisfying the sub additivity condition $a_{n+m} \le a_n + a_m$ .Then the $\lim_{n\to\infty} \frac{a_n}{n}$ exists as a real number, or, possibly, $-\infty$.

Proof: Let $m \in N$ be fixed, but arbitrary. Each $n \in N$ can be written as $n = lm + r$,

$where\ l \ge 0\ and\ 0 \le r < m. thus\ a_n \le la_m + a_r.$

Dividing by n we get that

$$\frac{a_n}{n} \le \frac{la_m}{n} + \frac{a_r}{n} = \frac{mla_m}{nm} + \frac{a_r}{n} = \frac{(n-r)}{nm} a_m + \frac{a_r}{n} = \frac{a_m}{m} - \frac{r a_m}{nm} + \frac{a_r}{n}$$

Whence $\lim_{n\to\infty} sup\frac{a_n}{n} \leq \frac{a_m}{m}\ for\ all\ m.$ and $\lim_{n\to\infty} sup\frac{a_n}{n} \leq \lim_{n\to\infty} inf\frac{a_n}{n}$

Hence,$\lim_{n\to\infty} sup\frac{a_n}{n} = \lim_{n\to\infty} inf\frac{a_n}{n}$. Therefore, $\lim_{n\to\infty}\frac{a_n}{n}$ exists

Note: Using proposition 2(c) and lemma1, the limit

$$H(\mathfrak{B},\varphi) = \lim_{n\to\infty}\frac{H(\mathfrak{B}^{(n)})}{n} \text{ exists.}$$

$H(\mathfrak{B},\varphi)$ be the time average for the entropies associated with the measurable partitions $\mathfrak{B}^{(n)}$ the quantity

$$h(\varphi) = sup\{H(\mathfrak{B},\varphi) : \mathfrak{B}\ a\ partition\ of\ (\Omega,\beta)\}$$

which can be viewed as the maximum amount of information that can be extracted from the dynamical system per time, is called the entropy of the measurable dynamical system$(\Omega,\beta,\mu.\varphi)$.

2.4.The Kolmogrove- Sinai Theorem

Proposition 3

Let $(\Omega,\beta,\mu,\varphi)$ be a measurable dynamical system and let $\mathfrak{B}$, $\boldsymbol{\mathcal{D}}$ and $\mathfrak{N}$ be measurable partition of(Ω,β). Then the following hold:

a) $H(\mathfrak{B}\backslash\boldsymbol{\mathcal{D}}) \leq \boldsymbol{H}(\mathfrak{B})$
b) $H(\mathfrak{B}\vee\mathcal{D}) = H(\mathcal{D}) + H(\mathfrak{B}\backslash\boldsymbol{\mathcal{D}})$
c) $H(\mathfrak{B}\vee\mathcal{D}\backslash\mathfrak{N}) \leq H(\mathfrak{B}\backslash\mathfrak{N}) + H(\mathcal{D}\backslash\mathfrak{N})$
d) $If\ \mathfrak{B}\ll\mathcal{D}\ \ then\ H(\mathfrak{B}\backslash\mathfrak{N}) \leq H(\mathcal{D}\backslash\mathfrak{N})$
e) $\boldsymbol{If\ \mathcal{D}\ll\mathfrak{N}\ then}\ H(\mathfrak{B}\backslash\mathfrak{N}) \leq H(\mathfrak{B}\backslash\boldsymbol{\mathcal{D}})$
f) $H(\varphi^{-1}\mathfrak{B}\backslash\varphi^{-1}\boldsymbol{\mathcal{D}}) = H(\mathfrak{B}\setminus\boldsymbol{\mathcal{D}})$
g) $H(\mathcal{D},\varphi) \leq H(\boldsymbol{\mathcal{D}}\setminus\mathfrak{B}) + H(\mathfrak{B},\varphi)$

Proof:

a) $H(\mathfrak{B} \setminus \boldsymbol{\mathcal{D}}) = -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(B)\mu(A\backslash B)\log\mu(A\backslash B)$

$$= -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)}$$

$$= -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(A\cap B)\log\mu(A\cap B) + \sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(A\cap B)\log\mu(B)$$

$$= -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(A\cap B)\log\mu(A\cap B) + \sum_{B\in\mathcal{D}}\mu(B)\log\mu(B)$$

$$= H(\mathfrak{B}\vee\mathcal{D}) - H(\mathcal{D}) \qquad (*)$$

$$\leq H(\mathfrak{B}) + H(\mathcal{D}) - H(\mathcal{D}) = H(\mathfrak{B})$$

$Therefore, H(\mathfrak{B} \setminus \boldsymbol{\mathcal{D}}) \leq H(\mathfrak{B})$

b) From $(*)$ in the proof of (a), we have

$H(\mathfrak{B} \setminus \boldsymbol{\mathcal{D}}) = H(\mathfrak{B}\vee\mathcal{D}) - H(\mathcal{D})$

