SOBRE LA CLASIFICACIÓN DE TRANSFORMACIONES ERGÓDICAS

Julián David Gutiérrez Pineda

Trabajo de grado para optar el título de Matemático

Director
Doctor Ernesto Acosta Gempeler
Profesor asociado

Escuela Colombiana de Ingeniería *Julio Garavito*
Programa de Matemáticas
Bogotá D.C., 2013 - 2

Agradecimientos

A mis profesores por enseñarme la mejor forma de estudiar matemáticas, y por incularme la capacidad de cuestionar hasta mis propias afirmaciones. A mi madre, por cada atención que tuvo conmigo. A mi incansable amigo por su compañia.

Índice general

Introducción

Este documento es la recopilación de varios resultados que permiten presentar una noción de los problemas de clasificación en Teoría Ergódica. Tuvo como motivación los cursos de Sistemas Dinámicos y Teoría de la Medida llevados a cabo el segundo semestre del año 2012, así como un especial interés en el estudio de los sistemas dinámicos bajo estructuras de medida. La clasificación de ciertos objetos es una cuestión usual en el estudio de estructuras matemáticas, y en el caso de Teoría Ergódica tiene relación con otras áreas, como el Análisis Funcional y la Topología. Una de las cosas que se muestra en este trabajo es cómo la clasificación puede hacerse bajo una estructura (espacio de Hilbert) que inicalmente no está presente en el contexto dado. Los problemas de clasificación se enmarcan además en un contexto mucho más general que lo aquí presentado, lo cual se nombra brevemente luego del capítulo 5, y cuya exposición va más allá de los alcances de este trabajo.

Se presentan inicialmente los preliminares necesarios para abordar las definiciones y resultados que se tienen en el contexto de los problemas de clasificación en Teoría Ergódica. Se requieren varias nociones del Análisis Funcional, las cuales no se presentan aquí, y de Teoría de la Medida, la mayoría contenidas en el capítulo 1. Posteriormente se introduce la noción de clasificación de transformaciones en espacios de medida y finalmente se muestra cómo la ergodicidad permite caracterizar ciertas transformaciones y de ahí obtener la clasificación deseada. Finalmente se comenta lo descrito en las primeras páginas del documento [17] que llevó a los interrogantes que condujeron al desarrollo de este trabajo. La exposición tiene el proposito de que cualquier lector, con las nociones básicas de las áreas de estudio nombradas anteriormente, pueda entender en menor tiempo la problemática de clasificación en este contexto y llegar a otros resultados, posteriores a lo contenido en este documento.

Notación

- El complemento de un conjunto se denota por A^c.
- El conjunto de partes de X se denota por $\mathcal{P}(X)$.
- Para $A, B \subset X$, $A \uplus B$ denota la unión disjunta, esto es, cuando $A \cap B = \emptyset$.
- La función $n \in \mathbb{N} \mapsto x_n \in X$, es la sucesión, que como función se denota por (x_n) y cuyo rango se denota por $\{x_n\}$.
- Una función $f : X \to Y$ biyectiva permite definir una función $f^! : Y \to X$ tal que $f \circ f^! = id = f^! \circ f$. A la función $f^!$ se le conoce como función inversa de f y se suele denotar igual que $f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$, la imagen recíproca mediante f, definida en subconjuntos. En el presente trabajo $f^!$ denota la función inversa definida mediante una biyección $f : X \to Y$. Cuando $A \subset X$, $f(A) := \{y \in Y : y = f(x) \text{ para algún } x \in A\}$. La imagen recíproca mediante f se denota como f^{-1} y para $B \subset Y$ es $f^{-1}(B) = \{x \in X : f(x) \in B\}$[1]. De esta forma, para $B \subset Y$ se tiene (cuando f es biyección) $f^{-1}(B) = f^!(B)$ y para $A \subset X$ se tiene que $f^{!^{-1}}(A) = f(A)$. También, $1/f$ se denota de esta forma y jamás como f^{-1}.
- Para $f : X \to Y$ y $C \subset \mathcal{P}(Y)$, $f^{-1}C := \left\{f^{-1}B : B \in C\right\} \subset \mathcal{P}(X)$.
- Para $A \subset X$ la función de pertenencia se denota como $\mathcal{X}_A$ y es tal que a $x \in X \mapsto 1$ cuando $x \in A$ y $x \in X \mapsto 0$ cuando $x \in A^c$.

[1] Cuando no haya lugar a dudas, fA denota igualmente $f(A)$ y $f^{-1}A$ denota $f^{-1}(A)$.

- En (X, d) espacio métrico, dados $0 < \epsilon$ y $x \in X$, $B_\epsilon(x) := \{y \in X : d(x, y) < \epsilon\}$ es la bola abierta con centro en x y radio ϵ.

- Si la función $f : X \to \mathbb{R}$ es idénticamente nula esto se denotará por $f \equiv 0$.

CAPÍTULO 1

Preliminares

1.1. Preliminares en Teoría de la Medida

A continuación se presentan los resultados de Teoría de la Medida que resultan más relevantes para este trabajo. Se introducen los conceptos básicos y se enuncian los resultados sin incluir demostraciones. Cualquier detalle adicional puede consultarse en [1] y [2].

1.1.1. Espacios y funciones medibles

En esta sección se consideran las construcciones básicas, a partir de conjuntos, hasta llegar a la estructura que permite definir las funciones medibles. La exposición se hace progresivamente de forma que se verifique cómo se definen cada vez condiciones más fuertes en las estructuras presentadas.

Definición 1.1. Para $X \neq \emptyset$, $\mathcal{S} \subset \mathcal{P}(X)$ con $\mathcal{S} \neq \emptyset$ es una semi-álgebra si satisface:

1. $\emptyset \in \mathcal{S}$.

2. Si $A, B \in \mathcal{S}$ entonces $A \cap B \in \mathcal{S}$.

3. Si $A \in \mathcal{S}$ entonces $A^c = \bigcup_{i=1}^{n} A_i$ para algunos $A_1, \ldots, A_n \in \mathcal{S}$.

Ejemplo 1.1. El principal ejemplo de una semi-álgebra es la colección

$$\tilde{\mathcal{I}} = \{(a,b], (-\infty, b], (a, +\infty), (-\infty, +\infty) \ : a, b \in \mathbb{R}\} \cup \{\emptyset\}.$$

A partir de esta colección se obtiene la medida de Lebesgue.

Definición 1.2. Para $X \neq \emptyset$, $\mathcal{A} \subset \mathcal{P}(X)$ con $\mathcal{A} \neq \emptyset$ es un álgebra si satisface:

1. Si $A \in \mathcal{A}$ entonces $A^c \in \mathcal{A}$.
2. Si $A_1, \ldots, A_n \in \mathcal{A}$ entonces $\bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$.

Definición 1.3. Para $X \neq \emptyset$, $\mathcal{X} \subset \mathcal{P}(X)$ con $\mathcal{X} \neq \emptyset$ es una σ-álgebra si satisface:

1. Si $A \in \mathcal{X}$ entonces $A^c \in \mathcal{X}$.
2. Si $A_1, A_2, \ldots \in \mathcal{X}$ entonces $\bigcup\limits_{i=1}^{+\infty} A_i \in \mathcal{X}$.

Si $A \in \mathcal{X}$ se dice que A es $\mathcal{X}$-medible, y se utiliza cuando se involucran dos o más σ-álgrebas. De lo contrario, cuando no hay lugar a dudas sobre la σ-álgebra considerada, se dice que A es medible. Toda σ-álgebra es un álgebra y una semi-álgebra. Toda álgebra es una semi-álgebra. Dado que $\mathcal{P}(X)$ es una σ-álgebra, se puede considerar la generación de estas estructuras como sigue.

Definición 1.4. Para $X \neq \emptyset$ y $B \subset \mathcal{P}(X)$, $B \neq \emptyset$:

- El álgebra generada por B es $\mathcal{A}(B) = \bigcap\limits_{\mathcal{A} \in C} \mathcal{A}$, donde $C = \{\mathcal{A} \text{ álgebra de } X : B \subset \mathcal{A}\}$.
- La σ-álgebra generada por B es $\Sigma(B) = \bigcap\limits_{\mathcal{X} \in D} \mathcal{X}$, donde $D = \{\mathcal{X}\ \sigma\text{-álgebra de } X : B \subset \mathcal{X}\}$.

Cuando B es contable se dice que $\Sigma(B)$ es contablemente generada. Por otro lado, el caso particular en el que la σ-álgebra está generada por una topología es de principal interés y la σ-álgebra así obtenida se llamará en adelante σ-álgebra de Borel (y todo abierto es medible).

Nota 1.1. El álgebra generada por una semi-álgebra $\mathcal{S}$ resulta ser la colección de uniones finitas de sus elementos. Dado que $\mathcal{S} \subset \mathcal{A}(\mathcal{S})$ implica que $\Sigma(\mathcal{S}) \subset \Sigma(\mathcal{A}(\mathcal{S}))$ y que $\mathcal{A}(\mathcal{S}) \subset \Sigma(\mathcal{S})$ se obtiene que $\Sigma(\mathcal{S}) = \Sigma(\mathcal{A}(\mathcal{S}))$.

Un espacio medible es una dupla $(X, \mathcal{X})$ en la que $\mathcal{X}$ es una σ-álgebra de subconjuntos de X. En este espacio se consideran las funciones medibles.

Definición 1.5. Para $(X, \mathcal{X})$, una función $f : X \to \mathbb{R}$ es $\mathcal{X}$-medible (o medible) si $f^{-1}(\alpha, +\infty) \in \mathcal{X}$ para todo $\alpha \in \mathbb{R}$.

Las partes positiva y negativa de una función $f : X \to \mathbb{R}$ son las funciones definidas como $f^+(x) = \sup\{f(x), 0\}$ y $f^-(x) = \sup\{-f(x), 0\}$ respectivamente. Se puede verificar que f es medible si y sólo si f^+ y f^- son medibles. El conjunto de funciones medibles se denota por $\mathcal{M}(X, \mathcal{X})$. Se considera también $\mathcal{M}^+(X, \mathcal{X}) = \{f \in \mathcal{M}(X, \mathcal{X}) : 0 \leq f\}$. Cuando no haya lugar a duda sobre X y $\mathcal{X}$ se denotarán solamente por $\mathcal{M}$ y $\mathcal{M}^+$ respectivamente.

Se amplía el conjunto de funciones medibles $\mathcal{M}$ a $\mathcal{M}_{\mathbb{C}}$, considerando $f : X \to \mathbb{C}$ y definiendo $f \in \mathcal{M}_{\mathbb{C}}$ si $\mathrm{Re}(f), \mathrm{Im}(f) \in \mathcal{M}$, donde $f = \mathrm{Re}(f) + i\,\mathrm{Im}(f)$. Más generalmente, si $(X, \mathcal{X})$ y $(Y, \mathcal{Y})$ son espacios medibles, $f : X \to Y$ es medible si $f^{-1}\mathcal{Y} \subset \mathcal{X}$.

Otros resultados que utilizaremos más adelante, cuando se considera una función $f : X \to Y$, son:

- $\{E \subset Y : f^{-1}E \in \mathcal{X}\}$ es una σ-álgebra de Y.
- $f^{-1}\mathcal{Y}$ es una σ-álgebra de X.
- Si (f_n) es una sucesión en $\mathcal{M}$ entonces las funciones definidas puntualmente como $\inf\limits_n f_n$, $\sup\limits_n f_n$, $\liminf\limits_n f_n$, y $\limsup\limits_n f_n$ pertenecen a $\mathcal{M}$.
- Si (f_n) es una sucesión en $\mathcal{M}$ y $\lim\limits_n f_n(x) = f(x)$ para todo $x \in X$ entonces $f \in \mathcal{M}$.
- Si $f \in \mathcal{M}^+$ entonces existe una sucesión (φ_n) en $\mathcal{M}^+$ de funciones medibles, cada una de las cuales toma un número finito de valores, tales que $\varphi_n \leq \varphi_{n+1}$ y $\lim\limits_n \varphi(x) = f(x)$ para todo $x \in X$.

1.1.2. Espacios de medida y funciones integrables

En la parte anterior se consideró y definió la noción de σ-álgebra, que se usa como dominio de la función de medida. En general se estudia, a partir de la definición de medida, cómo se obtienen (generación de medidas) y de que formas se representan (descomposición de medidas) las medidas. Como se muestra en esta sección, las condiciones pueden debilitarse para definir una función de la cual se obtenga una medida. La forma más general de la integral de una función se define en el contexto de Teoría de la Medida. Nótese que uno de los resultados de la sección anterior dice que cualquier función medible no negativa puede ser aproximada mediante funciones que toman solamente un número finito de valores. La definición de la integral de una función medible se hace progresivamente, partiendo de la motivación del resultado anterior, pues como se verá, una vez se tiene una medida, las funciones medibles que toman solamente un número finito de valores son de especial simplicidad pero permiten extender las definiciones de manera consistente.

Definición 1.6. Una pre-medida sobre $C \subset \mathcal{P}(X)$, con $\emptyset \in C$, es una función $\tilde{\mu} : C \to \mathbb{R}_0^+ \cup \{+\infty\}$ tal que:

1. $\tilde{\mu}(\emptyset) = 0$.
2. Si $\{A_n\} \subset C$ es una sucesión disjunta y $\biguplus_{n=1}^{+\infty} A_n \in C$ entonces $\tilde{\mu}\left(\biguplus_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \tilde{\mu}(A_n)$.

La última condición se conoce como σ-aditividad. Una pre-medida es una función σ-aditiva sobre uniones contables, siempre y cuando estas uniones contables pertenezcan a la colección donde está definida, esto es, su dominio.

Definición 1.7. Una medida sobre una σ-álgebra $\mathcal{X}$ de subconjuntos de X es una función $\mu : \mathcal{X} \to \mathbb{R}_0^+ \cup \{+\infty\}$ tal que:

1. $\mu(\emptyset) = 0$.
2. Si $\{A_n\} \subset \mathcal{X}$ es una sucesión disjunta entonces $\mu\left(\biguplus_{n=1}^{+\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n)$.

Una medida en una σ-álgebra $\mathcal{X}$ se puede obtener a partir de definir una pre-medida en una semi-álgebra $\mathcal{S}$ tal que $\Sigma(\mathcal{S}) = \mathcal{X}$, esto mediante el Teorema de Extensión de Carathéodory, el cual se presenta luego de las consideraciones a continuación.

Nota 1.2. Se dice que una medida μ definida en una σ-álgebra $\mathcal{X}$ de subconjuntos de X es:

- σ-finita si existe una sucesión de conjuntos medibles (A_n) tal que $X = \bigcup_{n=1}^{+\infty} A_n$ y $\mu(A_n) < +\infty$ para todo $n \in \mathbb{N}$,
- finita si $\mu(X) < +\infty$.
- de probabilidad si $\mu(X) = 1$,

Ejemplo 1.2. Algunos ejemplos de medidas son:

- En $\mathcal{P}(\mathbb{N})$ la medida μ que asigna el cardinal de $A \subset \mathbb{N}$ cuando A es finito y $+\infty$ cuando es contable, se denomina la medida del conteo.
- En $X \neq \emptyset$ y $\mathcal{X}$ σ-álgebra de X, con $x \in X$ fijo, la medida de probabilidad concentrada en x esta definida como $\mu(A) = 1$ cuando $x \in A$ y $\mu(A) = 0$ en caso contrario.

Teorema 1.1. Teorema de Extensión de Carathéodory. Si $\tilde{\mu}$ es una pre-medida sobre la semi-álgebra $C \subset \mathcal{P}(X)$ entonces existe una medida μ sobre $\Sigma(C)$ que coincide con $\tilde{\mu}$ en C. Si $\tilde{\mu}$ es finita entonces μ es única.

La demostración de este teorema (de Extensión de Carathéodory) puede consultarse en [1]. Se parte de extender la pre-medida al álgebra generada (que se obtiene de manera explicita como se presentó en la nota 1.1.) mediante σ-aditividad directamente, y de ahí definir una medida exterior[1] (a partir del álgebra) que genera una σ-álgebra que contiene a $\Sigma(C)$ (por definición). Finalmente se demuestra que esta medida exterior coincide con $\tilde{\mu}$ en C, y se obtiene la extensión deseada. La siguiente definición enriquece la estructura de σ-álgebra y es de principal interés cuando se trabajan proposiciones de la forma *casi en toda parte*, las cuales se introducirán más adelante.

Definición 1.8. La completación de una σ-álgebra $\mathcal{X}$ respecto a una medida μ en ella definida es la σ-álgebra

$$\hat{\mathcal{X}} = \{A \cup \hat{Z} : A \in \mathcal{X} \text{ y } \hat{Z} \subset Z \in \mathcal{X}, \mu(Z) = 0\}.$$

Nota 1.3. Obsérvese que $\mathcal{X} \subset \hat{\mathcal{X}}$ y que todos los subconjuntos de conjuntos $\mathcal{X}$-medibles de medida nula[2] son $\hat{\mathcal{X}}$-medibles. La completación de una medida μ respecto a la completación de la σ-álgebra $\mathcal{X}$ es la medida $\hat{\mu}$ definida en $\hat{\mathcal{X}}$ por $\hat{\mu}\left(A \cup \hat{Z}\right) = \mu(A)$, y como resultado satisface $\hat{\mu}|_{\mathcal{X}} \equiv \mu$.

