Pro Mathematica: Vol. VI, Nos. 11-12, 1992

INTRODUCCION A LA TEORIA ERGODICA

Carlos **Gutierrez***

La Teoría Ergódica está edificada con ejemplos, teoremas de convergencia, el estudio de varias propiedades de recurrencia y la Teoría de Entropía. Aquí tratamos de todos estos asuntos en un nivel introductorio. Al lector interesado en estudios más amplios y profundos le recomendamos los libros de R. Mañé [Mañ], K. Petersen [Pet] o P. Walters [Wal]. También, con la misma finalidad, nos gustaría recomendar la lectura de los libros de P. Billingsley [Bil] y W. Szlenk [Szl].

Son requisitos, para acompañar el texto, alguna familiaridad con los elementos de Teoría de la medida y de Topología General.

Damos a continuación una idea de lo que tratan las diversas secciones de este texto.

En la sección I revisamos las definiciones y resultados de la Teoría de la Medida que utilizaremos.

En la sección II introducimos el objeto fundamental de nuestro estudio: los sistemas dinámicos medibles $(X, \mathcal{A}, \mu, T)$ constituidos por un conjunto X, por una σ-álgebra de subconjuntos de X, una medida de probabilidad sobre $\mathcal{A}$ (o sea $\mu(X) = 1$) y una transformación $T: X \to X$ que preserva medida (i.e., para todo $A \in \mathcal{A}$, $T^{-1}A \in \mathcal{A}$ y $\mu(T^{-1}(A)) = \mu(A)$).

Fijemos, para lo que sigue de esta introducción, un sistema dinámico medible $(X, \mathcal{A}, \mu, T)$.

En la sección III encontramos el tópico de recurrencia que es de gran importancia en la Teoría Ergódica. El Teorema de Recurrencia de Poincaré dice que si $A \in \mathcal{A}$ y A_0 es el conjunto de los puntos $x \in A$ tales que $\{x, Tx, T^2x, \dots\} \cap A$ es un conjunto infinito, entonces $\mu(A) = \mu(A_0)$. Por

* Profesor del Instituto de Matemática Pura y Aplicada, Rio de Janeiro, Brasil.

recurrencia entenderemos el comportamiento cualitativo de los conjuntos de la forma $\{x, Tx, T^2x, \dots\}$.

En la sección IV tratamos del Teorema Ergódico de Birkhoff que es uno de los resultados trascendentales de la teoría. Él tiene que ver con la convergencia de los promedios

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x))$$

donde $f: X \to \mathbf{C}$ es una función integrable. El estudio de la convergencia de este tipo de promedios es otro de los asuntos básicos de la Teoría Ergódica.

En la sección V estudiamos las transformaciones ergódicas T, o sea aquéllas en que para casi todo punto $x \in X$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x)) = \int_X f d\mu,$$

donde $f: X \to \mathbf{C}$ es integrable. Recalcamos el hecho de que el límite existe y tiene el mismo valor para casi todo $x \in X$. Las transformaciones ergódicas también están caracterizadas por lo siguiente: Si $A \in \mathcal{A}$ y $T^{-1}A = A$ entonces $\mu(A) = 0$ o $\mu(A) = 1$. Luego, si suponemos que T no es ergódica entonces existirá $B \in \mathcal{A}$ tal que $T^{-1}(B) = B$ y $0 < \mu(B) < 1$, y, así, $T^{-1}(B^c) = B^c$ y $0 < \mu(B^c) < 1$, donde $B^c = X \setminus B$. En estas condiciones podremos estudiar $(X, \mathcal{A}, \mu, T)$ mediante los dos sistemas dinámicos más simples $(B, \mathcal{A}_B, \mu_B, T|_B)$ y $(B^c, \mathcal{A}_{B^c}, \mu_{B^c}, T|_{B^c})$, donde $\mathcal{A}_C = \{A \cap C : A \in \mathcal{A}\}$ y $\mu_C = \frac{1}{\mu(C)} \cdot \mu$. En este contexto, las transformaciones ergódicas son las indescomponibles.

En la sección VI tratamos de las transformaciones mezclantes T las cuales están caracterizadas por la siguiente propiedad de recurrencia: Para todo $A, B \in \mathcal{A}$

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A) \cdot \mu(B).$$

Las transformaciones mezclantes forman una subfamilia de las ergódicas; pues, como se verá, las ergódicas están caracterizadas también por lo siguiente: Para todo $A, B \in \mathcal{A}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}A \cap B) = \mu(A) \cdot \mu(B).$$

En la sección VII estudiamos las propiedades ergódicas y mezclantes de una clase de transformaciones conocidas como desplazamientos de Bernoulli y de Markov. Ellas constituyen una clase muy importante de ejemplos de sistemas dinámicos medibles.

La sección VIII tiene por finalidad dar una idea de las conexiones que existen entre la Teoría ergódica y la Dinámica Topológica.

En las secciones IX – XI tratamos del tercer asunto de gran relevancia en la Teoría Ergódica: El problema de clasificación. Decimos que dos sistemas dinámicos medibles $(X, \mathcal{A}, \mu, T)$ y $(Y, \mathcal{B}, \nu, S)$ son equivalentes si existen conjuntos de medida uno $X_1 \in \mathcal{A}$, $Y_1 \in \mathcal{B}$ y una transformación biyectiva $\phi: X_1 \to Y_1$ tal que $\phi \circ T = S \circ \phi$ en X_1 y $\mu(\phi^{-1}(B \cap Y_1)) = \nu(B \cap Y_1) = \nu(B)$, para todo $B \in \mathcal{B}$. De la manera clásica, para decidir si dos conjuntos son o no equivalentes estudiamos propiedades u objetos que son invariantes (i.e., preservadas) por equivalencia, como por ejemplo ergodicidad y mezcla. En la sección X introducimos la entropía $h(T)$ de T que viene a ser un invariante, por equivalencia, bastante sensible. Mostramos que ella puede ser usada para distinguir algunos sistemas no equivalentes. Aunque no será probado, mencionamos que la entropía es un invariante completo por equivalencia dentro de la clase de los desplazamientos de Bernoulli [Orn] los cuales son definidos en la sección II.

§I. Preliminares: Medida e Integración

En esta sección presentamos las definiciones y Teoremas de la Teoría de la Medida que necesitaremos. Las pruebas de los resultados pueden ser encontradas, por ejemplo, en [Fer], [Mun] y [Par].

A. Medidas y extensión de medidas

Una familia $\mathcal{A}$ de subconjuntos de un conjunto X es una semiálgebra si

(i) $\phi \in \mathcal{A}$,

(ii) $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$, y

iii) $A \in \mathcal{A} \Rightarrow X \setminus A = \sum_{i=1}^{n} A_i$, donde $A_1, A_2 \ldots, A_n \in \mathcal{A}$. El signo $\sum$ indica *unión disjunta*.

Una semiálgebra $\mathcal{A}$ de subconjuntos de un conjunto X es un *álgebra* si $A \in \mathcal{A} \Rightarrow A^c = X \setminus A \in \mathcal{A}$. Se sigue que las dos condiciones siguientes son satisfechas

(i) $A, B \in \mathcal{A} \Rightarrow A \cup B = (A^c \cap B^c)^c \in \mathcal{A}$;

(ii) $A, B \in \mathcal{A} \Rightarrow A \setminus B = A \cap B^c \in \mathcal{A}$.

Un álgebra $\mathcal{A}$ de subconjuntos de un conjunto X es una *σ-álgebra* si dada una sucesión $\{A_n\}$ de elementos de $\mathcal{A}$, la unión $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Se sigue que también la intersección $\bigcap_{n=1}^{\infty} A_n$ pertenece a $\mathcal{A}$.

Sea A una familia no vacía de subconjuntos de X. Observemos que existe una σ-álgebra de subconjuntos de X que contiene A y que es la menor de todas. Para ver esto, notemos que la familia 2^X de todos los subconjuntos de X es una σ-álgebra conteniendo A y que la intersección $\mathcal{A}$ de todas las σ-álgebras conteniendo A es también una σ-álgebra conteniendo A. Esta σ-álgebra $\mathcal{A}$ será llamada *σ-álgebra generada por A.*

Tenemos el siguiente resultado

Teorema I.1. *Sea S una semiálgebra de subconjuntos de X. EL* álgebra, *$\mathcal{A}(S)$, generada por S, consiste precisamente de aquellos subconjuntos de X que pueden ser escritos en la forma $E = \sum_{i=1}^{n} A_i$, donde cada $A_i \in S$.*

Si X es un espacio topológico, la *σ-álgebra de Borel* es la *σ-álgebra* $\mathcal{B}(X)$ generada por la familia de los conjuntos abiertos. Los elementos de $\mathcal{B}(X)$ son llamados *borelianos*.

Sea S una semiálgebra de **subconjuntos** de X. Una función $\tau: S \mapsto [0,\infty]$ (donde $[0,\infty]$ es la semirecta extendida $[0,\infty) \cup \{\infty\}$) es *finitamente aditiva* si se cumplen las dos condiciones siguientes: i) $\tau(0) = 0$; ii) $A = \sum_{i=1}^{n} A_i$, $\quad A_i \in S$, $A \in S \Rightarrow \tau(A) = \sum_{i=1}^{n} \tau(A_i)$. Una función $\tau: S \mapsto [0,\infty]$ es una *medida* si es finitamente aditiva y si $A = \sum_{i=1}^{\infty} A_i, \quad A_i \in S, \quad A \in S \Rightarrow \tau(A) = \sum_{i=1}^{\infty} \tau(A_i)$.

Una terna $(X, \mathcal{A}, \mu)$, en que X es un conjunto, $\mathcal{A}$ una σ-álgebra de subconjuntos de X y $\mu: \mathcal{A} \to [0,\infty]$ es una medida, es llamado *espacio de medida*. El par $(X, \mathcal{A})$ es llamado *espacio medible*. Decimos que X es *σ-finito* si lo podemos escribir como una unión enumerable $X = \bigcup_{n=1}^{\infty} A_n$ tal que $\mu(A_n) \leq \infty$ para todo n. Decimos que $(X, \mathcal{A}, \mu)$ es un *espacio de probabilidad* si $\mu(X) = 1$.

Sea $(X, \mathcal{A}, \mu)$ un espacio de medida. Decimos que un conjunto $A \subset X$ *es de medida cero* si existe $A_1 \in \mathcal{A}$ tal que $A \subset A_1$ y $\mu(A_1) = 0$. Una propiedad aplicable a puntos de un conjunto $S \subset X$ vale *para casi todo* $x \in S$ (abreviadamente c.t.p.) si el conjunto de los puntos de S donde ella es falsa es de medida cero.

Los siguientes teoremas de extensión serán de mucha utilidad cuando se estudien ejemplos particulares.

Teorema I.2. *Sean S una semiálgebra de subconjuntos de X y $\mathcal{A}(S)$ el álgebra generada por S. Si $\tau: S \mapsto [0,\infty)$ es finitamente aditiva, entonces existe una única función finitamente aditiva $\tau_1: \mathcal{A}(S) \mapsto [0,\infty)$ que extiende τ. Además, si τ es una medida, entonces τ_1 también lo es.*

Puede ocurrir que $X \notin S$; no obstante X es unión disjunta de un número finito de miembros $A_1, ..., A_n$ de S y así $\tau_1(X) = \sum_{i=1}^{n} \tau(A_i)$.

Teorema I.3. *(***Teorema de Extensión de Hahn-Kolmogorov***).*

Sean $\mathcal{A}_0$ una semiálgebra de subconjuntos de X y $\mathcal{A}$ la σ-álgebra generada por $\mathcal{A}_0$. Sea $\mu_0: \mathcal{A}_0 \mapsto [0,1]$ una medida. Entonces existe una única medida $\mu: \mathcal{A} \mapsto [0,1]$ que extiende μ_0.

Teorema I.4. *Sea $\mathcal{A}$ la σ-álgebra de los borelianos de $\mathbf{R}^n$. Existe una única medida $\lambda: \mathcal{A} \mapsto [0,\infty]$ tal que, si $A = \prod_{i=1}^{n}(a_i, b_i)$,*

$$\lambda(A) = \prod_{i=1}^{n}(b_i - a_i).$$

Esta medida denomínase medida de Lebesgue.

Teorema I.5. *(***Teorema de Aproximación***).*

Sea $(X, \mathcal{A}, \mu)$ un espacio de medida y sea $\mathcal{A}_0$ un álgebra que genera $\mathcal{A}$. Entonces, para todo $A \in \mathcal{A}$ tal que $\mu(A) < \infty$ y para todo $\epsilon > 0$, existe $A_0 \in \mathcal{A}_0$ tal que $\mu(A \triangle A_0) < \epsilon$.

Teorema I.6. *(***Criterio de Aditividad***)*

Sea S una semiálgebra y $\tau: S \mapsto [0,\infty)$ una función finitamente aditiva. Supongamos que existe una familia $\mathcal{C} \subset S$ tal que:

i) *$\mathcal{C}$ es una clase compacta; esto es, si $C_1 \supset C_2 \supset \dots C_i \supset \dots$ son elementos de $\mathcal{C}$, entonces $\bigcap_{i=1}^{\infty} C_i \neq \phi$*

ii) *$\mathcal{C}$ posee la propiedad de la aproximación; esto es, para todo $A \in S$, $\tau(A) = \sup\{\tau(C)/C \subset A, C \in \mathcal{C}\}$.*

Entonces τ es una medida.

B. Funciones Medibles

Sean $(X_1, \mathcal{A}_1)$ y $(X_2, \mathcal{A}_2)$ espacios medibles. Decimos que $T: X_1 \to X_2$ es *medible* si, para todo $A_2 \in \mathcal{A}_2$, $T^{-1}(A_2) \in \mathcal{A}_1$. Si X_1 y X_2 son espacios topológicos, decimos que $T: X_1 \to X_2$ es *medible* si lo es con respecto a las σ-álgebras de borelianos.

Proposición I.7. *Sean $(X_1, \mathcal{A}_1)$ y $(X_2, \mathcal{A}_2)$ espacios medibles, $T: X_1 \to X_2$, y $\mathcal{C}$ una familia de subconjuntos de X_2 tal que la σ-álgebra generada por $\mathcal{C}$ es $\mathcal{A}_2$. Si para todo $C \in \mathcal{C}$, $T^{-1}(C) \in \mathcal{A}_1$, entonces T es medible.*

Se sigue de esta proposición que toda función continua $T: X_1 \to X_2$, entre espacios topológicos X_1 y X_2, es medible.

Proposición I.8. *Sea $(X, \mathcal{A})$ un espacio medible.*

(i) *Si $\{f_n\}_{n\geq 1}$ es una sucesión de funciones medibles de X en $[-\infty, \infty]$, entonces $\sup f_n$, inf f_n $\limsup f_n$ y $\lim$ inf f_n son medibles.*

(ii) *Si $f, g: X \to \mathbf{C}$, son medibles y $c \in \mathbf{C}$, entonces $cf, f+g, f.g$ y $|f|$ son medibles.*

(iii) *Si $\{f_n\}_{n\geq 1}$ es una sucesión de funciones medibles de $X \to \mathbf{C}$ tal que $\lim_n f_n(x) = f(x)$ existe para todo $x \in X$, entonces f es medible.*

Observamos que (iii) se obtiene fácilmente de (i). En efecto, cada f_n puede ser escrita como $f_n = Re(f_n) + i\ \mathrm{Im}(f_n)$ donde $Re(f_n)$, $\mathrm{Im}(f_n): X \to \mathbf{R}$ son funciones medibles.

C.. Integración

En esta subsección $(X, \mathcal{A}, \mu)$ será un espacio de medida fijo y todas las funciones consideradas serán de X en $\mathbf{C}$.

Una función F es *simple* si es medible y toma valores en un conjunto finito. Ella puede ser representada de manera única en la forma

$$f = \sum_{i=1}^{n} a_i \chi_{A_i}$$

donde los $A_i \in \mathcal{A}$ son disjuntos dos a dos, los $a_i \in \mathbf{C}$ son distintos dos a dos y donde χ_{A_i} denota la función característica de A_i. Si cada $\mu(A_i) < \infty$, decimos que f *es integrable* y definimos

$$\int f = \sum_{i=1}^{n} a_i \mu(A_i).$$

Una función g es *integrable* si existe una sucesión de funciones simples integrables $\{f_n\}_{n \geq 1}$ tal que:

$$\lim f_n(x) = g(x) \qquad \mu - c.t.p. \quad x \in X, \quad \text{y}$$

$$\lim_{n,m \to \infty} \int |f_n - f_m| = 0.$$

En este caso $\lim_n \int f_n$ existe y su valor no depende de la sucesión particular $\{f_n\}$. Para indicar este valor límite usaremos indistintamente los símbolos

$$\int_X g d\mu = \int_X g = \int g.$$

Observamos que la función g puede no ser medible pero siempre coincide con una medible c.t.p. En efecto, existe $A \in \mathcal{A}$ tal que $\mu(X \setminus A) = 0$ y $\lim f_n(x) = g(x)$ para todo $x \in A$. Tenemos entonces, por la proposición I.8, que

$$\lim(f_n \cdot \chi_A) = g \cdot \chi_A$$

es medible y coincide con g en A.

Teorema I.9. (**Teorema de la Convergencia Monótona**). *Si $f_n: X \to \mathbf{R}$, $n = 1, 2, \ldots$, es una sucesión de funciones integrables tales que, para c.t.p. $x \in X$, $f_1(x) \leq f_2(x) \leq \ldots$ y $\sup_n \int f_n d\mu < \infty$, entonces:*

(i) $\lim_n f_n$ *existe c.t.p. y es integrable;*

(ii) $\int \lim_n f_n d\mu = \lim_n \int f_n d\mu.$

Teorema I.10. (**Lema de Fatou**).

Sea $f_n: X \to \mathbf{R}$ una sucesión de funciones integrables que son acotadas inferiormente por una función integrable. Si

$$\lim_n \inf \int f_n d\mu < \infty$$

entonces la función $x \to \lim_n \inf f_n(x)$ es integrable y

$$\int \lim_n \inf f_n d\mu \leq \lim_n \inf \int f_n d\mu.$$

Teorema I.11. (**Teorema de la Convergencia Dominada**).

Si $g: X \to \mathbf{R}$ es integrable y $f_n: X \to \mathbf{R}$, $n = 1, 2, \ldots$, es una sucesión de funciones integrables con $|f_n| \leq g$ c.t.p. $(n \geq 1)$ y $\lim_n f_n = f$ c.t.p. entonces f es integrable y

$$\lim_n \int f_n d\mu = \int f d\mu.$$

Sea $p \in [1, \infty)$. Consideremos el espacio de todas las funciones medibles $f: X \to \mathbf{C}$ con $|f|^p$ integrable. Este espacio es un espacio vectorial con respecto a la adición usual de funciones y a la multiplicación de funciones por escalares.

Si definimos una relación de equivalencia en este conjunto por $f \sim g$ si, y sólo si, $f = g$ c.t.p. entonces el espacio de clases de equivalencia es también un espacio vectorial. Denotaremos por $L^p(X, \mathcal{A}, \mu)$ (or por L^p) el espacio de estas clases de equivalencia, sin embargo, por abuso de notación, escribiremos $f \in L^p(X, \mathcal{A}, \mu)$ para significar que la función $f: X \to \mathbf{C}$ es medible y tiene $|f|^p$ integrable. La formula $\|f\|_p = [\int |f|^p \, d\mu]^{1/p}$ define una norma en $L^p(X, \mathcal{A}, \mu)$ la cual es completa; o sea $L^p(X, \mathcal{A}, \mu)$ es un espacio de Banach. Escribiremos $f \in L^p_{\mathbf{R}}(X, \mathcal{A}, \mu)$ para indicar que f es una función de valores reales.

Denotaremos por $L^\infty(X, \mathcal{A}, \mu)$ (o por L^∞) el espacio de todas las funciones medibles $f: X \to \mathbf{C}$ para las que existe $K > 0$ tal que $|f(x)| \leq K$ para c.t.p. $x \in X$ y donde una pareja de dichas funciones son identificadas si, y sólo si, son iguales c.t.p. El ínfimo de los K con esta propiedad denótase por $\|f\|_\infty$ y define una norma en $L^\infty(X, \mathcal{A}, \mu)$ que lo torna un espacio de Banach.

Proposición I.12. *Si* $\mu(X) < \infty$ *y* $1 \leq p < q \leq \infty$, *entonces* $L^q(X, \mathcal{A}, \mu)$ *está densamente contenido en* $L^p(X, \mathcal{A}, \mu)$.

Un *espacio de Hilbert* $\mathcal{H}$ es un espacio de Banach en el cual la norma está dada por un producto interno, i.e., $\mathcal{H}$ es un espacio de Banach y existe una aplicación $(\cdot,\cdot): \mathcal{H} \times \mathcal{H} \to \mathbf{C}$ tal que $(\cdot,\cdot)$ es bilineal, $(f, g) = \overline{(g, f)}$ para todo $f, g \in \mathcal{H}$, $(f, f) \geq 0$ para todo $f \in \mathcal{H}$, y $|f| = (f, f)^{1/2}$ es la norma en $\mathcal{H}$.

El espacio de Banach $L^p(X, \mathcal{A}, \mu)$ es un espacio de Hilbert si, y sólo si, $p = 2$. El producto interno en $L^2(X, \mathcal{A}, \mu)$ está dado por $(f, g) = \int f\bar{g} d\mu$. Sólo usaremos los siguientes hechos cuyas pruebas pueden ser encontradas en [Rud]:

(i) Si $f, g \in L^2(X, \mathcal{A}, \mu)$ y f_n es una sucesión que converge a f en $L^2(X, \mathcal{A}, \mu)$, entonces (f_n, g) converge a (f, g).

(ii) Sea $S^1 = \{z \in \mathbf{C}/|z| = 1\}$. Sea μ la medida de Haar de S^1, o sea la medida sobre los borelianos determinada por lo siguiente: si $0 \leq x < y < 1$ y $A = \{z \in S^1/2\pi x \leq \text{argumento de } z < 2\pi y\}$, entonces $\mu(A) = y - x$. Todo $f \in L^2(S^1, \mathcal{B}(S^1), \mu)$ determina una única sucesión de números complejos $\{b_n\}_{-\infty}^{+\infty}$ tal que

$$\sum_{n=-\infty}^{\infty} |b_n|^2 < \infty \quad \text{y} \quad f(z) = \sum_{n=-\infty}^{\infty} b_n z^n \quad \text{c.t.p.} \quad z \in S^1.$$

§II. Transformaciones que Preservan Medida

Sean $(X, \mathcal{A}, \mu)$ y $(Y, \mathcal{B}, \nu)$ espacios de probabilidad. Decimos que $T: X \to Y$ es una *transformación que preserva medida* si $\mu(T^{-1}(B)) = \nu(B)$ para todo $B \in \mathcal{B}$. Decimos que $T: X \to Y$ es una transformación *inversible que preserva medida* si T es biyectiva y T, T^{-1} preservan medida. Cuando una transformación $S: X \to X$ del mismo espacio medible $(X, \mathcal{A}, \mu)$ preserva medida, diremos que el sistema $(X, \mathcal{A}, \mu, S)$ es un *sistema dinámico medible*. Este sistema es el objeto fundamental de nuestro estudio.

Proposición II.1. *Supongamos que $(X, \mathcal{A}, \mu)$, $(Y, \mathcal{B}, \nu)$ son espacios de medida σ-finitos y que $T: X \to Y$ es una transformación. Sea $\mathcal{B}_1$ una semiálgebra que genera $\mathcal{B}$. Si para cada $B \in \mathcal{B}_1$ tenemos que $T^{-1}(B) \in \mathcal{A}$ y $\mu(T^{-1}(B)) = \nu(B)$, entonces T preserva medida.*

Demostración: T es medible por la proposición I.7.

Sea $\nu_0: \mathcal{B} \to [0, \infty]$ la medida definida como:

$$\nu_0(B) = \mu(T^{-1}(B))$$

por hipótesis $\nu_0|_{\mathcal{B}_1} = \nu$. Por el Teorema de Extensión de Hahn-Kolmogorov, ν es la única extensión de la medida $\nu_0|_{\mathcal{B}_1}$. Luego $\nu_0 = \nu$, y, así, T preserva medida. ∎

Damos algunos ejemplos de transformaciones que preservan medida.