$Therefore, H(\mathfrak{B}\vee\mathcal{D}) = H(\mathfrak{B} \setminus \boldsymbol{\mathcal{D}}) + H(\mathcal{D})$

c) By definition of common refinement, we get

$(\mathfrak{B}\vee\mathcal{D})\backslash\mathfrak{N} = \{(A\cap B)\backslash C: A\in\mathfrak{B}, B\in\mathcal{D}, C\in\mathfrak{N}\}$

$= \{(A\backslash C)\cap(B\backslash C): A\in\mathfrak{B}, B\in\mathcal{D}, C\in\mathfrak{N}\}$

$= (\mathfrak{B}\backslash\mathfrak{N})\vee(\mathcal{D}\backslash\mathfrak{N})$

Then, $H(\mathfrak{B}\vee\mathcal{D})\backslash\mathfrak{N} = H((\mathfrak{B}\backslash\mathfrak{N})\vee(\mathcal{D}\backslash\mathfrak{N}))$

By proposition (1), we have,

$H((\mathfrak{B}\vee\mathcal{D})\backslash\mathfrak{N}) \leq H(\mathfrak{B}\backslash\mathfrak{N}) + H(\mathcal{D}\backslash\mathfrak{N})$ ■

d) Suppose $\mathfrak{B}\ll\boldsymbol{\mathcal{D}}$, that is, each $A\in\mathfrak{B}$ is a disjoint union of members of $\boldsymbol{\mathcal{D}}$, for $\mu(C) > 0$, we have

$-\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)} = -\sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}$

$$= -\sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(A\cap C)}{\mu(C)} + \sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)} - \sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)}$$

$= -\sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)} + \sum_{B\in\mathcal{D}}\mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(A\cap C)}$

$\leq -\sum_{B\in\mathcal{D}} \mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)} + \sum_{B\in\mathcal{D}} \mu(B\cap C)\log\frac{\mu(A\cap C)}{\mu(A\cap C)}$

$= -\sum_{B\in\mathcal{D}} \mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)}$

Summing over $A \in \mathfrak{B}$ and $C \in \mathfrak{N}$

$$H(\mathfrak{B}\backslash\mathfrak{N}) = -\sum_{C\in\mathfrak{N}}\sum_{A\in\mathfrak{B}} \mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}$$

$$\leq -\sum_{C\in\mathfrak{N}}\sum_{A\in\mathfrak{B}}\sum_{B\in\mathcal{D}} \mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)}$$

$$\leq -\sum_{C\in\mathfrak{N}}\sum_{B\in\mathcal{D}} \mu(B\cap C)\log\frac{\mu(B\cap C)}{\mu(C)} = H(\mathcal{D}\backslash\mathfrak{N})$$

Therefore, $H(\mathfrak{B}\backslash\mathfrak{N}) \leq H(\mathcal{D}\backslash\mathfrak{N})$ ∎

e) Suppose $\boldsymbol{\mathcal{D}} \ll \boldsymbol{\mathfrak{N}}$, by definition, each element $B \in \boldsymbol{\mathcal{D}}$ is a disjoint union of elements of $\mathfrak{N}$. Define a concave function $g(t) = -t\log t$ and satisfy

$g(\sum_{i=1}^{n} c_i t_i) \geq \sum_{i=1}^{n} c_i g(t_i)$, where $\sum_{i=1}^{n} c_i = 1$

$g\left(\sum_{C\in\mathfrak{N}}\frac{\mu(C)}{\mu(B)}\frac{\mu(A\cap C)}{\mu(C)}\right) \geq \sum_{C\in\mathfrak{N}}\frac{\mu(C)}{\mu(B)} g\left(\frac{\mu(A\cap C)}{\mu(C)}\right)$

$g\left(\sum_{C\in\mathfrak{N}}\frac{\mu(A\cap C)}{\mu(B)}\right) \geq -\sum_{C\in\mathfrak{N}}\frac{\mu(C)}{\mu(B)}\frac{\mu(A\cap C)}{\mu(C)}\log\frac{\mu(A\cap C)}{\mu(C)}$

$g(\frac{1}{\mu(B)}\sum_{C\in\mathfrak{N}}\mu(A\cap C)) \geq -\sum_{C\in\mathfrak{N}}\frac{\mu(A\cap C)}{\mu(B)}\log\frac{\mu(A\cap C)}{\mu(C)}$

$-\frac{\mu(A\cap B)}{\mu(B)}\log\frac{\mu(A\cap B)}{\mu(B)} \geq -\frac{1}{\mu(B)}\sum_{C\in\mathfrak{N}}\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}$