Un espacio de medida es una tripla $(X, \mathcal{X}, \mu)$ tal que $\mathcal{X}$ es una σ-álgebra de subconjuntos de X y μ es una medida definida en $\mathcal{X}$. Cuando $\mu(X) = 1$ la tripla se denomina espacio de probabilidad. En adelante, siempre que se escriba $(X, \mathcal{X}, \mu)$, se tratará de un espacio de medida.

Nota 1.4. Se dice que una proposición se cumple casi en toda parte (o casi en todo X) si existe $N \in \mathcal{X}$ tal que la proposición se cumple en N y $\mu(N^c) = 0$. Cuando la σ-álgebra es completa, puede tomarse directamente como N al conjunto donde estrictamente la proposición es cierta, y en su complemento no lo es. Obsérvese además que dos proposiciones que se cumplen casi en toda parte definen subconjuntos medibles N y M de medida completa y que necesariamente satisfacen $N \cap M \neq \emptyset$. Se obtiene así el conjunto $N \cap M$ de medida completa en el que se satisfacen ambas proposiciones. Es por esto que, aunque las proposiciones casi en toda parte tienen un rol fundamental en Teoría de la Medida, pueden relajarse las condiciones y trabajar varias proposiciones a

[1]La definición de medida exterior puede consultarse igualmente en [1]

[2]Cuyos complementos se denominan de medida completa

la vez, teniendo en cuenta que siempre existirá un conjunto medible en el que todas las proposiciones se satisfacen, el cual no es otra cosa que la intersección de los conjuntos medibles en los que se satisface cada proposición individualmente.

Cuando μ es una medida en $\mathcal{X}$ y $A, B, A_n \in \mathcal{X}$ para todo $n \in \mathbb{N}$, con $A \subset B$, se tiene que:

- $\mu(A) \leq \mu(B)$.
- Si $\mu(A) < +\infty$ entonces $\mu(B \setminus A) = \mu(B) - \mu(A)$.
- Si $A_n \subset A_{n+1}$ para todo $n \in \mathbb{N}$ entonces $\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_n \mu(A_n)$.
- Si $A_{n+1} \subset A_n$ para todo $n \in \mathbb{N}$ y $\mu(A_1) < +\infty$ entonces $\mu\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_n \mu(A_n)$.

Ejemplo 1.3. El ejemplo principal de la construcción de una σ-álgebra y su completación, a partir de una semi-álgebra y una pre-medida en ésta, es la σ-álgebra de Borel cuya completación es la σ-álgebra de Lebesgue. La σ-álgebra de Borel $\mathscr{B}$ en $\mathbb{R}$ se define como la generada por la colección de intervalos abiertos de $\mathbb{R}$ y resulta ser contablemente generada por la colección $\{(t, +\infty) : t \in \mathbb{Q}\}$. Dado que

$$\tilde{\mathcal{I}} = \{(a, b], (-\infty, b], (a, +\infty), (-\infty, +\infty) : a, b \in \mathbb{R}\}$$

es una semi-álgebra, se considera el álgebra generada por esta colección. En esta álgebra se define una pre-medida que asigna a los intervalos su longitud: $\tilde{\lambda}((a, b)) := b - a$. Con $\tilde{\lambda}$ se considera una medida exterior λ (la medida de Lebesgue) que define una σ-álgebra (de Lebesgue) sobre $\mathbb{R}$, y que escribimos como $\mathscr{L}$. Por definición resulta que $\mathscr{B} \subset \mathscr{L}$ y se puede verificar que $\mathscr{L}$ es la completación de $\mathscr{B}$. Los detalles de esta construcción se encuentran en [1].

Nota 1.5. Si $\mathcal{X}$ es una σ-álgebra de X y $A \subset X$, $\mathcal{X}$ induce la estructura sobre A definida por $\mathcal{X}_A := \{A \cap B : B \in \mathcal{X}\}$, que es una σ-álgebra de subconjuntos de A. Para cualquier subconjunto A de $\mathbb{R}$ se considera entonces la σ-álgebra de Borel y la σ-álgebra de Lebesgue inducida en A. El caso más usual es cuando $A = [0, 1]$ y se denotan por $\mathscr{B}_{[0,1]}$ y $\mathscr{L}_{[0,1]}$ respectivamente.

En un espacio de medida $(X, \mathcal{X}, \mu)$ se consideran las funciones medibles y se definen progresivamente las funciones integrables, como sigue:

Definición 1.9. Dado $(X, \mathcal{X}, \mu)$, una función $\varphi : X \to \mathbb{R}$ es simple si toma finitos valores.

Nota 1.6. Obsérvese que cuando la función es medible, puede representarse como $\varphi = \sum_{j=1}^{n} a_j \mathcal{X}_{A_j}$, donde $A_j \in \mathcal{X}$.

Definición 1.10. Si $\varphi : X \to \mathbb{R}$ es una función medible y simple, su integral se define como el número real (o inclusive $+\infty$) dado por

$$\int \varphi \, d\mu := \sum_{j=1}^{n} a_j \mu\left(A_j\right).$$

Con la aproximació de funciones medibles mediante funciones simples se puede extender consistentemente la definición de integral a funciones no negativas mediante funciones simples.

Definición 1.11. Para $f \in \mathcal{M}^+$ se define la integral de f como el número real (o inclusive $+\infty$) dado por

$$\int f \, d\mu := \sup \left\{ \int \varphi \, d\mu : \varphi \in \mathcal{M}^+, \varphi \text{ función simple tal que } \varphi \leq f \right\}.$$

Si $A \in \mathcal{X}$ se define $\int_A f \, d\mu := \int f \mathcal{X}_A \, d\mu$.

Algunos resultados para funciones en $\mathcal{M}^+ (X, \mathcal{X})$ son:

- La integral es una funcional lineal en el espacio vectorial de funciones simples.
- Si $f, g \in \mathcal{M}^+$ y $f \leq g$ entonces $\int f \, d\mu \leq \int g \, d\mu$.
- Si $f \in \mathcal{M}^+$ y $A, B \in \mathcal{X}$ con $A \subset B$ entonces $\int_A f \, d\mu \leq \int_B f \, d\mu$.
- Teorema de Convergencia Monótona. Si (f_n) es una sucesión en $\mathcal{M}^+$ tal que $f_n \leq f_{n+1}$ y $\lim_n f_n(x) = f(x)$ para todo $x \in X$ entonces $\int f \, d\mu = \lim_n \int f_n \, d\mu$.
- Si $f \in \mathcal{M}^+$ y $0 \leq c$ entonces $cf \in \mathcal{M}^+$ y $\int cf \, d\mu = c \int f \, d\mu$.
- Si $f, g \in \mathcal{M}^+$ entonces $f + g \in \mathcal{M}^+$ y $\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu$.
- Si $f \in \mathcal{M}^+$ entonces la función $A \in \mathcal{X} \mapsto \int_A f \, d\mu$ es una medida en $\mathcal{X}$.
- Para $f \in \mathcal{M}^+$, $f(x) = 0$ casi en toda parte si y sólo si $\int f \, d\mu = 0$.
- Si (f_n) es una sucesión en $\mathcal{M}^+$ tal que $f_n \leq f_{n+1}$, $\lim_n f_n(x) = f(x)$ casi en toda parte y $f \in \mathcal{M}^+$ entonces $\int f \, d\mu = \lim_n \int f_n \, d\mu$.

- Si (f_n) es una sucesión en $\mathcal{M}^+$ entonces $\int \sum_{n=1}^{+\infty} f_n \, d\mu = \sum_{n=1}^{+\infty} \int f_n \, d\mu$.

Definición 1.12. Dado $(X, \mathcal{X}, \mu)$, se define el espacio de funciones integrables

$$\mathcal{L}(X, \mathcal{X}, \mu) := \left\{ f \in \mathcal{M} : \int f^+ \, d\mu, \int f^- \, d\mu < +\infty \right\}.$$

Para $f \in \mathcal{L}(X, \mathcal{X}, \mu)$, $\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$. Si $A \in \mathcal{X}$, $\int_A f \, d\mu := \int_A f^+ \, d\mu - \int_A f^- \, d\mu$. Cuando no haya lugar a duda sobre cuáles son X, $\mathcal{X}$ y μ, se denotará a $\mathcal{L}(X, \mathcal{X}, \mu)$ solamente por $\mathcal{L}$.

Así como con las funciones medibles se amplía el conjunto de funciones integrables, considerando $f : X \to \mathbb{C}$ y definiendo $f \in \mathcal{L}_{\mathbb{C}}$ siempre que $\mathrm{Re}(f), \mathrm{Im}(f) \in \mathcal{L}$.

Algunos resultados para funciones en $\mathcal{L}(X, \mathcal{X}, \mu)$ son:

- Cuando $f \in \mathcal{M}$, $f \in \mathcal{L}$ si y sólo si $|f| \in \mathcal{L}$, y en este caso $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$.
- Si $f \in \mathcal{M}$ y $g \in \mathcal{L}$ es tal que $|f| \leq |g|$ entonces $f \in \mathcal{L}$ y $\int |f| \, d\mu \leq \int |g| \, d\mu$.
- Si $f, g \in \mathcal{L}$ y $\alpha \in \mathbb{R}$ entonces αf, $f + g \in \mathcal{L}$ y $\int \alpha f \, d\mu = \alpha \int f \, d\mu$, $\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu$ [3].
- Teorema de Convergencia Dominada. Si (f_n) es una sucesión en $\mathcal{L}$ que satisface $\lim_n f_n(x) = f(x)$ casi en toda parte, con $f \in \mathcal{M}$ de valor real, y $g \in \mathcal{L}$ es tal que $|f_n| \leq g$ para todo $n \in \mathbb{N}$, entonces $f \in \mathcal{L}$ y $\int f \, d\mu = \lim_n \int f_n \, d\mu$.

1.1.3. Los espacios $\mathcal{L}^p$

Como consecuencia de las propiedades presentadas en la sección anterior se tiene que $\mathcal{L}$ es un espacio vectorial. Si $f : X \to \mathbb{R}$, $|f|$ puede expresarse como $|f| = f^+ + f^-$, con lo cual, siempre que f sea medible, se obtiene la equivalencia $f \in \mathcal{L}$ si y sólo si $|f| \in \mathcal{L}$. De esto y los resultados anteriores, la función $f \in \mathcal{L} \mapsto \int |f| \, d\mu$ es una seminorma. Para obtener una norma se establece la relación de equivalencia $\mathcal{R}$ en $\mathcal{L}$ definida por $f \mathcal{R} g$ si y sólo si $f(x) = g(x)$ casi en toda parte. La proposición $f(x) = g(x)$ casi en toda parte se denotará de ahora en adelante como $f \overset{\mu}{=} g$. Se define el espacio vectorial $\mathcal{L}^1 := \mathcal{L}/\mathcal{R}$, en el que la función $[f] \mapsto \int |f| \, d\mu$ es una norma. Más generalmente, para $1 \leq p < +\infty$, $\mathcal{L}^p := \left\{ [f] \in \mathcal{M}/\mathcal{R} : \int |f|^p \, d\mu < +\infty \right\}$ y es un espacio

[3] El mismo resultado se tiene en $\mathcal{L}_{\mathbb{C}}$.

vectorial en el que la función $[f] \mapsto \left\|[f]\right\|_p := \left(\int |f|^p \, d\mu\right)^{1/p}$ es una norma[4]. El espacio normado $\left(\mathcal{L}^p, \|\cdot\|_p\right)$ es un espacio de Banach, esto es, es completo respecto a la métrica inducida por la norma. Cuando no haya ambigüedad se escribirá $f \in \mathcal{L}^p$ y $\|f\|_p$, haciendo referencia a $[f] \in \mathcal{L}^p$ y $\left\|[f]\right\|_p$ respectivamente.

Algunos resultados en $\mathcal{L}^p$ son:

- (Desigualdad de Hölder) Si $f \in \mathcal{L}^p$ y $g \in \mathcal{L}^q$, con $\frac{1}{p} + \frac{1}{q} = 1$, entonces $fg \in \mathcal{L}^1$ y $\|fg\|_1 \leq \|f\|_p \, \|g\|_q$.
- Si (f_n) es una sucesión en $\mathcal{L}^p$ tal que $\lim_n f_n \overset{\mu}{=} f$, con $f \in \mathcal{M}$, y existe $g \in \mathcal{L}^p$ tal que $|f_n| \leq g$ para todo $n \in \mathbb{N}$, entonces $f \in \mathcal{L}^p$ y $\lim_n \|f_n - f\|_p = 0$.
- Si se denota por S al conjunto de funciones simples en $\mathcal{L}^p$ entonces, con la métrica inducida por $\|\cdot\|_p$, $\overline{S} = \mathcal{L}^p$, esto es, dados $f \in \mathcal{L}^p$ y $\epsilon > 0$ existe una función simple φ tal que $\|f - \varphi\|_p < \epsilon$. Este resultado permite extender ciertas propiedades a todo $\mathcal{L}^p$ partiendo de funciones características de conjuntos medibles, luego a funciones simples y finalmente a funciones arbitrarias en $\mathcal{L}^p$.

1.2. Preliminares en Teoría Ergódica

La Teoría Ergódica estudia las propiedades de transformaciones $T : X \to Y$ cuando X y Y tienen estructuras de medida. Nótese que si $Y = X$, se obtiene un sistema dinámico en una estructura de medida.

Si T es una función de X en Y y $\mathcal{Y}$ es una σ-álgebra de subconjuntos de Y entonces $T^{-1}\mathcal{Y}$ es una σ-álgebra de subconjuntos de X. Por otro lado, si $C \subset \mathcal{P}(Y)$, $T^{-1}C = \{T^{-1}B : B \in C\} \subset \mathcal{P}(X)$ y entonces $T^{-1}\Sigma(C) = \Sigma\left(T^{-1}C\right)$. En efecto:

- Dado que $C \subset \Sigma(C)$, por la definición de $\Sigma\left(T^{-1}C\right)$ se tiene $\Sigma\left(T^{-1}C\right) \subset T^{-1}\Sigma(C)$.
- $\mathcal{Y}^* = \left\{E \subset Y : T^{-1}E \in \Sigma\left(T^{-1}C\right)\right\}$ es una σ-álgebra de Y tal que $C \subset \mathcal{Y}^*$ (pues $T^{-1}C \subset \Sigma\left(T^{-1}C\right)$), entonces $\Sigma(C) \subset \mathcal{Y}^*$. También $T^{-1}\mathcal{Y}^* \subset \Sigma\left(T^{-1}C\right)$, con lo cual $T^{-1}\Sigma(C) \subset \Sigma\left(T^{-1}C\right)$

[4]Obsérvese que se puede definir un espacio vectorial análogo $\mathcal{L}^p_{\mathbb{C}}$ considerando $f : X \to \mathbb{C}$.

Cuando X y Y tienen estructuras de medida se definen transformaciones medibles de la siguiente manera:

Definición 1.13. Para $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ espacios de medida[5], se dice que $T : X \to Y$ es medible si $T^{-1}\mathcal{Y} \subset \mathcal{X}$. Cuando T es medible, se dice que T preserva la medida si para todo $A \in \mathcal{Y}$ se tiene que $\mu\left(T^{-1}A\right) = \lambda(A)$.

Cuando $T : X \to X$, se obtiene un sistema dinámico discreto $(X, \{T^n : n \in \mathbb{N}\})$ en el que la composición es operación de semi-grupo. En el caso en que T es invertible, el sistema discreto es $(X, \{T^n : n \in \mathbb{Z}\})$ y la composición es operación de grupo. En adelante, $(X, \mathcal{X}, \mu, T)$ denota un espacio de medida junto con una transformación $T : X \to X$, $\mathcal{X}$-medible (o medible)[6] que preserva la medida μ. Estas transformaciones se pueden caracterizar de las siguiente forma:

Proposición 1.1. Sea $T : X \to X$ una transformación medible. T preserva la medida si y sólo si para toda $f \in \mathcal{M}$: $\int f \, d\mu = \int f \circ T \, d\mu$.