(1) *Tranformación de Gauss.*

Sea $([0,1), \mathcal{A}, \mu)$ el espacio de probabilidad tal que $\mathcal{A}$ es la σ-álgebra de los borelianos de $[0,1)$ y μ es definida en un boreliano A, por:

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} d\lambda$$

donde λ es la medida de Lebesgue. En particular, si $[a, b)$ es un subintervalo de $[0,1)$ entonces.

$$\mu([a, b)) = \frac{1}{\log 2} \int_a^b \frac{dx}{1+x} = \frac{1}{\log 2} \log\left(\frac{1+b}{1+a}\right).$$

Sea $\varphi\colon [0,1) \to [0,1)$ la transformación definida por $\varphi(0) = 0$ y, si $x \neq 0$, $\varphi(x) = \frac{1}{x} - \left[\frac{1}{x}\right]$, donde $[z]$ denota parte entera de z. Esta φ es llamada *Transformación de Gauss*. Probaremos que φ es medible y que preserva la probabilidad μ.

Verifiquemos que si $I_n = \left(\frac{1}{n+1}, \frac{1}{n}\right]$, siendo n un entero positivo, entonces $\varphi|_{I_{n+1}}\colon I_{n+1} \to [0,1)$ es un difeomorfismo con pendiente negativa. En efecto, $x \in I_{n+1} \Rightarrow \varphi(x) = \frac{1}{x} - n$. Ver el gráfico de φ en la figura II.1. El conjunto de puntos de discontinuidad de φ es $\{0\} \cup \{\frac{1}{n} / n \in \mathbb{N}, n \geq 2\}$.

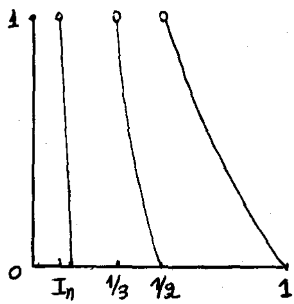


Figura II.1

Sea S la familia formada por (el conjunto vacío y) todos los intervalos de la forma $[a, b)$ con $0 \leq a < b \leq 1$. Ciertamente S es una semiálgebra de subconjuntos de $[0,1)$ que genera la σ-álgebra $\mathcal{A}$ de los borelianos. Como para todo $b \in (0,1]$,

$$\varphi^{-1}([0, b)) = \left[\bigcup_{n=1}^{\infty} \left(\frac{1}{b+n}, \frac{1}{n}\right]\right) \cup \{0\}$$

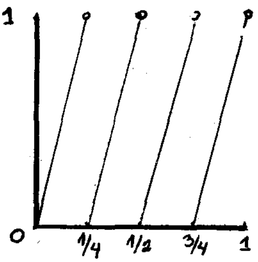


Figura II.2

Se sigue de la Proposición I.7 que φ es medible.

Probemos ahora que φ preserva μ. Sean $0 \leq a < b \leq 1$. Tenemos que:

$$\mu(\varphi^{-1}([a,b)) = \sum_{n=1}^{\infty} \mu((\frac{1}{b+n}, \frac{1}{a+n}]) =$$

$$= \sum_{n=1}^{\infty} \frac{1}{\log 2} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{dt}{1+t} =$$

$$= \sum_{n=1}^{\infty} \frac{1}{\log 2} \log \frac{(1+\frac{1}{a+n})}{(1+\frac{1}{b+n}} =$$

$$= \frac{1}{\log 2} \sum_{n=1}^{\infty} \log \left[\frac{(a+n+1)/(b+n+1)}{(a+n)/(b+n)} \right] =$$

$$\frac{1}{\log 2} \sum_{n=1}^{\infty} \left[\log \frac{a+n+1}{b+n+1} - \log \frac{a+n}{b+n} \right]$$

$$= -\frac{1}{\log 2} \log(\frac{a+1}{b+1}) =$$

$$= \frac{1}{\log 2} \int_a^b \frac{1}{t+1} d\lambda = \mu([a,b)).$$

Consecuentemente, por la Proposición II.1, φ preserva la medida μ.

(2) *La tranformación* $x \to mx - [mx]$ *de* $[0,1)$ *con* $m > 1$ *entero.*

Sea $([0,1], \mathcal{A}, \lambda)$ el espacio de probabilidad tal que $\mathcal{A}$ es la σ-álgebra de los borelianos de $[0,1)$ y λ es la medida de Lebesgue.

Sea $m > 1$ entero y $\varphi: [0,1) \to [0,1)$ definida por $\varphi(x) = mx - [mx]$. Ver el gráfico de φ, cuando $m = 4$, en la figura II.2.

Sea S la misma subálgebra de subconjuntos de $[0,1)$ considerada en el ejemplo anterior (1). Como para todo $[a,b) \in S$

$$\varphi^{-1}([a,b)) = \bigcup_{i=0}^{m-1} [\frac{a+i}{m}, \frac{b+i}{m})$$

Se sigue de las Proposiciones I.7 y II.1 (argumentando como en el ejemplo (1)) que φ es medible y que preserva la medida λ.

(3) *El endomorfismo $z \to z^m$ del círculo con $m \neq 0$.*

Sea $(S^1, \mathcal{A}, \mu)$ el espacio de probabilidad tal que $S^1 = \{z \in \mathbf{C}/|z| = 1\}$, $\mathcal{A}$ es la σ-álgebra de los borelianos de S^1 y μ es la medida de Haar de S^1, o sea la medida de probabilidad μ determinada por lo siguiente: si $0 \leq x < y < 1$ y $A = \{z \in S^1/2\pi x \leq$ argumento de $z < 2\pi y\}$ entonces $\mu(A) = y - x$. Es fácil ver que $\mu(A)$ es igual a $\frac{1}{2\pi}$ por la longitud (canónica) de A. La prueba de que $T: S^1 \to S^1$, definida por $T(z) = z^m$, con $m \neq 0$, preserva medida es la misma que la del ejemplo (2) anterior.

(4) *Desplazamientos ("Shifts").*

En este ítem no damos ningún ejemplo de transformación que preserva medida; solamente introducimos los espacios medibles $(B(n), \mathcal{A})$, $(B^+(n), \mathcal{A})$ y las transformaciones medibles, conocidas como desplazamientos, $\sigma: B(n) \to B(n)$ y $\sigma: B^+(n) \to B^+(n)$.

Sean $n > 1$ un entero y $Y = \{0, 1, \ldots, n-1\}$ un "alfabeto" de finitos símbolos. Definimos el *producto* $B(n) = \prod_{-\infty}^{\infty} Y$ como el conjunto de todas las sucesiones bilaterales $\theta: \mathbf{Z} \to Y$ con valores en los símbolos de Y. Considerando el espacio de medida $(Y, 2^Y)$, donde 2^Y denota la σ-álgebra (finita) de todos los subconjuntos de Y, vamos a proceder a definir el ***espacio de medida producto*** $(B(n), \mathcal{A}) = \prod_{-\infty}^{\infty}(Y, 2^Y)$ de la siguiente forma:

Sean $m \geq 0$ entero y $A_i \in 2^Y$, para $0 \leq i \leq m$. La colección S de los conjuntos, llamados ***cilindros***, de la forma

$$C(j, A_0, A_1, \ldots, A_m) = \{\theta \in B(n)/\theta(j+i) \in A_i, 0 \leq i \leq m\},$$

donde $j \in \mathbf{Z}$ y cada $A_i \in 2^Y$, forma una semiálgebra de subconjuntos de X.

La σ-álgebra $\mathcal{A}$ es aquella generada por S.

La transformación biyectiva $\sigma: B(n) \to B(n)$ dada por $\sigma(\theta)(k) = \theta(k+1)$, donde $\theta: \mathbf{Z} \mapsto Y$ y $k \in \mathbf{Z}$, llamada ***desplazamiento***, es medible de acuerdo con la proposición I.7, pues para cada cilindro $C \in S$, $\sigma^{-1}(C)$ es también un cilindro. Por el mismo motivo su inversa σ^{-1} también es medible.

Otra manera de presentar el espacio medible $(B(n), \mathcal{A})$ y la transformación bimedible $\sigma: B(n) \to B(n)$ es la siguiente: Si atribuimos a cada Y la topologia discreta, $B(n) = \prod_{-\infty}^{\infty} Y$ tiene una estrutura de espacio topológico producto donde los cilindros de S forman una base de abiertos. Por el Teorema de Tichonoff, $B(n)$ es un espacio topológico compacto y como el complemento de un cilindro es unión finita de cilindros resulta que cada cilindro, siendo cerrado, es compacto. En estas circunstancias $\sigma: B(n) \to B(n)$ es un homeomorfismo y $\mathcal{A}$ es precisamente la σ-álgebra de los borelianos del espacio topológico $B(n)$.

De manera análoga definimos:

(i) el espacio topológico producto $B^+(n) = \prod_0^{\infty} Y$;

(ii) la σ-álgebra $\mathcal{A}$ de subconjuntos de $B^+(n)$ como aquella generada por los cilindros

$$C(j; A_0, A_1, \ldots, A_n) = \{\theta \in B^+(n)/\theta(j+1) \in A_i, 0 \leq i \leq m\},$$

donde $j \in \mathbf{N}$ y cada $A_i \in 2^Y$; y finalmente
(iii) el desplazamiento $\sigma: B^+(n) \rightarrow B^+(n)$ dado por $\sigma(\theta)(k) = \theta(k+1)$, donde $\theta: \mathbf{N} \mapsto Y$ y $k \in \mathbf{N}$.

(5) *Desplazamientos de Bernoulli*

Sea $Y = \{0, 1, \ldots, n-1\}$. Sean $(B(n), \mathcal{A})$ el espacio medida producto, S la semiálgebra y $\sigma: B(n) \rightarrow B(n)$ el desplazamiento como se definieron en el ítem (4) anterior. Sea $(Y, 2^Y, \nu)$ un espacio de probabilidad. La probabilidad ν está determinada por los valores $\nu(\{i\}) = p_i$ y se verifica que $(p_0, p_1, \ldots, p_{n-1})$ es un *vector de probabilidad* (i.e., cada $p_i \geq 0$ y $\sum_{i=0}^{n-1} p_i = 1$).

Si definimos $\mu_0: S \mapsto [0,1]$ dando al cilindro $C(j, A_0, A_1, \ldots, A_n)$ el valor

$$\prod_{i=0}^{m} \nu(A_i),$$

no es difícil ver que μ_0 es finitamente aditiva. En estas condiciones se puede aplicar directamente el Teorema I.6 para extender μ_0 a una medida $\mu: \mathcal{A} \mapsto [0,1]$, pues la semiálgebra S es una clase compacta (como ya fue observado en (4)).

Es fácil ver que para todo $A \in S$, $\mu(\sigma^{-1}A) = \mu(A)$, y, así, por la proposición II.1, σ preserva la medida μ. El sistema dinámico medible $(B(n), \mathcal{A}, \mu, \sigma)$, que habitualmente denotaremos por $(B(n), p, \sigma)$, será llamado *desplazamiento (bilateral) de Bernoulli.* El *desplazamiento* σ es una transformación inversible que preserva medida.

De manera análoga definimos el *desplazamiento de Bernoulli unilateral* $(B^+(n), p, \sigma)$. Esta transformación preserva medida mas no es inversible.

En el siguiente resultado, que es un caso especial del Teorema de Consistencia de Daniell- Kolmogorov [Par, p 119], usamos las notaciones acabadas de introducir en (4).

Teorema II.2. *Sean $k \geq 1$ un entero fijo $Y = \{0, 1, \ldots, k-1\}$*

y $\quad (B(k),\mathcal{A}) = \prod_{-\infty}^{\infty}(Y,2^Y)$. *Si para cada* $n \in N$ *y cada sucesión* $a_0, a_1, \ldots, a_n \in Y$ *está dado un número real no negativo* $p_n(a_0, a_1, \ldots, a_n)$ *de tal forma que:*

(i) $\sum_{a_0 \in Y} p_0(a_0) = 1$,

(ii) $p_n(a_0, \ldots, a_n) = \sum_{a_{n+1} \in Y} p_{n+1}(a_0, \ldots, a_n, a_{n+1})$,

(iii) $p_n(a_1, \ldots, a_n) = \sum_{a_0 \in Y} p_{n+1}(a_0, a_1, \ldots, a_{n+1})$;

entonces:

(a) *existe una única medida de probabilidad* μ *en* $(B(k),\mathcal{A})$ *tal que*

$$\mu(C(j; a_0, a_1, \ldots, a_n)) = p_n(a_0, a_1, \ldots, a_n)$$

donde $j \in \mathbf{Z}$ *y* $a_i \in Y$ *denota (por abuso de notación) al conjunto unitario* $\{a_i\} \subset 2^Y$;

(b) *el desplazamiento bilateral* $\sigma: B(k) \to B(k)$ *preserva la medida* μ.

(6) Desplazamiento bilateral de Markov

Particularicemos el desplazamiento general σ del Teorema II.5. Sean $p = (p_0, p_1, \ldots, p_{k-1})$ un vector de probabilidad con entradas $p_i > 0$ y $P = (p_{ij})_{i,j \in Y}$ una *matriz estocástica* (i.e., $p_{ij} \geq 0$, y, para cada i, $\sum_{j=0}^{k-1} p_{ij} = 1$) tales que $\quad pP = p \quad$ (i.e., $\sum_{i=0}^{k-1} p_i p_{ij} = p_j$). Las funciones

$$p_n(a_0, \ldots, a_n) = p_{a_0} p_{a_0 a_1} p_{a_1 a_2} \cdots p_{a_{n-1} a_n}$$

satisfacen las condiciones de consistencia del Teorema II.5; la condición (i) porque p es un vector de probabilidad, la segunda condición porque p es estocástica, y la tercera porque $pP = p$. Por consiguiente, hay una única medida μ, preservada por el desplazamiento σ y tal que

$$\mu(C(j; a_0, a_1, \ldots, a_n)) = p_{a_0} p_{a_0 a_1} \cdots p_{a_{n-1} a_n}.$$

El sistema dinámico medible $(B(k), \mathcal{A}, \mu, \sigma)$, que denotaremos habitualmente por $(B(k), p, P, \sigma)$, será llamado *Desplazamiento de Markov* y μ será llamada *(p, P)-medida de Markov*.

Una demostración del siguiente resultado de Kriloff y Bogoluboff puede ser encontrada en el libro de P. Walters [Wal].

Teorema II.3. *Sea X un espacio métrico compacto; sea $T: X \to X$ una aplicación continua. Entonces existe una medida de probabilidad sobre los borelianos de X que es preservada por T.*

§III El Teorema de Recurrencia de Poincaré

Dada una transfomación que preserva medida y un subconjunto medible de su dominio, casi todo punto del referido subconjunto regresa infinitas veces a él mismo por iteraciones de la transformación. El enunciado preciso de esto, que lo hacemos a continuación, es la versión probabilística del Teorema de Recurrencia de Poincaré.

Teorema III.1. *Sea $(X, \mathcal{A}, \mu, T)$ un sistema dinámico medible. Dado $A \in \mathcal{A}$, sea A_0 el conjunto de los puntos $x \in A$ tales que $T^n(x) \in A$ para infinitos valores $n \geq 0$. Entonces $A_0 \in \mathcal{A}$ y $\mu(A_0) = \mu(A)$.*

Demostración:

Sea $m \in \mathrm{N}$ y $B_m = \{x \in X : T^n(x) \in A$ para algún $n \geq m\}$. Vemos que

(i) $B_0 \supset B_1 \supset B_2 \supset \ldots$,

(ii) $T^{-1}(B_m) = B_{m+1}$, y

(iii) $A_0 = A \cap (\bigcap_{m=0}^{\infty} B_m)$.

Luego, como T preserva medida, $\mu(B_m) = \mu(B_{m+1})$ para todo $m \geq 0$. Por consiguiente, como $\mu(X) < \infty$, concluimos que $\mu(\bigcap_{m=0}^{\infty} B_m) = \mu(B_0)$. Así, $\mu(A_0) = \mu(A \cap B_0) = \mu(A)$, pues $A \subset B_0$. ∎

Sean X un espacio topológico y $T: X \to X$ una transformación. Definimos el *ω-límite* de $x \in X$ como el conjunto de los $z \in X$ tales que para toda vecindad U de z y todo $m \in \mathrm{N}$, existe $n \geq m$ tal que $T^n x \in U$. Decimos que $x \in X$ es *recurrente* si $x \in \omega(x)$ y denotamos

$$Rec(T) = \{x \in X : x \in \omega(x)\}$$

Enunciamos ahora la versión topológica del Teorema de recurrencia de Poincaré.

Teorema III.2. *Sea* $(X, \mathcal{B}(X), \mu, T)$ *un sistema dinámico medible, donde* X *es un espacio topológico con base enumerable (y* $\mathcal{B}(X)$ *es la* σ*-álgebra de los borelianos de* X*). Entonces* μ*-c.t.p.* x *es recurrente.*

Demostración:

Sea $(U_n)_1^\infty$ una base enumerable de abiertos de la topologia de X y sea $F_n = \{x \in U_n / \quad \mathcal{O}^+(Tx) \cap U_n = \phi\}$, donde $\mathcal{O}^+(T(x) = \{Tx, T^2x, \dots\}$ es la órbita positiva de Tx.

Como cada U_n es abierto, $F_n \subset X \setminus Rec(T)$. Así, $\bigcup_{n=1}^{\infty} F_n \subset X \setminus Rec(T)$. Recíprocamente $x \in X \setminus Rec(T)$ implica que x pertence a algún F_n. En conclusión

$$X \setminus Rec(T) = \bigcup_{n=1}^{\infty} F_n.$$

Esto termina la prueba, pues, por la versión probabilistica (Teorema III.1), $\mu(F_n) = 0$, para cada $n \geq 1$ y de este modo $\mu(Rec(T)) = 1$. ∎

Sea X un espacio topológico y $T: X \to X$ una aplicación. Definimos como *conjunto no errante* $\Omega(T)$ de T al conjunto de los $x \in X$ tales que para toda vecindad V de x existe $n \geq 1$ tal que $T^n V \cap V \neq \phi$.

Veamos que $X \setminus \Omega(T)$ es abierto. Si $x \in X \setminus \Omega(T)$ entonces existe una vecindad abierta V de x tal que, para todo $n \geq 1$, $V \cap T^n V = \phi$. Luego $V \subset X \setminus \Omega(T)$. En otras palabras $\Omega(T)$ es cerrado y, en particular, boreliano.

Corolario III.2.1. $\mu(\Omega(T) = 1$.

Demostración: $Rec(T) \subset \Omega(T)$. ∎

Corolario III.2.2. *Si* μ *es positiva sobre todo abierto no vacío, entonces* $\Omega(T) = X$.

Demostración: Si $\Omega(T)$ estuviese propiamente contenido en X entonces, como $X \setminus \Omega(T)$ es abierto, $\mu(X \setminus \Omega(T) > 0$, contradiciendo al Corolario II.9. ∎

Definimos el *soporte de* μ como el conjunto supp $(\mu) = \{x$: para toda vecindad abierta V de x, $\mu(V) > 0\}$.

Corolario III.2.3. *supp* $(\mu) \subset \overline{Rec(T)}$

Demostración: Si $x \in$ supp (μ) y V es una vecindad abierta de x, $\mu(V) > 0$; luego $V \cap Rec(T) \neq \phi$, porque sino, $1 = \mu(X) \geq \mu(V \cup Rec(T)) = 1 + \mu(V) > 1$. ∎

§IV El Teorema de Birkhoff

El teorema más importante de la Teoría Ergódica es el Teorema Ergódico de Birkhoff. Para probarlo necesitaremos de algunos resultados preliminares.

Proposición IV.1: (Cambio de variables).

Sean $(X, \mathcal{B}, \mu)$ un espacio de medida, $(Y, \mathcal{C})$ un espacio medible y $T: X \to Y$ una transformación medible. *Definamos* $T_*\mu = \mu \circ T^{-1}: \mathcal{C} \to [0, \infty)$ por $T_*\mu(C) = \mu(T^{-1}(C))$. Entonces:

(i) $T_*\mu$ es una medida sobre $\mathcal{C}$,

(i) Para toda función medible $f: Y \to \mathbf{C}$

$$\int_Y f d(T_*\mu) = \int_X f \circ T d\mu,$$

en el sentido de que si una integral existe, lo mismo ocurre con la otra y las dos son iguales.

Demostración:

Es fácil ver que $T_*\mu$ es una medida.

Cuando $f = \chi_E$ es la función característica de $E \in \mathcal{C}$

$$\int_Y \chi_E d(T_* u) = \mu(T^{-1}(E)) \qquad y$$

$$\int_X \chi_E \circ T du = \int_X \chi_{T^{-1}(E)} d\mu = \mu(T^{-1} E),$$

lo que implica que el resultado es verdadero por definición de $T_*\mu$. En consecuencia, la fórmula vale para funciones simples (i.e., combinaciones lineales de funciones características). Si $f: Y \to [0, \infty)$ es una función medible, entonces f es límite puntual de una sucesión creciente de funciones simples, y el resultado sigue del Teorema de Convergencia Monótona. Además, cualquier función medible $f: Y \to \mathbf{R}$ puede ser escrita como la diferencia $f = f^{+} - f^{-}$ de dos funciones medibles no negativas sobre Y y de este modo la fórmula es verdadera también en este caso. Finalmente, cualquier función medible $f: Y \to \mathbf{C}$ puede ser escrita como $f = Re(f) + iIm(f)$ donde $Re(f), Im(f): Y \to \mathbf{R}$ son funciones medibles. Luego la fórmula es verdadera en general. ∎

Teorema IV.2. (**Teorema Ergódico Maximal**)

Sean $(X, \mathcal{A}, \mu)$ un espacio de medida σ-finito, T una transformación que preserva medida y $f \in L^1(X, \mathcal{A}, \mu)$ de valores reales. Si

$$f^*(x) = \sup_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x),$$

entonces $f^*: X \to (\infty, \infty]$ es medible y

$$\int_{\{f^*>0\}} f d\mu \geq 0,$$

donde $\{f^* > 0\}$ denota el conjunto $\{x \in X: f^*(x) > 0\}$.