$multiplying\ by\ \mu(B) and\ summing\ over\ A \in \mathfrak{B}, B \in \boldsymbol{\mathcal{D}}$

$$-\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)} \geq -\sum_{C\in\mathfrak{N}}\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}$$

$$-\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}}\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)} \geq -\sum_{C\in\mathfrak{N}}\sum_{A\in\mathfrak{B}}\sum_{B\in\mathcal{D}}\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)}$$

$$H(\mathfrak{B}\backslash\mathcal{D}) = \sum_{B\in\mathcal{D}}(-\sum_{C\in\mathfrak{N}}\sum_{A\in\mathfrak{B}}\mu(A\cap C)\log\frac{\mu(A\cap C)}{\mu(C)})$$

$$H(\mathfrak{B}\backslash\mathcal{D}) \ = \textstyle\sum_{B\in\mathcal{D}} H(\mathfrak{B}\backslash\mathfrak{N})$$

$$\geq \ H(\mathfrak{B}\backslash\mathfrak{N})$$

Hence,$H(\mathfrak{B}\backslash\mathcal{D}) \ \geq \ H(\mathfrak{B}\backslash\mathfrak{N})$, therefore , $H(\mathfrak{B}\backslash\mathfrak{N}) \leq \ H(\mathfrak{B}\backslash\mathcal{D})$ ■

f) $H(\varphi^{-1}\mathfrak{B}\backslash\varphi^{-1}\mathcal{D}) \ = -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}} \mu(\varphi^{-1}A\cap\varphi^{-1}B)\log\frac{\mu(\varphi^{-1}A\cap\varphi^{-1}B)}{\mu(\varphi^{-1}B)}$

$$= -\sum_{B\in\mathcal{D}}\sum_{A\in\mathfrak{B}} \mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)}$$

$$= H(\mathfrak{B}\backslash\mathcal{D})$$

Therefore,$H(\varphi^{-1}\mathfrak{B}\backslash\varphi^{-1}\mathcal{D}) = H(\mathfrak{B}\backslash\mathcal{D})$ ■

By proposition (1.1) and (b) and(c), we have

$$H\big(\mathcal{D}^{(n)}\big) = -\textstyle\sum_{B\in\mathcal{D}^{(n)}} \mu(B)\log\mu(B)\,, but\ A\cap B\subset B$$

$$\leq -\textstyle\sum_{B\in\mathcal{D}^{(n)}}\sum_{A\in\mathfrak{B}^{(n)}} \mu(B\cap A)\log\mu(B\cap A)$$

$$= H\big(\mathfrak{B}^{(n)}\vee\mathcal{D}^{(n)}\big)$$

Then, we have $H(\mathcal{D}^{(n)}) \leq H(\mathfrak{B}^{(n)}\vee\mathcal{D}^{(n)})$ (1)

And, $\mathcal{D}^{(n)}\backslash\mathfrak{B}^{(n)} = \mathcal{D}\vee\varphi^{-1}\mathcal{D}\vee\ldots\vee\varphi^{-(n-1)}\mathcal{D}\backslash\mathfrak{B}^{(n)}$

$$= \big\{A_1\cap\varphi^{-1}A_2\cap\varphi^{-2}A_3\cap\ldots\cap\varphi^{-(n-1)}A_n\backslash B\colon A_i\in\mathcal{D}\ and\ B\in\mathfrak{B}^{(n)}\big\}$$

$$= \big\{(A_1\cap\varphi^{-1}A_2\cap\varphi^{-2}A_3\cap\ldots\cap\varphi^{-(n-1)}A_n)\cap B^C\colon\ A_i\in\mathcal{D}, B\in\mathfrak{B}^{(n)}\big\}$$

$$= \big\{(A_1\cap B^c)\cap(\varphi^{-1}A_2\cap B^c)\cap\ldots\cap\big(\varphi^{-(n-1)}A_n\cap B^c\big)\colon\ A_i\in\mathcal{D}, B\in\mathfrak{B}^{(n)}\big\}$$

$$= \big\{(A_1\backslash B^c)\cap((\varphi^{-1}A_2\backslash B^c)\cap\ldots\cap((\varphi^{-(n-1)}A_n\backslash B^c)\big\}$$

$$= \big(\mathcal{D}/\mathfrak{B}^{(n)}\big)\vee\big(\varphi^{-1}\mathcal{D}/\mathfrak{B}^{(n)}\big)\vee\ldots\vee\big(\varphi^{-(n-1)}\mathcal{D}/\mathfrak{B}^{(n)}\big)$$