Demostración:

Si para toda $f \in \mathcal{M}$ se satisface $\int f \, d\mu = \int f \circ T \, d\mu$, entonces tomando $A \in \mathcal{X}$, $\mathcal{X}_A \in \mathcal{M}$ y $\mu(A) = \int \mathcal{X}_A \, d\mu = \int \mathcal{X}_A \circ T \, d\mu = \int \mathcal{X}_{T^{-1}A} \, d\mu = \mu\left(T^{-1}A\right)$. Recíprocamente, si T preserva la medida, para todo $A \in \mathcal{X}$ se tiene $\int \mathcal{X}_A \, d\mu = \int \mathcal{X}_A \circ T \, d\mu$, y por linealidad de la integral se tiene también para toda función simple φ: $\int \varphi \, d\mu = \int \varphi \circ T \, d\mu$. Para $f \in \mathcal{M}$ sean (φ_n) y (β_n) sucesiones monótonas de funciones simples, no negativas, tales que $\lim_n \varphi_n(x) = f^+(x)$ y $\lim_n \beta_n(x) = f^-(x)$ para todo $x \in X$. Por el Teorema de Convergencia Monótona, $\int f^+ \, d\mu = \lim_n \int \varphi_n \, d\mu$ y $\int f^- \, d\mu = \lim_n \int \beta_n \, d\mu$. Dado que $\lim_n (\varphi_n \circ T)(x) = (f^+ \circ T)(x)$ y $\lim_n (\beta_n \circ T)(x) = (f^- \circ T)(x)$ para todo $x \in X$, nuevamente por Teorema de Convergencia Monótona $\int f^+ \circ T \, d\mu = \lim_n \int \varphi_n \circ T \, d\mu$ y $\int f^- \circ T \, d\mu = \lim_n \int \beta_n \circ T \, d\mu$. Se obtiene entonces que

$$\int f \, d\mu = \lim_n \int \varphi_n \circ T \, d\mu - \lim_n \int \beta_n \circ T \, d\mu = \int (f^+ - f^-) \circ T \, d\mu = \int f \circ T \, d\mu.$$

Si $\mu(X) < +\infty$, $\int f \, d\mu = \int f \circ T \, d\mu$ para toda $f \in \mathcal{M}$ es lo mismo que para toda $f \in \mathcal{L}^1$ o $f \in \mathcal{L}^2$.

Definición 1.14. La transformación T en $(X, \mathcal{X}, \mu, T)$, donde $(X, \mathcal{X}, \mu)$ es un espacio de probabilidad, es ergódica respecto a μ si para todo $A \in \mathcal{X}$ se tiene que $T^{-1}A = A$ implica $\mu(A) \in \{0, 1\}$.

[5]En este contexto, λ denota una medida arbitaria, no se debe confundir con la medida de Lebesgue

[6]Aquí la estructura de media es la misma en el dominio y el codominio.

Obsérvese que la transformación T puede ser ergódica con una medida pero no neceseriamente con otra. Cuando T preserva la medida μ se dice que μ es T-invariante, y cuando T es además ergódica se dice también que μ es una medida ergódica. Si fijamos la estructura medible $(X, \mathcal{X})$ y la transformación, el conjunto de medidas satisface propiedades de caracterización importantes. Por ejemplo, las medidas T-invariantes son un conjunto convexo y las medidas ergódicas resultan ser los puntos extremos de ese conjunto. Estas observaciones dan lugar a otros resultados, los cuales pueden ser consultados en [13].

Las transformaciones ergódicas se pueden caracterizar de varias formas (Ver [9]):

Proposición 1.2. Sea $(X, \mathcal{X}, \mu, T)$, donde $(X, \mathcal{X}, \mu)$ es de probabilidad. Las siguientes afirmaciones son equivalentes:

1. T es ergódica.
2. Para todo $A \in \mathcal{X}$, si $\mu\left(T^{-1}A \triangle A\right) = 0$ entonces $\mu(A) \in \{0, 1\}$.
3. Para todo $A \in \mathcal{X}$, si $\mu(A) > 0$ entonces $\mu\left(\bigcup_{n=1}^{+\infty} T^{-n}A\right) = 1$.
4. Para todos $A, B \in \mathcal{X}$, si $\mu(A), \mu(B) > 0$, entonces existe $n \in \mathbb{Z}^+$ tal que $\mu\left(T^{-n}A \cup B\right) > 0$.
5. Para toda $f \in \mathcal{M}$, si $f \circ T \overset{\mu}{=} f$ entonces $f \overset{\mu}{=} c$ (es constante casi en toda parte), para algún $c \in \mathbb{R}$.

Demostración:

(1. implica 2.) Sea $A \in \mathcal{X}$ tal que $\mu\left(T^{-1}A \triangle A\right) = 0$. Sea

$$C = \bigcap_{n \in \mathbb{N}} \bigcup_{k=n}^{+\infty} T^{-k}A$$

Entonces, $T^{-1}C = C$ y por 1. esto implica que $\mu(C) \in \{0, 1\}$. Por otro lado

$$\begin{aligned}
\mu(C \triangle A) &= \mu\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k=n}^{+\infty} T^{-k}A \cap A^c\right) + \mu\left(\bigcup_{n \in \mathbb{N}} \bigcap_{k=n}^{+\infty} T^{-k}A^c \cap A\right) \\
&\leq \mu\left(\bigcup_{k=1}^{+\infty} T^{-k}A \cap A^c\right) + \mu\left(\bigcup_{k=1}^{+\infty} T^{-k}A^c \cap A\right) \\
&\leq \sum_{k=1}^{+\infty} \mu\left(T^{-k}A \triangle A\right)
\end{aligned}$$

Dado que en general $E\triangle F \subset E\triangle G \cup G\triangle F$ para arbitrarios $E, F, G \subset X$, por inducción se obtiene que $\mu\left(T^{-k}A\triangle A\right) = 0$ para todo $k \in \mathbb{N}$. Entonces $\mu(C\triangle A) = 0$, y siendo $\mu(C) = \mu(C\cap A) + \mu(C\cap A^c)$ y $\mu(A) = \mu(A\cap C) + \mu(A\cap C^c)$, $\mu(C) = \mu(A)$, de donde $\mu(A) \in \{0,1\}$.

(2. implica 3.) Sea $A \in \mathcal{X}$ tal que $\mu(A) > 0$. Sea $B = \bigcup\limits_{n\in\mathbb{N}} T^{-n}A$. Entonces $T^{-1}B \subset B$, y dado que T preserva la medida, necesariamente $\mu(B) > 0$, pues en caso contrario se tendría $\mu\left(T^{-1}A\right) \leq \mu(B) = 0$, contrario a la escogencia de A. Por otro lado

$$\mu\left(T^{-1}B\triangle B\right) = \mu\left(B\setminus T^{-1}B\right) = \mu(B) - \mu\left(T^{-1}B\right) = 0,$$

por lo que 2. permite concluir que $\mu(B) = 1$.

(3. implica 4.) Sean $A, B \in \mathcal{X}$ con $\mu(A), \mu(B) > 0$. Por la hipótesis

$$\mu(B) = \mu\left(B\cap\bigcup_{n=1}^{+\infty}T^{-n}A\right) = \mu\left(\bigcup_{n=1}^{+\infty}B\cap T^{-n}A\right) > 0.$$

Si para todo $k \in \mathbb{N}$ se tuviera que $\mu\left(B\cap T^{-k}A\right) = 0$ se llegaría a contradecir lo anterior. De esto se sigue el resultado.

(4. implica 1.) Sea $A \in \mathcal{X}$ tal que $T^{-1}A = A$. Si $\mu(A) = 0$ el resultado se tiene. Supóngase que $\mu(A) > 0$. Si $\mu(A^c) > 0$, por 4., existe $k \in \mathbb{N}$ tal que $\mu\left(A^c\cap T^{-k}A\right) > 0$. Dado que la hipótesis implica que $T^{-k}A = A$, se segiría que $\mu(A^c\cap A) > 0$, contradiciendo la escogencia de A. Se concluye que $\mu(A) = 1$, por lo tanto T es ergódica.

(1. implica 5.) Sea $f \in \mathcal{M}$ tal que $f\circ T \overset{\mu}{=} f$. Para cada $n \in \mathbb{N}$ y $k \in \mathbb{Z}$ sea

$$X_{k,n} = \left\{x\in X:\ \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\}.$$

Entonces $T^{-1}X_{k,n}\triangle X_{k,n} \subset \{x\in X:\ f\circ T(x) \neq f(x)\}$ implica que

$$\mu\left(T^{-1}X_{k,n}\triangle X_{k,n}\right) = 0.$$

Dado que 1. implica 2. se obtiene que $\mu(X_{k,n}) \in \{0,1\}$ para cada $k \in \mathbb{Z}$. Por otro lado, para $n \in \mathbb{N}$ fijo, $X = \biguplus\limits_{k\in\mathbb{Z}} X_{k,n}$, por lo que para cada $n \in \mathbb{N}$ existe un único $k_n \in \mathbb{Z}$ tal que $\mu(X_{k_n,n}) = 1$, y necesariamente $X_{k_1,1} \supseteq X_{k_2,2} \supseteq \ldots$, de donde $\left(\frac{k_n}{2^n}\right)$ es una sucesión monótona creciente y acotada, por lo tanto convergente a $\lim\limits_n \frac{k_n}{2^n}$. Sea $Y = \bigcap\limits_{n\in\mathbb{N}} X_{k_n,n}$. Entonces $\mu(Y) = 1$ y si $x \in Y$ se tiene que $0 \leq \left|f(x) - \frac{k_n}{2^n}\right| < 2^{-n}$ para todo $n \in \mathbb{N}$, de donde $f(x) = \lim\limits_n \frac{k_n}{2^n}$. Se concluye entonces que f es constante casi en toda parte.

(5. implica 1.) Sea $A \in \mathcal{X}$ tal que $T^{-1}A = A$. Si $\mu(A) =$ el resultado se tiene. Supóngase entonces que $\mu(A) > 0$. Dado que $\mathcal{X}_A \circ T = \mathcal{X}_{T^{-1}A} = \mathcal{X}_A$, por 5. resulta que $\mathcal{X}_A$ es constante casi en toda parte, esto es $\mathcal{X}_A \overset{\mu}{=} 1$, de donde $\mu(A) = 1$.

Para verificar ergodicidad se tiene una condición más fuerte:

Definición 1.15. La transformación T en $(X, \mathcal{X}, \mu, T)$, donde $(X, \mathcal{X}, \mu)$ es de probabilidad, es una mezcla fuerte respecto a μ si para todos $A, B \in \mathcal{X}$ se tiene que $\lim_n \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$.

Toda mezcla fuerte es ergódica, como se verifica al tomar el límite:

$$\mu(A) = \lim_n \mu(T^{-n}A \cap A) = \mu(A)\mu(A).$$

Las transformaciones que son una mezcla fuerte también se puede caracterizar de la siguiente forma:

Proposición 1.3. En $(X, \mathcal{X}, \mu, T)$, con $(X, \mathcal{X}, \mu)$ de probabilidad, T es una mezcla fuerte respecto a μ si y sólo si para todas $f, g \in \mathcal{M}$ se tiene que

$$\lim_n \int f(g \circ T^{-n})\, d\mu = \int f\, d\mu \int g\, d\mu$$

Demostración:

Si para todas $f, g \in \mathcal{M}$ se tiene que $\lim_n \int f(g \circ T^{-n})\, d\mu = \int f\, d\mu \int g\, d\mu$, entonces considerando las funciones características de $A, B \in \mathcal{X}$ arbitrarios:

$$\begin{aligned}
\lim_n \mu(T^{-n}A \cap B) &= \lim_n \int \mathcal{X}_{T^{-n}A \cap B}\, d\mu \\
&= \lim_n \int \mathcal{X}_B (\mathcal{X}_A \circ T^{-n})\, d\mu = \int \mathcal{X}_A\, d\mu \int \mathcal{X}_B\, d\mu \\
&= \mu(A)\mu(B)
\end{aligned}$$

Por otro lado, si T es una mezcla fuerte respecto a μ entonces, reescribiendo la identidad del límite, se obtiene que

$$\lim_n \int \mathcal{X}_B (\mathcal{X}_A \circ T^{-n})\, d\mu = \int \mathcal{X}_A\, d\mu \int \mathcal{X}_B\, d\mu,$$

para cualquier par $A, B \in \mathcal{X}$. Por linealidad de la integral se sigue el resultado para funciones simples. Con la aproximación de funciones medibles mediante sucesiones de funciones simples se obtiene el resultado para cualquier par $f, g \in \mathcal{M}$.

Las transformaciones ergódicas tiene propiedades de interés en el espacio $\mathcal{L}^2_{\mathbb{C}}$ respectivo. Obsérvese que el espacio $\mathcal{L}^2_{\mathbb{C}}$ es un espacio de Hilbert[7] con el producto escalar $\langle f, g\rangle := \int f\,\overline{g}\,d\mu$. La norma inducida por $\langle\cdot,\cdot\rangle$ es $\|\cdot\|_2$. De la caracterización de transformaciones $T : X \to X$ medibles que preservan la medida, se introduce el Operador de Koopman U_T que es una transformación lineal en el espacio vectorial $\mathcal{L}^2$ asociado a $(X, \mathcal{X}, \mu)$. El estudio del operador U_T en el contexto usual del Análisis Funcional permite definir propiedades para T en términos de propiedades de U_T.

Definición 1.16. El Operador de Koopman asociado a T en $(X, \mathcal{X}, \mu, T)$ es la transformación lineal $U_T : \mathcal{L}^2 \to \mathcal{L}^2$ definida por $U_T(f) = f \circ T$.

Nota 1.7. El operador esta bien definido. En efecto, si $[f] = [g]$, sea $A \in \mathcal{X}$ tal que $f(x) = g(x)$ para todo $x \in A$ y $\mu(A^c) = 0$. Entonces, dado que necesariamente $T^{-1}A \neq \emptyset$, $f(T(x)) = g(T(x))$ para todo $x \in T^{-1}A$ y $\mu\left(T^{-1}A\right)^c = 0$, siendo $U_T(f) = [f \circ T] = [g \circ T] = U_T(g)$. Por otro lado, $U_T(f) \in \mathcal{L}^2$ pues, dado que $|f|^2 \in \mathcal{M}$, $\int |f \circ T|^2\, d\mu = \int |f|^2 \circ T\, d\mu = \int |f|^2\, d\mu < +\infty$. De esta misma observación se obtiene que U_T es una isometría con la métrica inducida por la norma: $\|f \circ T\|_2 = \|f\|_2$ y por lo tanto U_T es un operador acotado.

Con el operador U_T se tiene una caracterización adicional de transformaciones ergódicas:

Proposición 1.4. En $(X, \mathcal{X}, \mu, T)$, T es ergódica si, y sólo si, $U_T(f) \overset{\mu}{=} f$ si y sólo si $f \overset{\mu}{=} c \in \mathbb{R}$. Esto quiere decir que las únicas funciones asociadas al valor propio 1 son las funciones constantes casi en toda parte.

Demostración:

Si T es ergódica, por la proposición 1.2., cada $f \in \mathcal{L}^2$ tal que $f \circ T \overset{\mu}{=} f$ satisface $f \overset{\mu}{=} c$, y si $f \overset{\mu}{=} c$ entonces $f \circ T \overset{\mu}{=} c$, de donde f es un vector propio asociado al valor propio 1. Recíprocamente, para $A \in \mathcal{X}$ tal que $T^{-1}A = A$ se tiene que $U_T(\mathcal{X}_A) = \mathcal{X}_A \circ T = \mathcal{X}_{T^{-1}A} = \mathcal{X}_A$ por lo que $\mathcal{X}_A \overset{\mu}{=} c$, luego $\mu(A) \in \{0, 1\}$ de donde T ergódica.

1.3. Preliminares en Esperanza Condicional

Se consideran en X dos σ-álgebras $\mathcal{Y}$ y $\mathcal{X}$ con $\mathcal{Y} \subset \mathcal{X}$ junto con una medida μ definida en $\mathcal{X}$, que al restringirla define una medida en $\mathcal{Y}$, $\mu|_{\mathcal{Y}}$. Dada $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$ con $0 \leq f$, se puede definir en

[7] Considerando la forma más general $f : X \to \mathbb{C}$. Cuando $f : X \to \mathbb{R}$ también se tiene el mismo resultado.