Demostración: (Debida a A. Garsía). Sean $f_0 = 0$, $f_n = f + f\circ T + \cdots + f \circ T^{n-1}$ y $F_N = \max_{0\leq n\leq N} f_n$. Observemos que

$$\{f^* > 0\} = \bigcup_{N=1}^{\infty} \{F_N > 0\}.$$

Así, basta probar que, para todo $N \geq 1$.

$$(1) \qquad \int_{\{F_N>0\}} f d\mu \geq 0.$$

Para cada $0 \leq n \leq \mathbf{N}$ se tiene que $F_N \geq f_n$ y de este modo $F_N \circ T + f \geq f_n \circ T + f = f_{n+1}$. Por consiguiente, si $N \geq 1$ y $y \in \{F_N > 0\}$,

$$\begin{aligned} F_N \circ T(y) + f(y) &\geq \max_{1 \leq n \leq N} f_n(y) \\ &= \max_{0 \leq n \leq \mathbf{N}} f_n(y) \qquad \text{pues} \quad y \in \{F_N > 0\} \\ &= F_N(y); \end{aligned}$$

i.e.,

$$f \geq F_N - F_N \circ T \quad \text{en} \quad \{F_N > 0\}. \tag{2}$$

Como $F_N, F_N \circ T \in L^1(X_1, \mathcal{A}, \mu)$, se sigue de (2) que:

$$\begin{aligned} \int_{\{F_N > 0\}} f &\geq \int_{\{F_N > 0\}} F_N - \int_{\{F_N > 0\}} F_N \circ T \\ &= \int_X F_N - \int_{\{F_N > 0\}} F_N \circ T \qquad \text{pues} \quad F_N \geq 0 \\ &\geq \int_X F_N - \int_X F_N \circ T \qquad \text{pues} \quad F_N \geq 0, \\ &= 0 \qquad \text{por la Proposición IV.1.} \end{aligned}$$

Esto termina la prueba de (1) y del Teorema. ∎

Corolario IV.2.1. *Si $\mu(X) < \infty$, se tiene que: para todo $\alpha \in \mathbf{R}$*

$$\alpha \mu(\{f^* > \alpha\}) \leq \int_{\{f^* > \alpha\}} f d\mu;$$

en particular,

$$\mu(\{x : f^*(x) = \infty\}) = 0.$$

Demostración: Como $\mu(X) < \infty$, $f - \alpha \in L^1(X, \mathcal{A}, \mu)$. Haciendo $g = f - \alpha$ vemos que $\{g^* > 0\} = \{f^* > \alpha\}$. Así, por el Teorema Ergódico Maximal IV.2,

$$0 \leq \int_{\{f^* > \alpha\}} (f - \alpha) d\mu = \int_{\{f^* > \alpha\}} f d\mu - \int_{\{f^* > \alpha\}} \alpha d\mu.$$

Luego

$$\int_{\{f^*>\alpha\}} f d\mu \geq \alpha u(\{f^* > \alpha\}). \qquad \blacksquare$$

Teorema IV.3. *(***Teorema Ergódico de Birkhoff***)*

Sean $(X, \mathcal{A}, \mu)$ un espacio de medida σ-finito, $T: X \to X$ una transformación que preserva medida y $f \in L^1(X, \mathcal{A}, \mu)$. Entonces:

(i) $\lim_{n\to\infty} \frac{1}{n} \left(\sum_{j=0}^{n-1} f \circ T(x) \right) = \tilde{f}(x)$ *existe c.t.p.;*

(ii) $\tilde{f}$ *es T-invariante i.e., $\tilde{f} \circ T = \tilde{f}$ c.t.p.; y*

(iii) $\tilde{f} \in L^1$ *y* $\left\|\tilde{f}\right\|_1 \leq \|f\|_1$.

Cuando $\mu(X) < \infty$ también se cumplen las dos condiciones siguientes:

(iv) $\frac{1}{n} \left(\sum_{j=0}^{n-1} f \circ T^j \right)$ *converge a $\tilde{f}$ en L^1;*

(v) $\int \tilde{f} d\mu = \int f d\mu$.

Demostración:

Basta probar este Teorema cuando $f \in L^1(X, \mathcal{A}, \mu)$ es de valores reales pues, una vez hecho esto, podemos escribir $g \in L^1(X, \mathcal{A}, \mu)$ como $g = g_1 + ig_2$ donde $g_1, g_2: X \to \mathbf{R}$.

Supondremos, de ahora en adelante, que $f: X \to \mathbf{R}$.

(i):

Para cada $\alpha, \beta \in \mathbf{R}$, con $\alpha < \beta$, sea

$$E_{\alpha,\beta} = \{x \in X: \liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) < \alpha < \beta <$$

$$< \limsup_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)\}$$

Probaremos que para cada α, β (con $\alpha < \beta$), $\mu(E_{\alpha,\beta}) = 0$. Luego la unión sobre todos los racionales $\alpha < \beta$ tendrá también medida cero y de este modo el límite de (i) existirá c.t.p. $x \in X$.

Cada $E_{\alpha\beta}$ es medible porque $\limsup$ y $\liminf$ de funciones medibles resultan siendo funciones medibles.

Afirmamos que

$$\mu(E_{\alpha,\beta}) < \infty. \tag{1}$$

En efecto, sea $C \subset E_{\alpha,\beta}$ medible de medida finita. Vamos a suponer que $\beta > 0$ pues, caso contrario, argumentaríamos con $-f$ y $-\alpha$.

Observemos que $x \in E_{\alpha,\beta}$ implica que, para algún n, $\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x) >$ β, de este modo $(f - \beta\chi_C)^*(x) := \sup_{n\geq 1} \frac{1}{n} \sum_{j=0}^{n-1} (f - \beta\chi_C) \circ T^j(x) > 0$, o sea

$$E_{\alpha,\beta} \subset \{(f - \beta\chi_C)^* > 0\}.$$

En consecuencia, aplicando el Teorema Ergódico Maximal IV.2 a la función $f - \beta\chi_C$, tenemos que

$$\begin{aligned} \int_{\{(f-\beta\chi_C)^*>0\}} f d\mu &\geq \beta \int_{\{(f-\beta\chi_C)^*>0\}} \chi_C \\ &\geq \beta \int_{E_{\alpha,\beta}} \chi_C = \beta\mu(C) \end{aligned}$$

En conclusión, si $C \subset E_{\alpha,\beta}$ tiene medida finita y $\beta > 0$ entonces

$$\mu(C) \leq \frac{1}{\beta} \int_{\{(f-\beta\chi_C)^*>0\}} f d\mu \leq \frac{1}{\beta} \int_X |f| \, d\mu.$$

Ahora, como X es σ-finito podemos escribir $E_{\alpha,\beta} = \bigcup_{n\geq 1} C_n$ de forma que $C_1 \subset C_2 \subset \ldots$ y $\mu(C_n) \leq \frac{1}{\beta} \int |f| \, d\mu$, lo que implica que $\mu(E_{\alpha,\beta}) \leq \frac{1}{\beta} \int |f| < \infty$. Esto prueba (1).

Es fácil ver que $T(E_{\alpha,\beta}) \subset E_{\alpha,\beta}$. Así, podemos aplicar el Corolario IV.2.1 a
$(f-\beta)\,|_{E_{\alpha,\beta}} \in L^1$ y $T\,|_{E_{\alpha,\beta}}: E_{\alpha,\beta} \to E_{\alpha,\beta}$ obteniendo

$$\beta\mu(E_{\alpha,\beta}) \leq \int_{E_{\alpha,\beta}} f d\mu$$

ya que $E_{\alpha,\beta} = \{x \in E_{\alpha,\beta}: f^*(x) > \beta\}$.

Argumentando con la pareja $(-f, -\alpha)$ de la manera como lo hicimos con (f, β), vemos que $E_{\alpha,\beta} \subset \{(-f)^* > -\alpha\}$ y que

$$\int_{E_{\alpha,\beta}} -f d\mu \geq -\alpha\mu(E_{\alpha,\beta}),$$

o sea

$$\int_{E_{\alpha,\beta}} f d\mu \leq \alpha\mu(E_{\alpha,\beta}).$$

Así

$$\beta\mu(E_{\alpha,\beta}) \leq \int_{E_{\alpha,\beta}} f d\mu \leq \alpha\mu(E_{\alpha,\beta}),$$

y como $\alpha < \beta$, esto último es posible sólo si $\mu(E_{\alpha,\beta}) = 0$. Esto termina la prueba de (i).

(ii):

Para c.t.p. $x \in X$ tenemos que:

$$\begin{aligned}
\tilde{f} \circ T(x) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^{j+1}(x)) \\
&= \lim_{n\to\infty} \left(\frac{1}{n} \sum_{j=0}^{n} f \circ T^j(x) - \frac{1}{n} f(x) \right) \\
&= \lim_{n\to\infty} \left(\frac{n+1}{n} . \frac{1}{n+1} \sum_{j=0}^{n} f \circ T^j(x) \right) \\
&= \tilde{f}(x).
\end{aligned}$$

(iii):

Sea $\widetilde{|f|}$ la función obtenida aplicando i) a $|f| \in \mathcal{L}'(X, \mathcal{A}, \mu)$. Como

$$\left|\frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j\right| \leq \frac{1}{n}\sum_{j=0}^{n-1} |f| \circ T^j$$

tenemos que $\left|\tilde{f}\right| \leq \widetilde{|f|}$. También, como T preserva μ (Ver Proposición IV.1), $\int \left|f \circ T^j\right| d\mu = \int |f| d\mu$ para $j \geq 0$. En consecuencia, usando el Lema de Fatou obtenemos

$$\begin{aligned}
\int \left|\tilde{f}\right| \leq \int \widetilde{|f|} &\leq \liminf_{n\to\infty} \int \frac{1}{n}\sum_{j=0}^{n-1} \left|f \circ T^j\right| \\
&= \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int \left|f \circ T^j\right| \\
&= \liminf_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \int |f| \\
&= \liminf_{n\to\infty} \int |f| \\
&= \int |f| < \infty.
\end{aligned}$$

En otras palabras (iii) es verdadero.

(iv):

En esta parte de la prueba usaremos la Proposición I.12. Como $\mu(X) < \infty$, $L^\infty(X, \mathcal{A}, \mu) \subset L^1(X, \mathcal{A}, \mu)$. Supongamos primero que $f \in L^\infty$; se sigue de (i) que

$$\left|\tilde{f}(x) - \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j(x)\right| \to 0 \quad \text{c.t.p.} \tag{2}$$

También, para c.t.p. x,

$$\begin{aligned}\left|\tilde{f}(x)\right| &\leq \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\left|f\circ T^j(x)\right| \\ &\leq \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\|f\|_\infty \\ &= \|f\|_\infty ,\end{aligned}$$

y así, para c.t.p. x,

$$\begin{aligned}&\left|\tilde{f}(x)-\frac{1}{n}\sum_{k=0}^{n-1}(f\circ T^j)(x)\right| \\ &\leq \|f\|_\infty+\frac{1}{n}\sum_{j=0}^{n-1}\left\|f\circ T^j\right\|_\infty \leq 2\,\|f\|_\infty .\end{aligned}$$

Por tanto la sucesión (2) está dominada (μ-c.t.p.) por una constante, luego podemos aplicar el Teorema de la convergencia dominada para concluir que $f\in L^\infty\subset L^1$ implica que

$$\text{(3)}\qquad \frac{1}{n}\sum_{j=0}^{n-1}(f\circ T^j) \text{ converge a } \tilde{f} \text{ en } L^1.$$

Sean $f\in L^1$ y $\varepsilon>0$. Por la densidad de L^∞ en L^1 existe $f_0\in L^\infty$ tal que $\|f-f_0\|_1<\varepsilon/3$; luego como $\tilde{f}-\tilde{f}_0=(\widetilde{f-f_0})$ se sigue de (ii) que

$$\left\|\tilde{f}-\tilde{f}_0\right\|_1\leq\|f-f_0\|_1<\varepsilon/3.$$

También, por (3), existe un entero N tal que, para todo $n\geq N$,

$$\left\|\tilde{f}_0-\frac{1}{n}\sum_{j=0}^{n-1}f_0\circ T^j\right\|_1<\varepsilon/3.$$

En consecuencia, para todo $n \geq N$,

$$\left\|\tilde{f} - \frac{1}{n}\sum_{j=0}^{n-1} f \circ T^j\right\|_1 \leq \left\|\tilde{f} - \tilde{f}_0\right\|_1 + \left\|\tilde{f}_0 - \frac{1}{n}\sum_{j=0}^{n-1} f_0 \circ T^j\right\|_1 + \left\|\frac{1}{n}\sum_{j=0}^{n-1}(f - f_0) \circ T^j\right\|_1 < \varepsilon,$$

pues

$$\left\|\frac{1}{n}\sum_{j=0}^{n-1}(f - f_0) \circ T^j\right\|_1 \leq \frac{1}{n}\sum_{j=0]}^{n-1} \|f - f_0\|_1 = \|f - f_0\|_1 < \varepsilon/3.$$

Como $\varepsilon > 0$ es arbitrario, esto termina la prueba de (iv).

(v):

Usando (iv) y el hecho que T preserva μ tenemos que:

$$\begin{aligned}
\int \tilde{f} d\mu &= \lim_{n\to\infty} \sum_{j=0}^{n-1} \int f \circ T^j d\mu \\
&= \lim_{n\to\infty} \sum_{j=0}^{n-1} \int f d\mu \qquad \text{(Prop. IV.1)} \\
&= \int f d\mu \qquad \blacksquare
\end{aligned}$$

Corolario IV.3.1. (**Teorema Ergódico L^p de Von Neumann**).

Sean $(X, \mathcal{A}, \mu)$ un espacio de probabilidad, $T: X \to X$ una transformación que preserva medida y $1 \leq p < \infty$. Si $f \in \mathcal{L}^p(X, ,\mu)$ entonces:

(i) $\lim_{n\to\infty} \frac{1}{n}\left(\sum_{j=0}^{n-1} f \circ T(x)\right) = \tilde{f}(x)$ *existe c.t.p.;*

(ii) $\tilde{f} \in L^p$ *y* $\left\|\tilde{f}\right\|_p \leq \|f\|_p$; *y*

(iii) $\frac{1}{n}\left(\sum_{j=0}^{n-1} f \circ T^j\right)$ *converge a* $\tilde{f}$ *en* L^p.

Demostración: Basta probar este resultado para el caso $f: X \to \mathbf{R}$. Como X tiene medida finita, $L^p \subset L^1$ y asi (i) del Teorema de Birkhoff implica (i) de este corolario. También $L^p \subset L^1$ implica que las pruebas de (iii) y (iv) del Teorema de Birkhoff pueden ser fácilmente adaptadas para probar (ii) y (iii), respectivamente, de este corolario. ∎

Corolario IV.3.2. *Para todo* $A, B \in \mathcal{A}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) = \int \tilde{\chi}_A \cdot \chi_B d\mu$$

Demostración: Para cada j,

$$\mu(T^{-j}(A) \cap B) = \int_X \chi_{T^{-j}(A)} \cdot \chi_B d\mu = \int_X \chi_A \circ T^j \cdot \chi_B d\mu.$$

Como $\chi_A \in L^2(X)$ se sigue del Teorema Ergódico L^p de Von Neumann que, en L^2,

$$\frac{1}{n} \sum_{j=0}^{n-1} \chi_A \circ T^j \to \tilde{\chi}_A.$$

Luego

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) \to \int_X \tilde{\chi}_A \cdot \chi_B d\mu. \quad ∎$$

En el Teorema de Birkhoff, la función $\tilde{f}$ llámase *promedio orbital* de f y $\int f d\mu$ llámase *promedio espacial de* f. Si χ_A es la función característica de un conjunto medible $A \in \mathcal{A}$ y $x \in A$, el promedio orbital de χ_A denótase por τ_A y $\tau_A(x)$ es llamado *promedio temporal de estadía de x en A.* La función τ_A es particularmente expresiva pues

$$\tau_A(x) = \lim_{n \to \infty} \frac{1}{n} \#\{0 \leq j \leq n-1 : T^j(x) \in A\}$$

y

$$\int_X \tau_A d\mu = \mu(A).$$

§V Transformaciones Ergódicas

Sea $(X, \mathcal{A}, \mu, T)$ un sistema dinámico medible. Decimos que $A \in \mathcal{A}$ es T-invariante si $\mu(A \triangle T^{-1}A) = 0$. Decimos que T es ergódica, o también que μ es ergódica, si $A \in \mathcal{A}$ T-invariante implica que $\mu(A)$ es cero o uno.

Lema V.1. *Sea $T: X \to X$ una transformación que preserva medida de un espacio de probabilidad $(X, \mathcal{A}, \mu)$. Sean $\mathcal{A}_0$ un álgebra que genera $\mathcal{A}$ y $A, B \in \mathcal{A}$. Para todo $\varepsilon > 0$ existen $A_0, B_0 \in \mathcal{A}_0$ tales que para todo entero $j \geq 0$,*

(i) $|\mu(A) \cdot \mu(B) - \mu(A_0) \cdot \mu(B_0)| < \varepsilon$

(ii) $\left|\mu(T^{-j}A \cap B) - \mu(T^{-j}A_0 \cap B_0)\right| < \varepsilon$

Demostración: Sea $\varepsilon > 0$ dado y sean $A, B \in \mathcal{B}$. Usando el Teorema de Aproximación I.5, escojamos $A_0, B_0 \in \mathcal{A}_0$ con $\mu(A \triangle A_0) < \varepsilon/2$, $\mu(B \triangle B_0) < \varepsilon/2$ y $|\mu(A) \cdot \mu(B) - \mu(A_0)\mu(B_0)| < \varepsilon$. Para todo $j \geq 0$, $(T^{-j}A \cap B) \triangle (T^{-j}A_0 \cap B_0) \subset (T^{-j}A \triangle T^{-j}A_0) \cup (B \triangle B_0)$, de este modo tenemos que $\mu(T^{-j}A \cap B) \triangle (T^{-j}A_0 \cap B_0) < \varepsilon$ y, como consecuencia, lo requerido en el lema. ∎

Teorema V.2. *Sea $(X, \mathcal{A}, \mu, T)$ un sistema dinámico medible. Sean $1 \leq p \leq \infty$ y S una semiálgebra que genera $\mathcal{A}$. Las siguientes afirmaciones son equivalentes:*

(i) *T es ergódica;*

(ii) *Para toda función medible $f: X \to \mathbf{R}$ tal que $f(Tx) \geq f(x)$ c.t.p., se tiene que f es constante c.t.p.;*

(iii) *Para toda función medible $f: X \mapsto \mathbf{C}$ que sea T-invariante (i.e., $f(Tx) = f(x)$, c.t.p.) se tiene que f es constante c.t.p.;*

(iv) *$f \in L^p$ implica que el promedio orbital $\tilde{f}$ es constante c.t.p.;*

(v) *$f \in L^p$ implica que el promedio orbital $\tilde{f} = \int f d\mu$ c.t.p.;*

(vi) para todo $A \in \mathcal{A}$ y para c.t.p. $x \in X$

$$\mu(A) = \tilde{\chi}_A(x) = \lim_n \frac{1}{n} \#\{0 \leq j \leq n-1 : T^j(x) \in A\};$$

(vii) Para todo $A, B \in S$ vale

$$\lim_n \frac{1}{n} \sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) = \mu(A) \cdot \mu(B).$$

Demostración:

$(i) \Rightarrow (ii)$:

Sea $f: X \to \mathbf{R}$ medible tal que $f \circ T(x) \geq f(x)$ c.t.p.; si f no es constante c.t.p. entonces existe $c \in \mathbf{R}$ tal que el conjunto medible

$$B = \{x: c \geq f(x)\}$$

satisface $0 < \mu(B) < 1$. Sin embargo $y \in T^{-1}(B)$ implica por hipótesis que

$$c \geq f(T(y)) \geq f(y) \quad \text{c.t.p. } y \in T^{-1}(B);$$

Así, $\mu(T^{-1}(B) \setminus B) = 0$ y como también $\mu(T^{-1}(B) = \mu(B)$ (pues T preserva μ), concluimos que $\mu(T^{-1}(B) \Delta B) = 0$. Por ergodicidad $\mu(B)$ es cero o uno, lo que es una contradicción.

(ii)$\Rightarrow$ (iii): Obvio.

(iii)$\Rightarrow$ (iv): Obvio

(iv) $\Rightarrow$ (v):

Dado $f \in L^p$, porque $L^p \subset L^1$ y por el Teorema Ergódico de Birkhoff,

$$\tilde{f} = \int \tilde{f} d\mu = \int f d\mu \quad \text{c.t.p.}$$

(v) $\Rightarrow$ (vi):

Se sigue directamente de la hipótesis pues si $A \in \mathcal{A}$ y $x \in A$ entonces $\chi_A \in L^p$ y

$$\frac{1}{n}\#\{0 \leq j \leq n-1 : T^j x \in A\} = \frac{1}{n}\sum_{j=0}^{n-1} \chi_A \circ T^j(x).$$

(vi) $\Rightarrow$ (vii):

Por hipótesis y por el Corolario IV.3.2

$$\begin{aligned}
\lim_n \frac{1}{n}\sum_{j=0}^{n-1} \mu(T^{-j}(A) \cap B) &= \int \tilde{\chi}_A \cdot \chi_B \\
&= \mu(A) \cdot \int \chi_B \\
&= \mu(A) \cdot \mu(B).
\end{aligned}$$

(vii) $\Rightarrow$ (i):

Como cada elemento del álgebra, $\mathcal{A}_0$ generada por S puede ser escrita como unión disjunta finita de miembros de S (Teorema I.1), se sigue que la propiedad de convergencia (vii) vale también para todos los elementos de $\mathcal{A}_0$.

Sea $\varepsilon > 0$ y sean $A, B \in \mathcal{A}$. Usando el Lema V.1 escojamos $A_0, B_0 \in \mathcal{A}_0$ tales que para todo $j \geq 0$

$$\begin{aligned}
&\left|\mu(T^{-j}A \cap B) - \mu(T^{-j}A_0 \cap B_0)\right| < \varepsilon \quad \text{y} \\
&|\mu(A) \cdot \mu(B) - \mu(A_0) \cdot \mu(B_0)| < \varepsilon.
\end{aligned}$$

Como por el Corolario IV.3.2, $\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)$ existe, tenemos que

$$\begin{aligned}
&\left|\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)-\mu(A_0)\cdot\mu(B_0)\right| \\
&\quad=\left|\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)-\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A_0\cap B_0)\right| \\
&\quad\leq \lim_n \frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(T^{-j}A\cap B)-\mu(T^{-j}A_0\cap B_0)\right|\leq\varepsilon,
\end{aligned}$$

y así

$$\left|\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)-\mu(A)\cdot\mu(B)\right|\leq 2\varepsilon.$$

Como $\varepsilon>0$ es arbitrario, esto implica que la propiedad de convergencia (vii) vale para todos los elementos de $\mathcal{A}$.

Si $A\in\mathcal{A}$ es T-invariante entonces, para todo $j\geq 0$, $\mu(T^{-j}A\triangle A)=0$ y de este modo $\mu(T^{-j}A\cap A^c)=0$, de donde se concluye que

$$\mu(A)\cdot\mu(A^c)=\lim_n \frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap A^c)=0,$$

luego $\mu(A)\in\{0,1\}$. ∎

Ahora presentamos algunos ejemplos de transformaciones ergódicas. Estos ejemplos fueron introducidos en § II.

(1) *Los desplazamientos de Bernoulli son ergódicos*:

Esto sigue del Teorema V.2,(vii) porque la condición

$$\frac{1}{n}\sum_{j=0}^{n-1}\mu(T^{-j}A\cap B)\to\mu(A)\mu(B)$$

es fácilmente verificada cuando A y B son cilindros.

(2) *Las rotaciones irracionales del círculo son ergódicas*

Sea $(S^1, \mathcal{B}, \mu)$ el espácio de probabilidad donde $S' = \{z \in \mathbf{C} / |z| = 1\}$, $\mathcal{B}$ es el álgebra de los borelianos y μ es la medida de Haar.

Veamos que la rotación $T(z) = az$ en S^1 es ergódica si, y sólo si, a no es raíz de la unidad. En efecto. Sea a una raíz de la unidad, entonces para algún entero $p \neq 0$, $a^p = 1$. Si $f(z) = z^p$ entonces $f \circ T = f$ pero f no es constante c.t.p. Luego, por el Teorema V.2,(iii), T no es ergódica. Recíprocamente, supongamos que para todo entero $n \neq 0$ $a^n \neq 1$ y que $f \circ T = f$ donde $f \in L^2(\mu)$. Usando las series de Fourier,

$$f(z) = \sum_{n=-\infty}^{\infty} b_n z^n \quad \text{y} \quad f \circ T(z) = f(az) = \sum_{n=-\infty}^{\infty} b_n a^n z^n,$$

lo que implica que $b_n(a^n - 1) = 0$ para todo n. Si $n \neq 0$ entonces $b_n = 0$ y de este modo $f = b_0$ es constante c.t.p. Luego, por el Teorema V.2,(iv), T es ergódica.

(3) *Los endomorfismos $z \to z^p$ del círculo, con $|p| > 1$, son ergódicos.*

Sea $p \neq 0$ un entero y sea $T: S^1 \to S^1$ dado por $Tz = z^p$. Sea μ la medida de Haar de S' (ver el ejemplo anterior). Ya vimos que T preserva μ. Probemos que si $|p| > 1$ entonces T es ergódica. Supongamos que $f \in L^2(\mu)$ y $f \circ T = f$. Si $f(z)$ tiene serie de Fourier $f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$ entonces $f(Tz) = \sum_{n=-\infty}^{\infty} a_n z^{pn}$. Por consiguiente $a_n = a_{pn} = a_{p^2 n} = a_{p^3 n} = \dots$ y así si $n \neq 0$ debemos tener que $a_n = 0$ porque los coeficientes de Fourier satisfacen $\sum_{j=-\infty}^{\infty} |a_j|^2 < \infty$. Luego $f(z) = a_0$ constante c.t.p y por el Teorema V.2,(iii) T es ergódico.