$$H\big(\mathcal{D}^{(n)}/\mathfrak{B}^{(n)}\big) = H(\big(\mathcal{D}/\mathfrak{B}^{(n)}\big)\vee\big(\varphi^{-1}\mathcal{D}/\mathfrak{B}^{(n)}\big)\vee\ldots\vee\big(\varphi^{-(n-1)}\mathcal{D}/\mathfrak{B}^{(n)}\big))$$

$$\leq H\big(\mathcal{D}/\mathfrak{B}^{(n)}\big) + H\big(\varphi^{-1}\mathcal{D}/\mathfrak{B}^{(n)}\big) + \cdots + H\big(\varphi^{-(n-1)}\mathcal{D}/\mathfrak{B}^{(n)}\big)$$

$$= \sum_{j=0}^{n-1} H(\varphi^{-j}\, \mathcal{D}/\mathfrak{B}^{(n)}) \qquad (2)$$

From (1) and (2), we get

$$H(\mathcal{D}^{(n)}) \leq H(\mathfrak{B}^{(n)} \vee \mathcal{D}^{(n)}) = H(\mathfrak{B}^{(n)}) + H(\mathcal{D}^{(n)}/\mathfrak{B}^{(n)})$$

$$\leq H(\mathfrak{B}^{(n)}) + \sum_{j=0}^{n-1} H(\varphi^{-j}\, \mathcal{D}/\mathfrak{B}^{(n)})$$

$$\leq H(\mathfrak{B}^{(n)}) + \sum_{j=0}^{n-1} H(\varphi^{-j}\, \mathcal{D}/\mathfrak{B}^{(n)})$$

$$\leq H(\mathfrak{B}^{(n)}) + \sum_{j=0}^{n-1} H(\varphi^{-j}\, \mathcal{D}/\varphi^{-j}\, \mathfrak{B})$$

$$= H(\mathfrak{B}^{(n)}) + \sum_{j=0}^{n-1} H(\mathcal{D}\backslash\mathfrak{B})$$

$$= H(\mathfrak{B}^{(n)}) + nH(\mathcal{D}\backslash\mathfrak{B})$$

But $H(\mathcal{D},\varphi) = \lim_{n\to\infty} \frac{H(\mathcal{D}^{(n)})}{n} \leq \lim_{n\to\infty} \frac{1}{n}(H(\mathfrak{B}^{(n)}) + n\, H(\mathcal{D}\backslash\mathfrak{B}))$

$$= \lim_{n\to\infty} \frac{H(\mathfrak{B}^{(n)})}{n} + H(\mathcal{D}\backslash\mathfrak{B})$$

$$= H(\mathfrak{B},\varphi) + H(\mathcal{D}\backslash\mathfrak{B})$$

Therefore, $H(\mathcal{D},\varphi) \leq H(\mathfrak{B},\varphi) + H(\mathcal{D}\backslash\mathfrak{B})$. ■

Lemma 2:

Let (Ω, β, μ) be a probability space, $\mathcal{F} \subset \beta$ an algebra, and ε the smallest σ-algebra containing $\mathcal{F}$, then given $E \in \varepsilon$ *and* $\delta > 0$, *there exsts an* $F \in \mathcal{F}$ *such that* $\mu(E \bigtriangleup F) < \delta$.

Proof: Let $\mathcal{G}$ denote the collection of all $G \in \beta$ having the property that there is a sequence $\{F_n\}_{n=1}^{\infty} \subset \mathcal{F}$ such that $\lim_{n\to\infty} \mu(G \bigtriangleup F_n) = 0$.

Claim 1:

I) $\mathcal{G}$ is an algebra of a set.

Let $G \in \mathcal{G}$ hence β $is\ a\ \sigma$- algebra and $\mathcal{F} \subset \beta$. Then $G^c \in \beta$ and $F_n \in \mathcal{F}$ for each n such that $\lim_{n\to\infty} \mu(G \,\Delta\, F_n) = 0$.

Let $G_n = {F_n}^c$ then $\{G_n\}_{n=1}^{\infty} = \{{F_n}^c\}_{n=1}^{\infty} \subset \mathcal{F}$. Since $\mathcal{F}$ is an algebra.

Now, $\mu(G^c \vartriangle G_n) = \mu((G^c/G_n) \cup (G_n/G^c))$

$$= \mu\big((G^c \cap {G_n}^c) \cup (G_n \cap G)\big)$$

$$= \mu((G^c \cap F_n) \cup ({F_n}^c \cap G) \ = \mu(G \vartriangle F_n)$$

$\lim_{n\to\infty} \mu(G^c \vartriangle G_n) = \lim_{n\to\infty} \mu(G \vartriangle F_n) = 0$

Hence, for $G^c \in \beta$,there is a sequence $\{G_n\} \subset \mathcal{F}$ such that

$\lim_{n\to\infty} \mu(G^c \vartriangle G_n) = 0$, Thus $G^c \in \mathcal{G}$

II) Let $A, B \in \mathcal{G}$. That is, for all $A, B \in \beta$ there is a sequence $\{F_n\}_{n=1}^{\infty}, \{E_n\}_{n=1}^{\infty} \subset \mathcal{F}$ such that $\lim_{n\to\infty} \mu(A \vartriangle F_n) = 0$ and $\lim_{n\to\infty} \mu(B \vartriangle E_n) = 0$. Since β is algebra. So,$A \cup B \in \beta$.