$\mathcal{Y}$ la medida μ_f por $\mu_f(A) = \int_A f\,d\mu$ para todo $A \in \mathcal{Y}$. La medida μ_f es absolutamente continua respecto a $\mu|_{\mathcal{Y}}$ en el sentido en que para $A \in \mathcal{Y}$, si $\mu|_{\mathcal{Y}}(A) = 0$ entonces $\mu_f(A) = \int_A f\,d\mu = 0$. El teorema de Radon-Nikodym[8] establece la existencia de una función $g \in \mathcal{M}^+(X, \mathcal{Y})$ tal que $\int_A f\,d\mu = \mu_f(A) = \int_A g\,d\,\mu|_{\mathcal{Y}}$ para todo $A \in \mathcal{Y}$. Esta función g está determinada casi en todas partes, pues si $h \overset{\mu}{=} g$ entonces $\int_A h\,d\,\mu|_{\mathcal{Y}} = \int_A g\,d\,\mu|_{\mathcal{Y}}$. Para hallar $g \in \mathcal{M}^+(X, \mathcal{Y})$ se requiere de f y de $\mathcal{Y}$, por lo que la función g se denota como $E(f|\mathcal{Y})$ y se denomina la esperanza condicional de f respecto a $\mathcal{Y}$. Cuando el contexto lo permita, $\mu|_{\mathcal{Y}}$ se denotará simplemente como μ. Obsérvese que:

- De $\mathcal{Y} \subset \mathcal{X}$, se sigue que $\mathcal{M}(X, \mathcal{Y}) \subset \mathcal{M}(X, \mathcal{X})$ y que $E(f|\mathcal{Y}) \in \mathcal{M}(X, \mathcal{X})$, sin que necesariamente f sea $\mathcal{Y}$-medible.
- Como $X \in \mathcal{Y}$ y $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$, entonces $\int E(f|\mathcal{Y})\,d\,\mu|_{\mathcal{Y}} = \int f\,d\mu < +\infty$, por lo que $E(f|\mathcal{Y}) \in \mathcal{L}^1\left(X, \mathcal{Y}, \mu|_{\mathcal{Y}}\right)$.
- Para $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$ arbitraria se define $E(f|\mathcal{Y}) = E(f^+|\mathcal{Y}) - E(f^-|\mathcal{Y})$, y entonces se puede considerar el operador lineal

$$E(\cdot|\mathcal{Y}) : \mathcal{L}^1(X, \mathcal{X}, \mu) \to \mathcal{L}^1(X, \mathcal{Y}, \mu).$$

Algunas propiedades de la esperanza condicional son:

- Si $g \in \mathcal{M}(X, \mathcal{Y})$ es finita ($g(X) \subset \mathbb{R}$) entonces $E(fg|\mathcal{Y}) = gE(f|\mathcal{Y})$ para toda $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$.
- Si $T : X \to X$ es $\mathcal{X}$-medible y preserva la medida μ entonces $E(f|\mathcal{Y}) \circ T = E\left(f \circ T | \Sigma\left(T^{-1}\mathcal{Y}\right)\right)$ para toda $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$.
- $E(f|\mathcal{X}) = f$.
- Si $\mathcal{N} := \{N \in \mathcal{X} : \mu(N) \in \{0, 1\}\}$[9] entonces $E(f|\mathcal{N}) = \int f\,d\mu$ para toda $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$.
- Si $0 \leq f$ entonces $0 \leq E(f|\mathcal{Y})$. En particular, si $A \in \mathcal{X}$, $0 \leq E(\mathcal{X}_A|\mathcal{Y})$.

Definición 1.17. Con la misma notación anterior, para $A \in \mathcal{X}$, la probabilidad condicional de A dada $\mathcal{Y}$ es $E(\mathcal{X}_A|\mathcal{Y})$.

[8] Ver [2], teorema 4.2.2.

[9] El cuál es una σ-álgebra de X.

Definición 1.18. Sea $(\mathcal{Y}_n)$ sucesión de σ-álgebras de X, y $\mathcal{X}$ σ-álgebra de X. Se dice que la sucesión $(\mathcal{Y}_n)$ converge a $\mathcal{X}$ si $\bigcup_{n\in\mathbb{N}} \mathcal{Y}_n$ genera $\mathcal{X}$, esto es $\Sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{Y}_n\right) = \mathcal{X}$.

En el caso en que se considera un número contable de σ-álgebras encajadas se tiene el siguiente resultado sobre los operadores de esperanza condicional correspondientes:

Proposición 1.5. Convergencia de Martingalas. En $(X, \mathcal{X}, \mu)$, si $(\mathcal{Y}_n)$ es una sucesión de σ-álgebras de X que satisface $\mathcal{Y}_n \subset \mathcal{Y}_{n+1}$ y converge a una σ-álgebra $\mathcal{Y}$ de subconjuntos de X, con $\mathcal{Y} \subset \mathcal{X}$, entonces para toda $f \in \mathcal{L}^1(X, \mathcal{X})$:

- $\lim_n E(f|\mathcal{Y}_n)(x) = E(f|\mathcal{Y})(x)$ casi en toda parte.
- $\lim_n \|E(f|\mathcal{Y}_n) - E(f|\mathcal{Y})\|_1 = 0$.

CAPÍTULO 2

Introducción a los problemas de clasificación en Teoría Ergódica

En el contexto de Teoría de la Tedida es de interés clasificar los espacios de medida. En el contexto de Teoría Ergódica se busca clasificar las transformaciones entre espacios de medida bajo ciertos criterios. En el primer caso, dos espacios de medida $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ pueden identificarse, de tal forma que algunas de las propiedades de uno de los espacios de medida también se verifican en el otro espacio de medida. Por ejemplo, si $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ están identificados y $\mathcal{X}$ es la completación de una σ-álgebra, se puede establecer que $\mathcal{Y}$ también es la completación de una σ-álgebra de conjuntos de Y. En el segundo caso, se busca identificar la acción de una transformación $T : X \to X$ (medible que preserva la medida) con la acción de otra transformación $S : Y \to Y$, cuando $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ son isomoformos.

Definición 2.1. Se dice que los espacios de medida $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ son isomorfos si existen:

1. $M \in \mathcal{X}$ tal que $\mu(M^c) = 0$,
2. $N \in \mathcal{Y}$ tal que $\lambda(N^c) = 0$,
3. $\phi : M \to N$ biyección medible ($\phi^{-1}\mathcal{Y} \subset \mathcal{X}$) que preserva la medida ($\lambda = \mu \circ \phi^{-1}$), con $\phi^!$ medible ($\phi^{!^{-1}}\mathcal{X} \subset \mathcal{Y}$) que preserva la medida ($\lambda \circ \phi^{!^{-1}} = \mu$).

Nota 2.1. Cuando $B \subset N^c$, $\phi^{-1}B = \emptyset \in \mathcal{X}$, y cuando $B \cap N \neq \emptyset$, $\phi^{-1}B = \phi^{-1}B \cap N$, por lo que $\phi^{-1}\mathcal{Y} \subset \mathcal{X}$ tiene sentido. Igualmente con $\phi^!$. Ahora, $\lambda = \mu \circ \phi^{-1}$ si y solamente si $\lambda \circ \phi^{!^{-1}} = \mu$, pues cuando $A \in \mathcal{X}$ con $A \subset M^c$, $\phi^{!^{-1}} A = \emptyset$ siendo $\lambda\left(\phi^{!^{-1}} A\right) = 0 = \mu(A)$. Cuando $A \cap M \neq \emptyset$, $\phi^{!^{-1}} A = \phi^{!^{-1}} A \cap M \in \mathcal{Y}$, luego

$$\lambda\left(\phi^{!^{-1}} A \cap M\right) = \mu\left(\phi^{-1}\left(\phi^{!^{-1}} A \cap M\right)\right) = \mu\left(\phi^{-1}\left(\phi A \cap M\right)\right) = \mu\left(A \cap M\right) = \mu(A),$$

pues en M se tiene que $\phi^{!^{-1}} = \phi$ y $\phi^{-1} \circ \phi = \phi^{!} \circ \phi = id$ vistas como funciones de subconjuntos.

Cuando $\left(X, \hat{\mathcal{X}}, \mu\right)$ y $\left(Y, \hat{\mathcal{Y}}, \lambda\right)$ son isomorfos las propiedades de $\left(X, \hat{\mathcal{X}}, \mu\right)$ se satisfacen también en $\left(Y, \hat{\mathcal{Y}}, \lambda\right)$ y viceversa. Por ejemplo:

Proposición 2.1. Sean $\left(X, \hat{\mathcal{X}}, \mu\right)$ y $\left(Y, \hat{\mathcal{Y}}, \lambda\right)$ espacios de medida que son isomorfos. Entonces:

- Si μ es σ-finita entonces λ es σ-finita.
- Si $\hat{\mathcal{X}}$ es la completación de $\mathcal{X}$ contablemente generada entonces $\hat{\mathcal{Y}}$ es la completación de una σ-álgebra $\mathcal{Y}$ de Y contablemente generada.

Demostración:

Sean $M \in \hat{\mathcal{X}}$, $N \in \hat{\mathcal{Y}}$ de medida completa y $\phi : M \to N$ biyección medible que preserva la medida con inversa medible que preserva la medida.

- Sea (A_n) una sucesión de conjuntos $\hat{\mathcal{X}}$-medibles tal que $\mu\left(A_n\right) < +\infty$ y $X = \bigcup_{i=1}^{+\infty} A_i$. Entonces $\left(\phi^{!^{-1}} A_n\right)$ es una sucesión de $\hat{\mathcal{Y}}$-medibles tal que $\lambda\left(\phi^{!^{-1}} A_n\right) = \mu\left(A_n\right) < +\infty$ y $Y = N^c \cup \bigcup_{i=1}^{+\infty} \phi^{!^{-1}} A_i$.
- Sea $C \subset \mathcal{P}(X)$, C contable y tal que $\Sigma(C) = \mathcal{X}$. Entonces $\phi^{!^{-1}} C$ es una sub-colección contable de $\mathcal{P}(Y)$ tal que $\phi^{!^{-1}} C \subset \phi^{!^{-1}} \hat{\mathcal{X}} \subset \hat{\mathcal{Y}}$. Obsérvese que $\phi^{!^{-1}} \mathcal{X}$ es una σ-álgebra de subconjuntos de N y no de X, pero que sí es una sub-colección de subconjuntos de X. Así podemos considerar $\mathcal{Y} = \Sigma\left(\phi^{!^{-1}} C \cup \{N\}\right)$, que es σ-álgebra de Y contablemente generada. Sea $A \in \hat{\mathcal{Y}}$. Entonces $\phi^{-1} A \in \hat{\mathcal{X}}$ y es de la forma $\phi^{-1} A = B \cup Z$ donde $B \in \mathcal{X}$ y $Z \subset Z^* \in \mathcal{X}$ con $\mu\left(Z^*\right) = 0$. Dado que $A = A \cap N \uplus A \cap N^c = \phi^{!^{-1}}\left(\phi^{-1}\left(A \cap N\right)\right) \uplus A \cap N^c = \phi^{!^{-1}}\left(\phi^{-1} A\right) \uplus A \cap N^c = \phi^{!^{-1}} B \cup \phi^{!^{-1}} Z \uplus A \cap N^c$, con:
  - $\phi^{!^{-1}} B \in \phi^{!^{-1}} \mathcal{X} = \phi^{!^{-1}} \Sigma(C) = \Sigma\left(\phi^{!^{-1}} C\right) \subset \Sigma\left(\phi^{!^{-1}} C \cup \{N\}\right) = \mathcal{Y}$.
  - $\phi^{!^{-1}} Z \uplus A \cap N^c \subset \phi^{!^{-1}} Z^* \uplus N^c$, donde $\phi^{!^{-1}} Z^* \in \phi^{!^{-1}} \mathcal{X} \subset \mathcal{Y}$, $N^c \in \mathcal{Y}$ y $\lambda\left(\phi^{!^{-1}} Z^* \uplus N^c\right) \leq \mu\left(Z^*\right) + \lambda\left(N^c\right) = 0$.

Se obtiene entonces que $\hat{\mathcal{Y}}$ es la completación de $\mathcal{Y}$ que es contablemente generada.

Se considera ahora una relación de equivalencia que permite identificar transformaciones en espacios de medida.

Definición 2.2. En $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$, se dice que T y S son isomorfas si existen:

1. $M \in \mathcal{X}$ tal que $\mu(M^c) = 0$ y $T(M) \subset M$,

2. $N \in \mathcal{Y}$ tal que $\lambda(N^c) = 0$ y $S(N) \subset N$,

3. $\phi : M \to N$ biyección medible que preserva la medida, con inversa medible que preserva la medida,

4. $\phi \circ T|_M = S|_N \circ \phi$ [1].

El isomorfismo de transformaciones que preservan la medida define una relación de equivalencia mediante $T \; \mathcal{R} \; S$ si y sólo si T y S son isomorfas. Los isomorfismo de transformaciones que preservan la medida requieren de un isomorfismo entre sus correspondientes espacios de medida. El problema de clasificación consiste en establecer cuándo T y S son isomorfas. Se considera ahora una relación de equivalencia más débil, motivada por la caracterización de transformaciones $T : X \to X$ que preservan la medida, mediante el Operdor de Koopman.

Definición 2.3. En $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$, se dice que T y S son espectralmente isomorfas si existe un operador lineal $\mathcal{W} : \mathcal{L}^2(X, \mathcal{X}, \mu) \to \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ tal que:

1. $\mathcal{W}$ es biyectivo,

2. $\langle \mathcal{W}(f), \mathcal{W}(g) \rangle = \langle f, g \rangle$ para todos $f, g \in \mathcal{L}^2(X, \mathcal{X}, \mu)$,

3. $U_S \circ \mathcal{W} = W \circ U_T$.

Un isomorfismo espectral de transformaciones que preservan la medida requiere entonces de un isomorfismo de espacios de Hilbert que sea a su vez una conjugación entre U_T y U_S. Los isomorfismos espectrales definen una relación de equivalencia en el conjunto de transformaciones que preservan la medida definidas en espacios de medida. Además:

[1]En Sistemas Dinámicos ([12], capítulo 1) se dice que ϕ es una conjugación entre T y S. Según el contexto se imponen condiciones a esta conjugación, como en éste caso que preserva la estructura de medida.

Proposición 2.2. Si en $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$ se tiene que T y S son isomorfas, entonces T y S son espectralmente isomorfas.

Demostración:

Sean $M \in \mathcal{X}$, $N \in \mathcal{Y}$ y $\phi : M \to N$ mediante los cuales T y S son isomorfas. Dada $f \in \mathcal{L}^2(X, \mathcal{X}, \mu)$ se obtiene mediante $\phi^!$ la función $f \circ \phi^!$ definida en Y salvo en un conjunto de medida nula, que puede extenderse a todo Y como $f \circ \phi^!\big|_{N^c} \equiv c \in \mathbb{R}$. Ahora, $\left(f \circ \phi^!\right)^{-1} = \phi^{!-1} \circ f^{-1}$ por lo que $f \circ \phi^! \in \mathcal{M}(Y, \mathcal{Y})$.

- Si $[f] = [g]$, sea $A \in \mathcal{X}$ tal que $f(x) = g(x)$ para todo $x \in A$ y $\mu(A^c) = 0$. Nótese que necesariamente $A \cap M \neq \emptyset$, y entonces $\phi^{!-1} A \cap M \in \mathcal{Y}$ con $\lambda\left(\phi^{!-1}\ (A \cap M)^c\right) = \mu((A \cap M)^c) = 0$ y es tal que $f \circ \phi^! = g \circ \phi^!$ en $\phi^{!-1} A \cap M$, esto es $\left[f \circ \phi^!\right] = \left[g \circ \phi^!\right]$.
- Para ver que $\int \left|f \circ \phi^!\right|^2 d\lambda = \int |f|^2 \circ \phi^!\, d\lambda < +\infty$ tenemos que verificar que para toda $h \in \mathcal{M}(X, \mathcal{X})$ se satisface $\int h\, d\mu = \int h \circ \phi^!\, d\lambda$:
  - Si $h = \mathcal{X}_B$ con $B \in \mathcal{X}$, $\int h \circ \phi^!\, d\lambda = \int_N \mathcal{X}_{\phi^{!-1}B}\, d\lambda = \lambda\left(N \cap \phi^{!-1}B\right) = \mu(M \cap B) = \int \mathcal{X}_B\, d\mu = \int h\, d\mu$.
  - Si $h = \sum\limits_{i=1}^{n} a_i\, \mathcal{X}_{A_i}$ con $A_i \in \mathcal{X}$,
  $$\int h \circ \phi^!\, d\lambda = \sum_{i=1}^{n} a_i \int \mathcal{X}_{A_i}\, d\mu = \int \sum_{i=1}^{n} a_i\, \mathcal{X}_{A_i}\, d\mu = \int h\, d\mu.$$
  - Para h^+ y h^- sean $\left(\varphi_j\right)$ y $\left(\beta_j\right)$ sucesiones de funciones simples en $\mathcal{M}(X, \mathcal{X})$ tales que $\int h^+\, d\mu = \lim\limits_{j} \int \varphi_j\, d\mu$ y $\int h^-\, d\mu = \lim\limits_{j} \int \beta_j\, d\mu$ (Teorema de Convergencia Monótona y aproximación de medibles mediante funciones simples). Entonces, para todo j se cumple que $\int \varphi_j \circ \phi^!\, d\lambda = \int \varphi_j\, d\mu$ y $\int \beta_j \circ \phi^!\, d\lambda = \int \beta_j\, d\mu$, entonces por Teorema de Convergencia Monótona: $\int h^+ \circ \phi^!\, d\lambda = \int h^+\, d\mu$ y $\int h^- \circ \phi^!\, d\lambda = \int h^-\, d\mu$, así $\int h \circ \phi^!\, d\lambda = \int h\, d\mu$.

  Como $|f|^2 \in \mathcal{M}(X, \mathcal{X})$ entonces $\int |f|^2 \circ \phi^!\, d\lambda = \int |f|^2\, d\mu < +\infty$, por lo que $f \circ \phi^! \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$.