Sean X un espacio topológico y $T: X \to X$ una aplicación. Decimos que T *es transitiva* si existe $x \in X$ tal que la órbita positiva de x $\mathcal{O}^+(x) =$

$\{x, Tx, T^2x, \dots\}$ es densa en X. En el caso de espacios topológicos, el resultado siguiente establece una conexión entre ergodicidad y transitividad.

Teorema V.3. *Sea $(X, \mathcal{B}(X), \mu, T)$ un sistema dinámico medible, donde X es un espacio topológico. Si T es ergódica y μ es positiva en todo abierto (no vacío), entonces T es transitiva.*

Demostración: Sea $\{U_i\}_{i=1}^{\infty}$ una base enumerable de abiertos. Sea

$$A_i = \bigcup_{n=0}^{\infty} T^{-n}(U_i).$$

Como $T^{-1}(A_i) \subset A_i$ y $\mu(T^{-1}A_i) = \mu(A_i), \mu(T^{-1}A_i \Delta A_i) = 0$; i.e., cada A_i es T-invariante y así, por la ergodicidad de T, $\mu(A_i) = 1$. Esto implica que si

$$A := \bigcap_{i=1}^{\infty} A_i = \bigcap_{i=1}^{\infty} \bigcup_{n \geq 0} T^{-n}(U_i)$$

entonces $\mu(A) = 1$ y también que si $x \in A$ entonces, para cada i, existe un entero $n_i \geq 0$ tal que $T^{n_i}(x) \in U_i$, o sea la órbita de x es densa. ∎

Observaciones:

(1) Si X es de Baire,

$$\bigcap_{i=1}^{\infty} \bigcup_{n \geq 0} T^{-n}(A_i)$$

es residual.

(2) El recíproco del Teorema V.3 es falso. Furstemberg [Fur] construyó un difeomorfismo analítico del toro T^2, minimal (i.e. toda órbita densa), preservando la medida de Haar de T^2, positiva en abiertos pero que no es ergódica. La medida de Haar de $T^2 = S^1 \times S^1$ es la medida producto $\mu \times \mu$, donde μ es la medida de Haar de S^1.

Un par de condiciones equivalentes a la transitividad son presentadas en el siguiente resultado.

Proposición V.4. *Sea $T: X \to X$ una aplicación continua de un espacio topológico de Baire X con base enumerable. Las siguientes afirmaciones son equivalentes.*

(i) T es transitiva;

(ii) Existe un subconjunto residual $U \subset X$ tal que $x \in U$ implica que el conjunto ω-límite $\omega(x)$ es igual a X;

(iii) Para todo abierto $U \subset X$ el conjunto $\bigcup_{j \geq 1} T^{-j}(U)$ es denso en X.

Demostración:

$(i) \Rightarrow (iii)$:

Sean $x \in X$ tal que $\omega(x) = X$, $z \in X$ y $U \subset X$ un conjunto abierto. Como $\{z\} \cup U \subset \omega(x)$ tenemos en primer lugar, que dada una vecindad V de z existe un entero $n > 0$ tal que $T^n x \in V$ y en segundo lugar, que existe un entero $m > n$ tal que $T^m(x) \in U$. Se sigue de esto que

$$T^{-(m-n)}(U) \supset T^{-(m-n)}(\{T^m(x)\}) \supset \{T^n(x)\}.$$

Luego, $T^{-(m-n)}(U)$ interseca V. Como el punto z y la vecindad V de z son arbitrarios, $\bigcup_{j \geq 1} T^{-j}(U)$ es denso en X.

$(iii) \Rightarrow (ii)$:

Sea $\{U_n\}_{n=1}^{\infty}$ una base enumerable de abiertos de X. Como cada $S_n = \bigcup_{j \geq 1} T^{-j}(U_n)$ es abierto y denso, $S = \bigcap_{n \geq 1} S_n$ es residual. Resta probar que $x \in S$ implica que $\omega(x) = X$. En efecto, dado U_n, tenemos que $x \in S_n$ y de este modo existe $m \geq 1$ tal que $T^m x \in U_n$. Como U_n es arbitrario, esto prueba que $\{x, Tx, T^2x, \dots\}$ es denso en X y que $\omega(x) = X$.

$(ii) \Rightarrow (i)$ Obvio. ∎

§VI Mezcla

Sea $(X, \mathcal{A}, \mu, T)$ un sistema dinámico medible. Decimos que T es *mezclante* si, para todo $A, B \in \mathcal{A}$,

$$\lim_{j\to\infty} \mu(T^{-j}A \cap B) = \mu(A)\mu(B)$$

Proposición VI.1. *Sean $(X, \mathcal{B}, \mu)$ un espacio de probabilidad, S una semiálgebra que genera $\mathcal{A}$ y $T: X \to X$ una transformación que preserva medida. Entonces T es mezclante si, y sólo si, para todo $A, B \in S$*

(i) $$\lim_{j\to\infty} \mu(T^{-j}A \cap B) = \mu(A)\mu(B)$$

Demostración: La necesidad es obvia. Probemos la suficiencia. Como cada elemento del álgebra $\mathcal{A}_0$, generada por S, puede ser escrito como una unión disjunta finita de miembros de S (Teorema I.1) se sigue que la propiedad de convergencia (i) vale también para todos los elementos de $\mathcal{A}_0$.

Sea $\varepsilon > 0$ dado y sean $A, B \in \mathcal{A}$. Usando el Lema V.1 tenemos que existen $A_0, B_0 \in \mathcal{A}_0$ tales que, para todo $j \geq 0$,

$$|\mu(A) \cdot \mu(B) - \mu(A_0) \cdot \mu(B_0)| < \varepsilon \quad \text{y}$$
$$\left|\mu(T^{-j}A \cap B) - \mu(T^{-j}A_0 \cap B_0)\right| < \varepsilon.$$

Luego, si j es suficientemente grande,

$$\left|\mu(T^{-j}A \cap B)) - \mu(A) \cdot (B)\right| < 4\varepsilon.$$

Como A, B e $\varepsilon > 0$ son arbitrarios, T es mezclante. ∎

Sean X un espacio topológico y T una aplicación. Decimos que T es *topológicamente mezclante* si para todo par de abiertos no vacíos U, V existe $N \in \mathbf{N}$ tal que, para todo $n \geq N$, $\quad T^{-n}(U) \cap V \neq \phi$.

Las nociones de mezclante y topológicamente mezclante relaciónanse, en espacios topológicos, por el siguiente resultado.

Teorema VI.2. *Sea $(X, \mathcal{B}(X), \mu, T)$ un sistema dinámico medible donde X es un espacio topológico. Supongamos que μ es positiva en todo abierto (no vacío) y que T es una transformación mezclante. Entonces T es topológicamente mezclante.*

Demostración: Sean U y V abiertos no vacíos. Como T es mezclante

$$\lim_{n} \mu(T^{-n}U \cap V) = \mu(U) \cdot \mu(V) > 0.$$

Entonces para todo n suficientemente grande $\mu(T^{-n}U \cap V) > 0$ y de este modo $T^{-n}(U) \cap V \neq \phi$. ∎

Como en el caso de ergodicidad, podemos usar la Proposición VI.1 para concluir que los desplazamientos de Bernoulli (unilaterales y bilaterales) son mezclantes.

Es claro que toda transformación mezclante es ergódica; sin embargo la recíproca no es verdadera y un ejemplo de ello son las rotaciones irracionales del círculo (verifíquese).

El endomorfismo $z \to z^m$ del círculo, con $m \geq 2$, es mezclante. La prueba será hecha más adelante, en §IX, mostrando que él, como sistema dinámico medible, es el mismo objeto que el desplazamiento unilateral $B^+(m, p, \sigma)$ donde $p = \left(\frac{1}{m}, \frac{1}{m}, \ldots, \frac{1}{m}\right)$.

Transformación de Gauss

En la sección II probamos que la Transformación de Gauss $\varphi: [0,1) \to [0,1)$ definida por

$$\varphi(x) = \frac{1}{x} - [\frac{1}{x}] \quad \text{si } x \neq 0 \tag{1}$$
$$\varphi(0) = 0$$

preserva la medida μ dada por

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{d\lambda}{1+x}.$$

La transformación φ es mezclante (Ver [Bil]). Nosotros sólo probaremos que φ es ergódica siguiendo de manera bastante próxima a la presentación de Billingsley [Bil]. Para esto necesitaremos de algunos resultados de la Teoría de Números.

Decimos que una sucesión $\{x_n\}_{n\geq 0}$ de elementos de la semirecta extendida $(0,\infty] = (0,\infty) \cup \{\infty\}$ es *normal* si $x_j = \infty$ implica que $x_n = \infty$ para todo $n \geq j$. Definimos

$$|x_0| x_1 |x_2 \ldots | x_k = \cfrac{1}{x_0 + \cfrac{1}{x_1 + \cfrac{}{\ddots \cfrac{}{+x_{k-1} + \cfrac{1}{x_k}}}}}$$

donde usamos el convenio habitual $1/\infty = 0$. Cuando el límite $\lim\limits_{n\to\infty} |x_0| x_1 |x_2 \ldots | x_n$ exista, lo denotaremos por

$$|x_0| x_1 |x_2 \ldots$$

Asociemos a cada $x \in [0,1)$ la sucesión $a(x) = \{a_n\}_0^\infty$ en $\mathbf{N} \cup \{\infty\}$ cuyos términos están dados por

$$a_n = a_n(x) = \infty \quad \text{si } \varphi^n(x) = 0 \tag{2}$$

$$a_n = a_n(x) = \left[\frac{1}{\varphi^n(x)}\right] \quad \text{si } \varphi^n(x) \neq 0.$$

Ciertamente $a(x)$ es normal y la llamaremos *sucesión normal asociada* a x. Se sigue
de (1) que si $0 \leq x < 1$, entonces

(3) $$x = (a_0 + \varphi(x)) = \frac{1}{a_0 + \varphi(x)}.$$

Repitiendo el procedimiento obtenemos

$$x = |a_0|\,(a_1 + \varphi^2(x)) = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \varphi^2(x)}}$$

Más generalmente,

(4) $$x = |a_0|\,a_1 \ldots |a_{n-2}|\,(a_{n-1} + \varphi^n(x)).$$

Dada una sucesión normal $b = \{b_n\}_0^\infty$ de números naturales ≥ 1, definamos por inducción en n, los números naturales $p_n = p_n(b)$, $q_n = q_n(b)$ mediante las fórmulas

(5) $$p_{-1} = 1, \quad p_0 = 0, \quad p_n = b_{n-1}p_{n-1} + p_{n-2}, \quad n \geq 1;$$

$$q_{-1} = 0, \quad q_0 = 1, \quad q_n = b_{n-1}q_{n-1} + q_{n-2}, \quad n \geq 1.$$

Argumentando por inducción es fácil ver que

(6) $$p_{n-1}q_n - p_nq_{n-1} = (-1)^n, \quad n \geq 0,$$

(7) $$|b_0|\,b_1 \ldots |b_{n-2}|\,(b_{n-1} + t) = \frac{p_n + tp_{n-1}}{q_n + tq_{n-1}}, \quad t \in [0,1], \quad n \geq 1,$$

(8)
$$0 = \frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \frac{p_{2n-1}}{q_{2n-1}} < \cdots < \frac{p_3}{q_3} < \frac{p_1}{q_1} \leq 1, \quad n \geq 1,$$

y

$$p_n \geq 2^{(n-2)/2}, \quad q_n \geq 2^{(n-1)/2}, \quad n \geq 1. \tag{9}$$

Usando (6) obtenemos

$$\left| \frac{p_{n+1} + tp_n}{q_{n+1} + tq_n} - \frac{p_n}{q_n} \right| \leq \frac{1}{q_{n+1}q_n}, \quad n \geq 0, \quad t \in [0,1] \tag{10}$$

y de este modo por (9)

$$\left| \frac{(p_{n+1} + tp_n)/(q_{n+1} + tq_n)}{p_n/q_n} - 1 \right| \leq 2^{-n+1}, \quad n \geq 0, \quad t \in [0,1]. \tag{11}$$

Como la derivada de la función $x \to \log x$, en $x = 1$, es 1, se sigue de (11) y del teorema del Valor Medio aplicado a

$$\log \left(\frac{(p_{n+1} + tp_n)/(q_{n+1} + tq_n)}{p_n/q_n} \right) - \log(1)$$

que

$$\left| \log \frac{(p_{n+1} + tp_n)/(q_{n+1} + tq_n)}{p_n/q_n} \right| \leq 2^{-n+2}, \quad n \geq 0, \quad t \in [0,1] \tag{12}$$

Teorema VI.3.

(i) *Para toda sucesión normal* $b = \{b_n\}_0^\infty$ *de elementos de* $\mathrm{N} \cup \{\infty\}$, *el límite* $|b_0|\, b_1 |b_2 \ldots$ *existe, pertenece a* $[0,1)$ *y*

$$\big| |b_0|\, b_1 |b_2 \cdots - |b_0|\, b_1 \ldots |b_n| \big| \leq 2^{-n+2}, \quad n \geq 2;$$

(ii) *Para todo* $0 \leq x < 1$, *la sucesión normal* $b = \{b_n\}_0^\infty$ *asociada a* x *es la única que satisface*

$$x = |b_0|\, b_1 |b_2 \ldots$$

Además, x es racional $\Leftrightarrow$ $\varphi^n(x) = 0$ para algún $n \in \mathbf{N}$ $\Leftrightarrow$ $x = |b_0|\, b_1 \ldots |b_n$ para algún $n \in \mathbf{N}$.

Demostración: (i) Probemos este ítem solamente cuando todo $b_n \in \mathbf{N}$. Sea $p_n = p_n(b)$ y $q_n = q_n(b)$, $n = -1, 0, 1, \ldots$. Como, para $n \geq 2$ y $k \in \mathbf{N}$.

$$\left|\frac{p_n}{q_n} - \frac{p_{n+k}}{q_{n+k}}\right| \leq \sum_{j=0}^{k-1} \left|\frac{p_{n+j}}{q_{n+j}} - \frac{p_{n+1+j}}{q_{n+1+j}}\right|,$$

usando (9) y (10), podemos obtener

$$\left|\frac{p_n}{q_n} - \frac{p_{n+k}}{q_{n+k}}\right| \leq \frac{1}{2^{n-2}}, \quad n \geq 2, \quad k \geq 0.$$

En consecuencia, por (7), concluimos que el límite $|b_0|\, b_1 |b_2 \ldots$ existe y satisface la desigualdad requerida. Se sigue de (8) que $|b_0|\, b_1 |b_2 \ldots$ es un elemento de $[0, 1)$.

(ii) Supongamos que para todo $n \in \mathbf{N}$, $\varphi^n(x) \neq 0$ (i.e., $b_n \neq 0$). Se sigue de (4) y (7) que

$$(13) \qquad x = |b_0|\, b_1 \ldots |b_{n-2}|\, (b_{n-1} + \varphi^n(x)) = \frac{p_n + \varphi^n(x) p_{n-1}}{q_n + \varphi^n(x) q_{n-1}}, \quad n \geq 1,$$

donde $p_n = p_n(b)$ y $q_n = q_n(b)$. Así, usando (9) y (10) concluimos que

$$x = |b_0|\, b_1 |b_2 \ldots$$

Vamos a probar ahora que si $\{c_n\}_0^\infty$ es una sucesión normal de elementos de $\mathbf{N} \cup \{\infty\}$ tal que $x = |c_1|\, c_1 |c_2 \ldots$ entonces, para todo $n \in \mathbf{N}$, $b_n = c_n$. Se sigue de (i) que $s_1 = |b_1|\, b_2 |b_3 \ldots$ y $t_1 = |c_1|\, c_2 |c_3 \ldots$ son elementos de $[0, 1)$; luego

$$|b_0|\, b_1 |b_2 \cdots = \frac{1}{b_0 + s_1} = \frac{1}{c_0 + t_1} = |c_0|\, c_1 |c_2 \ldots$$

con $b_0 \in \mathbf{N}\setminus\{0\}$ y $c_0 \in \mathbf{N}\cup\{\infty\}$, implica que $b_0 = c_0$ y $s_1 = t_1$. Repitiendo el proceso obtenemos que $|b_1|b_2|b_3\cdots = |c_1|c_2|c_3\ldots$ implica que $b_1 = c_1$ y $|b_2|b_3|b_4\cdots = |c_2|c_3|c_4\ldots$. Así, por inducción, concluimos que $b_n = c_n$ para todo n.

Lo que resta por probar de (ii) es dejado al lector. ∎

Sean $b_0, b_1, \ldots, b_{n-1}$ enteros positivos y sea $\Delta(b_0, b_1, \ldots, b_{n-1})$ el conjunto de los $x \in [0,1)$ tales que $a_0(x) = b_0$, $a_1(x) = b_1, \ldots, a_{n-1}(x) = b_{n-1}$. El conjunto $\Delta(b_0, b_1, \ldots, b_{n-1})$ será llamado intervalo fundamental de rango n. Observemos que $\Delta(b_0, b_1, \ldots, b_{n-1})$ es la imagen de $[0,1)$ mediante la función $\psi_{(b_0,b_1,\ldots,b_{n-1})}$ definida por

$$\psi_{(b_0,b_1,\ldots,b_{n-1})}(t) = |b_0|b_1\ldots|(b_{n-1}+t), \quad 0 \le t < 1. \tag{14}$$

Por la forma de la función (ver (8)) se sigue que $\psi_{(b_0,b_1,\ldots,b_{n-1})}$ es creciente para n par y decreciente para n impar. Por (7) tenemos

$$\psi_{(b_0,b_1,\ldots,b_{n-1})}(t) = \frac{p_n + tp_{n-1}}{q_n + tq_{n-1}},$$

donde p_n y q_n son definidos inductivamente, en término de los b_k, por las fórmulas (5). Por consiguiente

$$\Delta(b_0, b_1, \ldots, b_{n-1}) = \left[\frac{p_n}{q_n}, \frac{p_n + p_{n-1}}{q_n + q_{n-1}}\right) \quad \text{si } n \text{ es par,}$$

$$\Delta(b_0, b_1, \ldots, b_{n-1}) = \left[\frac{p_n + p_{n-1}}{q_n + q_{n-1}}, \frac{p_n}{q_n}\right) \quad \text{si } n \text{ es impar.}$$

Se sigue de (6) que

$$\lambda(\Delta(b_0, b_1, \ldots, b_{n-1})) = \frac{1}{q_n(q_n + q_{n+1})}$$

donde λ denota la medida de Lebesgue. Es fácil ver que los intervalos fundamentales de rango n determinan una partición de $[0,1)$ en intervalos cuya longitud (por (9) y (15)) es de, a lo más, 2^{-n+1}. En particular la clase de intervalos fundamentales genera la σ-álgebra $\mathcal{A}$ de los borelianos de $[0,1)$.

Dada una probabilidad $m: \mathcal{A} \to [0,1]$ y dados $A, B \in \mathcal{A}$ con $m(B) > 0$, denotaremos

$$m(A|B) = \frac{m(A \cap B)}{m(B)}.$$

Probaremos ahora que φ es ergódica con respecto a μ. Fijemos $b_0, b_1, \ldots, b_{n-1}$ y escribamos ψ por $\psi_{(b_0,b_1,\ldots,b_{n-1})}$ y Δ_n por $\Delta(b_0, b_1, \ldots, b_{n-1})$. El intervalo Δ_n tiene longitud $|\psi(1) - \psi(0)|$ y, si $0 \leq y < z \leq 1$, el intervalo

$$\{x : y \leq T^n x < z\} \cap \Delta_n$$

tiene longitud $|\psi(z) - \psi(y)|$. En consecuencia

$$\lambda(\varphi^{-n}[y,z)|\Delta_n) = \frac{\psi(z) - \psi(y)}{\psi(1) - \psi(0)}.$$

Por (14) y (6)

$$\lambda(\varphi^{-n}[y,z)|\Delta_n) = (z-y)\frac{q_n(q_n + q_{n-1})}{(q_n + yq_{n-1})(q_n + zq_{n-1})}. \tag{16}$$

Como

$$2 \geq \frac{q_n + q_{n-1}}{q_n} > 0, \quad n \geq 0,$$

la expresión de la derecha en (16) está entre 1/2 y 2 y así

$$\frac{1}{2}\lambda(A) \leq \lambda(\varphi^{-n}A|\Delta_n) \leq 2\lambda(A), \tag{17}$$

donde escribimos A por $[y,z)$. Pero entonces (17) vale también si A es una unión disjunta de intervalos y, consecuentemente, para cualquier $A \in \mathcal{A}$.

Como, para todo $x \in [0,1)$, $1/2 \leq 1/(1+x) \leq 1$, se sigue de la definición de la medida de Gauss μ que

$$\frac{\lambda(M)}{2\log 2} \leq \mu(M) \leq \frac{\lambda(M)}{\log 2}, \quad M \in \mathcal{A}. \tag{18}$$

Se sigue de (17) y (18) que

$$C^{-1}\mu(A) \leq \mu(\varphi^{-n}A|\Delta_n) \leq C\mu(A), \quad \forall A \in \mathcal{A}, \tag{19}$$

donde $C = 4/\log 2$.

Supongamos que A es invariante por φ. Entonces

$$C^{-1}\mu(A) \leq \mu(A|\Delta_n), \qquad \text{o sea, } C^{-1}\mu(A) \leq \frac{\mu(A \cap \Delta_n)}{\mu(\Delta_n)},$$

y en consecuencia, caso $\mu(A) > 0$,

$$C^{-1}\mu(\Delta_n) \leq \frac{\mu(A \cap \Delta_n)}{\mu(A)} = \mu(\Delta_n|A).$$

Luego,

$$C^{-1}\mu(E) \leq \mu(E|A) \tag{20}$$

vale para uniones disjuntas finitas E de intervalos fundamentales; como estos conjuntos forman un álgebra generando $\mathcal{A}$, (20) vale para cualquier $E \in \mathcal{A}$. Tomando $E = A^c$, vemos que $\mu(A)$ debe ser 1. Luego φ es ergódica con respecto a μ.

Se sigue del Teorema Ergódico de Birkhoff que si f es una función integrable en el intervalo unitario, entonces

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(\varphi^k(x)) = \frac{1}{\log 2}\int_0^1 \frac{f(x)}{1+x}dx \quad c.t.p.$$

Aquí "integrable" y "c.t.p." se refieren tanto a μ como λ gracias a la relación (18).

Para obtener una aplicación a aproximación Diofántica, mostraremos que

$$\lim_{n\to\infty} \frac{1}{n} q_n(a(x)) = \frac{\pi^2}{12\log 2}, \quad c.t.p. \quad x. \tag{21}$$

Probemos primeramente que

$$\frac{1}{q_n(a(x))} = \prod_{k=0}^{n-1} (|a_k(x)|\, a_{k+1}(x) \dots |a_{n-1}(x)). \tag{22}$$

Se sigue inductivamente por (5) que $p_{j+1}(a(x)) = q_j(a(\varphi(x)))$, y por consiguiente

$$\frac{1}{q_n(a(x))} = \prod_{k=1}^{n} \frac{p_{n+1-k}(a(\varphi^{k-1}(x)))}{q_{n+1-k}(a(\varphi^{k-1}(x)))}$$

lo que es exactamente (22).