Now, $\big((A \cup B)\Delta(F_n \cup E_n)\big) = ((A \cup B) \cap ({F_n}^c \cap {E_n}^c)) \cup ((A^c \cap B^c) \cap (F_n \cup E_n)$

$$= \Big(\big(A \cap ({F_n}^c \cap {E_n}^c)\big) \cup \big(B \cap ({F_n}^c \cap {E_n}^c)\big)\Big) \cup$$

$$(((A^c \cap B^c) \cap F_n) \cup ((A^c \cap B^c) \cap E_n)$$

$$\subseteq (A \cap {F_n}^c) \cup (B \cap {E_n}^c) \cup (A^c \cap F_n) \cup (B^c \cap E_n)$$

$$= (A \backslash F_n) \cup (F_n \backslash A) \cup (B \backslash E_n) \cup (E_n \backslash B)$$

$$= (A \,\Delta\, F_n) \cup (B \,\Delta\, E_n)$$

$=> \mu\big((A \cup B) \vartriangle (F_n \cup E_n)\big) \leq \mu(A \,\Delta\, F_n) + \mu(B \,\Delta\, E_n)$

Therefore,$\lim_{n\to\infty} \mu\big((A \cup B) \vartriangle (F_n \cup E_n)\big) = 0$

Therefore, for $A \cup B \in \beta$ there is a sequence $\{F_n\} \cup \{E_n\} \subset \mathcal{F}$ such that

$\lim_{n\to\infty} \mu\big((A \cup B) \vartriangle (F_n \cup E_n)\big) = 0$. Hence, $A \cup B \in \mathcal{G}$.

Therefore, $\mathcal{G}$ is an algebra.

Let $\{G_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{G}$. We must prove that $\bigcup_{n=1}^{\infty} G_n \in \mathcal{G}$ First we disjointize the $G_n's$. Let $E_1 = G_1$ *and, for* $n \geq 2$, let $E_n = G_n \setminus \bigcup_{k=1}^{n-1} G_k$. Because $\mathcal{G}$ is an algebra, $\{E_n\}_{n=1}^{\infty} \subset \mathcal{G}$ moreover, we have $\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} G_n$.Let $E = \bigcup_{n=1}^{\infty} E_n$.

Because $\mathcal{G}$ is an algebra, $\bigcup_{j=1}^{n} E_j \in \mathcal{G}$. It follows that for each $n \in N$, there is an $F_n \in \mathcal{F}$ such that $\mu((\bigcup_{j=1}^{n} E_j)\Delta F_n) < \frac{1}{n}$. Now, we have

$$E \,\Delta\, F_n \subset \left(\bigcup_{j=n+1}^{\infty} E_j\right) \cup \left(\left(\bigcup_{j=1}^{n} E_j\right)\Delta\, F_n\right)$$

Hence, $\mu(E \,\Delta\, F_n) \leq \mu\left(\bigcup_{j=n+1}^{\infty} E_j\right) + \mu\left(\left(\bigcup_{j=1}^{n} E_j\right)\Delta\, F_n\right)$

$$\leq \sum_{j=n+1}^{\infty} \mu(E_j) + \frac{1}{n}.$$

Since $\sum_{n=1}^{\infty} \mu(E_j) \leq 1$, we conclude that $\lim_{n\to\infty} \mu(E\Delta\, F_n) = 0$. Consequently, $E \in \mathcal{G}$

Remark:

For $A, B \in \beta$, the expiration $\mu(A\backslash B)\log \mu(A\backslash B)$ will be close to zero $\mu(A\backslash B)$ is either close to zero or close to one. In other words, $\mu(A\backslash B)\log \mu(A\backslash B)$ will be close to zero if A *and* B are either nearly disjoint or nearly equal.

Lemma 3: Let $\boldsymbol{\mathcal{F}} \subset \beta$ be an algebra of sets ε the smallest σ-algebra containitg $\boldsymbol{\mathcal{F}}$, and $\mathfrak{B} \subset \varepsilon$ a measurable partition. Then for each $\epsilon > 0$, there is a measurable partition $\mathcal{D} \subset \mathcal{F}$ such that $H(\mathfrak{B}\backslash\mathcal{D}) < \epsilon$.