De igual manera, dada $\hat{f} \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ se obtiene mediante ϕ la función $\hat{f} \circ \phi \in \mathcal{M}(X, \mathcal{X})$ que define la clase $\left[\hat{f} \circ \phi\right] \in \mathcal{L}^2(X, \mathcal{X}, \mu)$. Se define entonces $\mathcal{W} : \mathcal{L}^2(X, \mathcal{X}, \mu) \to \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ por $\mathcal{W}(f) = f \circ \phi^!$, que satisface:

- $\mathcal{W}$ es inyectiva: Si $\left[f \circ \phi^!\right] = \left[g \circ \phi^!\right]$, sea $B \in \mathcal{Y}$ tal que $f \circ \phi^! = g \circ \phi^!$ en B y $\lambda(B^c) = 0$. Nótese que $B \cap N \neq \emptyset$ y de $\phi^{-1}(B \cap N) \in \mathcal{X}$, con $\mu\left(\phi^{-1}(B \cap N)^c\right) = \lambda((B \cap N)^c) = 0$, se obtiene que si $x \in \phi^{-1}(B \cap N)$ entonces $f(x) = f \circ \phi^!(\phi(x)) = g \circ \phi^!(\phi(x)) = g(x)$. Así $[f] = [g]$.

- $\mathcal{W}$ es sobreyectiva: Si $\hat{f} \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$, $\hat{f} \circ \phi^! \in \mathcal{L}^2(X, \mathcal{X}, \mu)$ y es tal que $\left[\left(\hat{f} \circ \phi\right) \circ \phi^!\right] = \left[\hat{f}\right]$.

- $\langle \mathcal{W}(f), \mathcal{W}(g) \rangle = \int \left(f \circ \phi^!\right) \overline{g \circ \phi^!} \, d\lambda = \int (f \overline{g}) \circ \phi^! \, d\lambda = \int f \overline{g} \, d\mu = \langle f, g \rangle$, pues $f \overline{g} \in \mathcal{M}(X, \mathcal{X})$.

- Si $[f] \in \mathcal{L}^2(X, \mathcal{X}, \mu)$, $(U_S \circ \mathcal{W})[f] = \left[\left(f \circ \phi^!\right) \circ S\right] = \left[f \circ \left(\phi^! \circ S\right)\right] = \left[f \circ \left(T \circ \phi^!\right)\right] = \left[(f \circ T) \circ \phi^!\right] = (\mathcal{W} \circ U_T)[f]$.

La contrarecíproca de la proposición anterior, si T y S no son espectralmente isomorfas entonces T y S no son isomorfas, permite descartar algunos casos de clasificación. Sin embargo, determinar la existencia de isomorfismos y de isomorfismos espectrales presenta gran dificultad. Se buscan entonces objetos o propiedades, denominados *invariantes*, que comparten las transformaciones según el tipo de isomorfismo que se esté considerando.

Definición 2.4. Un invariante de isomorfismos es un objeto o una propiedad tal que, si T y S son isomorfas entonces T y S tienen el mismo objeto o la misma propiedad. Un invariante completo de isomorfismos es un invariante de isomorfismos que además, si T y S tienen el mismo objeto o la misma propiedad entonces T y S son isomorfas. De igual manera se define un invariante de isomorfismo espectral y un invariante completo de isomorfismo espectral.

Nótese que todo invariante de isomorfismos espectrales es un invariante de isomorfismos. En los capítulos 3 al 5 se presentan la entropía y el conjunto de valores propios de una transformación como dos invariantes de interés, y se verifican los casos en que estos invariantes son completos. Según el caso, es conveniente restringir los espacios de medida en los que se trabaja, imponiendo condiciones y estructuras sobre X o sobre $\mathcal{X}$. Una restricción importante, que se mantiene a lo largo de este documento, es la siguiente:

Definición 2.5. Un espacio polaco[2] es un espacio métrico (X, d) completo (toda sucesión de Cauchy converge) y separable(existe un subconjunto denso y contable). Un espacio de probabilidad estándar es un espacio de probabilidad $(X, \mathcal{B}, \mu)$ donde X es polaco, $\mathcal{B}$ es la σ-álgebra de

[2]*A short history of Polish mathematics* por W. Zelazko es una breve nota sobre el término "polaco", disponible en la web: http://www.impan.pl/Sixty/polmat.pdf

Borel generada por los abiertos de la topología con la cual X es polaco y μ es una medida de probabilidad sobre $\mathcal{B}$ (denominada también medida de Borel).

En adelante se consideran solamente espacios de probabilidad estándar. Al obtener un espacio de medida mediante la σ-álgebra de Borel a partir de un espacio métrico compacto se tienen propiedades de aproximación que permiten extender resultados directamente al espacio de medida a partir de su verificación en clases de subconjuntos más pequeñas que la σ-álgebra. Los siguientes resultado se han extraído de [6].

Proposición 2.3. En todo espacio polaco (X, d) se satisfacen:

- 2-enumerabilidad. Existe una colección contable de abiertos $\mathcal{U}$ tal que cada abierto V de X es unión de una sub-colección de $\mathcal{U}$.
- Lindelöf. Todo cubrimiento por abiertos del espacio tiene un sub-cubrimiento contable.
- La intersección de una sucesión de bolas cerradas encajadas cuyo radio tiende a cero es un solo punto.

Demostración:

- Dado que X es separable, existe un subconjunto $\{x_n\}$ de X denso y numerable. Sea $\mathcal{U} = \{B_r(x_n) : n \in \mathbb{N}, r \in \mathbb{Q}\}$. $\mathcal{U}$ es contable y para todo conjunto abierto A se tiene que $A = \bigcup\limits_{i \in \mathbb{N}} B_{r_i}(x_{n_i})$, por la densidad de $\{x_n\}$ y la densidad de $\mathbb{Q}$ en $\mathbb{R}$, con $\{B_{r_i}(x_{n_i}) : i \in \mathbb{N}\} \subset \mathcal{U}$.
- Sea $\{U_\alpha\}$ con $\alpha \in I$ un cubrimiento por abiertos de X y $\{B_n\}$ la colección $\mathcal{U}$ de la parte anterior. Dado que, para todo $x \in X$ existen U_{α_x} y B_{n_x} tales que $x \in B_{n_x} \subset U_{\alpha_x}$, sean $\tilde{B} = \{B_n : B_n \subset U_{\alpha_n} \text{ para algún } U_{\alpha_n}\}$ y $V = \left\{U_{\alpha_n} : B_n \subset U_{\alpha_n} \text{ para algún } B_n \in \tilde{B}\right\}$. Entonces la sub-colección V cubre a X y es a lo más contable.
- Si $B_n = \overline{B_{r_n}(x_n)}$ es una sucesión de bolas cerradas tal que $B_{n+1} \subset B_n$ y $r_n \to 0$, entonces para cada $n \in \mathbb{N}$ fijo y $n < k$ se tiene que $x_k \in B_n$, por lo que (x_n) es una sucesión de Cauchy que convergente entonces a un punto x. Para cada $n \in \mathbb{N}$ la sub-sucesión $(x_k)_{n<k}$ de B_n también converge a x. Como B_n es un conjunto cerrado, $x \in B_n$ para cada $n \in \mathbb{N}$, por lo tanto $x \in \bigcap\limits_n B_n$. Ahora, para $\epsilon > 0$ se tiene que $\operatorname{diam}\left(\bigcap\limits_{n \in \mathbb{N}} B_n\right) \leq \operatorname{diam}(B_N) < \epsilon$ para algún $N \in \mathbb{N}$. Así, $\operatorname{diam}\left(\bigcap\limits_{n \in \mathbb{N}} B_n\right) = 0$, de donde $\bigcap\limits_{n \in \mathbb{N}} B_n = \{x\}$.

Con esto se pueden demostrar:

Proposición 2.4. En un espacio de probabilidad estándar se satisfacen:

- Regularidad. Para $E \in \mathcal{B}$ y $\epsilon > 0$ existe un conjunto abierto U y un conjunto cerrado F que satisfacen $F \subset E \subset U$ y $\mu(U \setminus F) < \epsilon$.
- Separabilidad. Existe una colección contable $\{E_n\} \subset \mathcal{B}$ tal que para cada $E \in \mathcal{B}$ y $\epsilon > 0$ $\mu(E \triangle E_n) < \epsilon$, para algún $n \in \mathbb{N}$.

Demostración:

- Las bolas abiertas $B_r(x)$ son medibles y satisfacen la propiedad de regularidad con $U = B_r(x)$ y $F = \overline{B_{r-1/n}(x)}$ para n tal que $\mu(U \setminus F) < \epsilon$, ya que $\lim_n \mu\left(\overline{B_{r-1/n}(x)}\right) = \mu(B_r(x))$. Los abiertos arbitrarios A también satisfacen la propiedad. En efecto, tomemos $U = A$, y escribamos $A = \bigcup_{i\in\mathbb{N}} B_i$ donde $\{B_i : i \in \mathbb{N}\}$ es una sub-colección de bolas abiertas de la colección $\mathcal{U} = \{B_r(x_n) : n \in \mathbb{N}, r \in \mathbb{Q}\}$. Para cada $i \in \mathbb{N}$ se obtiene F_i cerrado tal que $\mu\left(B_i \cap F_i^c\right) < \epsilon/2^{i+1}$. Entonces

$$\begin{aligned} \mu\left(A \cap \left(\bigcup_{i\in\mathbb{N}} F_i\right)^c\right) &= \mu\left(\bigcup_{i\in\mathbb{N}} B_i \cap \bigcap_{i\in\mathbb{N}} F_i^c\right) \\ &\leq \mu\left(\bigcup_{i\in\mathbb{N}} B_i \cap F_i^c\right) \\ &\leq \sum_{i\in\mathbb{N}} \mu\left(B_i \cap F_i^c\right) \\ &\leq \sum_{i\in\mathbb{N}} \epsilon/2^{i+1} \\ &= \epsilon/2, \end{aligned}$$

  y dado que

$$\mu\left(A \cap \left(\bigcup_{i\in\mathbb{N}} F_i\right)^c\right) = \mu\left(\bigcap_{i\in\mathbb{N}} A \cap F_i^c\right) = \lim_i \mu\left(A \cap F_i^c\right),$$

  basta tomar $N \in \mathbb{N}$ tal que

$$\mu\left(A \cap F_N^c\right) - \mu\left(\bigcap_{i\in\mathbb{N}} A \cap F_i^c\right) < \epsilon/2,$$

con lo que $\mu\left(A \cap F_N^c\right) < \epsilon$ y así, de $A \cap \left(\bigcap_{i=1}^{N} F_i^c\right) \subset A \cap F_N^c$ se obtiene que $\mu\left(A \setminus \bigcup_{i=1}^{N} F_i\right) < \epsilon$, con $\bigcup_{i=1}^{N} F_i \subset A$ cerrado.

Ahora, la colección $\tilde{\mathcal{B}} = \{E \in \mathcal{B} : E$ satisface la propiedad de regularidad $\}$ contiene a los abiertos y es una σ-álgebra. En efecto, si $E \in \tilde{\mathcal{B}}$ y $\epsilon > 0$, se tienen un conjunto cerrado F y un conjunto abierto U tales que $F \subset E \subset U$ y $\mu(U \cap F^c) < \epsilon$. Entonces $U^c \subset E^c \subset F^c$ y $\mu((U^c)^c \cap F^c) < \epsilon$, por lo que $E^c \in \tilde{\mathcal{B}}$. Si (E_n) es una sucesión de elementos de $\tilde{\mathcal{B}}$ y $\epsilon > 0$, entonces para todo $n \in \mathbb{N}$ existen conjuntos cerrados F_n y conjuntos abiertos U_n tales que $F_n \subset E_n \subset U_n$, con $\mu(U_n \cap F_n^c) < \epsilon/2^{n+1}$. Sea $A_k = \bigcup_{n=1}^{k} F_n$ para cada $k \in \mathbb{N}$. La sucesión (A_k) es creciente, $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{n\in\mathbb{N}} F_n$ y $\bigcup_{n\in\mathbb{N}} F_n \subset \bigcup_{n\in\mathbb{N}} E_n \subset \bigcup_{n\in\mathbb{N}} U_n$. Además

$$
\begin{aligned}
\mu\left(\left(\bigcup_{n\in\mathbb{N}} U_n\right) \cap \left(\bigcup_{n\in\mathbb{N}} F_n\right)^c\right) = & \\
= & \quad \mu\left(\left(\bigcup_{n\in\mathbb{N}} U_n\right) \cap \left(\bigcap_{n\in\mathbb{N}} F_n^c\right)\right) \\
\leq & \quad \mu\left(\bigcup_{n\in\mathbb{N}} U_n \cap F_n^c\right) \\
\leq & \quad \sum_{n\in\mathbb{N}} \mu(U_n \cap F_n^c) \\
< & \quad \sum_{n\in\mathbb{N}} \epsilon/2^{n+1} \\
= & \quad \epsilon/2.
\end{aligned}
$$

Dado que $A_{k+1}^c \subset A_k^c$, $\bigcup_{n\in\mathbb{N}} U_n \cap A_{k+1}^c \subset \bigcup_{n\in\mathbb{N}} U_n \cap A_k^c$ y

$$
\bigcap_{k\in\mathbb{N}}\left(\bigcup_{n\in\mathbb{N}} U_n \cap A_k^c\right) = \bigcap_{k\in\mathbb{N}}\left(\bigcup_{n\in\mathbb{N}} U_n \cap \left(\bigcap_{n=1}^{k} F_n^c\right)\right),
$$

entonces

$$
\mu\left(\bigcap_{k\in\mathbb{N}}\left(\bigcup_{n\in\mathbb{N}} U_n \cap \left(\bigcap_{n=1}^{k} F_n^c\right)\right)\right) = \lim_{k} \mu\left(\bigcup_{n\in\mathbb{N}} U_n \cap \left(\bigcap_{n=1}^{k} F_n^c\right)\right),
$$

por lo que para algún $N \in \mathbb{N}$ se tiene

$$
\mu\left(\bigcup_{n\in\mathbb{N}} U_n \cap \left(\bigcap_{n=1}^{N} F_n^c\right)\right) - \mu\left(\bigcap_{k\in\mathbb{N}}\left(\bigcup_{n\in\mathbb{N}} U_n \cap \left(\bigcap_{n=1}^{k} F_n^c\right)\right)\right) < \epsilon/2.
$$

Así $\mu\left(\left(\bigcup_{n\in\mathbb{N}} U_n\right)\cap\left(\bigcup_{n=1}^{N} F_n\right)^c\right) < \epsilon$, con $\bigcup_{n=1}^{N} F_n$ cerrado y $\bigcup_{n=1}^{N} F_n \subset \bigcup_{n\in\mathbb{N}} E_n \subset \bigcup_{n\in\mathbb{N}} U_n$ abierto, por lo que $\bigcup_{n\in\mathbb{N}} E_n \in \tilde{\mathcal{B}}$.

Entonces por definición de $\mathcal{B}$ se obtiene que $\tilde{\mathcal{B}} = \mathcal{B}$. Así, todo conjunto $\mathcal{B}$-medible satisface la propiedad de regularidad.

- Consideremos la colección contable de conjuntos medibles

$$\mathcal{V} = \left\{\bigcup_{i=1}^{M} B_{n_i} \,:\, B_{n_i} \in \mathcal{U} \text{ y } M \in \mathbb{N}\right\}.$$

Dado un conjunto abierto A existen conjuntos $B_i \in \mathcal{U}$ tales que $A = \bigcup_{i\in\mathbb{N}} B_i$ y $\mu(A) = \lim_i \mu\left(\bigcup_{j=1}^{i} B_j\right)$. Dado $\epsilon > 0$, existe $N \in \mathbb{N}$ tal que $\mu(A) - \mu\left(\bigcup_{j=1}^{N} B_j\right) < \epsilon$, con $\bigcup_{j=1}^{N} B_j \in \mathcal{V}$. Para $E \in \mathcal{B}$ y $\epsilon > 0$ existe un conjunto cerrado F y un conjunto abierto U tales que $F \subset E \subset U$ y $\mu(U \setminus F) < \epsilon$. Al usar el resultado anterior sobre U se obtiene la propiedad de separabilidad sobre E.

Proposición 2.5. Sea $(X, \mathcal{B}, \mu)$ un espacio de probabilidad estándar. Si $E \in \mathcal{B}$ y $\mu(E) > 0$ entonces existe $x \in X$ tal que para todo $r > 0$ se satisface $\mu(E \cap B_r(x)) > 0$.