Ahora, por (13)

$$x = \frac{p_n(a(x)) + \varphi^n(x) p_{n-1}(a(x))}{q_n(a(x)) + \varphi^n(x) q_{n-1}(a(x))}, \quad n \geq 1, \quad x \in [0,1),$$

y de este modo por (12) y (7)

$$\left|\log \varphi^{k-1}(x) - \log(|a_{k-1}(x)|\, a_k(x) \dots |a_{n-1}(x))\right| \leq 2^{-n+k+1} \quad 1 \leq k \leq n,$$

y, en consecuencia, por (22),

$$\log \frac{1}{q_n(a(x))} = \sum_{k=1}^{n} \log(\varphi^{k-1}(x)) + \sum_{k=1}^{n} \theta \cdot 2^{-n+k-1},$$

donde aquí, y en lo que sigue, θ es un número, cuyo valor no es el mismo en cada ocurrencia, satisfaciendo $|\theta| \leq 1$. Por consiguiente,

$$\frac{1}{n} \log q_n(a(x)) = -\frac{1}{n} \sum_{k=0}^{n-1} \log(\varphi^k(x)) + \frac{4\theta}{n}. \tag{23}$$

Por el Teorema Ergódico

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\sum_{k=0}^{n-1}\log(\varphi^k(x))\right\} = -\frac{1}{\log 2}\int_0^1 \frac{\log x dx}{1+x} \qquad c.t.p.$$

Integración por partes reduce esta integral a

$$\begin{aligned}\frac{1}{\log 2}\int_0^1 \frac{\log(1+x)}{x}dx &= \frac{1}{\log 2}\sum_{k=0}^{\infty}(-1)^k\int_0^1 \frac{x^k}{k+1}dx \\ &= \frac{1}{\log 2}\sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)^2} = \frac{\pi^2}{12\log 2}\end{aligned}$$

Así, (21) sigue de (23).

Veamos algunas consecuencias de (21). Por (13) y (10)

$$\left|x - \frac{p_n(a(x))}{q_n(a(x))}\right| = \frac{1}{q_{n+1}(a(x))q_n(a(x))},$$

así, por (21),

$$\lim_{n\to\infty}\frac{1}{n}\log\left|x - \frac{p_n(a(x))}{q_n(a(x))}\right| = -\frac{\pi^2}{6\log 2}, \quad \text{c.t.p.}$$

Esto quiere decir que la discrepancia entre x y su aproximante $\frac{p_n(a(x))}{q_n(a(x))}$ es c.t.p. x del orden de $e^{-n\pi^2/(6\log 2)}$. Similarmente se sigue, por (15), que si $\Delta_n(x)$ es el intervalo fundamental de rango n que contiene x, entonces

$$\lim_{n\to\infty}\frac{1}{n}\log\lambda(\Delta_n(x)) = -\frac{\pi^2}{6\log 2} \quad \text{c.t.p. } x.$$

Finalmente, por (18), tenemos

$$\lim_{n\to\infty}\frac{1}{n}\log\mu(\Delta_n(x)) = -\frac{\pi^2}{6\log 2} \quad \text{c.t.p. } x.$$

§VII Desplazamientos de Markov: Mezcla y ergodicidad

Sea $A = (a_{ij})$ una matriz. Decimos que A es *no negativa* (resp. *positiva*) si $a_{ij} \geq 0$ (resp. $a_{ij} > 0$) para todo i, j. Una matriz cuadrada $A = (a_{ij})_0^{n-1}$ no negativa es irreducible si para todo par i, j existe algún entero $k > 0$ tal que $a_{ij}^{(k)} > 0$ donde $a_{ij}^{(k)}$ es el (i, j)-ésimo elemento de A^k. Una matriz irreducible A es *aperiódica* si existe $k > 0$ tal que $a_{ij}^{(k)} > 0$ para todo i, j.

Necesitaremos del siguiente resultado (ver [Gau]).

Teorema VII.1. *(Teorema de Perron-Frobenius). Sea $A = (a_{ij})_0^{n-1}$ una matriz $n \times n$ no negativa. Entonces:*

(i) existe un autovalor $\lambda \geq 0$ tal que ningún otro autovalor de A tiene valor absoluto mayor que λ;

(ii) $\min_i \left(\sum_{j=0}^{n-1} a_{ij}\right) \leq \lambda \leq \max_i \left(\sum_{j=0}^{n-1} a_{ij}\right)$;

(iii) asociados a λ existen un autovector fila a izquierda $u = (u_0, u_1, \ldots, u_{n-1})$ y un autovector columna a derecha

$$v = \begin{pmatrix} v_0 \\ v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_{n-1} \end{pmatrix}$$

ambos no negativos;

(iv) si A es irreducible entonces λ es un autovalor simple y los autovectores correspondientes son positivos;

(v) si A es aperiódica entonces $\lambda > |\mu|$ para todo autovalor $\mu \neq \lambda$ de A.

Es fácil probar, por inducción en ℓ, lo siguiente:

Lema VII.2. *Sea* $A = (a_{ij})_0^{n-1}$ *una matriz* $n \times n$. *Entonces para todo entero* $\ell \geq 2$, *la* (i,j)*-ésima entrada* $a_{ij}^{(\ell)}$ *de* A^ℓ *satisface*

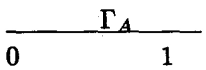
$$a_{ij}^{(\ell)} = \sum_{k_i=0}^{n-1} a_{ik_1} a_{k_1k_2} \cdot a_{k_{\ell-1}j}.$$

Sea $A = (a_{ij})_0^{n-1}$ una matriz $n \times n$ no negativa. Asociémosle un gráfico Γ_A formado por los n puntos del conjunto $\{0, 1, \ldots, n-1\}$ y exactamente por un arco orientado $[i; j]$ de i hacia j siempre y cuando $a_{ij} > 0$. Observemos que Γ_A es un gráfico conexo por caminos orientados si, y sólo si, para todo (i,j), existe una sucesión finita $j_1, j_2, \ldots, j_\ell$, con $j_1 = 1$ y $j_\ell = j$, tal que

$$a_{j_1j_2} a_{j_2j_3} \ldots a_{j_{\ell-1}j_\ell} > 0,$$

pues el camino orientado $[j_1; j_2] * \cdots * [j_{\ell-1}; j_\ell]$ (obtenido por yuxtaposición) conecta i con j. Observemos que si

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$



entonces Γ_A es conexo, mas no conexo por caminos orientados.

Para obtener una matriz irreducible o para reconocerla tenemos:

Proposición VII.3. *Sea* $A = (a_{ij})_0^{n-1}$ *una matriz* $n \times n$ *no negativa. Son equivalentes:*

(i) A es irreducible;

(ii) El gráfico Γ_A *es conexo por caminos orientados.*

Demostración:

(i) $\Rightarrow$ (ii):

Como A es irreducible tenemos que para todo i, j existe un entero $N \geq 1$ tal que

$$0 < a_{ij}^{(N)} = \sum_{k_i=0}^{n-1} a_{ik_1} a_{k_1k_2} \ldots a_{k_{N-1}j}.$$

Así, algún $a_{ik_1}a_{k_1k_2}\dots a_{k_{N-1}j} \neq 0$, pues (siendo A no negativa) todos los sumandos son ≥ 0.

(ii) $\Rightarrow$ (i):

Por hipótesis, dado (i,j) arbitrario existe un camino orientado

$$[i;k_1] * [k_1;k_2] * \cdots * [k_{N-1};j]$$

en Γ_A; luego $a_{ij}^{(N)} \neq 0$. Así, A es irreducible. ∎

Recordamos que los desplazamientos de Markov fueron introducidos en (5) de §II. Tenemos lo siguiente:

Teorema VII.4. *Sea $(B(n),p,P,\sigma)$ un desplazamiento de Markov con p positivo. Entonces:*

(i) $(B(n),p,P,\sigma)$ es ergódico si, y sólo si, la matriz estocástica P es irreducible;

(ii) $(B(n),p,P,\sigma)$ es mezclante si, y sólo si, P es aperiódica.

La prueba de este teorema está contenida en las proposiones VII.6 y VII.7 siguientes.

Lema VII.5. *Sea P una matriz estocástica que tiene un vector de probabilidad p positivo tal que $pP = p$. Entonces*

$$Q = \lim_N \frac{1}{N}(I + P + \cdots + P^N)$$

existe. La matriz Q también es estocástica, $QP = PQ = Q$, y todo autovector de P asociado al autovalor 1 es también autovector de Q. Además, $Q^2 = Q$.

Demostración: Seja χ_j la función característica del cilindro $C(0;j)$. Por el Teorema de Birkhoff

$$\frac{1}{N}\sum_{k=0}^{N-1} \chi_j(\sigma^{-k}x) \rightarrow \tilde{\chi}_j(x) \quad \mu - \text{c.t.p.}$$

Entonces, para todo i,

$$\frac{1}{N}\sum_{k=0}^{N-1}\chi_i\cdot\chi_j\circ\sigma^{-k}\to\chi_i\tilde{\chi}_j \quad \mu-\text{c.}t.p.$$

Por el Teorema de la Convergencia Dominada,

$$\frac{1}{N}\sum_{k=0}^{N-1}\int\chi_i\cdot\chi_j(\sigma^{-k})d\mu\mapsto\int\chi_i\tilde{\chi}_j d\mu.$$

Ahora bien, si $Y=\{0,1,\ldots,n-1\}$ entonces

$$\begin{aligned}\frac{1}{N}\sum_{k=0}^{N-1}\int\chi_i\cdot\chi_j(\sigma^{-k})d\mu &= \frac{1}{N}\sum_{k=0}^{N-1}\int\chi_i\cdot\chi_{C(k,j)}d\mu\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\mu(C(0;i)\cap C(k;j))\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{\ell_i\in Y}p_ip_{i\ell_1}p_{\ell_1\ell_2}\cdots p_{\ell_{k-1}j}\\ &= p_i\left(\frac{1}{N}\sum_{k=0}^{N-1}p_{ij}^{(k)}\right)\end{aligned}$$

siendo que la expresión dentro del paréntesis es la (i,j) entrada de la matriz

$$\frac{1}{N}\sum_{k=0}^{N-1}P^k,$$

donde P^0 denota la matriz identidad. Luego, existe una matriz

$$Q=\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}P^K$$

con entradas

$$q_{ij} = \frac{1}{p_i} \int \chi_i \cdot \tilde{\chi}_j d\mu \geq 0.$$

Si $R = (r_{ij})$ as una matriz estocástica, entonces $PR = (x_{ij})$ es también estocástica. En efecto,

$$\begin{aligned}
\sum_{j=0}^{n-1} x_{ij} &= \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} p_{ik} r_{kj} \\
&= \sum_{k=0}^{n-1} p_{ik} \sum_{j=0}^{n-1} r_{kj} \\
&= \sum_{k=0}^{n-1} p_{ik} = 1.
\end{aligned}$$

Análogamente $P + R$ es estocástica y de este modo $\frac{1}{N} \sum_{j=0}^{N-1} P^j$ y (consecuentemente) Q son estocásticas.

En estas circunstancias es fácil ver que $QP = Q = PQ$, que $Q^2 = Q$ y que $pQ = p$. ∎

Proposición VII.6. *Sea* $(B(n), p, P, \sigma)$ *un desplazamiento de Markov (con* p *positivo). Sea* Q *la matriz obtenida en el Lema VII.5. Las siguientes afirmaciones son equivalentes:*

(i) La (p, P)*-medida de Markov* μ *es ergódica;*

(ii) Todas las filas de la matriz $Q = (q_{ij})_0^{n-1}$ *son idénticas. Cuando este es el caso, para todo* (i, j), $p_j = q_{ij}$;

(iii) Toda entrada de $Q = (q_{ij})$ *es positiva;*

(iv) P *is irreducible;*

(v) 1 es un autovalor simple de P.

Demostración: Denotemos $Y = \{0, 1, \ldots, n-1\}$ y probemos primeramente que cuando las filas de la matriz $Q = (q_{ij})$ son idénticas, digamos a

$q = (q_0, q_1, \ldots, q_{n-1})$, entonces $p = q$. En efecto, se sigue del Lema VII.5 que $pQ = p$, o sea

$$p_j = \sum_{i=0}^{n-1} p_i q_{ij} = \sum_{i=0}^{n-1} p_i q_j = q_j,$$

para todo j. Esto prueba la implicación enunciada en (ii).

(i) $\Rightarrow$ (ii):

Por el Teorema V.2,(vii),

$$\frac{1}{N} \sum_{k=0}^{N-1} \mu(C(0;i) \cap C(k;j)) \to \mu(C(0;i)) \cdot \mu(C(0;j) = p_i \cdot p_j$$

Además, ya vimos en la prueba del Lema VII.5 que

$$\frac{1}{N} \sum_{k=0}^{N-1} \mu(C(0;i) \cap C(k;j)) \to p_i \cdot q_{ij}.$$

Luego $q_{ij} = p_j$.

(ii) $\Rightarrow$ (iii):

Se sigue del hecho de que p es positivo.

(iii) $\Rightarrow$ (iv):

Sea $(i,j) \in Y \times Y$ fijo. Como $\frac{1}{N} \sum_{k=0}^{N-1} p_{ij}^{(k)} \mapsto q_{ij} > 0$, existe $k = k(i,j)$ tal que $p_{ij}^{(k)} > 0$.

(iv) $\Rightarrow$ (iii):

Fijemos $i \in Y$ y sea $S_i = \{k \in Y | q_{ik} > 0\}$. Como $Q = QP^m$ para todo $m \geq 0$ y también $q_{st} \geq 0$ para todo (s,t), obtenemos que para todo $m \geq 0$ y para todo $\ell \in Y$,

$$q_{ij} = \sum_{k=0}^{n-1} q_{ik} p_{kj}^{(m)} \geq q_{i\ell} p_{\ell j}^{(m)}. \tag{1}$$

Se tiene que $S_i \neq \phi$, pues Q estocástica implica que $\sum_{k=0}^{n-1} q_{ik} = 1$ y $q_{ik} \geq 0$; así, para algún k, $q_{ik} > 0$. Afirmamos que $S_i = Y$; en efecto, si $(\ell, j) \in S_i \times Y$ entonces en primer lugar, por definición de S_i, $q_{i\ell} > 0$ y en segundo lugar, por ser P irreducible, existe $m \geq 0$ tal que $p_{\ell j}^{(m)} > 0$, de donde (como consecuencia de (1)) $q_{ij} > 0$, o sea $j \in S_i$.

(iii) $\Rightarrow$ (ii):

Fijemos $j \in Y$ y sea $q_j = \max_i q_{ij}$. Si para algún $i \in Y$ $q_{ij} < q_j$ entonces

(2) $$\text{para todo } \ell \in Y, \quad q_{\ell j} < q_j$$

pues, como $Q^2 = Q$,

$$q_{\ell j} = \sum_{i=0}^{n-1} q_{\ell i} q_{ij} < q_j \sum_{i=0}^{n-1} q_{\ell i} = q_j.$$

No obstante (2) es imposible $\forall \ell \in Y$, pues $q_j = \max_\ell q_{\ell j}$. Así, para todo $(i, j) \in Y \times Y$, $q_{ij} = q_j > 0$.

(ii) $\Rightarrow$ (i): Por el Teorema V.2,(vii), basta probar que para toda pareja C_1, C_2 de cilindros

$$\frac{1}{N} \sum_{k=0}^{N-1} (\sigma^{-k} C_1 \cap C_2) \to \mu(C_1)\mu(C_2).$$

Sean $C_1 = C(a; i_0, \ldots, i_r)$ $C_2 = C(b; j_0, \ldots, j_s)$. Como para todo $k > b + s - a$ tenemos que

$$\mu(\sigma^{-k} C_1 \cap C_2) = p_{j_0} p_{j_0 j_1} \cdots p_{j_{s-1} j_s} p_{j_s i_0}^{(a+k-b-s)} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{r-1}} \cdot i_r$$

y también, como sabemos que (ii) implica $q_{ij} = p_j$, obtenemos

$$\begin{aligned}
\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} \mu(\sigma^{-k} C_1 \cap C_2) &= \lim_{N \to \infty} \frac{1}{N} \sum_{k > |b+s-a|}^{N-1} \mu(\sigma^{-k} C_1 \cap C_2) \\
&= p_{j_0} \cdot p_{j_0 j_1} \cdot \ldots \cdot p_{j_{s-1} j_s} (p_{i_0}) \cdot p_{i_0 i_1} \cdot \ldots \cdot p_{i_{r-1} i_r} \\
&= \mu(C_1)\mu(C_2)
\end{aligned}$$

(ii) $\Rightarrow$ (v):

Como (ii) implica que la matriz Q es aperiódica, por el Teorema de Perron-Frobenius y el Lema VII.5 concluimos que los únicos autovectores a izquierda de Q para el autovalor 1 son los múltiplos de p. En consecuencia, aun por el Lema VII.5, estos múltiplos de p son los únicos autovectores de P para el autovalor 1.

(v) $\Rightarrow$ (ii):

Supongamos que 1 es un autovalor simple de P. Como $Q = QP$ (Lema VII.5) cada fila de Q es un autovector a izquierda y en consecuencia todas ellas son idénticas entre sí.

Proposición VII.7. *Sea $(B(n), p, P, \sigma)$ un desplazamiento de Markov con p positivo. Sea μ la (p, P)-medida de Markov. Las siguientes afirmaciones son equivalentes:*

(i) P es aperiódica;

(ii) Para todo (i, j), $\lim\limits_N p_{ij}^{(N)} = p_j$;

(iii) μ es mezclante.

Demostración: (i) $\Rightarrow$ (ii). Como P es estocástica, se sigue del Teorema de Perron-Frobenius que $\lambda = 1$ es un autovalor simple de P y los otros autovalores de P tienen módulo menor que 1. Así, podemos escribir $\mathbf{R}^n = F + p\mathbf{R}$, donde $p\mathbf{R} = \{p\alpha / \alpha \in \mathbf{R}\}$, de tal forma que lo siguiente es satisfecho:

(1) Para todo (vector fila) $f \in F$, $fP \in F$; además, para algún $0 < b < 1$, todos los autovalores de la aplicación lineal $f \to fP$, de F en sí mismo, tienen módulo menor que b.

Sea e_j la j-ésima fila de la matriz identidad $n \times n$ y escribamos $e_j = f_j + p\alpha_j$, donde $f_j \in F$ e $\alpha_j \in \mathbf{R}$. Se sigue de (1) que $f_j P^N \to 0$ cuando $N \to \infty$. Así, para cada j, $\lim\limits_N e_j \cdot P^N = p\alpha_j$ existe y consecuentemente, para todo (i, j)

$$\lim_N p_{ij}^{(N)} \quad \text{existe.}$$

Como, por la Proposición VII.6,

$$\lim_N \frac{1}{N} \sum_{k=0}^{N-1} p_{ij}^{(k)} = p_j$$

concluimos que $\lim_N p_{ij}^{(N)} = p_j$.

(ii) $\Rightarrow$ (i):

Como $\lim_N p_{ij}^{(N)} = p_j > 0$, para todo (i,j), existe $N_{ij} > 0$ tal que, para todo $N \geq N_{ij}$, $p_{ij}^{(N)}$ está próximo de $p_j > 0$ y como consecuencia P es aperiódica.

(ii) $\Rightarrow$ (iii):

Por la Proposición VI.1 basta probar que para toda pareja C_1, C_2 de cilindros

$$\lim_k \mu(\sigma^{-k} C_1 \cap C_2) = \mu(C_1)\mu(C_2).$$

Sean $C_1 = C(a; i_0, \ldots, i_r)$ y $C_2 = C(b; j_0, \ldots, j_s)$. Tenemos que

$$\begin{aligned}
&\lim_k (\sigma^{-k} C_1 \cap C_2) \\
&\quad = \lim_k p_{j_0} p_{j_0 j_1} \cdot \ldots \cdot p_{j_{s-1} j_s} \cdot p_{j_s i_0}^{(a+k-b-s)} \cdot p_{i_0 i_1} \cdot \ldots \cdot p_{i_{r-1} i_r} \\
&\quad = p_{j_0} p_{j_0 j_1} \cdot \ldots \cdot p_{j_{s-1} j_s} p_{i_0} p_{i_0 i_1} \cdot p_{i_{r-1} i_r} \\
&\quad = \mu(C_1)\mu(C_2).
\end{aligned}$$

(iii) $\Rightarrow$ (ii):

Como μ es mezclante

$$\lim_N \mu(C(0;i) \cap C(N;j)) = p_i p_j.$$

Además,

$$\mu(C(0;i) \cap C(N;j)) = \sum_{\ell_i \in Y} p_i p_{i\ell_1} \cdot p_{\ell_1 \ell_2} \ldots p_{\ell_{N-1} j} = p_i \cdot p_{ij}^{(N)},$$

donde $Y = \{0, 1, \ldots, n-1\}$. Luego $p_{ij}^{(N)} \to p_j$ cuando $N \to \infty$. ∎

§VIII Desplazamientos: Aspectos topológicos y medibles

Esta sección tiene por finalidad dar una idea de la conexión existente entre la Teoría Ergódica y la Dinámica Topológica. Para un estudio más profundo de este tópico recomendamos el libro de R. Mañé [Mañ] y los artículos [Bw1], [Bw2] y [Bw3] de R. Bowen.

Sea (X, d) un espacio métrico y $T: X \to X$ un homeomorfismo. Decimos que T es *expansivo* si existe $\epsilon > 0$, llamado *una constante de expansividad* de T, tal que si $x, y \in X$ y $d(T^k x, T^k y) \leq \epsilon$ para todo $k \in \mathbf{Z}$ entonces $x = y$.

Sea X un espacio metrizable compacto cuya topología la denotamos por τ. Sea $T: X \to X$ un homeomorfismo. Supongamos que T es *expansivo*, o sea que T lo es con relación a alguna métrica d en X que es compatible con la topología τ. Afirmamos que T también es expansivo con relación a cualquier otra métrica $\tilde{d}$ en X que induzca la topología τ (i.e., la propiedad para T de ser expansivo no depende de la métrica particular d). En efecto, por hipótesis, la aplicación identidad

$$Id: (X, \tilde{d}) \to (X, d)$$

es un homeomorfismo y por la compacidad de X, Id es uniformemente continua; así, para todo $\epsilon > 0$ existe $\delta > 0$ tal que $x, y \in X$, $\tilde{d}(x, y) \leq \delta \Rightarrow d(x, y) \leq \epsilon$. Consecuentemente, si tomamos $\epsilon > 0$ como siendo una contante de expansividad para $T: (X, d) \to (X, d)$, tendremos que $\delta > 0$ será una constante de expansividad para $T: (X, \tilde{d}) \to (X, \tilde{d})$.

Decimos que un espacio topológico es *totalmente disconexo* si cada uno de sus puntos posee una base de vecindades abiertas y cerradas simultáneamente.

Recordemos (ver (4) de § II) que el desplazamiento $\sigma: B(n) \mapsto B(n)$ es un homeomorfismo del espacio topológico $B(n)$. Este espacio es compacto y tiene como base de abiertos a los cilindros que son también compactos; consecuentemente $B(n)$ es totalmente disconexo.

Lema VIII.1. *$B(n)$ es un espacio metrizable y el desplazamiento $\sigma: B(n) \to B(n)$ es expansivo.*

Demostración: Sean τ la topología de $B(n)$ y sea $d(.,.)$ la métrica en $B(n)$ dada por

$$d(\alpha, \beta) = \sum_{k \in \mathbf{Z}} \frac{1}{2^{|k|}} |\alpha(k) - \beta(k)|.$$

Demostremos que τ coincide con la topología asociada a d.

Sea $\alpha \in B(n)$ y sea C una vecindad de α en la topología τ. Se puede suponer que C es de la forma

$$C = C(j; \alpha(j), \alpha(j+1), \ldots, \alpha(j+\ell))$$

y de este modo tendremos que

$$U = \{\beta \in B(n): d(\alpha, \beta) < \frac{1}{2^{|j|+\ell+1}}\} \subset C,$$

pues $\beta \in U$ implica que, para todo $0 \leq s \leq \ell$, $|\alpha(j+s) - \beta(j+s)| = 0$, o sea $\beta \in C$.

Recíprocamente, sea $\alpha \in B(n)$. Fijemos $r > 0$ y mostremos que existe una vecindad C de α para la topología τ tal que

$$C \subset U = \{\beta \in B(n): d(\alpha, \beta) < r\}.$$

Tomemos C de la forma

$$C = C(-j; \alpha(-j), \alpha(-j+1), \ldots, \alpha(j))$$

donde $j \geq 2$ es un entero tal que $n \cdot 2^{1-j} < r$. De este modo $\beta \in C$ implica que, para todo $|k| \leq j$, $\alpha(k) = \beta(k)$. Luego

$$
\begin{aligned}
d(\alpha, \beta) &= \sum_{|k|>j} \frac{1}{2^{|k|}} |\alpha(k) - \beta(k)| \\
&\leq 2n \sum_{k>j} \frac{1}{2^k} = \frac{n}{2^{j-1}} < r.
\end{aligned}
$$

Esto termina la prueba de que $\beta(n)$ es metrizable.