Proof: Let $\mathfrak{B}$=$\{A_1, A_2, \dots, A_n\}$ and δ be a smallest positive number. By lemma 2, we can find, for each j, a set $C_j \in \boldsymbol{\mathcal{F}}$ such that $\mu\left(A_j \,\Delta\, C_j\right) < \delta$. we will use the $C_j's$ to construct a measurable partition of Ω. first, we disjointize the $C_j s$ by defining $B_j = C_j \setminus \bigcup_{k\neq j} C_k$. Then we obtain a measurable partition $\mathcal{D} = \{B_1, B_2, .. B_n, B_{n+1}\}$ by letting $B_{n+1} = \Omega \setminus \bigcup_{j=1}^{n} B_j$ Because $\boldsymbol{\mathcal{F}}$ is an algebra, it follows that $B_j \in \boldsymbol{\mathcal{F}}$ for all j.

Now we consider the conditional entropy

$$H(\mathfrak{B}\backslash\mathcal{D}) = -\sum_{j=1}^{n}\sum_{k=1}^{n+1}\mu(B_k)\mu(A_j\backslash B_k)\log\mu(A_j\backslash B_k).$$

On the right- hand side of the previous equation, the sum of the terms for which $k = n + 1$ is dominated by $n\mu(B_{n+1})\log 2/2$.this expression can be made by choosing δ appropriately, because $|\mu(A_j) - \mu(B_j)|$ is small for $1 \leq j \leq n$ and $\sum_{j=1}^{n}\mu(A_j) = 1$. We use

$$-\mu(B_k)\mu(A_j\backslash B_k)\log\mu(A_j\backslash B_k) \leq -\mu(A_j\backslash B_k)\log\mu(A_j\backslash B_k).$$

And the observation that $\mu(A_j\backslash B_k)$ is close to 0, when $j \neq k$, and close to 1, when $j = k$, to assert that the sum of the remaining terms of $H(\mathfrak{B}\backslash\mathcal{D})$ is small when δ is sufficiently small.

■

Lemma 4:

Let $\mathfrak{B}$ be a measurable partition. Then $H(\mathfrak{B}^{(k)}, \varphi) = H(\mathfrak{B}, \varphi)\ for\ all\ k \geq 1.$

Proof: $(\mathfrak{B}^{(k)})^{(n)} = \mathfrak{B}^{(k+n-1)}$. Hence,

$$H(\mathfrak{B}^{(k)}, \varphi) = \lim_{n\to\infty}\frac{H((\mathfrak{B}^{(k)})^{(n)})}{n} = \lim_{n\to\infty}\frac{H(\mathfrak{B}^{(k+n+1)})}{n}$$

$$= \lim_{n\to\infty}\frac{H(\mathfrak{B}^{(m)})}{m} = H(\mathfrak{B}, \varphi).$$ ■

If φ is a 1-1 correspondance and $(\Omega, \beta, \mu, \varphi^{-1})$ is a measurable dynamical system, then we say that φ is invertible. In such cases, the notation

$$\mathfrak{B}^{(m,n)} = \varphi^{-m}\mathfrak{B} \vee \varphi^{-m-1}\mathfrak{B} \vee \dots\ \varphi^{-n}\mathfrak{B}$$

is meaningful for each pair of integers $n, m\ with\ m \leq n$.

Lemma 5:

If φ is invertible and $\mathfrak{B}$ is a measurable partition, then

$$H(\mathfrak{B}^{(m,n)}, \varphi) = H(\mathfrak{B}, \varphi)$$

For each pair of integers n,m with $m \leq n$.

$$\mathfrak{B}^{(m,n)} = \varphi^{-m}\mathfrak{B} \vee \varphi^{-m-1}\mathfrak{B} \vee \dots\ \varphi^{-n}\mathfrak{B}$$

$$= \varphi^{-m}\left(\mathfrak{B} \vee \varphi^{-1}\mathfrak{B} \vee \dots \vee \varphi^{-(n-m)}\mathfrak{B}\right) = (\varphi^{-m}\mathfrak{B})^{n-m+1}$$

Hence by lemma 4, we have $H\left(\ \mathfrak{B}^{(m,n)}, \varphi\right) = H((\varphi^{-m}\mathfrak{B})^{n-m+1}, \varphi)$.Since μ is invariant with respect to both $\varphi\ and\ \varphi^{-1}$, it follows that

$$H(\varphi^{-m}\mathfrak{B}, \varphi) = H(\mathfrak{B}, \varphi)$$

Therefore, $H\left(\mathfrak{B}^{(m,n)}, \varphi\right) = H(\mathfrak{B}, \varphi)$. ■

If φ is invertible and $\mathfrak{B}$ is a measurable partition, then for each $n \in N$, the collection

$\beta_n(\mathfrak{B}) = \left\{B \in\ \beta : B\ is\ a\ union\ of\ elements\ of\ \mathfrak{B}^{(-n,n)}\right\}$ is an algebra of subset of Ω.