Demostración:

Tomemos $E \in \mathcal{B}$ con $\mu(E) > 0$. Sea (r_n) una sucesión de números reales que converge a 0. Dado que $X = \bigcup_{x\in X} B_{r_1}(x)$, por la propiedad de Lindelöf de la proposición 2.3., $X = \bigcup_{i\in\mathbb{N}} B_{r_1}(x_i)$. Si para todo $i \in \mathbb{N}$ se tuviera $\mu(E \cap B_{r_1}(x_i)) = 0$ entonces $\mu(E) = \mu\left(E \cap \left(\bigcup_{i\in\mathbb{N}} B_{r_1}(x_i)\right)\right) \leq \sum_{i\in\mathbb{N}} \mu(E \cap B_{r_1}(x_i)) = 0$, lo que contradice la escogencia de E. Así, necesariamente alguna de las bolas abiertas $B_{r_1}(x_{i_1})$ satisface que $\mu(E \cap B_{r_1}(x_{i_1})) \neq 0$. Pongamos $B_1 = B_{r_1}(x_{i_1})$ y escribámosla como unión contable de bolas abiertas de radio r_2, una de las cuales, B_2 satisface $\mu(E \cap B_1 \cap B_2) \neq 0$. Continuando de esta manera se obtiene una sucesión (B_n) de bolas cuyos diámetros convergen a 0, encajadas $B_{n+1} \subset B_n$, y tales que $0 < \mu\left(E \cap \bigcap_{i=1}^{n} B_i\right) \leq \mu(E \cap B_n)$. La sucesión (x_{i_n}) de los centros de las bolas B_n es de Cauchy, pues son encajadas y su radio r_n converge a 0, y definen para cada $\overline{B_n}$ una sub-sucesión convergente al mismo límite $x \in X$. Por lo tanto, $x \in \overline{B_n}$ para todo $n \in \mathbb{N}$ y $d(x, x_{i_n}) \leq r_n$. Dado $r > 0$, sea $N \in \mathbb{N}$ tal que $r_N < r/2$. Entonces $B_N = B_{r_N}(x_{i_N}) \subset B_r(x)$, siendo $0 < \mu(E \cap B_N) \leq \mu(E \cap B_r(x))$.

Obsérvese que todo espacio métrico compacto es un espacio polaco con la σ-álgebra generada por la topología de la métrica. Para establecer una propiedad importante de los espacios de probabilidad estándar se introduce el siguiente concepto:

Definición 2.6. $A \in \mathcal{X}$ es un átomo de $(X, \mathcal{X}, \mu)$ si $\mu(A) > 0$ y para todo $B \subset A$ tal que $B \in \mathcal{X}$ se tiene que $\mu(B) \in \{\mu(A), 0\}$. Un espacio de medida sin átomos se denomina no-atómico, y es tal que, para todo $A \in \mathcal{X}$ con $\mu(A) > 0$ existe $B \subset A$, $B \in \mathcal{X}$, tal que $0 < \mu(B) < \mu(A)$.

Nota 2.2. Obsérvese que dados dos espacios de medida $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$ que son isomorfos mediante $M \in \mathcal{X}$, $N \in \mathcal{Y}$ y $\phi : M \to N$, si $A \in \mathcal{Y}$ es un átomo de $(Y, \mathcal{Y}, \lambda)$ necesariamente $A \cap N \neq \emptyset$ y $\lambda(A \cap N) = \lambda(A)$. Entonces se tiene que $\phi^{-1}A \cap N$ es un átomo de $(X, \mathcal{X}, \mu)$. Así, la propiedad de un espacio de medida de ser no-atómico se preserva mediante isomorfismos de espacios de medida.

Una caracterización de los átomos en un espacio de probabilidad estándar permite entender la proposición que se enuncia a continuación

Proposición 2.6. En un espacio de probabilidad estándar $(X, \mathcal{B}, \mu)$ todo átomo es de la forma $\{x\} \cup Z$, donde $\mu(Z) = 0$ y $\mu(\{x\}) > 0$.

Demostración:

Consideremos un átomo $A \in \mathcal{B}$. Dado que $X = \bigcup_{x \in X} B_{r_1}(x)$ para $r_1 = 1$ y, por la propiedad de Lindelöf de la proposición 2.3., resulta que $X = \bigcup_{i \in \mathbb{N}} B_{r_1}(x_i)$, $A = \bigcup_{i \in \mathbb{N}} A_i$, donde $A_i = B_{r_1}(x_i) \cap A \in \mathcal{B}$ y $\operatorname{diam}(A_i) \leq r_1$. Necesariamente $\mu(A_{i_1}) > 0$ para algún conjunto medible de la colección $\{A_i\}$. Entonces $\mu(A_{i_1}) = \mu(A)$ y A_{i_1} es también un átomo. Continuando de la misma forma sobre A_{i_1}, con $r_2 = 1/2$ y así sucesivamente, se obtiene la sucesión (A_{i_m}) de átomos encajados $A_{i_{m+1}} \subset A_{i_m}$ que tienen la misma medida y con diámetro convergente a 0. $\bigcap_{m \in \mathbb{N}} A_{i_m} \neq \emptyset$ pues $\lim_m \mu(A_{i_m}) = \mu(A) \neq 0$ y entonces $\bigcap_{m \in \mathbb{N}} A_{i_m} = \{x\} \in \mathcal{B}$, con $x \in A$ y $\mu(\{x\}) = \mu(A)$. El conjunto Z es $A \cap \{x\}^c$.

Proposición 2.7. En un espacio de probabilidad estándar no atómico $(X, \mathcal{X}, \mu)$, si $r > 0$ y $E \in \mathcal{B}$ es de medida positiva, entonces $E = \biguplus_{i \in \mathbb{N}} F_i \biguplus N$ donde $\mu(N) = 0$, y F_i es un conjunto cerrado con $\operatorname{diam}(F_i) < r$ y $\mu(F_i) > 0$ para cada $i \in \mathbb{N}$.

Demostración:

Sean $r > 0$ y $E \in \mathcal{B}$ de medida positiva. Como $X = \bigcup_{x \in X} B_{r/2}(x)$, se tiene que $E = \bigcup_{i \in \mathbb{N}} E \cap B_{r/2}(x_i)$, donde $\{x_i\}$ es un sunconjunto denso contable de X. Consideremos la sucesión disjunta (E_i) definida por $E_i = E \cap B_{r/2}(x_i) \setminus \left(\bigcup_{k=1}^{i-1} E \cap B_{r/2}(x_i)\right)$. Obsérvese que $\operatorname{diam}(E_i) < r$ y que $\biguplus_{i \in \mathbb{N}} E_i = \bigcup_{i \in \mathbb{N}} E \cap B_{r/2}(x_i)$. Sin pérdida de generalidad, se considera $E \in \mathcal{B}$ con $\operatorname{diam}(E) < r$. Por la propiedad de regularidad, sea F_1 cerrado con $F_1 \subset E$ y $\mu(E \setminus F_1) < 1/2$, pues si U_1 es el abierto tal que $E \subset U_1$ y $\mu(U_1 \setminus F_1) < 1/2$ entonces $E \setminus F_1 \subset U_1 \setminus F_1$. Igualmente, sea $F_2 \subset E \setminus F_1$ tal que $\mu((E \setminus F_1) \setminus F_2) = \mu(E \setminus (F_1 \cup F_2)) < 1/2^2$. De esta manera se obtiene la sucesión (F_i) de subconjuntos de E disjuntos y cerrados tal que $\mu\left(E \setminus \biguplus_{i=1}^{n} F_i\right) < 1/2^n$ para todo $n \in \mathbb{N}$. Si para ningún $n \in \mathbb{N}$ se tiene $\mu\left(E \setminus \biguplus_{i=1}^{n} F_i\right) = 0$, se toma $N = E \setminus \biguplus_{i \in \mathbb{N}} F_i$ que es tal que $\mu(N) < 1/2^n$ para todo $n \in \mathbb{N}$, por lo que $\mu(N) = 0$. Por construcción, $E = \biguplus_{i \in \mathbb{N}} F_i \uplus N$. Si en cambio para algún $n \in \mathbb{N}$ se tiene que $\mu\left(E \setminus \biguplus_{i=1}^{n} F_i\right) = 0$, sea F_n el primer conjunto cerrado que satisface esta propiedad. Por la proposición 2.5., sea $x \in X$ tal que $0 < \mu(F_n \cap B_r(x))$ para todo $r > 0$. Dado que para todo $j \in \mathbb{N}$ se tiene $B_{1/j}(x) \in \mathcal{B}$ entonces $\{x\} \in \mathcal{B}$ y como el espacio no es atómico, necesariamente $\mu(\{x\}) = 0$. Sea (r_m) una sucesión estrictamente decreciente de números reales positivos que converge a 0. Entonces, $\lim_m \mu(F_n \cap B_{r_m}(x)) \le \lim_m \mu(B_{r_n}(x)) = 0$. Se definen la sucesión (C_k) de subconjuntos de F_n cerrados disjuntos con medida positiva por $C_k = F \cap \left(\overline{B_{r_k}(x)} \setminus B_{r_{k+1}}(x)\right)$ que son también disjuntos de los $F_1, \ldots, F_{n-1}$. Tomando $\tilde{E} = E \setminus \left(\bigcup_{i=1}^{n-1} F_i \cup \bigcup_{k \in \mathbb{N}} C_k\right)$ y descomponiendolo de la misma forma se obtienen finitos o contables subconjuntos cerrados disjuntos y un conjunto N de medida nula. Con esta colección y $\{F_1 \ldots, F_{n-1}\} \cup \{C_k\}$ se obtiene la descomposición de E.

Los espacios de probabilidad estándar cuentan con la propiedad que enunciamos en el siguiente teorema:

Teorema 2.1. Todo espacio de probabilidad estándar no-atómico es isomorfo a $\left([0,1], \mathscr{B}_{[0,1]}, \lambda\right)$, donde λ es la medida de Lebesgue. Cuando el espacio es atómico, es isomorfo al mismo espacio de medida junto con una colección finita o contable de átomos.

Demostración:

Sea $(X, \mathcal{X}, \mu)$ un espacio de probabilidad estándar no-atómico. Tómese una sucesión decreciente de números reales no negativos (r_n) que converge a 0. Por la proposición 2.7., se descompone a X como $X = \biguplus_{j_1 \in \mathbb{N}} F(j_1) \uplus N(1)$, donde la sucesión $(F(j_1))$ es de conjuntos cerrados disjuntos de medida positiva, $\operatorname{diam}(F(j_1)) < r_1$, y $\mu(N_1) = 0$. Cada $F(j_1)$ puede descomponerse de la misma manera como $F(j_1) = \biguplus_{j_2 \in \mathbb{N}} F(j_1, j_2) \uplus N(1, j_1)$. Nótese que, dado que los $F(j_1)$ son disjuntos se obtiene que $F(j_1, i_2) \cap F(j_1, k_2) = \emptyset$ para $i_2 \neq k_2$. Continuando de esta manera, se obtienen las colecciones contables $\{F(j_1,\ldots,j_n) : n \in \mathbb{N}\}$ de conjuntos cerrados disjuntos, y $\{N(1, j_1, \ldots, j_n) : n \in \mathbb{N}\}$ tales que:

- Para todo $n \in \mathbb{N}$ la sucesión $(F(j_1,\ldots,j_n))$ es disjunta y tal que

$$\operatorname{diam}(F(j_1,\ldots,j_n)) < r_n$$

  para todo $j_n \in \mathbb{N}$.

- $F(j_1,\ldots,j_{n-1}) = \biguplus_{j_n\in\mathbb{N}} .F(j_1,\ldots,j_{n-1},j_n) \uplus N(1, j_1,\ldots,j_{n-1})$ para todo $j_n \in \mathbb{N}$ con $1 < n$.
- $\mu(N(1, j_1,\ldots,j_n)) = 0$ para todo $n \in \mathbb{N}$.

Sea $\widehat{X} = \bigcap_{n\in\mathbb{N}} \biguplus_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)$. Por construcción,

$$\biguplus_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n) \subset \biguplus_{j_1,\ldots,j_{n-1}\in\mathbb{N}} F(j_1,\ldots,j_{n-1}),$$

por lo que la sucesión definida para cada $n \in \mathbb{N}$ por

$$n \mapsto \bigcap_{j_1,\ldots,j_n\in\mathbb{N}} \biguplus_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)^c$$

es creciente y entonces

$$\mu\left(\bigcup_{n\in\mathbb{N}} \bigcap_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)^c\right) = \lim_n \mu\left(\bigcap_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)^c\right).$$

Ahora, para todo $n \in \mathbb{N}$ se tiene que

$$\mu(X) = \mu\left(\bigcap_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)^c\right),$$

por lo que

$$\mu\left(\bigcup_{n\in\mathbb{N}} \bigcap_{j_1,\ldots,j_n\in\mathbb{N}} F(j_1,\ldots,j_n)^c\right) = 0.$$

Entonces $\mu\left(X \setminus \widehat{X}\right) = 0$. Si $x \in \widehat{X}$, entonces para todo $n \in \mathbb{N}$, existen $i_1,\ldots,i_n \in \mathbb{N}$, tales que $x \in F(i_1,\ldots,i_n)$, y necesariamente $x \in F(i_1,\ldots,i_n,j_{n+1})$ para algún $j_{n+1} \in \mathbb{N}$. De esto, la sucesión $(F(j_1,\ldots,j_n))$ definida por cada $x \in \widehat{X}$ es tal que $\bigcap_{n\in\mathbb{N}} F(j_1,\ldots,j_n) = \{x\}$, pues son conjuntos compactos de intersección finita no vacía (medida no nula) y de diámetro convergente a 0 en un espacio métrico. Se asocia a cada $x \in \widehat{X}$ la sucesión $j_1, j_2, \ldots$ que se obtiene mediante los $F(j_1,\ldots,j_n)$ a los cuales x pertenece.

Se define $\phi : \widehat{X} \to [0,1]$ por:

$$\phi(x) = \cfrac{1}{j_1 + \cfrac{1}{j_2 + \cfrac{1}{j_3 + \dots}}}$$

donde el lado derecho de la igualdad es la fracción continua $[j_1, j_2, \dots]$ definida por la sucesión $j_1, j_2, \dots$. Obsérvese que j_1 es la parte entera de $1/\phi(x)$, j_2 es la parte entera de $1/\left(1/\phi(x) - j_1\right)$ y así sucesivamente. De esto último, si $\phi(x) = \phi(y)$ y $j_1, j_2, \dots$ y $i_1, i_2, \dots$ son las sucesiones asociadas a x y y respectivamente, entonces $j_n = i_n$ para todo $n \in \mathbb{N}$. Así $x, y \in F(j_1, \dots, j_n)$ para todo $n \in \mathbb{N}$, de donde $d(x,y) \leq \operatorname{diam}(F(i_1, \dots, i_n)) < r_n$ y entonces $x = y$. Se concluye que ϕ es inyectiva. Cada $\tilde{x} \in [0,1] \setminus \mathbb{Q}$ tiene una representación única como fracción continua $[j_1, j_2, \dots]$ que define el elemento $x \in \widehat{X}$ por $\{x\} = \bigcap_{n\in\mathbb{N}} F(j_1, \dots, j_n)$, de donde $\phi(x) = \tilde{x}$.

Para verificar que ϕ es un isomorfismo entre espacios de medida se debe observar que:

- Para cada $n \in \mathbb{N}$, los índices $j_1, j_2, \dots, j_n$ definen el subconjunto de los $\tilde{x}$ en $[0,1] \setminus \mathbb{Q}$ tales que su expansión en fracción continua comienza por $j_1, j_2, \dots, j_n$. Estos subconjuntos son intervalos cuya longitud tiende a 0 a medida que n tiende a $+\infty$, y la colección de estos intervalos genera $\mathscr{B}_{[0,1]\setminus\mathbb{Q}}$.

- La imagen recíproca mediante ϕ del subconjunto de los $\tilde{x}$ tales que su expansión en fracción continua comienza por $j_1, j_2, \dots, j_n$ es precisamente $F(j_1, \dots, j_n) \in \mathcal{X}$. Dado que $\left\{E \in \mathscr{B}_{[0,1]\setminus\mathbb{Q}} : \phi^{-1}E \in \mathcal{X}\right\}$ es una σ-álgebra de $[0,1] \setminus \mathbb{Q}$ que contiene intervalos que generan $\mathscr{B}_{[0,1]\setminus\mathbb{Q}}$ entonces $\left\{E \in \mathscr{B}_{[0,1]\setminus\mathbb{Q}} : \phi^{-1}E \in \mathcal{X}\right\} = \mathscr{B}_{[0,1]\setminus\mathbb{Q}}$.

- La colección $\left\{E \in \mathcal{X} : \phi^{!-1}(E) \in \mathscr{B}_{[0,1]\setminus\mathbb{Q}}\right\}$ es una σ-álgebra de $\mathcal{X}$ que contiene a los subconjuntos cerrados de la forma $F(j_1, \dots, j_n)$, por lo que $\left\{E \in \mathcal{X} : \phi^{!-1}(E) \in \mathscr{B}_{[0,1]\setminus\mathbb{Q}}\right\} = \mathcal{X}$.

- $\mu \circ \phi^{-1}$ define una medida en $\mathscr{B}_{[0,1]\setminus\mathbb{Q}}$ con extensión idénticamente nula a subconjuntos de $\mathbb{Q}$. Llamemos m a esta extensión.

Los espacios $(X, \mathcal{X}, \mu)$ y $\left([0,1], \mathscr{B}_{[0,1]}, m\right)$ resultan entonces isomorfos. Nótese que, dado que $(X, \mathcal{X}, \mu)$ es no-atómico, $\left([0,1], \mathscr{B}_{[0,1]}, m\right)$ es no-atómico.