Afirmamos que $\epsilon = \frac{1}{2}$ es una constante de expansividad de σ. En efecto, supongamos que $\alpha, \beta \in B(n)$ y $d(\sigma^\ell \alpha, \sigma^\ell \beta) \leq \frac{1}{2}$, para todo $\ell \in \mathbf{Z}$; entonces, si fijamos ℓ, tendremos que

$$
\begin{aligned}
\frac{1}{2} \geq d(\sigma^\ell \alpha, \sigma^\ell \beta) &= \sum_{k \in \mathbf{Z}} \frac{1}{2^k} |\alpha(k+\ell) - \beta(k+\ell)| \\
&\geq |\alpha(\ell) - \beta(\ell)|,
\end{aligned}
$$

y de este modo $\alpha(\ell) = \beta(\ell)$. Como ℓ es arbitrario, $\alpha = \beta$. ∎

Sea X un espacio topológico y $T: X \to X$ un homeomorfismo. La *órbita* de $x \in X$ es el conjunto $\mathcal{O}(x) = \{T^n x : n \in \mathbf{Z}\}$. Cuando $\mathcal{O}(x) = \{x\}$ es un punto decimos que x es un *punto fijo de* T. Si $\mathcal{O}(x)$ es finito decimos que x es *periódico* y el menor entero $n > 0$ tal que $T^n(p) = p$ es llamado el período de p.

La equivalencia entre las estruturas de órbitas de los homeomorfismos es expresada mediante equivalencia topológica:

Sean $T: X \to X$, $S: Y \to Y$ homeomorfismos entre espacios topológicos. Decimos que T y S son *topológicamente equivalentes* si existe un homeomorfismo $h: X \to Y$ tal que $h \circ T = S \circ h$; cuando éste es el caso, escribiremos $T \sim S$. La equivalencia topológica es una relación de equivalencia en el conjunto de los pares (X, T) donde X es un espacio topológico y $T: X \to X$ es un homeomorfismo. Si $T \sim S$ entonces $h \circ T^n \sim S^n \circ h$, para todo entero n, y así $h(\mathcal{O}_T(x)) = \mathcal{O}_S(z)$ si $z = h(x)$; esto es, h lleva órbitas de T sobre órbitas de S y, en particular, lleva puntos periódicos de T a puntos periódicos de S del mismo período. La propiedad $T \sim S$ indica que T y S son esencialmente el mismo homeomorfismo módulo un "cambio de cordenadas".

Decimos que un subconjunto $\Lambda \subset B(n)$ es un ***subdesplazamiento*** si es compacto e invariante por σ (i.e., $\sigma(\Lambda) = \Lambda$).

Teorema VIII.2. *Sea X un espacio topológico y $T: X \to X$ un homeomorfismo. Son equivalentes:*

(i) T es topológicamente conjugado a un subdesplazamiento.
(ii) X es un espacio metrizable compacto totalmente disconexo y T es expansivo.

Demostración:
(i) $\Rightarrow$ (ii):

Este es el contenido del lema VIII.1 y de las observaciones que le precedieron.

(ii) $\Rightarrow$ (i):

Sea d una métrica en X compatible con su topología y sea $\epsilon > 0$ una constante de expansividad para (X, d). Como X es totalmente disconexo, podemos escoger para cada $x \in X$ una vecindad V_x de x que sea abierta y cerrada, simultáneamente, y que tenga diámetro $< \epsilon$ (i.e., $d(y, z) < \epsilon$, para todo $y, z \in V_x$). Por la compacidad de X,

$$X = \bigcup_{i=0}^{n-1} V_x;$$

para algunos $x_0, x_1, \ldots, x_{n-1} \in X$. Denotemos $U_0 = V_{x_0}$, $U_1 = V_{x_1} \setminus V_{x_0}$, $U_2 = V_{x_2} \setminus (V_{x_0} \cup V_{x_1}), \ldots, U_{n-1} = V_{x_{n-1}} \setminus \bigcup_{i=0}^{n-2} V_{x_i}$. Ciertamente, los U_j, $j = 0, 1, 2, \ldots, n-1$, son abiertos y cerrados simultáneamente, disjuntos dos a dos, tienen diámetro $< \epsilon$ y cubren X.

Sea $I: X \to B(n)$ dada por

$$I(x)(k) = e(T^k(x)), \quad \text{para todo } k \subset \mathbf{Z}.$$

donde $e: X \to \{0, 1, \dots, n-1\}$ está definida por $e(x) = j$ si, y sólo si, $x \in U_j$.

Veremos que I es continua e inyectiva; como X es compacto, de esto seguirá que I es un homeomorfismo de X sobre el subdesplazamiento $I(X)$ de $B(n)$.

Mostremos pues que I es continua en $x \in X$. Si V es una vecindad dada de $I(x)$, escogemos un cilindro $C \subset V$ que tiene la forma siguiente

$$C = C(-N; e(T^{-N}x), e(T^{-N+1}x), \dots, e(T^N x)),$$

donde $N \in \mathrm{N}$. Podemos encontrar una vecindad W_x de x en X tal que, $\forall k \in \{-N, -N+1, \dots, N\}$ y para todo $z \in W_x$, $e(T^k x) = e(T^k z)$ (o sea, $T^k x$ y $T^k z$ pertenecen al mismo U_j). Esto implica que si $z \in W_x$ entonces $I(z) \in C$; i.e., I es continua en x.

Probemos que I es inyectiva. Sean $x, z \in X$ tales que $I(x) = I(z)$; o sea, para todo $k \in \mathbf{Z}$, $e(T^k x) = e(T^k z)$ y de este modo $d(T^k x, T^k z) < \epsilon$, pues todo U_k tiene diámentro menor que ϵ. Como ϵ es una constante de expansividad para $T: (X, d) \to (X, d)$, concluimos que $x = z$. ∎

Decimos que el subdesplazamiento $\Lambda \subset B(n)$ es de *tipo finito* si existe una matriz $A = (a_{ij})_0^{n-1}$ con entradas en $\{0, 1\}$ y tal que $\theta \in \Lambda$ si, y sólo si,

$$a_{\theta(k)\theta(k+1)} = 1$$

para todo $k \in \mathbf{Z}$. En este caso denotamos $\Lambda = B_A(n)$. Recíprocamente, dada una matriz $A = (a_{ij})_0^{n-1}$ con entradas en $\{0, 1\}$ definimos

$$\Lambda = \{\theta \in B(n): a_{\theta(k)\theta(k+1)} = 1, \text{ para todo } k \in \mathbf{Z}\}.$$

Obviamente Λ es invariante por σ. Es fácil ver que $B(n) \setminus \Lambda$ es abierto y de este modo concluir que Λ es compacto.

Veamos algunos ejemplos:

Si A es la matriz identidad $n \times n$ entonces $B_A(n)$ tiene exactamente n elementos θ_i, $i = 0, 1, \ldots, n-1$, que son puntos fijos de σ; ellos están definidos por

$$\theta_i(k) = i,$$

para todo $k \in \mathbf{Z}$.

Si $A = (a_{ij})_0^{n-1} = (1)_0^{n-1}$ entonces $B_A(n) = B(n)$.

Si $A = (a_{ij})_0^{n-1}$ satisface $a_{ij} = 0$ para todo $0 \leq i \leq j \leq n-1$, entonces $B_A(n) = \phi$.

Sea $(X, \mathcal{B}(X), \mu, T)$ un sistema dinámico medible donde X es un espacio topológico. El *Soporte de* μ, $\mathrm{supp}(\mu)$, es el conjunto de los $x \in X$ tales que para toda vecindad abierta V de x, $\mu(V) > 0$.

Teorema VIII. 3. *Sea $(B(n), p, P, \sigma)$ un desplazamiento de Markov con p positivo. El soporte $\mathrm{supp}(\mu)$ de la (p, P)-medida de Markov μ es el subdesplazamiento de tipo finito $B_A(n)$ determinado por la matriz $A = (a_{ij})_0^{n-1}$ definida por:*

$$\begin{array}{lll} a_{ij} = 1 & \textit{si} & p_{ij} > 0, \quad \textit{y} \\ a_{ij} = 0 & \textit{si} & p_{ij} = 0, \end{array}$$

donde $P = (p_{ij})_0^n$.

Demostración: Sean $\theta \in B_A(n)$ y V una vecindad abierta de θ en $B(n)$. Afirmamos que $\mu(V) > 0$. En efecto, V contiene, un cilindro C de la forma

$$C = C(k; \theta(k), \theta, (k+1), \ldots, \theta(k+\ell)),$$

así

$$\mu(V) \geq \mu(C) = p_{\theta(k)} p_{\theta(k)\theta(k+1)} \cdots p_{\theta(k+\ell-1)\theta(k+\ell)} > 0,$$

pues p positiva $\Rightarrow p_{\theta(k)} > 0$ y $\beta \in B_A(n) \Rightarrow p_{\theta(s)\theta(s+1)} > 0$, para $s = k, k+1, \ldots, k+\ell-1$. Esto prueba que $B_A(n) \subset$ supp (μ).

Recíprocamente si $\beta \in B(n) \setminus B_A(n)$, existe $k \in \mathbf{Z}$ tal que $a_{\theta(k)\theta(k+1)} = 0$, lo que implica que $p_{\theta(k)\theta(k+1)} = 0$ y de este modo $\mu(C(k; \theta(k), \theta(k+1)) = 0$. En otras palabras $\beta \notin \mathrm{supp}(\mu)$. En conclusión $B_A(n) = \mathrm{supp}(\mu)$. ∎

§IX Isomorfismos de estructura

Para poder clasificar los sistemas dinámicos medibles es necesario establecer una relación de equivalencia entre ellos. De dos sistemas relacionados, uno de ellos siempre puede ser obtenido del otro por un "cambio de coordenadas". En esta sección definimos con precisión esta equivalencia y damos ejemplos de cómo ella puede ser construida.

Sean $(X_1, \mathcal{A}_1, \mu_1)$ y $(X_2, \mathcal{A}_2, \mu_2)$ espacios de probabilidad. Supongamos que $T: X_1 \to X_2$ es una ***transformación inversible que preserva medida***, o sea: T es biyectiva, T, T^{-1} son medibles, y T preserva medida. En estas condiciones:

Proposición IX.1. *$T^{-1}: X_2 \to X_1$ preserva medida.*

Demostración: Si $A_1 \in \mathcal{A}_1$ entonces $TA_1 \in \mathcal{A}_2$ y, como T preserva medida $\mu_1(T^{-1}TA_1) = \mu_2(TA_1)$. Luego $\mu_2((T^{-1})^{-1}A_1) = \mu_2(TA_1) = \mu_1(T^{-1}TA_1) = \mu_1(A_1)$. ∎

Decimos que los espacios de probabilidad $(X_1, \mathcal{A}_1, \mu_1)$ y $(X_2, \mathcal{A}_2, \mu_2)$ ***son equivalentes*** si existen $M_1 \in \mathcal{A}_1$, $M_2 \in \mathcal{A}_2$, con $\mu_1(M_1) = 1 = \mu_2(M_2)$, y existe una transformación $\phi: M_1 \to M_2$ inversible que preserva medida. Estamos suponiendo que el espacio M_i está equipado con la σ-álgebra

$$M_i \cap \mathcal{A}_i = \{M_i \cap A : A \in \mathcal{A}_i\}$$

y la restricción de μ_i a esta σ-álgebra. Supongamos que $T_1: X_1 \to X_1$ y $T_2: X_2 \to X_2$ son transformaciones que preservan medida. Decimos que T_1 ***es equivalente a*** T_2 (o que $(X_1, \mathcal{A}_1, \mu_1, T_1)$ ***es equivalente a*** $(X_1, \mathcal{A}_2, \mu_2, T_2)$ si existen $M_1 \in \mathcal{A}_1$, $M_2 \in \mathcal{A}_2$ con $\mu_1(M_1) = 1 = \mu_2(M_2)$ tales que
(i) $T_1(M_1) \subset M_1$, $T_2(M_2) \subset M_2$, y

(ii) Hay una transformación inversible que preserva medida $\Phi\colon M_1 \to M_2$ con $\Phi \circ T_1(x) = T_2 \circ \Phi(x)$, $\forall x \in M_1$.

Cuando queramos dar énfasis al papel de los conjuntos M_1, M_2 y de la función ϕ diremos que T_1 es equivalente a T_2 según la terna (M_1, M_2, ϕ).

Es fácil ver que lo que acabamos de definir es una relación de equivalencia entre los sistemas dinámicos medibles. También es fácil ver que si un sistema dinámico medible es ergódico (resp. mezclante), entonces cualquier otro equivalente a él tendrá la misma propiedad.

Nos gustaría observar que, de lo que probaremos más adelante y del hecho de que el desplazamiento unilateral $B^+(2), p, \sigma)$ es mezclante (como sigue fácilmente de la Proposición VI.1 aplicada a la semiálgebra de los cilindros), concluiremos que la aplicación del intervalo $[0,1)$ en sí mismo dada por $x \to 2x$ (mod (1)) también es mezclante.

Tenemos el siguiente resultado sobre equivalencia de espacios de medida. Una prueba puede ser encontrada en [Roy, pg. 327].

Teorema IX.2. *Sean X un espacio métrico separable completo y $\mathcal{B}(X)$ el álgebra de sus borelianos. Sea μ una medida de probabilidad en $\mathcal{B}(X)$ con $\mu(\{x\}) = 0$ para todo subconjunto unitario $\{x\} \subset X$. Sean $\mathcal{B}([0,1])$ el álgebra de los borelianos de $[0,1]$ y $\lambda\colon \mathcal{B}([0,1]) \to [0,1]$ la medida de Lebesgue. Entonces los espacios de medida $(X, \mathcal{B}(X), \mu)$ y $([0,1], \mathcal{B}([0,1], \lambda)$ son equivalentes:*

Ejemplos de transformaciones equivalentes

(1) *La aplicación $T\colon z \to z^m$ del círculo y la aplicación $S\colon x \to mx - [mx]$ del intervalo $[0,1)$ son equivalentes.*

Aquí los espacios de medida son los canónicos (ver ejemplos (2) e (3) de §II). Una transformación inversible que preserva medida y que puede ser usada para confirmar esta afirmación es la aplicación $\Phi\colon M_1 \to M_2$ dada por $\Phi(x) = e^{2\pi i x}$, donde

$$M_1 = \{x \in [0,1)/S^k(x) \neq 0, \forall k \in \mathbb{N}\}$$
$$M_2 = \{z \in S^1/T^k(z) \neq 0, \forall k \in \mathbb{N}\}$$

(2) *El desplazamiento unilateral* $B^+(2,p,\sigma)$ *y la aplicación* $\varphi\colon x \to 2x$ *(mod 1) del intervalo* $[0,1)$ *son equivalentes*, donde $p = (\frac{1}{2}, \frac{1}{2})$.

Estas dos transformaciones fueron introducidas en §II. Procedamos a probar esta afirmación.

Primeramente veamos que si $x \in [0,1)$ tiene una representación diádica de la forma

$$x = \sum_{n=1}^{\infty} \frac{\alpha_n}{2^n},$$

entonces

$$\varphi(x) = \sum_{n=1}^{\infty} \frac{\alpha_{n+1}}{2^n}.$$

En efecto, si $\alpha_1 = 0$ entonces

$$\varphi(x) = \sum_{n=2}^{\infty} \frac{\alpha_n}{2^{n-1}} = \sum_{n=1}^{\infty} \frac{\alpha_{n+1}}{2^n};$$

si $\alpha_1 = 1$ entonces

$$\varphi(x) = 2x - 1 = 2\left(\sum_{n=2}^{\infty} \frac{\alpha_n}{2^n} + \frac{1}{2}\right) - 1 = \sum_{n=2}^{\infty} \frac{\alpha_n}{2^{n-1}} = \sum_{n=1}^{\infty} \frac{\alpha_{n+1}}{2^n}.$$

Observemos que un número real tiene por lo menos una representación diádica y a lo sumo dos. Pero $t \in [0,1)$ tiene dos representaciones si, y sólo si, t es de la forma

$$t = \sum_{n=1}^{\infty} \frac{\alpha_n}{2^n} = \sum_{n=1}^{m} \frac{\alpha_n}{2^n}$$

siendo $\alpha_n = 0$ para $n > m$ y $\alpha_m = 1$, con $m \geq 1$ entero. La otra representación es entonces

$$t = \sum_{n=1}^{m-1} \frac{\alpha_n}{2^n} + \sum_{n=m+1}^{\infty} \frac{1}{2^n}.$$

Sea M el conjunto de los $\theta \in B^+(2)$ tales que ambos $\{n \in \mathbf{N} : \theta(n) = 1\}$ y $\{n \in \mathbf{N} : \theta(n) = 0\}$ son conjuntos infinitos; entonces $\sigma(M) = M$. Además, si $\Phi: M \to [0,1)$ es definido por

$$\Phi(\theta) = \sum_{n=1}^{\infty} \frac{\theta(n)}{2^n},$$

entonces (como todo elemento de $\Phi(M)$ tiene representación diádica única) Φ es una biyección entre M y $\Phi(M)$.

Ciertamente

$$\varphi(\Phi(M)) = \Phi(M),$$

$$\varphi \circ \Phi = \Phi \circ \sigma|_M.$$

Como $B(2) \setminus M$ y $[0,1] \setminus \Phi(M)$ son conjuntos enumerables, M y $\Phi(M)$ son borelianos tales que

$$\mu(M) = \lambda(\Phi(M)) = 1$$

donde μ es la probabilidad de $B^+(2, \varphi, \sigma)$ y λ es la medida de Lebesgue de $[0,1]$.

Para probar la requerida equivalencia de (2), resta verificar que $\Phi: M \to \Phi(M)$ es una transformación inversible que preserva medida.

Dados $a_1, a_2, \ldots, a_m \in \{0,1\}$, com $m \geq 1$. Denotemos por $D(0; a_1, a_2, \ldots, a_m)$ el intervalo semiabierto $[x, y)$ de extremos

$$x = \sum_{j=1}^{m} \frac{a_j}{2^j}, \quad y = \sum_{j=1}^{m} \frac{a_j}{2^j} + \frac{1}{2^{m+1}}.$$

La colección de todos estos intervalos (junto con el conjunto vacío) forma una semiálgebra que genera la σ-álgebra de los borelianos de $[0,1)$; pues, los extremos de estos intervalos forman un subconjunto denso de $[0,1)$. Observando que los cilindros

$$C(0; a_1, a_2, \ldots, a_m) \text{ de } B^+(2), \text{ com } m \geq 1,$$

forman una semiálgebra que genera la σ-álgebra de los borelianos y

$$\Phi(M \cap C(0; a_1, a_2, \ldots, a_m)) = \Phi(M) \cap D(0; a_1, a_2, \ldots, a_m),$$

$$\mu(C(0; a_1, \ldots, a_m)) = \frac{1}{2^m} = \lambda(D(0, a_1, \ldots, a_m),$$

para todo cilindro $C(0; a_1, a_2, \ldots, a_m)$, conchuimos que $\Phi: M \to \Phi(M)$ es una transformación inversible que preserva medida (Ver Proposición II.1).
(3) *El desplazamiento unilateral* $B^+(n, p, \sigma)$ *y la aplicación* $\varphi: x \to nx(\mathit{mod}\ 1)$ *del intervalo* $[0, 1)$ *son equivalentes donde* $p = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ *y* $n \geq 2$ *entero.*

La prueba es análoga a la del ejemplo (2) y será omitida.

§X Entropía

Afirmamos en la sección anterior que ergodicidad y mezcla son invariantes (i. e., preservados) por equivalencia de los sistemas dinámicos medibles. En esta sección asociaremos a cada sistema dinámico medible $(X, \mathcal{A}, \mu, T)$ un número real no negativo denotado por $h(T)$, llamado entropía de T, que será un invariante por equivalencia mucho mas sensible que ergodicidad y mezcla. Él nos servira para diferenciar bastantes sistemas dinámicos medibles que no son equivalentes.

Sea $(X, \mathcal{A}, \mu)$ un espacio de probabilidad. Un subconjunto finito

$$\mathcal{P} = \{A_1, A_2, \dots, A_m\} \subset \mathcal{A}$$

es una *partición* de $(X, \mathcal{A}, \mu)$ si

(i) $i \neq j \Rightarrow \mu(A_i \cap A_j) = 0$;

(ii) $\mu(X - \bigcup_{i=1}^{m} A_i) = 0$;

(iii) $\mu(A_i) > 0$ para todo i.

Cada A_i será llamado *átomo de la partición* $\mathcal{P}$.

Si $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_N$ son particiones finitas, definimos $\bigvee_{n=1}^{N} \mathcal{P}_n = \mathcal{P}_1 \vee \mathcal{P}_2 \vee \cdots \vee \mathcal{P}_N$ como la partición cuyos átomos son de la forma $A_1 \cap A_2 \cap \cdots \cap A_N$ con $A_i \in \mathcal{P}_i$ y $\mu(A_1 \cap A_2 \cap \cdots \cap A_N) > 0$.

Si $T: X \to X$ es una transformación que preserva medida en un espacio de probabilidad $(X, \mathcal{A}, \mu)$, la entropía $h(T)$ de T es definida en tres etapas: Sea $\mathcal{P}$ una partición finita de $(X, \mathcal{A}, \mu)$. Definimos la entropía $H(\mathcal{P})$ de $\mathcal{P}$ por:

$$H(\mathcal{P}) = -\sum_{A \in \mathcal{P}} \mu(A) \log(\mu(A)).$$

A continuación definimos la entropía de T relativamente a $\mathcal{P}$ por

$$h(T,\mathcal{P}) = \limsup_{k} \frac{1}{k} H\left(\bigvee_{j=0}^{k-1} T^{-j}\mathcal{P}\right)$$

donde $T^{-j}\mathcal{P} = \{T^{-j}A : A \in \mathcal{P}\}$. Probaremos más adelante que este límite superior es de hecho un límite ordinario. Finalmente, la entropía de T es

$$h(T) = \sup h(T,\mathcal{P})$$

donde el supremo es tomado sobre todas los particiones finitas $\mathcal{P}$ de $(X,\mathcal{A},\mu)$.

Proposición X.1. *Supongamos que los sistema dinámicos medibles $(X,\mathcal{A},T,\mu)$ y $(\tilde{X},\tilde{\mathcal{A}},\tilde{T},\tilde{\mu})$ son equivalentes. Entonces $h(T) = h(\tilde{T})$.*

Demostración: Supongamos que T es equivalente a $\tilde{T}$ según la terna $(M,\tilde{M},\phi)$. Así a cada partición $\mathcal{P}$ de $(X,\mathcal{A},\mu)$ corresponde una partición $\tilde{\mathcal{P}} = \{\phi(A \cap M) : A \in \mathcal{P}\}$ de $\tilde{X}$ y recíprocamente. Ciertamente $H(\mathcal{P}) = H(\tilde{\mathcal{P}})$ y también $h(T,\mathcal{P}) = h(\tilde{T},\tilde{\mathcal{P}})$. Tomando supremos obtenemos $h(T) = h(\tilde{T})$. ∎

Kolmogorov y Sinai probaron el siguiente resultado fundamental que necesitaremos aquí: Si $T\colon X \to X$ es una transformación inversible que preserva medida y $\mathcal{P}$ es una partición finita de $(X,\mathcal{A},\mu)$ tal que $\mathcal{A}$ coincide con la σ-algebra $\bigvee_{j=-\infty}^{\infty} T^j\mathcal{P}$ generada por $\bigcup_{j=-\infty}^{\infty} T^j(\mathcal{P})$, entonces $h(T) = h(T,\mathcal{P})$.

Más adelante probaremos la validez de este resultado junto con algunas variaciones. Por ahora veamos cómo puede ser utilizado.