Because $\beta_n(\mathfrak{B}) \subset \beta_{n+1}(\mathfrak{B})$ the collection $\beta_\infty(\mathfrak{B}) = \bigcup_{n=1}^{\infty} \beta_n(\mathfrak{B})$ is also an algebra of subsets

Of Ω.

THEOREM 3: Kolmogorove- Sinai Theorem

Let $(\Omega, \beta, \mu, \varphi)$ be a measurable dynamical system and assume that φ is invertable. Suppose that $\mathfrak{B}$ is a measurable partition of (Ω, β) such that β is the smallest σ- algebra containing $\beta_\infty(\mathfrak{B})$. Then $h(\varphi) = H(\mathfrak{B}, \varphi)$.

Proof: By the definition of $h(\varphi)$, to prove that

$$H(\mathcal{D}, \varphi) \leq H(\mathfrak{B}, \varphi)$$

for each measurable partition $\boldsymbol{\mathcal{D}}$. It follows from proposition 3(g),

$$H(\mathcal{D}, \varphi) \leq H(\mathcal{D}\backslash\mathfrak{B}^{(-n,n)}, \varphi) + H(\mathfrak{B}^{(-n,n)}, \varphi)$$

for all $n \in N$. Hence, by lemma 5, we have

$$H(\mathcal{D}, \varphi) \leq H\left(\mathcal{D}\backslash\mathfrak{B}^{(-n,n)}\right) + H(\mathfrak{B}, \varphi) \qquad (*)$$

Given $\varepsilon > 0$, we can apply lemma 3, to find a measurable partition $\mathfrak{N}\ such\ that$ $\mathfrak{N} \subset \beta_\infty(\mathfrak{B})$ and $H(\mathcal{D}\backslash\mathfrak{N}) < \varepsilon$. Since $\mathfrak{N}$ is a finite collection, it implies that

$\mathfrak{N} \subset \beta_n(\mathfrak{B})\ for\ some\ n$. in particular, we have $\mathfrak{N} \ll \mathfrak{B}^{(-n,n)}$.

Apply proposition 3(e), we get $H\left(\mathcal{D}\backslash\mathfrak{B}^{(-n,n)}\right) \leq H(\mathcal{D}\backslash\mathfrak{N}) < \varepsilon$.hence by (*) , we have $H(\mathcal{D},\varphi) \leq \varepsilon + H(\mathfrak{B},\varphi)$. Since ε is an arbitrary positive number,

so, $H(\mathcal{D},\varphi) \leq H(\mathfrak{B},\varphi)$. ■

The Kolmogrove- Sinai Theorem is valid when φ is not necessary invertible. For a measurable partition $\mathfrak{B}$, Let

$$\widetilde{\beta_n}(\mathfrak{B}) = \left\{B \in \beta : B\ is\ a\ union\ of\ members\ of\ \mathfrak{B}^{(n)}\right\}$$

And let $\widetilde{\beta_\infty}(\mathfrak{B}) = \bigcup_{n=1}^{\infty} \widetilde{\beta_n}(\mathfrak{B})$.

Theorem 4:

Let $(\Omega,\beta,\mu,\varphi)$ be a measurable dynamical system. Suppose that $\mathfrak{B}$ is a measurable partition of (Ω,β) such that β is the smallest σ- algebra containing $\widetilde{\beta_\infty}(\mathfrak{B})$. $then$

$$h(\varphi) = H(\mathfrak{B},\varphi).$$

Proof: In the proof of Kolmogrove-Sinai theorem, we replace $\beta_n(\mathfrak{B})\ and\ \beta_\infty(\mathfrak{B})\ by\ \widetilde{\beta_n}(\mathfrak{B})\ and\ \widetilde{\beta_\infty}(\mathfrak{B})$, respectively. ■

Example 5: Entropy of a Bernoulli Scheme

In this example, we apply the Kolmogrove- Sinai theorem to obtain the entropy of the Bernoulli $B(p_1, p_2, \dots, p_N)$. In example 4, consider the measurable

Partition (Ω,β) given by $\mathfrak{B}$=$\left\{C_{\{0\},k} : k = 1,2,\dots,N\right\}$.the entropy of $\mathfrak{B}$ is

$$H(\mathfrak{B}) = -\sum_{k=1}^{N} \mu(C_{\{0\},k}) \log C_{\{0\},k} = -\sum_{k=1}^{N} p_k \log p_k.$$

We will show that $\mathfrak{B}$ satisfies the hypothesis of the Kolmogrove –Sinai Theorem.

i) φ is invertible.