Se verifica ahora que $\left([0,1], \mathscr{B}_{[0,1]}, m\right)$ y $\left([0,1], \mathscr{B}_{[0,1]}, \lambda\right)$ son isomorfos. Considerese la función definida en $[0,1]$ por $s \mapsto m[0,s)$. Obsérvese que la función así definida es monónota, pues si $s < t$ entonces $[0, s) \subset [0, t)$ de donde $m[0, s) \leq m[0, t)$. También es contínua, pues si no lo fuera tendría una discontinuidad en $\tilde{x}_0 \in [0, 1]$, en el cual para algún $\epsilon > 0$ se tiene que

$$\epsilon < |m[0, \tilde{x}_0 + \delta) - m[0, \tilde{x}_0)|$$

para todo $\delta > 0$. La identidad $m([0, \tilde{x}_0 + \delta) \setminus [0, \tilde{x}_0)) = m[\tilde{x}_0, \tilde{x}_0 + \delta)$ permitiría concluir que $m\{\tilde{x}_0\} > \epsilon$, lo cual no es posible dado que $\left([0,1], \mathscr{B}_{[0,1]}, m\right)$ es no-atómico. Lo anterior permite definir la función $\vartheta : [0, 1] \to [0, 1]$ por $t \mapsto \text{mín}\{s \geq 0 : m[0, s) = t\}$, que es monótonamente creciente. La función ϑ es la biyección requerida para el isomorfismo entre espacios de medida, En efecto, obsérvese que:

- $m([0, 1] \setminus \vartheta[0, 1]) = 0$. En general $\vartheta[0, 1] \neq [0, 1]$. Si $m\left[s', s\right) = 0$ para algún $s' < s$, se tiene que, por las identidades $m[0, s) - m\left[0, s'\right) = m\left([0, s) \setminus \left[0, s'\right)\right) = m\left[s', s\right)$, $m[0, s) = m\left[0, s'\right)$, $\vartheta(m[0, s)) \leq s' < s$, entonces $s \notin \vartheta[0, 1]$. La recíproca también es cierta, esto es, si $s \notin \vartheta[0, 1]$ entonces existe algún $s' < s$ tal que $m[0, s) = m\left[0, s'\right)$, de donde $m\left[s', s\right) = 0$.

  Se define para cada $s \notin \vartheta[0, 1]$, $I(s) := [s_1, s_2]$, donde s_1 Y s_2 se definen por

  $$s_1 = \text{mín}\left\{s' < s : m\left[0, s'\right) = m[0, s)\right\}$$
  $$s_2 = \text{máx}\left\{s' > s : m\left[0, s'\right) = m[0, s)\right\},$$

  entonces $s \in I(s)$, $|I(s)| \geq \left|s - s'\right| > 0$ y, dado que m es no-atómica, $mI(s) = m(s_1, s_2) = 0$. Más aún, para $s', s'' \notin \vartheta[0, 1]$, o bien $I(s') = I(s'')$ o $I(s') \cap I(s'') = \emptyset$. En efecto, si $s' < s''$ entonces $m\left[0, s'\right) \leq m\left[0, s''\right)$. Si $m\left[0, s'\right) = m\left[0, s''\right)$, por la definición anterior sea $s_1' = \text{mín}\left\{s < s' : m[0, s) = m\left[0, s'\right)\right\}$ y entonces $s_1' < s' < s''$ y $m\left[0, s_1'\right) = m\left[0, s'\right) = m\left[0, s''\right)$, de donde $s_1'' \leq s_1'$. Igualmente, sea $s_1'' = \text{mín}\left\{s < s'' : m[0, s) = m\left[0, s''\right)\right\}$, entonces $s_1'' < s'$ y $m\left[0, s_1''\right) = m\left[0, s''\right) = m\left[0, s'\right)$, de donde $s_1' \leq s_1''$. Así, $s_1' = s_1''$. Con un razonamiento análogo para $s_2'' = \text{máx}\left\{s > s'' : m[0, s) = m\left[0, s''\right)\right\}$ y $s_2' = \text{máx}\left\{s > s' : m[0, s) = m\left[0, s'\right)\right\}$ se obtiene que $s_2' = s_2''$. Se concluye entonces que $I(s') = I(s'')$. Por otro lado, si $m\left[0, s'\right) = m\left[0, s''\right)$, para s_2' tal que $s' < s_2'$ y $m\left[0, s_2'\right) = m\left[0, s'\right) < m\left[0, s''\right)$, y para s_1'' tal que $s_1'' < s''$ y $m\left[0, s_1''\right) = m\left[0, s''\right)$, supóngase que $s_1'' \leq s_2'$, entonces $m\left[0, s_1''\right) \leq m\left[0, s_2'\right) < m\left[0, s''\right)$, contradiciendo la definición de s_1''. Entonces $s_2' < s_1''$ y por esto $I(s') \cap I(s'') = \emptyset$. Como la colección de intervalos disjuntos de medida positiva es a lo más contable, se tiene que la colección $\{I(s) : s \in [0, 1] \setminus \vartheta[0, 1]\}$ es a lo más contable. Se sigue entonces que $[0, 1] \setminus \vartheta[0, 1]$

está cubierto por una colección contable de conjuntos con medida m nula. Se concluye que $m([0,1] \setminus \vartheta[0,1]) = 0$.

- ϑ es uno a uno en $[0,1]$. En efecto, basta observar que, dado que m es no-atómica, $m[0,s) = m[0,s]$ para todo $s \in [0,1]$.

- ϑ es medible, con inversa medible, pues dado que ϑ es monónota, la imagen mediante ϑ de un intervalo es otro intervalo. Usando este hecho y la σ-álgebra generada por la imagen mediante ϑ de los intervalos de $[0,1]$, así como la σ-álgebra generada por la imagen recíproca mediante ϑ de los intervalos de $[0,1]$ se obtiene el resultado.

- $m \circ \vartheta = \lambda$. En efecto, la ecuación $m[0,\vartheta(t)) = t$ implica que, dado que ϑ es monótona, $m[\vartheta(s),\vartheta(t)] = t - s$ para todos $0 < s < t < 1$. Considerando la pre-medida respectiva sobre intervalos y por el Teorema de Extensión de Carathéodory, lo anterior implica que $m \circ \vartheta = \lambda$.

De lo anterior, los espacios de medida $(X, \mathcal{X}, \mu)$ y $\left([0,1], \mathscr{B}_{[0,1]}, \lambda\right)$ son isomorfos, mediante la composición $\vartheta \circ \phi$. El resultado cuando el espacio de medida es atómico puede consultarse en [6].

Al considerar solamente espacios de probabilidad estándar se obtienen propiedades, a partir del resultado anterior, que se verifican en $\left([0,1], \mathscr{B}_{[0,1]}, \lambda\right)$, como se presentó en la proposición 2.1., junto con las propiedades de aproximación.

CAPÍTULO 3

La entropía en el contexto de Teoría de la Medida

En un espacio de medida $(X, \mathcal{X}, \mu)$ se pueden tomar subconjuntos medibles para particionar a X. Por ejemplo, si $A \in \mathcal{X}$ es distinto de X, una partición de X sería de la forma $\{A, A^c\}$. La medida tanto de A como de A^c permite cuantificar la incertidumbre que se elimina cuando se quiere encontrar un elemento $x \in X$ pero solo se conoce a cual de los subconjuntos medibles de la partición pertenece. Cuando el espacio de medida es un espacio de probabilidad este ejemplo es más ilustrativo. Desde el punto de vista de teoría de la medida, cuando $\mu(A) \lll \mu(A^c)$, encontrar un elemento x que pertenece a A elimina más incertidumbre que encontrar un elemento y que pertenece a A^c. Estas nociones, propias de Teoría de la Información, guían las construcciones que se presentan en esta sección. Se define una función de información que modela lo introducido en este párrafo sobre la incertidumbre y particiones mediante conjuntos medibles. La entropía se define primero para particiones medibles del espacio y resulta ser el promedio de la función de información sobre el espacio. Posteriormente se define la entropía de una transformación medible que preserva la medida, y el procedimiento es análogo a la definición de la integral de Riemman de una función de valor real. El lector puede referirse a [6] para estudiar una noción más completa de lo que se ha mencionado en esta introducción.

Definición 3.1. En un espacio medible $(X, \mathcal{X})$:

- Una partición contable de X es una colección contable $\alpha = \{A_i\}$ de conjuntos $\mathcal{X}$-medibles disjuntos tal que $X = \biguplus_{i\in\mathbb{N}} A_i$.

- Si α y β son particiones contables, se dice que β es un refinamiento de α y se denota por $\alpha \leq \beta$, si cada $A_i \in \alpha$ es la unión de algunos B_{j_i} en β.

- Si α y β son particiones contables, $\alpha \vee \beta := \left\{A_i \cap B_j : A_i \in \alpha, B_j \in \beta\right\}$, que es una partición contable de X, denominada partición conjunta de α y β. $\alpha \vee \beta$ es un refinamiento de α y β.

- Se dice que las particiones α y β son independientes si

$$\mu\left(A_i \cap B_j\right) = \mu\left(A_i\right)\mu\left(B_j\right)$$

para todos $A_i \in \alpha$, $B_j \in \beta$.

Sea $\alpha \subset \mathcal{X}$ una partición contable de X y $f \in \mathcal{L}^1(X, \mathcal{X}, \mu)$. Entonces, la esperanza condicional de f respecto a $\Sigma(\alpha)$ tiene la forma:

$$E\left(f|\Sigma(\alpha)\right) = \sum_{i\in\mathbb{N}} \mathcal{X}_{A_i} \frac{\int_{A_i} f\, d\mu}{\mu\left(A_i\right)}$$

En particular, la probabilidad condicional de $A \in \mathcal{X}$ dada $\Sigma(\alpha)$ tiene la forma

$$E\left(\mathcal{X}_A|\Sigma(\alpha)\right) = \sum_{i\in\mathbb{N}} \mathcal{X}_{A_i} \frac{\int_{A_i} \mathcal{X}_A\, d\mu}{\mu\left(A_i\right)} = \sum_{i\in\mathbb{N}} \mathcal{X}_{A_i} \frac{\mu\left(A \cap A_i\right)}{\mu\left(A_i\right)}$$

Se introduce ahora la función de información y la entropía de una partición contable.

Definición 3.2. Para cada partición contable α de X se definen:

- La función de información $I(\alpha) : X \to \mathbb{R}^+$ por

$$I(\alpha)(x) = -\sum_{i\in\mathbb{N}} \mathcal{X}_{A_i}(x) \log\left(\mu\left(A_i\right)\right)$$

- La entropía $H(\alpha)$, que es el promedio de $I(\alpha)$ sobre el espacio X, es decir,

$$H(\alpha) = \int I(\alpha)\, d\mu = -\sum_{\substack{A_i\in\alpha, \\ \mu(A_i)\neq 0}} \mu\left(A_i\right) \log \mu\left(A_i\right)$$

Nota 3.1. La función de información puede tomar el valor $+\infty$, esto cuando $x \in A_i$ con $\mu\left(A_i\right) = 0$. Se adopta la convención $0 \log 0 = 0$. Obsérvese que en la expresión de la entropía solo cuentan los elementos de la partición que no tienen medida nula.

Definición 3.3. Sean $\mathcal{Y}$ y $\mathcal{X}$ σ-álgebras de X, con $\mathcal{Y} \subset \mathcal{X}$, y $\alpha \subset \mathcal{X}$ una partición contable de X.

- La función de información condicional $I(\alpha|\mathcal{Y}) : X \to \mathbb{R}^+$ definida por α dada $\mathcal{Y}$ se define como

$$I(\alpha|\mathcal{Y})(x) = -\sum_{i\in\mathbb{N}} \mathcal{X}_{A_i}(x) \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})(x)$$

- La entropía condicional $H(\alpha|\mathcal{Y})$ es el promedio de $I(\alpha|\mathcal{Y})$ sobre el espacio

$$H(\alpha|\mathcal{Y}) = \int I(\alpha|\mathcal{Y})\, d\mu$$

Nota 3.2. La función de información condicional también puede tomar el valor $+\infty$ e igualmente se adopta la convención $0 \log 0 = 0$. Por otro lado, con la misma notación de la definición anterior, para todo $A_i \in \alpha$ se tiene que $-\log \circ E(\mathcal{X}_{A_i}|\mathcal{Y}) : X \to \mathbb{R}^+$ es una función $\mathcal{Y}$-medible, por lo que la integral de la función de información tiene sentido. Si $A_i \in \alpha$ se tiene

$$\int \mathcal{X}_{A_i} \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})\, d\mu = \int E(\mathcal{X}_{A_i}|\mathcal{Y}) \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})\, d\mu$$

pues siendo $\log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})$ una función $\mathcal{Y}$-medible y finita, por propiedades de la esperanza condicional

$$\begin{aligned}\int E(\mathcal{X}_{A_i}|\mathcal{Y}) \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})\, d\mu &= \int E(\mathcal{X}_{A_i} \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})|\mathcal{Y})\, d\mu \\ &= \int \mathcal{X}_{A_i} \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})\, d\mu.\end{aligned}$$

Entonces, la entropía condicional toma la forma

$$H(\alpha|\mathcal{Y}) = -\int \sum_{i\in\mathbb{N}} E(\mathcal{X}_{A_i}|\mathcal{Y}) \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})\, d\mu$$

Proposición 3.1. Sean $\mathcal{Y}$ y $\mathcal{X}$ σ-álgebras de X, con $\mathcal{Y} \subset \mathcal{X}$, y $\alpha \subset \mathcal{X}$ una partición contable de X. Se tiene que $H(\alpha|\mathcal{Y}) = 0$ si y sólo si $\alpha \subset \mathcal{Y}$. En particular, $H(\alpha|\mathcal{X}) = 0$.

Demostración:

Si $A_i \in \alpha \subset \mathcal{Y}$ se tiene que $E(\mathcal{X}_{A_i}|\mathcal{Y}) = \mathcal{X}_{A_i}$ y entonces $\mathcal{X}_{A_i} \log \circ \mathcal{X}_{A_i} \equiv 0$, por lo cual $I(\alpha|\mathcal{Y}) \equiv 0$. Recíprocamente, si $H(\alpha|\mathcal{Y}) = 0$, entonces para cada $A_i \in \alpha$, dado que $0 \leq -\mathcal{X}_{A_i} \log \circ E(\mathcal{X}_{A_i}|\mathcal{Y})$, necesariamente $\mathcal{X}_{A_i} \log \circ E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \equiv 0$. En A_i se tiene que $\mathcal{X}_{A_i} \equiv 1$ por lo cual $\log \circ E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \equiv 0$ siendo $E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \equiv 1$ en A_i. También, de

$$
\begin{aligned}
\mu\left(A_i\right) &= \int \mathcal{X}_{A_i}\, d\mu \\
&= \int E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right)\, d\mu \\
&= \int_{A_i} E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right)\, d\mu + \int_{A_i^c} E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right)\, d\mu \\
&= \int_{A_i} 1\, d\mu + \int \mathcal{X}_{A_i^c} E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right)\, d\mu \\
&= \mu\left(A_i\right) + \int \mathcal{X}_{A_i^c} E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right)\, d\mu,
\end{aligned}
$$

se obtiene que $\mathcal{X}_{A_i^c} E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \equiv 0$ (pues es no negativa) y dado que en A_i^c $\mathcal{X}_{A_i^c} \equiv 1$ necesariamente $E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \equiv 0$ en A_i^c. Así $E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) = \mathcal{X}_{A_i}$ (pues coinciden en A_i y en A_i^c) y de $E\left(\mathcal{X}_{A_i}|\mathcal{Y}\right) \in \mathcal{M}(A, \mathcal{Y})$ se obtiene que $A_i \in \mathcal{Y}$. Entonces $\alpha \subset \mathcal{Y}$.

La siguiente proposición permite relacionar particiones, sus particiones conjuntas respectivas y las σ-álgebras generadas por estas.

Proposición 3.2. Sean $(X, \mathcal{X})$ y α, β, γ particiones contables de X. Entonces

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

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

En particular, si $\gamma = \{X\}$ entonces:

- $I(\alpha \vee \beta) = I(\alpha) + I(\beta|\Sigma(\alpha))$.

- $H(\alpha \vee \beta) = H(\alpha) + H(\beta|\Sigma(\alpha))$.