(1) Sea $B(n,p,\sigma)$ el desplazamiento de Bernoulli con $p = (p_0,p_1,\ldots,p_{n-1})$ y σ-álgebra asociada $\mathcal{A}$. Mostraremos que

$$h(\sigma) = -\sum_{i=0}^{n-1} p_i \log p_i.$$

En efecto, tomemos la partición $\mathcal{P} = \{C(0;i) : i = 0, 1, \ldots, n-1\}$ de $B(n,p)$. Como los cilindros generan $\mathcal{A}$,

$$\bigvee_{j=-\infty}^{\infty} \sigma^j \mathcal{P} = \mathcal{A}.$$

Observemos que los átomos de $\bigvee_{j=0}^{k-1} \sigma^{-j}(\mathcal{P})$ son los cilindros de forma

$$C(0; i_0, i_1, \ldots, i_{k-1})$$

los cuales tienen medida

$$p_{i_0} \cdot p_{i_1} \cdot \ldots \cdot p_{i_{k-1}},$$

luego

$$\begin{aligned}
H\left(\bigvee_{j=0}^{k-1} \sigma^{-j}(\mathcal{P})\right) &= -\sum_{i_0, i_1, \ldots, i_{k-1}=0}^{n-1} (p_{i_0} p_{i_1} \cdot \ldots \cdot p_{i_{k-1}}) \log(p_{i_0} p_{i_1} \cdot \ldots \cdot p_{i_{k-1}}) \\
&= -\sum_{i_0, i_1, \ldots, i_{k-1}=0}^{n-1} (p_{i_0} \cdot \ldots \cdot p_{i_{k-1}})[\log p_{i_0} + \cdots + \log p_{i_{k-1}}] \\
&= -k \sum_{i=0}^{n-1} p_i \log p_i.
\end{aligned}$$

Usando el Teorema de Kolmogorov-Sinai obtenemos

$$h(\sigma) = h(\sigma, \mathcal{P}) = -\sum_{i=0}^{n-1} p_i \log p_i$$

(2) Sea $B(n, p, P, \sigma)$ un desplazamiento de Markov. Mostremos que

$$h(\sigma) = -\sum_{i,j=0}^{n-1} p_i p_{ij} \log p_{ij}$$

donde $p = (p_i)$ y $P = (p_{ij})$.

En efecto, usando la notación del ejemplo anterior y procediendo de manera análoga tenemos que

$$\begin{aligned}
H\left(\bigvee_{j=0}^{k-1} \sigma^{-j}(\mathcal{P})\right) \\
&= -\sum_{i_0 \ldots i_{k-1}=0}^{n-1} p_{i_0} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{k-2} i_{k-1}} \log(p_{i_0} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{k-2} i_{k-1}}) \\
&= -\sum_{i_0, \ldots, i_{k-1}=0}^{n-1} (p_{i_0} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{k-2} i_{k-1}})[\log p_{i_0} + \log p_{i_0 i_1} + \cdots + \log p_{i_{k-2} i_{k-1}}] \\
&= -\sum_{i_0=0}^{n-1} p_{i_0} \log p_{i_0} - (k-1) \sum_{i,j=0}^{n-1} p_i p_{ij} \log p_{ij}
\end{aligned}$$

donde fueron usadas las relaciones

$$\sum_{i=0}^{n-1} p_i p_{ij} = p_j \text{ y } \sum_{j=0}^{n-1} p_{ij} = 1.$$

De aquí se sigue

$$h(\sigma) = -\sum_{i,j=0}^{n-1} p_i p_{ij} \log p_{ij}$$

Mencionaremos algunos resultados muy importantes. En 1971 Ornstein [Orn] probó lo siguiente:

Teorema X.2. *Dos desplazamientos de Bernoulli son equivalentes si, y sólo si, poseen la misma entropía.*

En 1972 Friedman y Ornstein [F-O] complementaron este teorema con un criterio que permite reconocer cúando una transformación es Bernoulli (o sea, equivalente a un desplazamiento de Bernoulli) y probaron en particular que:

Teorema X.3. *Un desplazamiento de Markov $B(n, p, P, \sigma)$ es Bernoulli si, y sólo si, P es aperiódica.*

§XI El Teorema de Kolmogorov-Sinai

Esta sección está dedicada a las pruebas del Teorema de Kolmogorov-Sinai y de variaciones de él.

En toda esta sección la *siguiente notación será empleada*:

(i) $(X, \mathcal{A}, \mu, T)$ denotará un sistema dinámico medible tal que μ es una probabilidad.

Siempre usaremos log para referirmos a logaritmo natural.

Proposición XI.1. *La función* η: $[0, \infty) \to \mathbf{R}$ *definida por*

$$\eta(x) = \begin{cases} x \log x & si\ x > 0 \\ 0 = 0 \log 0 & si\ x = 0 \end{cases}$$

tiene las siguientes propiedades:

(i) η *es continua y no positiva en el intervalo* $[0, 1]$;

(ii) η *es estrictamente convexa, i.e.,*

$$\eta(\alpha x + \beta y) \leq \alpha\eta(x) + \beta\eta(y)$$

si $x, y \in [0, 1]$, $\alpha \geq 0$, $\beta \geq 0$ *y* $\alpha + \beta = 1$; *siendo que la igualdad vale sólo cuando* $x = y$ *o* $\alpha \in \{0, 1\}$. *Por inducción se obtiene:*

$$\eta\left(\sum_{i=1}^{k} \alpha_i x_i\right) \leq \sum_{i=1}^{k} \alpha_i \eta(x_i)$$

si $x_i \in [0, \infty)$, $\alpha_i \geq 0$ *y* $\sum_{i=1}^{k} \alpha_i = 1$; *donde la igualdad vale sólo cuando* *todos los x_i, para los cuales el correspondiente $\alpha_i \neq 0$, son iguales.*

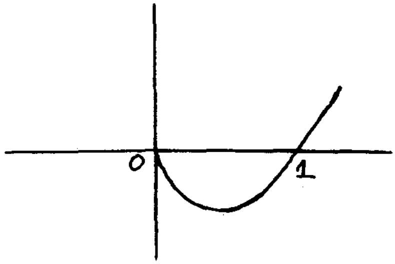


Fig. XI.1

Demostración: Primeramente veamos que

$$\eta'(x) = 1 + \log x,$$

$$\eta''(x) = \frac{1}{x} > 0 \text{ en } (0, \infty).$$

Fijemos α, β con $\alpha > 0$, $\beta > 0$. Supongamos que $y > x$. Por el Teorema del Valor Medio, $\eta(y) = \eta(\alpha x + \beta y) = \eta'(z)\alpha(y - x)$ para algún z con $\alpha x + \beta y < z < y$ y

$$\eta(\alpha x + \beta y) - \eta(x) = \eta'(w)\beta(y - x)$$

para algún w con $x < w < \alpha x + \beta y$. Como $\eta'' > 0$, se tiene que $\eta'(z) > \eta'(w)$ y así

$$\begin{aligned} \beta[\eta(y) - \eta(\alpha x + \beta y)] &= \eta'(z)\alpha\beta(y - x) \\ &> \eta'(w)\alpha\beta(y - x) \\ &= \alpha(\eta(x + \beta y) - \eta(x)). \end{aligned}$$

En consecuencia $\eta(\alpha x + \beta y) < \alpha\eta(x) + \beta\eta(y)$ si $x, y > 0$; ciertamente esta desigualdad también vale si $x, y \geq 0$ y $x \neq y$. ∎

Dados $A, B \in \mathcal{A}$, decimos que $A = B \bmod (0)$ (resp. $A \subset B \bmod (0)$) si $\mu(A \Delta B) = 0$ (resp. $\mu(B \setminus A) = 0$).

Sean $\mathcal{P} = \{P_1, P_2, \ldots, P_p\}$ y $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_q\}$ particiones del espacio de probabilidad $(X, \mathcal{A}, \mu)$. Definimos la entropía relativa de $\mathcal{P}$ con respecto a $\mathcal{Q}$ como

$$H(\mathcal{P}/\mathcal{Q}) = -\sum_{j=1}^{q}\sum_{i=1}^{p} \mu(P_i \cap Q_j) \log\left(\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}\right),$$

donde usamos el convenio

$$0 = 0 \log 0.$$

Observamos que

$$H(\mathcal{P}/\mathcal{Q}) \geq 0, \text{ y que}$$

$$H(\mathcal{P}) = H(\mathcal{P}/1),$$

donde $1 = \{X\}$ es la partición trivial. Decimos que $\mathcal{P} \leq \mathcal{Q}$ o sea que $\mathcal{Q}$ es más fina que $\mathcal{P}$ si $A \in \mathcal{Q} \Rightarrow \exists B \in \mathcal{P}$ tal que $A \subset B$ mod (0). Definimos $\mathcal{P} \vee \mathcal{Q}$ como la partición de $(X, \mathcal{A}, \mu)$ cuyos átomos son todos los conjuntos $P_i \cap Q_j$, con $P_i \in \mathcal{P}$, $Q_j \in \mathcal{Q}$, tales que $\mu(P_i \cap Q_j) > 0$. Si $T: X \to X$ preserva medida, $T^{-1}\mathcal{P} = \{T^{-1}P_1, \ldots, T^{-1}P_p\}$ es también una partición de $(X, \mathcal{A}, \mu)$.

Proposición XI.2. *Sean $\mathcal{P} = \{P_1, P_2, \ldots, P_p\}$, $\mathcal{Q} = \{Q_1, Q_2, \ldots, Q_q\}$ y $\mathcal{M} = \{M_1, M_2, \ldots, M_m\}$ particiones del espacio de probabilidad $(X, \mathcal{A}, \mu)$. Entonces*

(i) $H(\mathcal{P} \vee \mathcal{Q}/\mathcal{M}) = H(\mathcal{P}/\mathcal{M}) + H(\mathcal{Q}/\mathcal{P} \vee \mathcal{M})$;
(ii) $\mathcal{P} \leq \mathcal{Q} \Rightarrow H(\mathcal{P}/\mathcal{M}) \leq H(\mathcal{Q}/\mathcal{M})$;
(iii) $\mathcal{P} \leq \mathcal{Q} \Rightarrow H(\mathcal{M}/\mathcal{P}) \geq H(\mathcal{M}/\mathcal{Q})$;
(iv) $H(\mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q}/\mathcal{P})$;
(v) $H(\mathcal{P} \vee \mathcal{Q}/\mathcal{M}) \leq H(\mathcal{P}/\mathcal{M}) + H(\mathcal{Q}/\mathcal{M})$;
(vi) Si $T: X \to X$ preserva medida, $H(T^{-1}\mathcal{P}/T^{-1}\mathcal{Q}) = H(\mathcal{P}/\mathcal{Q})$;
(vii) $H(\mathcal{P}/\mathcal{Q}) = 0 \Leftrightarrow \mathcal{P} \leq \mathcal{Q}$.

Demostración: Usaremos los convenios

$$0 = 0 \log 0 = 0 \log(\frac{0}{0}).$$

(i):

$$
\begin{aligned}
H(\mathcal{P} \vee \mathcal{Q}/\mathcal{M}) &= -\sum_{i,j,k} \mu(P_i \cap Q_j \cap M_k) \log \left[\frac{\mu(P_i \cap Q_j \cap M_k)}{\mu(M_k)}\right] \\
&= -\sum_{i,j,k} \mu(P_i \cap Q_j \cap M_k) \log \left[\frac{\mu(P_i \cap Q_j \cap M_k)}{\mu(P_i \cap M_k)}\right] \\
&\quad -\sum_{i,j,k} \mu(P_i \cap Q_j \cap M_k) \log \left[\frac{\mu(p_i \cap M_k)}{\mu(M_k)}\right] \\
&= H(\mathcal{Q}/\mathcal{P} \vee \mathcal{M}) - \sum_{i,k} \mu(P_i \cap M_k) \log \left[\frac{\mu(P_i \cap M_k)}{\mu(M_k)}\right] \\
&= H(\mathcal{Q}/\mathcal{P} \vee \mathcal{M}) + H(\mathcal{P}/\mathcal{M}).
\end{aligned}
$$

(ii):
Si $\mathcal{P} \leq \mathcal{Q}$ entonces $\mathcal{Q} = \mathcal{P} \vee \mathcal{Q}$. Luego,

$$
\begin{aligned}
H(\mathcal{Q}/\mathcal{M}) &= H(\mathcal{P} \vee \mathcal{Q}/\mathcal{M}) \\
&= H(\mathcal{P}/\mathcal{M}) + H(\mathcal{Q}/\mathcal{P} \vee \mathcal{M}) \\
&\geq H(\mathcal{P}/\mathcal{M}).
\end{aligned}
$$

(iii):
Sea η: $[0, 1] \to \mathbf{R}$ dada por $\eta(x) = x \log(x)$ (ver el Lema XI.1). Entonces

$$
\begin{aligned}
H(\mathcal{M}/\mathcal{Q}) &= -\sum_{i,j} \mu(M_i \cap Q_j) \log \left[\frac{\mu(M_i \cap Q_j)}{\mu(Q_j)}\right] \\
&= -\sum_{i,j} \frac{\mu(Q_j)\mu(M_i \cap Q_j)}{\mu(Q_j)} \log \left[\frac{\mu(M_i \cap Q_j)}{\mu(Q_j)}\right] \\
&= \sum_{i,j} \mu(Q_j) \eta \left(\frac{\mu(M_i \cap Q_j)}{\mu(Q_j)}\right).
\end{aligned}
$$

Como $\mu(Q_j) = \sum\limits_k \mu(Q_j \cap P_k)$, se sigue que

$$\begin{aligned}
\sum_i \mu(Q_j) &= \sum_{j,k} \mu(Q_j \cap P_k) \\
&= \sum_{j,k} \mu(P_k) \frac{\mu(Q_j \cap P_k)}{\mu(P_k)} \\
&= \sum_k \mu(P_k) \sum_j \frac{\mu(Q_j \cap P_k)}{\mu(P_k)}.
\end{aligned}$$

Luego

$$H(\mathcal{M}/\mathcal{Q}) = \sum_{i,k} \mu(P_k) \sum_j \frac{\mu(Q_j \cap P_k)}{\mu(P_k)} \eta \left(\frac{\mu(M_i \cap Q_i)}{\mu(Q_j)} \right).$$

Observemos que $\sum\limits_j \dfrac{\mu(Q_j \cap P_k)}{\mu(P_k)} = 1$; así, usando el Lema XI.1, obtenemos

$$H(\mathcal{M}/\mathcal{Q}) \leq \sum_{i,k} \mu(P_k) \eta \left(\sum_j \frac{\mu(Q_j \cap P_k)}{\mu(P_k)} \cdot \frac{\mu(M_i \cap Q_j)}{\mu(Q_j)} \right).$$

Ahora, $\mathcal{P} \leq \mathcal{Q}$ implica que $\mu(Q_j \cap P_k) \in \{0, \mu(Q_j)\}$ para todo par j, k. Luego, fijados i, k:

$$\begin{aligned}
\sum_j \frac{\mu(Q_j \cap P_k)}{\mu(P_k)} \cdot \frac{\mu(M_i \cap Q_j)}{\mu(Q_j)} &= \sum_{Q_j \subset P_k} \frac{\mu(Q_j \cap P_k)}{\mu(P_k)} \cdot \frac{\mu(M_i \cap Q_j)}{\mu(Q_j)} \\
&= \sum_{Q_j \subset P_k} \frac{\mu(M_i \cap Q_j)}{\mu(P_k)} \\
&= \frac{\mu(M_i \cap P_k)}{\mu(P_k)}
\end{aligned}$$

de donde se sigue que

$$\begin{aligned}
H(\mathcal{M}/\mathcal{Q}) &\leq \sum_{i,k} \mu(P_k)\eta\left(\frac{\mu(M_i\cap P_k)}{\mu(P_k)}\right)\\
&= -\sum_{i,k} \mu(P_k)\cdot \frac{\mu(M_i\cap P_k)}{\mu(P_k)}\cdot \log\left(\frac{\mu(M_i\cap P_k)}{\mu(P_k)}\right)\\
&= -\sum_{i,k} \mu(M_i\cap P_k)\cdot \log\left(\frac{\mu(M_i\cap P_k)}{\mu(P_k)}\right)\\
&= H(\mathcal{M}/\mathcal{P}).
\end{aligned}$$

(iv):

$$\begin{aligned}
H(\mathcal{Q}) &\leq H(\mathcal{Q}\vee\mathcal{P})\\
&= H(\mathcal{Q}\vee\mathcal{P}/1)\\
&= H(\mathcal{P}/1) + H(\mathcal{Q}/\mathcal{P}\vee 1)\\
&= H(\mathcal{P}) + H(\mathcal{Q}/\mathcal{P}).
\end{aligned}$$

(v):

$$\begin{aligned}
H(\mathcal{P}\vee\mathcal{Q}/\mathcal{M}) &= H(\mathcal{P}/\mathcal{M}) + H(\mathcal{Q}/\mathcal{P}\vee\mathcal{M})\\
&\leq H(\mathcal{P}/\mathcal{M}) + H(\mathcal{Q}/\mathcal{M}),
\end{aligned}$$

pues $\mathcal{M}\leq\mathcal{P}\vee\mathcal{M}$.

(vi):

$$\begin{aligned}
H(T^{-1}\mathcal{P}/T^{-1}\mathcal{Q}) &= -\sum_{i,j}\mu(T^{-1}P_i\cap T^{-1}Q_j)\log\left[\frac{\mu(T^{-1}P_j\cap T^{-1}Q_j)}{\mu(T^{-1}Q_j)}\right]\\
&= -\sum_{i,j}\mu(P_i\cap Q_j)\log\left[\frac{\mu(P_i\cap Q_j)}{\mu(Q_j)}\right]\\
&= H(\mathcal{P}/\mathcal{Q}).
\end{aligned}$$

(vii):

$$\text{(vii)} \qquad H(\mathcal{P}/\mathcal{Q}) = -\sum_{i,j} \mu(P_i \cap Q_j) \log \left[\frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \right].$$

Así, $H(\mathcal{P}/\mathcal{Q}) = 0$ implica que cuando $\mu(P_i \cap Q_j) > 0$ entonces $\mu(P_i \cap Q_j) = \mu(Q_j)$, o sea $Q_j \subset P_i$ mod(0); luego $\mathcal{P} \leq \mathcal{Q}$. Recíprocamente, $\mathcal{P} \leq \mathcal{Q}$ implica que, para todo i, j, $\dfrac{\mu(P_i \cap Q_j)}{\mu(Q_j)} = 1$, o sea $H(\mathcal{P}/\mathcal{Q}) = 0$. ∎

Lema XI.3. *Si* $\{a_n\}$ *es una sucesión de números reales no negativos tal que*

$$a_{n+p} \leq a_n + a_p$$

para todo $n, p \in \mathbb{N}$, *entonces* $\lim\limits_{n\to\infty} \dfrac{a_n}{n}$ *existe y es igual a* $\inf\limits_n \dfrac{a_n}{n}$.

Demostración: Fijemos $p > 0$. Cada $n > 0$ se escribe como $n = kp + i$, con $0 \leq i < p$. Entonces

$$\frac{a_n}{n} = \frac{a_{i+kp}}{i+k_p} \leq \frac{a_i}{k_p} + \frac{a_{kp}}{kp} \leq \frac{a_i}{kp} + \frac{ka_p}{kp} = \frac{a_i}{kp} + \frac{a_p}{p}.$$

Cuando $n \to \infty$ entonces $k \to \infty$ y así

$$\limsup_n \frac{a_n}{n} \leq \frac{a_p}{p}$$

y por consiguiente

$$\limsup_n \frac{a_n}{n} \leq \inf \frac{a_p}{p}.$$

Pero

$$\inf_p \frac{a_p}{p} \leq \liminf_n \frac{a_n}{n},$$

de donde $\lim \frac{a_n}{n}$ existe y es igual a $\inf \frac{a_n}{n}$. ∎

Proposición XI.4. *Sea $T: X \to X$ una transformación que preserva medida en un espacio de probabilidad $(X, \mathcal{A}, \mu)$. Si $\mathcal{P}$ es una partición finita de A. Entonces*

$$h(T,\mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right)$$
$$= \inf \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right).$$

Demostración: Sea $a_n = H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right)$. Entonces

$$a_{n+p} = H\left(\bigvee_{j=0}^{n+p-1} T^{-j}\mathcal{P}\right)$$
$$\leq H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) + H\left(\bigvee_{j=n}^{n+p-1} T^{-j}\mathcal{P}\right)$$
$$= a_n + H\left(\bigvee_{j=0}^{p-1} T^{-j}\mathcal{P}\right)$$
$$= a_n + a_p.$$

En estas circunstancias podemos aplicar el Lema XI.3. ∎

Proposición XI.5. *Sean $(X, \mathcal{A}, \mu)$ un espacio de probabilidad y $T: X \to X$ una transformación que preserva medida. Sean $\mathcal{P}$ y $\mathcal{Q}$ particiones finitas de $\mathcal{A}$. Entonces*

(i) $h(T,\mathcal{P}) - h(T,\mathcal{Q}) \leq H(\mathcal{P}/\mathcal{Q})$;

(ii) $\mathcal{Q} \leq \mathcal{P} \Rightarrow h(T,\mathcal{Q}) \leq h(T,\mathcal{P})$;

(iii) $h(T, T^{-1}\mathcal{P}) = h(T,\mathcal{P})$;

(iv) Para todo $n \geq 0$;

$$h(T, \bigvee_{j=0}^{n} T^{-j}\mathcal{P}) = h(T, \mathcal{P})$$

(v) La sucesión

$$\frac{1}{n} H(\mathcal{P} \vee T^{-1}\mathcal{P} \vee \cdots \vee T^{-n+1}\mathcal{P})$$

es decreciente.

(vi) $h(T, \mathcal{P}) = \lim_n H\left(\mathcal{P} / \bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right)$.

Demostración:

(i):

Como

$$\begin{aligned} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) - H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{Q}\right) &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P} / \bigvee_{j=0}^{n-1} T^{-j}\mathcal{Q}\right) \\ &\leq \sum_{i=0}^{n-1} H\left(T^{-i}\mathcal{P} / \bigvee_{j=0}^{n-1} T^{-j}\mathcal{Q}\right) \\ &\leq \sum_{i=0}^{n-1} H(T^{-i}\mathcal{P} / T^{-i}\mathcal{Q}) \\ &= nH(\mathcal{P}/\mathcal{Q}), \end{aligned}$$

tenemos que

$$\begin{aligned} h(T, \mathcal{P}) - h(T, \mathcal{Q}) &= \lim \frac{1}{n}\left[H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) - H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{Q}\right)\right] \\ &\leq H(\mathcal{P}/\mathcal{Q}). \end{aligned}$$

(ii):

Por XI.2,(vii), $\mathcal{Q} \leq \mathcal{P}$ implica que $H(\mathcal{Q}/\mathcal{P}) = 0$. Así, por (i), $h(T, \mathcal{Q}) \leq h(T,\mathcal{P})$.