Define

$$\varphi: \Omega \to \Omega \; by \; \varphi(f)(j) = f(j+1)$$

Let $f, g \in \Omega \; and \; \varphi(f)(j) = \varphi(g)(j) => f(j+1) = g(j+1)$

The effect of φ is to move each term of a sequence f one place to the left.

Hence, $f = g \; for \; each \; j$. therefore, φ is one to one.

For each $f \in \Omega, there \; is \; g \in \Omega, such \; that \; \varphi(f)(j) = g$

Choose $g = f(j+1)$,

$$\varphi(f)(j) = f(j+1) = g$$

Hence, φ is onto. Therefore, φ is 1-1 correspondence.

ii) $\mu \; is \; invariant \; with \; respet \; to \; \varphi^{-1}$

Let $v(B) = \mu(\varphi(B))$. Then $v\left(C_{F,^*, a^*}\right) = \mu\left(\varphi\left(C_{F,^*, a^*}\right)\right) = \prod_{j \in F} \mu_0\left(a(j)\right)$

$$= \prod_{j \in F} \mu_0\left(a^*(j+1)\right) = \mu(C_{F,^*, a^*})$$

Hence, $\mu \; is \; invariant \; with \; respet \; to \; \varphi^{-1}$

By (i) and (ii), φ is invertible. We have $\varphi^{-1}\{C_{\{0\},k}\} = C_{\{1\},k}$ and more generally

$\varphi^{-l}\{C_{\{0\},k}\} = C_{\{l\},k} \; for \; every \; integer \; l$.therefore, a typical element of $\mathfrak{B}^{(-m,m)}$ is of the form $\bigcap_{l=-m}^{m} C_{\{0\},k} = C_{\{-m,-m+1,\ldots m\},b}$ where $b(l) = k_l$for $-m \leq l \leq m$.

β is the σ- algebra generated by sets of the form $C_{F,a}$ where F is a set of integers and a is a function from F in to $\{1,2,\ldots,N\}$. By choosing m large enough, we can assume that

$F \subset \{-m, \ldots m\}$. Hence, we can write

$C_{F,a} = \bigcup_b C_{\{-m,\ldots,m\},b}$ Where the union is over all function

$$b : \{-m, \ldots, m\} \to \{1,2, \ldots, N\} \; such \; that \; b(l) = a(l) for \; all \; l \in F.$$

It follows that $C_{F,a}$ belongs to $A_m(\mathfrak{B})$ and this in turn implies that the algebra $A_\infty(\mathfrak{B})$, containsall sets of the form $C_{F,a}$. Thus, β is the smallest σ- algebra containing $A_\infty(\mathfrak{B})$.

Next we calculate $H(\mathfrak{B}, \varphi)$. The entropy of $\mathfrak{B}^{(m)}$ is

$$H(\mathfrak{B}^{(m)}) = -\sum_{k_{0=1}}^{N} \sum_{k_1=1}^{N} \ldots \sum_{k_{m-1=1}}^{N} \prod_{l=0}^{m-1} \mu(C_{\{l\},k_l}) \log \prod_{l=0}^{m-1} \mu(C_{\{l\},k_l})$$

$$= -\sum_{k_{0=1}}^{N} \sum_{k_1=1}^{N} \ldots \sum_{k_{m-1=1}}^{N} \prod_{l=0}^{m-1} p_k \; \log \prod_{l=0}^{m-1} p_k$$

Using $\sum_{k_1=}^{N} p_k = 1$, we get

$$-\sum_{k_{0=1}}^{N} \sum_{k_1=1}^{N} \ldots \sum_{k_{m-1=1}}^{N} \prod_{l=0}^{m-1} p_k \log \prod_{l=0}^{m-1} p_k = m \sum_{k=1}^{N} p_k \log p_k$$

Apply the Kolmogrove- Sinai theorem, we conclude that

$$h(\varphi) = H(\mathfrak{B}, \varphi) = \lim_{m \to \infty} \frac{H(\mathfrak{B}^{(m)})}{m} = -\sum_{k=1}^{N} p_k \log p_k.$$

Thus we see that the entropy of the Bernoulli scheme $B(p_1, p_2, \ldots, p_N)$ equals $-\sum_{k=1}^{N} p_k \log p_k$.

■

REFERENCE

1. Deumlich R (1997) Functional Analysis I, Addis Ababa University.

2.John N. McDonald and Niel A. Weiss (2005) A course in Real Analysis -3rd ed. Elsevier (Singapore) Pte. Ltd

3. Richard R. Golderberg (1976) Method of Real Analysis, 2nd edition. John Wiley & Sons, Inc.