(iii):

Como

$$H\left(T^{-1}\left[\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right]\right) = H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right),$$

tenemos que

$$\begin{aligned} h(T,T^{-1}\mathcal{P}) &= \lim \frac{1}{n} H\left(\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) \\ &= \lim \frac{1}{n} H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) \\ &= h(T,\mathcal{P}). \end{aligned}$$

(iv):

$$\begin{aligned} h\left(T, \bigvee_{i=0}^{n} T^{-i}\mathcal{P}\right) &= \lim_{m\to\infty} \frac{1}{m} H\left(\bigvee_{j=0}^{m-1} T^{-j}\left(\bigvee_{i=0}^{n} T^{-j}\mathcal{P}\right)\right) \\ &= \lim_{m\to\infty} \frac{1}{m} H\left(\bigvee_{i=0}^{n+m-1} T^{-j}\mathcal{P}\right) \\ &= \lim_{m\to\infty} \frac{m+n-1}{m} \frac{1}{m+n-1} H\left(\bigvee_{i=0}^{n+m-1} T^{-j}\mathcal{P}\right) \\ &= h(T,\mathcal{P}). \end{aligned}$$

Esto prueba (iv). Para las demonstraciones de (v) y (vi) necesitaremos de lo siguiente:

Para todo $n \geq 1$

$$H\left(\bigvee_{j=0}^{n} T^{-1}\mathcal{P}\right) = H(\mathcal{P}) + \sum_{k=1}^{n} +H\left(\mathcal{P}/\bigvee_{j=1}^{k} T^{-j}\mathcal{P}\right). \tag{1}$$

En efecto, si $n = 1$ entonces

$$\begin{aligned} H(\mathcal{P} \vee T^{-1}\mathcal{P}) &= H(T^{-1}\mathcal{P}) + H(\mathcal{P}/T^{-1}\mathcal{P}) \\ &= H(\mathcal{P}) + H(\mathcal{P}/T^{-1}\mathcal{P}). \end{aligned}$$

Supongamos que (1) es válido para $n-1$, entonces

$$\begin{aligned} H\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) &= H\left(\bigvee_{j=1}^{n} T^{-j}\mathcal{P} \vee \mathcal{P}\right) \\ &= H\left(\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) + H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) \\ &= H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) + H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) \\ &= H(\mathcal{P}) + \sum_{k=1}^{n-1} H\left(\mathcal{P}/\bigvee_{j=1}^{k} T^{-j}\mathcal{P}\right) + H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) \\ &= H(\mathcal{P}) + \sum_{k=1}^{n} H\left(\bigvee_{j=1}^{k} T^{-j}\mathcal{P}\right). \end{aligned}$$

Esto prueba la afirmación (1).

(v):

Se sigue de (1) y de la Proposición XI.2,(iii) que

$$H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) \geq nH\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right).$$

Así,

$$\begin{aligned}
nH\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) &= n\left[H\left(\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) + H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right)\right] \\
&\leq nH\left(\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) + H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) \\
&= (n+1)H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right).
\end{aligned}$$

Luego,

$$\frac{1}{n+1}H\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) \leq \frac{1}{n}H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right),$$

lo que prueba (v).

(vi):

Como $\bigvee_{j=1}^{n} T^{-j}\mathcal{P} \geq \bigvee_{j=1}^{n+1} T^{-j}\mathcal{P}$, se tiene que para todo $n \geq 1$,

$$H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) \leq H\left(\mathcal{P}/\bigvee_{j=1}^{n+1} T^{-j}\mathcal{P}\right) \geq 0$$

o sea

$$\lim_n H\left(\mathcal{P}/\bigvee_{j=1}^{n} T^{-j}\mathcal{P}\right) = c \quad \text{existe.}$$

y así, por el Teorema de Cesaro (que dice que si una sucesión de numeros reales $x_n \to c$ entonces $\frac{1}{n}\sum_{j=0}^{n-1} x_j \to c$),

$$\lim_n \frac{1}{n+1}\sum_{k=1}^{n} H\left(\mathcal{P}/\bigvee_{j=1}^{k} T^{-j}\mathcal{P}\right) = c.$$

En estas condiciones, usando (v) tenemos que:

$$\begin{aligned} h(T,\mathcal{P}) &= \lim_n \frac{1}{n+1} H\left(\bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) \\ &= \lim_n \frac{1}{n+1} H(\mathcal{P}) + \lim_n \frac{1}{n+1} H\left(\mathcal{P}/\bigvee_{j=1}^{k} T^{-j}\mathcal{P}\right) = c. \ \blacksquare \end{aligned}$$

Sea $(X,\mathcal{A},\mu)$ un espacio de probabilidad y $\mathcal{P},\mathcal{P}_1,\mathcal{P}_2,\ldots,\mathcal{P}_n$, una sucesión de particiones. Denotaremos por $\bigvee_{n=1}^{\infty}\mathcal{P}_n$ la σ-álgebra generada por $\bigcup_{n=1}^{\infty}\mathcal{P}_n$. Diremos que $\mathcal{P}\subset\bigvee_{n=1}^{\infty}\mathcal{P}_n$ mod(0) si para todo $A\in\mathcal{P}$ existe $B\in\bigvee_{n=1}^{\infty}\mathcal{P}_n$ tal que $A\subset B$ mod(0).

Lema XI.6. *Sea $\mathcal{P}=\{P_1,P_2,\ldots,P_n\}$ una partición de $(X,\mathcal{A},\mu)$. Dado $\varepsilon>0$ existe $\delta>0$ tal que si $\mathcal{Q}=\{Q_1,Q_2,\ldots,Q_r\}$ es una partición de $(X,\mathcal{A},\mu)$ satisfaciendo $\Sigma_{i=1}^{r}\ \mu(P_i\triangle Q_i)<\delta$, entonces $H(\mathcal{P}/\mathcal{Q})<\varepsilon$*

Demostración: Dada una función $\varphi\colon\mathcal{D}(\varphi)\mapsto\mathbf{R}$ tal que $\mathcal{D}(\varphi)$ consiste de particiones de $(X,\mathcal{A},\mu)$ que tienen r elementos y $\mathcal{P}\in\mathcal{D}(\varphi)$, escribiremos

$$\lim_{\mathcal{Q}\to\mathcal{P}}\varphi(\mathcal{Q})=\varphi(\mathcal{P})$$

para significar que para todo $\varepsilon > 0$ existe $\delta > 0$ tal que si $\mathcal{Q} = (Q_1, Q_2, \ldots, Q_r)$ es una partición de $(X, \mathcal{A}, \mu)$ satisfaciendo $\Sigma_{i=1}^r \mu(P_i \Delta Q_i) < \delta$ entonces $\mathcal{Q} \in \mathcal{D}(\varphi)$ y $|\varphi(\mathcal{P}) - \varphi(\mathcal{Q})| < \varepsilon$.

Procedamos a probar el lema. Observemos que, para cada i,

$$P_i = (P_i \cap Q_i) \cup (P_i \setminus Q_i) \qquad \text{y} \qquad Q_i = (Q_i \cap P_i) \cup (Q_i \setminus P_i).$$

Como por hipótesis

$$\lim_{\mathcal{Q} \to \mathcal{P}} \mu(Q_i \setminus P_i) = 0 = \lim_{\mathcal{Q} \to \mathcal{P}} (P_i \setminus Q_i),$$

tenemos que

$$\lim_{\mathcal{Q} \to \mathcal{P}} \mu(Q_i) = \lim_{\mathcal{Q} \to \mathcal{P}} \mu(P_i \cap Q_i) = \mu(P_i) > 0$$

y así

$$(1) \qquad \lim_{\mathcal{Q} \to \mathcal{P}} \frac{\mu(P_i \cap Q_i)}{\mu(Q_i)} = 1,$$

$$\lim_{\mathcal{Q} \to \mathcal{P}} \mu(\bigcup_{i=1}^{n} P_i \cap Q_i) = \sum_{i=1}^{n} \mu(P_i) = 1.$$

De esto último y del hecho de que $\mathcal{P} \vee \mathcal{Q}$ es una partición de $(X, \mathcal{A}, \mu)$, se sigue que

$$\lim_{\mathcal{Q} \to \mathcal{P}} \mu\left(\bigcup_{i \neq j} P_i \cap Q_j\right) = \mu(X) - \lim_{\mathcal{Q} \to \mathcal{P}} \mu\left(\bigcup_{i=1}^{n} P_i \cap Q_i\right) = 0$$

luego

$$(2) \qquad i \neq j \Rightarrow \lim_{\mathcal{Q} \to \mathcal{P}} \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} = 0.$$

Observemos ahora que

$$H(\mathcal{P}/\mathcal{Q}) = -\sum_{i,j} \mu(Q_i)\ \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \log\left[\frac{\mu(P_i \cap Q_j)}{\mu(Q_i)}\right]$$

Por consiguiente, usando (1), (2) y la Proposición XI.1, concluimos que

$$\lim_{\mathcal{Q}\to\mathcal{P}} H(\mathcal{P}/\mathcal{Q}) = 0 \qquad \blacksquare$$

Teorema XI.7. *Sean $\mathcal{P}, \mathcal{P}_1, \mathcal{P}_2, \ldots$ particiones finitas de un espacio de probabilidad $(X, \mathcal{A}, \mu)$ tales que $\mathcal{P}_1 \leq \mathcal{P}_2 \leq \cdots \leq \mathcal{P}_n \leq \ldots$ Si $\mathcal{P} \subset \bigvee_{n=1}^{\infty} \mathcal{P}_n$ mod (0), entonces $\lim_n H(\mathcal{P}/\mathcal{P}_n) = 0$.*

Demostración: Sea $\mathcal{C}_n$ el álgebra (finita) generada por $\mathcal{P}_n$.

Sea $\mathcal{P} = \{A_1, A_2, \ldots, A_r\}$. Dado $\varepsilon > 0$ sea $\delta = \delta(\mathcal{P})$ como en el Lema XI.5. Como $\bigvee_{n=1}^{\infty} \mathcal{C}_n$ es un álgebra y

$$\mathcal{P} \subset \bigvee_{n=1}^{\infty} \mathcal{C}_n \mod (0)$$

podemos usar el Teorema de Aproximación I.5 para concluir que existe $N \geq 1$ entero tal que, para $i = 1, 2, \ldots, r$, $\tilde{B}_i \in \mathcal{C}_N$ y

$$\sum_{i=1}^{r} \mu(\tilde{B}_i \Delta A_i) < \frac{\delta}{4r^2},$$

lo que, por la suposición: $\mathcal{P}_N \leq \mathcal{P}_{N+1} \leq \ldots$, implica que:

(2) $$\text{para todo } n \geq N, \quad \tilde{B}_i \in \mathcal{C}_n.$$

Veamos **que**

(3) $$\mu\Big(\bigcup_{i \neq j} \tilde{B}_i \cap \tilde{B}_j\Big) < \frac{\delta}{2}$$

En efecto, $\tilde{B}_i \subset A_i \cup (\tilde{B}_i \Delta A_i)$, $\tilde{B}_j \subset A_j \cup (\tilde{B}_j \Delta A_j)$, y así $\tilde{B}_i \cap \tilde{B}_j \subset (A_i \cup (\tilde{B}_i \Delta A_i)) \cup (\tilde{A}_j \cup (\tilde{B}_j \Delta A_j)) \subset (\tilde{B}_i \Delta A_i) \cup (\tilde{B}_j \Delta A_j)$.

Esto y (1) implican (3).

Sean

$$
\begin{aligned}
B_1 &= \tilde{B}_1 \\
B_2 &= \tilde{B}_2 \setminus B_1 \\
&\vdots \\
B_{r-1} &= \tilde{B}_{r-1} \setminus \bigcup_{j=1}^{r-2} \tilde{B})j \\
B_r &= X \setminus \bigcup_{j=1}^{r-1} \tilde{B}_j.
\end{aligned}
$$

Observemos que

$$
B_i \setminus A_i = \left(\tilde{B}_i \setminus \bigcup_{j=1}^{i-1} \tilde{B}_j \right) \setminus A_i \subset \tilde{B}_i \setminus A_i,
$$

y que

$$
\begin{aligned}
A_i \setminus B_i = A_i \setminus \left(\tilde{B}_i \setminus \bigcup_{j=1}^{i-1} \tilde{B}_j \right) \\
= A_i \setminus \left(\tilde{B}_i \setminus \bigcup_{j=1}^{i-1} \tilde{B}_i \cap \tilde{B}_j) \right) \\
= (A_i \setminus \tilde{B}_i) \cup \left(A_i \cap \left(\bigcup_{j=1}^{i-1} \tilde{B}_i \cap \tilde{B}_j \right) \right).
\end{aligned}
$$

En consecuencia, por (1) y (3),

$$
\mu(A_i \Delta B_i) \leq \mu(\tilde{B}_i \setminus A_i) + \mu(A_i \setminus \tilde{B}_i) + \mu \left(\bigcup_{j=1}^{i-1} \tilde{B}_i \cap \tilde{B}_j \right) < \delta.
$$

Así, por el Lema XI.6, si $\mathcal{Q} = (B_1, \ldots, B_r)$,

$$H(\mathcal{P}/\mathcal{Q}) < \varepsilon.$$

Como por (2) $\mathcal{Q} \leq \mathcal{P}_n$, para todo $n \geq N$, podemos usar la Proposición XI.2(iii) para concluir que

$$\text{para todo} \quad n \geq N, \quad H(\mathcal{P}/\mathcal{P}_n) \leq H(\mathcal{P}/\mathcal{Q}) < \varepsilon. \qquad \blacksquare$$

Corolario XI.7.1. *Si $\mathcal{P}_1 \leq \mathcal{P}_2 \leq \ldots$ es una sucesión de particiones finitas tal que $\mathcal{A} = \bigvee_{n=1}^{\infty} \mathcal{P}_n \mod 0$, entonces*

$$h(T) = \sup_n h(T, \mathcal{P}_n).$$

Demostración: Sea $\mathcal{P}$ una partición finita de $(X, \mathcal{A}, \mu)$. Como $\mathcal{A} = \bigvee_{n=1}^{\infty} \mathcal{P}_n \mod 0$, se sigue del Teorema XI.7 que

$$\lim_n H(\mathcal{P}/\mathcal{P}_n) = 0. \tag{1}$$

Por la Proposición XI.5,(i), para todo n,

$$h(T, \mathcal{P}) - h(T, \mathcal{P}_n) \leq H(\mathcal{P}/\mathcal{P}_n),$$

o sea

$$h(T, \mathcal{P}_n) \geq h(T, \mathcal{P}) - H(\mathcal{P}/\mathcal{P}_n).$$

Luego, usando (1),

$$\sup_n h(T, \mathcal{P}_n) \geq h(T, \mathcal{P}).$$

Como $\mathcal{P}$ es arbitrario el corolario se sigue immediatamente. $\blacksquare$

Sea T una transformación inversible que preserva medida en un espacio de probabilidad $(X, \mathcal{A}, \mu)$. Decimos que una partición finita $\mathcal{P}$ de $(X, \mathcal{A}, \mu)$ es un T-*generador* si

$$\bigvee_{n=-\infty}^{\infty} T^n\mathcal{P} = \mathcal{A} \mod (0).$$

Corolario XI.7.2. *(Kolmogorov-Sinai).*

Si T es inversible y $\mathcal{P}$ es un T-generador, entonces

$$h(T, \mathcal{P}) = h(T).$$

Demostración: Sea $\mathcal{P}_n = \bigvee_{j=-n}^{n} T^{-j}\mathcal{P}$; tenemos entonces que $\mathcal{P}_1 \leq \mathcal{P}_2 \leq \cdots \leq \mathcal{P}_n \leq \ldots$ y que $\bigvee_{-\infty}^{\infty} T^{-n}\mathcal{P} = A \mod (0)$. Luego por el Corolario XI.7.1 y la Proposición XI.5,(iii)–(iv),

$$\begin{aligned} h(T) &= \sup_{n\geq 0} h(T, \bigvee_{j=-n}^{n} T^{-j}\mathcal{P}) \\ &= \sup_{n} h(T, T^n \left(\bigvee_{j=0}^{2n} T^{-j}\mathcal{P}\right) \\ &= \sup_{n} h\left(T, \bigvee_{j=0}^{2n} T^{-j}\mathcal{P}\right) \\ &= h(T, \mathcal{P}). \quad \blacksquare \end{aligned}$$

Corolario XI.7.3. *Si $\mathcal{P}$ es una partición finita y*

$$\bigvee_{j=0}^{\infty} T^{-j}\mathcal{P} = \mathcal{A} \text{ mod}(0)$$

entonces

$$h(T,\mathcal{P}) = h(T).$$

Demostración: Usando el Corolario XI.7.1 y la Proposición XI.5 obtenemos

$$h(T) = \sup_{n\geq 0} h\left(T, \bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) = h(T,\mathcal{P}). \qquad \blacksquare$$

Corolario XI.7.4. *Si T es inversible y existe una partición finita $\mathcal{P}$ tal que*

$$\bigvee_{j=0}^{\infty} T^{-j}\mathcal{P} = \mathcal{A} \quad mod\ (0),$$

entonces

$$h(T) = 0.$$

Demostración: Como T es inversible,

$$\begin{aligned}\bigvee_{i=1}^{\infty} T^{-i}\mathcal{P} &= T^{-1}\left(\bigvee_{i=1}^{\infty} T^{-i}\mathcal{P}\right) \quad \mathrm{mod}(0)\\ &= T^{-1}(\mathcal{A}) \ \mathrm{mod}(0)\\ &= \mathcal{A} \quad \mathrm{mod}(0).\end{aligned}$$

Así, si $\mathcal{P}_n = \bigvee_{i=1}^{n} T^{-i}\mathcal{P}$ entonces $\mathcal{P}_1 \leq \mathcal{P}_2 \leq \ldots$ y $\bigvee_{n=1}^{\infty} \mathcal{P}_n = \mathcal{A} \ \ \mathrm{mod}(0)$. En estas condiciones, se sigue del Teorema XI.7 que

$$\lim_n H(\mathcal{P}/\mathcal{P}_n) = 0.$$

Consecuentemente, usando el Corolario XI.7.3 y la Proposición XI.5 (vi),

$$h(T) = h(T,\mathcal{P}) = \lim_n H(\mathcal{P}/\mathcal{P}_n) = 0. \qquad \blacksquare$$

Proposición XI.8. *Sea T una transformación que preserva medida de un espacio de probabilidad $(X, \mathcal{A}, \mu)$. Entonces:*

(i) $m \in \mathbf{N} \Rightarrow h(T^m) = mh(T)$.

(ii) *T inversible y* $m \in \mathbf{Z} \Rightarrow h(T^m) = |m|\, h(T)$.

Demostración:

(i):

Sea $\mathcal{P}$ una partición finita de $(X, \mathcal{A}, \mu)$ y $m \geq 1$. Tenemos que

$$\begin{aligned}
h(T, \mathcal{P}) &= \lim_n \frac{1}{nm} H\left(\bigvee_{j=0}^{nm-1} T^{-j}\mathcal{P}\right) \\
&\geq \lim_n \frac{1}{nm}\ H\left(\bigvee_{j=0}^{m(n-1)} T^{-j}\mathcal{P}\right) \\
&\geq \frac{1}{m} \lim \frac{1}{m} H\left(\bigvee_{j=0}^{n-1} T^{-jm}\mathcal{P}\right) \\
&= \frac{1}{m}\ h(T^m, \mathcal{P}),
\end{aligned}$$

o sea

$$mh(T) \geq h(T^m, \mathcal{P})$$

y así

$$m\ h(T) \geq h(T^m).$$

Recíprocamente,

$$h\left(T^m, \bigvee_{j=0}^{n} T^{-j}\mathcal{P}\right) = \lim_n \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-im}\left(\bigvee_{j=0}^{m} T^{-j}\mathcal{P}\right)\right)$$

$$= \lim_n \frac{1}{n} H\left(\bigvee_{j=0}^{(n-1)m} T^{-j}\mathcal{P})\right)$$

$$= \lim_n \frac{(n-1)m}{n} \cdot \frac{1}{(n-1)m)} H\left(\bigvee_{j=0}^{(n-1)m} T^{-j}\mathcal{P}\right)$$

$$= mh(T, \mathcal{P}),$$

i.e., para toda partición finita $\mathcal{P}$

$$h(T^m) \geq mh(T, \mathcal{P})$$

lo que implica que

$$h(T^m) \geq mh(T);$$

luego $h(T^m) = mh(T)$ si $m \geq 1$. Como $h(T^0) = 0$, donde T^0 denota la transformación identidad, tenemos que (i) es verdadero.

(ii):

Por el ítem (i), basta probar que para toda partición finita $\mathcal{P}$,

$$h(T^{-1}, \mathcal{P}) = h(T, \mathcal{P}).$$

En efecto,

$$H\left(\bigvee_{j=0}^{n-1} T^{j}\mathcal{P}\right) = H\left(T^{-(n-1)}\Big(\bigvee_{j=0}^{n-1} T^{j}\mathcal{P}\Big)\right)$$

$$= H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right),$$

y así,

$$
\begin{aligned}
h(T^{-1}, \mathcal{P}) &= \lim \frac{1}{n}\, H\left(\bigvee_{j=0}^{n-1} T^{j}\mathcal{P}\right) \\
&= \lim \frac{1}{n}\, H\left(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}\right) \\
&= h(T, \mathcal{P}). \quad \blacksquare
\end{aligned}
$$

BIBLIOGRAFÍA

[A-A] V.I. Arnold, A. Avez – Problèmes Ergodiques de la Mécanique Classique. Gauthier-Villars (1967).

[Bir] G.D. Birkhoff – Proof of the Ergodic Theorem. Proc. Nat. Acad. of Sciences U.S.A., 17 (1931), 656–660.

[Bil] P. Billingsley – Ergodic Theory and Information. Wiley Series in Prob. and Math. Stat. (1965).

[Bw1] R. Bowen – Topological Entropy and Axiom A. Proc. of Symp. on Pure Math. XIV (1968), 23–41.

[Bw2] R. Bowen – Markov Partitions for Axiom A Diffeomorphisms. Amer. Journal of Math. 92 (1970), 907–918.

[Bw3] R. Bowen – Equilibrium States and Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Mathematics 470 (1975).

[Bro] J.R. Brown – Ergodic Theory and Topological Dynamics. Academic Press, New York, 1976.

[Din] E.I. Dinaburg – The relation between topological entropy and metric entropy. Soviet Math. 11, 13–16 (1970).

[Fer] P. Fernandez – Medida e Integração, Projeto Euclides, IMPA (1976).

[Fra] J. Franks – Anosov Diffeomorphisms. Proc. of Symposia in Pure Mathematics Vol. XIV (1970).

[Fri] N.A. Friedman – Introduction to Ergodic Theory. Van Nostrand (1970).

[F-O] N.A. Friedman, D.S. Ornstein – An isomorphism of weak Bernouilli transformations, Advances in Mathematics, 5 (1970), 365–374.

[Fur] H. Furstenberg – Strict ergodicity and transformations of the torus, Amer. J. of Math. 83 (1961), 573–601.

[Gan] F.R. Gantmacher – Applications of the Theory of Matrices. Interscience, New York (1959).

[H-R] E. Hewitt and K.A. Ross – Abstract of Harmonic Analysis Vol 1. Springer-Verlag (1963).

[K-B] A. Katok – Bernouilli diffeomorphisms on surfaces. Ann. of Math. 110 (1977), 529–574.

[Kat] A. Katok – Smooth non-Bernouilli diffeomorphisms on surfaces. Inv. Math. 61 (1980), 271–300.

[Ktz] Y. Katznelson – Ergodic automorphisms of T^n are Bernouilli shifts. Israel Journal of Mathematics 10 (1975), 186–195.

[Mañ] R. Mañé – Introdução à Teoria Ergódica. Projeto Euclides, IMPA. Rio de Janeiro (1983).

[Mun] M.E. Munroe – Introduction to Measure and Integration, Addison e Wesley (1953).

[New] S. Newhouse – Dynamical Systems, Progress in Mathematics 8, Birkhauser (1980).

[Orn] D.S. Ornstein – Bernouilli shifts with the same entropy or isomorphic, Adv. in Math. 4 (1970), 337–352.

[Oxt] J.C. Oxtoby – Ergodic sets. Bull. Amer. Math. Soc. 58, 116–136 (1952).

[Par] K.R. Parthasarathy – Introduction to probability and Measure. MacMillan, London (1977).

[Pet] K. Petersen – Ergodic Theory. Cambridge, Univ. Press (1983).

[Roy] H.L. Royden – Real Analysis (2nd Ed.) MacMillan, New York (1968).

[Rud] W. Rudin – Functional analysis. New York, McGraw-Hill (1973).

[Sin] Ya. G. Sinai – Introduction to Ergodic Theory. Princenton Univ. Press (1970).

[Sma] S. Smale – Differentiable Dynamical Systems, Bull. Amer. Math. Soc. 73 (1967), 97–116.

[Szl] W. Szlenk – Introducción a la Teoria Ergódica. Publicaciones del Centro de Investigación y de Estudios Avanzados del I.P.N. (1979).

[Tot] H. Totoki – Ergodic Theory. Aarhus Univ. Lecture Notes (1969).

[Wal] P. Walters – An Introduction to Ergodic Theory. Springer Verlag (1982).