Demostración:

Para $x \in \mathcal{X}$ existen $A_x \in \alpha$, $B_x \in \beta$ y $G_x \in \gamma$ únicos tales que $x \in A_x \cap B_x \cap G_x$. Entonces

$$
I(\alpha \vee \beta|\Sigma(\gamma))(x) = -\sum_{i,j \in \mathbb{N}} \mathcal{X}_{A_i \cap B_j}(x) \log \circ E\left(\mathcal{X}_{A_i \cap B_j}|\Sigma(\gamma)\right)(x)
$$

$$
\begin{aligned}
&= -\log \circ E\left(\mathcal{X}_{A_x \cap B_x} | \Sigma(\gamma)\right)(x) \\
&= -\log \circ \sum_{k \in \mathbb{N}} \mathcal{X}_{G_k} \frac{\mu\left(A_x \cap B_x \cap G_k\right)}{\mu\left(G_k\right)}(x) \\
&= -\log \left(\frac{\mu\left(A_x \cap B_x \cap G_x\right)}{\mu\left(G_x\right)}\right)
\end{aligned}
$$

De la misma manera se tienen las dos indentidades:

$$
I(\alpha|\Sigma(\gamma))(x) = -\log\left(\frac{\mu\left(A_x \cap G_x\right)}{\mu\left(G_x\right)}\right)
$$

$$
I(\beta|\Sigma(\alpha \vee \gamma))(x) = -\log\left(\frac{\mu\left(A_x \cap B_x \cap G_x\right)}{\mu\left(A_x \cap G_x\right)}\right),
$$

con lo cual, $I(\alpha|\Sigma(\gamma))(x) + I(\beta|\Sigma(\alpha \vee \gamma))(x) = I(\alpha \vee \beta|\Sigma(\gamma))(x)$. La identidad para las entropías se obtiene por integración. Cuando $\gamma = \{X\}$, hay que notar que:

- $\alpha \vee \gamma = \alpha$.
- Para todo $A_i \in \alpha$ se tiene por definición de esperanza condicional que $E\left(\mathcal{X}_{A_i}|\Sigma(\gamma)\right) = \mu(A_i)$, por lo que $I(\alpha|\Sigma(\gamma)) = I(\alpha)$.
- Igualmente $I(\alpha \vee \beta|\Sigma(\gamma)) = I(\alpha \vee \beta)$ y $I(\beta|\Sigma(\alpha \vee \gamma)) = I(\beta|\Sigma(\gamma))$.

Las identidades de la proposición anterior se conocen como identidades básicas y permiten obtener resultados como el siguiente:

Proposición 3.3. En $(X, \mathcal{X})$, sean α, β, γ particiones contables de X.

- Si $\alpha \leq \beta$ entonces $I(\alpha|\Sigma(\beta)) \equiv 0$.
- Si $\beta \leq \gamma$ entonces $I(\alpha \vee \beta|\Sigma(\gamma)) \equiv I(\alpha|\Sigma(\gamma))$, y en consecuencia

$$
H(\alpha \vee \beta|\Sigma(\gamma)) = H(\alpha|\Sigma(\gamma)).
$$

- Si $\beta \leq \alpha$ entonces $I(\beta|\Sigma(\gamma)) \leq I(\alpha|\Sigma(\gamma))$, y en consecuencia $H(\beta|\Sigma(\gamma)) \leq H(\alpha|\Sigma(\gamma))$. Particularmente, cuando $\gamma = \{X\}$ se obtiene $I(\beta) \leq I(\alpha)$, y en consecuencia $H(\beta) \leq H(\alpha)$.

Demostración:

- Si $\alpha \leq \beta$ entonces $\alpha \subset \Sigma(\beta)$ y, de la proposición 3.1., esto implica que $I(\alpha|\Sigma(\beta)) \equiv 0$.

- Si $\beta \leq \gamma$ entonces $\beta \leq \gamma \leq \gamma \vee \alpha$ y esto implica que $I(\beta|\Sigma(\alpha \vee \gamma)) \equiv 0$, y $H(\beta|\Sigma(\alpha \vee \gamma)) = 0$. El resultado se obtiene por las identidades básicas.

- Si $\beta \leq \alpha$ entonces

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

La siguiente proposición es una consecuencia inmediata de la Desigualdad de Jensen (Ver [7]):

Proposición 3.4. Desigualdad de Jensen para esperanzas condicionales. Sea $\varphi : [0,1] \to \mathbb{R}^+$ una función continua y cóncava, es decir, tal que para todo $p \in [0,1]$, $p\varphi(x) + (1-p)\varphi(y) \leq \varphi(px + (1-p)y)$ para todos $x, y \in [0,1]$. Sea $f : X \to [0,1]$ una función $\hat{\mathcal{X}}$-medible y $\hat{\mathcal{Y}}$ una σ-álgebra de X con $\hat{\mathcal{Y}} \subset \hat{\mathcal{X}}$. Entonces $E\left(\varphi \circ f|\hat{\mathcal{Y}}\right) \leq \varphi \circ \left(E\left(f|\hat{\mathcal{Y}}\right)\right)$ casi en toda parte.

Proposición 3.5. Si $\beta \leq \gamma$ entonces $H(\alpha|\Sigma(\gamma)) \leq H(\alpha|\Sigma(\beta))$.

Demostración:

La función $\varphi(t) = -t \log t$ para $t \in (0,1]$ y $\varphi(0) = 0$ es cóncava y continua en $[0,1]$. Sea $A_i \in \alpha$. Tomando $f = E(\mathcal{X}_{A_i}|\Sigma(\gamma))$, $\hat{\mathcal{Y}} = \Sigma(\beta)$ y $\hat{\mathcal{X}} = \Sigma(\gamma)$ en la proposición anterior, se obtiene que

$$E(\varphi \circ f|\Sigma(\beta)) \leq \varphi \circ (E(f|\Sigma(\beta))).$$

Ahora, como $\beta \leq \gamma$, por propiedad de esperanza condicional:

$$E(f|\Sigma(\beta)) = E(E(\mathcal{X}_{A_i}|\Sigma(\gamma))|\Sigma(\beta)) = E(\mathcal{X}_{A_i}|\Sigma(\beta)).$$

Entonces se obtiene

$$\begin{aligned} E(-E(\mathcal{X}_{A_i}|\Sigma(\gamma)) \log \circ E(\mathcal{X}_{A_i}|\Sigma(\gamma))|\Sigma(\beta)) &\leq \varphi \circ E(\mathcal{X}_{A_i}|\Sigma(\beta)) \\ &= -E(\mathcal{X}_{A_i}|\Sigma(\beta)) \log \circ E(\mathcal{X}_{A_i}|\Sigma(\beta)). \end{aligned}$$

Dado que

$$\begin{aligned} &\int E(-E(\mathcal{X}_{A_i}|\Sigma(\gamma)) \log \circ E(\mathcal{X}_{A_i}|\Sigma(\gamma))|\Sigma(\beta))\, d\mu \\ &\qquad = \int -E(\mathcal{X}_{A_i}|\Sigma(\gamma)) \log \circ E(\mathcal{X}_{A_i}|\Sigma(\gamma))\, d\mu \end{aligned}$$

se obtiene entonces que

$$\int -E\left(\mathcal{X}_{A_i}|\Sigma(\gamma)\right)\log\circ E\left(\mathcal{X}_{A_i}|\Sigma(\gamma)\right)\,d\mu \le \int -E\left(\mathcal{X}_{A_i}|\Sigma(\beta)\right)\log\circ E\left(\mathcal{X}_{A_i}|\Sigma(\beta)\right)\,d\mu.$$

Tomando las respectivas sumas sobre $i \in \mathbb{N}$ se obtiene $H(\alpha|\Sigma(\gamma)) \le H(\alpha|\Sigma(\beta))$.

Para definir la entropía de una transformación se hace un proceso análogo al de la integral de una función mediante particiones. Se requiere el siguiente resultado:

Proposición 3.6. Si (a_n) es una sucesión de números reales no negativos tal que $a_{n+m} \le a_n + a_m$ entonces la sucesión (a_n/n) es convergente[1].

En $(X, \mathcal{X}, \mu, T)$ sea $\alpha \subset \mathcal{X}$ una partición contable de X. Obsérvese que $T^{-1}\alpha = \left\{T^{-1}A_i : A_i \in \alpha\right\} \subset \mathcal{X}$ es una partición contable de X. Además, por la caracterización de transformaciones que preservan la medida, se tiene que

$$H\left(T^{-1}\alpha\right) = \int -\sum_{i\in\mathbb{N}} \mathcal{X}_{A_i}\circ T\log\mu\left(T^{-1}A_i\right)\,d\mu = \int -\sum_{i\in\mathbb{N}} \mathcal{X}_{A_i}\log\mu(A_i)\,d\mu = H(\alpha)$$

Se define la sucesión $(H_n(\alpha))$, donde $H_n(\alpha) := H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)$. $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ denota la partición conjunta de las particiones contables $\alpha, T^{-1}\alpha, \ldots, T^{n-1}\alpha$. De las identidades básicas, con $\gamma = \{X\}$ y teniendo en cuenta que $\gamma \le \bigvee_{i=0}^{n-1} T^{-i}\alpha$:

$$\begin{aligned}
H_{n+m}(\alpha) &= H\left(\bigvee_{i=0}^{n+m-1} T^{-i}\alpha\right)\\
&= H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \vee \bigvee_{i=n}^{n+m-1} T^{-i}\alpha\right)\\
&= H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}\alpha|\Sigma\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)\right)\\
&\le H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}\alpha\right)\\
&= H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(T^{-n}\bigvee_{i=0}^{m-1} T^{-i}\alpha\right)\\
&= H_n(\alpha) + H_m(\alpha)
\end{aligned}$$

[1] La propiedad $a_{n+m} \le a_n + a_m$ se conoce como sub-aditividad. La convergencia de la sucesión (a_n/n) queda garantizada por el lema de Fekete.

Nótese que, por lo anterior, $H_n(\alpha) \leq H_{n-1}(\alpha) + H(\alpha) \leq \ldots \leq nH(\alpha)$. La sucesión $(H_n(\alpha))$ satisface entonces la proposición 3.6., y cuando $H(\alpha) < +\infty$ se tiene que $\lim_n \frac{1}{n} H_n(\alpha) \leq H(\alpha) < +\infty$. Teniendo en cuenta lo anterior podemos definir la entropía de T con respecto a α.

Definición 3.4. En $(X, \mathcal{X}, \mu, T)$ sea $\alpha \subset \mathcal{X}$ una partición contable de X. La entropía de T con respecto a α, siempre que $H(\alpha) < +\infty$, es

$$h(T, \alpha) := \lim_n \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)$$

La definición de la entropía de T se independiza ahora de las particiones:

Definición 3.5. En $(X, \mathcal{X}, \mu, T)$, con μ de probabilidad, la entropía de T respecto a μ es:

$$h(T) := \sup\left\{h(T, \alpha) : \alpha \subset \mathcal{X} \text{ es una partición contable de } X, \text{ con } H(\alpha) < +\infty\right\}.$$

Dos resultados de importancia siguen a continuación. El primero es que $h(\cdot)$ es un invariante de isomorfismos. El segundo consiste en encontrar una partición α tal que $h(T) = h(T, \alpha)$, lo cual facilita establecer que $h(\cdot)$ es un invariante completo de isomorfismos para algunos casos.

Teorema 3.1. Si T y S en $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$ son isomorfas entonces $h(T) = h(S)$.

Demostración:

Sean $M \in \mathcal{X}$, $N \in \mathcal{Y}$ y $\phi : M \to N$ mediante los cuales T y S son isomorfas. Cada partición contable $\alpha \subset \mathcal{Y}$ de Y define la partición contable $\phi^{-1}\alpha$ de M. Dado que $\mu(M^c) = 0$, se obtiene una partición contable de X dada por $\phi^{-1}\alpha \cup \{M^c\}$ y es tal que, por la definición de entropía de una partición, $H\left(\phi^{-1}\alpha\right) = H\left(\phi^{-1}\alpha \cup \{M^c\}\right)$. Además, por el isomorfismo

$$H\left(\phi^{-1}\alpha\right) = -\sum_{i\in\mathbb{N}} \mu\left(\phi^{-1}A_i\right) \log \mu\left(\phi^{-1}A_i\right) = -\sum_{i\in\mathbb{N}} \lambda(A_i) \log \lambda(A_i) = H(\alpha).$$

Más generalmente, dado que $\phi \circ T|_M = S|_N \circ \phi$, se tiene que para todo $n \in \mathbb{N}$:

$$H\left(\bigvee_{j=0}^{n-1} T^{-j}\left(\phi^{-1}\alpha\right)\right) = H\left(\phi^{-1}\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right)\right) = H\left(\bigvee_{j=0}^{n-1} S^{-j}\alpha\right).$$

Por lo anterior $h(S, \alpha) = h\left(T, \phi^{-1}\alpha\right)$, y entonces:

$$
\begin{aligned}
h(S) &= \sup\{h(S,\alpha) : \alpha \subset \mathcal{Y} \text{ partición cont. de } Y, H(\alpha) < +\infty\} \\
&= \sup\left\{h\left(T,\phi^{-1}\alpha\right) : \alpha \subset \mathcal{Y} \text{ partición cont. de } Y, H(\alpha) < +\infty\right\} \\
&= \sup\left\{h\left(T,\phi^{-1}\alpha\right) : \phi^{-1}\alpha \subset \mathcal{X} \text{ partición cont. de } X, H\left(\phi^{-1}\alpha\right) < +\infty\right\} \\
&\leq \sup\{h(T,\beta) : \beta \subset \mathcal{X} \text{ partición cont. de } X, H(\beta) < +\infty\} \\
&= h(T)
\end{aligned}
$$

Igualmente con $\phi^!$ y $\beta \subset \mathcal{X}$ partición contable X se obtiene:

$$
\begin{aligned}
h(T) &= \sup\{h(T,\beta) : \beta \subset \mathcal{X} \text{ partición cont. de } X, H(\beta) < +\infty\} \\
&= \sup\left\{h\left(S,\phi^{!^{-1}}\beta\right) : \beta \subset \mathcal{X} \text{ partición cont. de } X, H(\beta) < +\infty\right\} \\
&= \sup\left\{h\left(S,\phi^{!^{-1}}\beta\right) : \phi^{!^{-1}}\beta \subset \mathcal{Y} \text{ partición cont. de } Y, H\left(\phi^{!^{-1}}\beta\right) < +\infty\right\} \\
&\leq \sup\{h(S,\alpha) : \alpha \subset \mathcal{Y} \text{ partición cont. de } Y, H(\alpha) < +\infty\} \\
&= h(S)
\end{aligned}
$$

Para calcular la entropía se tienen resultados que permiten, a partir de una partición contable con ciertas características, encontrar el supremo que define su valor:

Teorema 3.2. De Abramov. En $(X,\mathcal{X},\mu,T)$ sea (α_n) una sucesión de particiones contables y $\mathcal{X}$-medibles de X tales que $\alpha_n \leq \alpha_{n+1}$. Si $(\Sigma(\alpha_n))$ converge a $\mathcal{X}$ y $H(\alpha_n) < +\infty$ para todo $n \in \mathbb{N}$ entonces $\lim\limits_{n} h(T,\alpha_n) = h(T)$.

Demostración:

Sea $\alpha \subset \mathcal{X}$ una partición contable de X con $H(\alpha) < +\infty$. Para $n > 0$ $\bigvee_{j=0}^{k-1} T^{-j}\alpha \leq \bigvee_{j=0}^{k-1} T^{-j}\alpha \vee \bigvee_{j=0}^{k-1} T^{-j}\alpha_n$ por lo que, de las identidades básicas, se tiene que:

$$
\begin{aligned}
H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right) &\leq H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha \vee \bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right) \\
&= H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right) + H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha | \Sigma\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right).
\end{aligned}
$$

Dado que $\alpha_n \leq \bigvee_{j=0}^{k-1} T^{-j}\alpha_n$, se tiene que

$$
\begin{aligned}
& H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha|\Sigma\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right) = \\
= \quad & H\left(\alpha \vee \bigvee_{j=i}^{k-1} T^{-j}\alpha|\Sigma\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right) \\
= \quad & H\left(\alpha|\Sigma\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right) + H\left(\bigvee_{j=1}^{k-1} T^{-j}\alpha|\Sigma\left(\alpha \vee \bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right) \\
\leq \quad & H\left(\alpha|\Sigma\left(\alpha_n\right)\right) - H\left(\bigvee_{j=1}^{k-1} T^{-j}\alpha|\Sigma\left(\bigvee_{j=1}^{k-1} T^{-j}\alpha_n\right)\right) \\
= \quad & H\left(\alpha|\Sigma\left(\alpha_n\right)\right) - H\left(\bigvee_{j=0}^{k-2} T^{-j}\alpha|\Sigma\left(\bigvee_{j=0}^{k-2} T^{-j}\alpha_n\right)\right).
\end{aligned}
$$

Inductivamente se obtiene entonces que

$$
H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha|\Sigma\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right)\right) \leq kH\left(\alpha|\Sigma\left(\alpha_n\right)\right).
$$

Como $\bigvee_{j=0}^{k-1} T^{-j}\alpha \leq \bigvee_{j=0}^{k-1} T^{-j}\alpha_n \vee \bigvee_{j=0}^{k-1} T^{-j}\alpha$, según la proposición 3.3. se tiene que

$$
H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right) \leq H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n \vee \bigvee_{j=0}^{k-1} T^{-j}\alpha\right).
$$

Así, $H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right) \leq H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right) + kH\left(\alpha|\Sigma\left(\alpha_n\right)\right)$. Entonces

$$
\begin{aligned}
h\left(T, \alpha\right) &= \lim_k \frac{1}{k} H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha\right) \\
&= \lim_k \frac{1}{k} H\left(\bigvee_{j=0}^{k-1} T^{-j}\alpha_n\right) + H\left(\alpha\Sigma\left(\alpha_n\right)\right) \\
&= h\left(T, \alpha_n\right) + H\left(\alpha|\Sigma\left(\alpha\right)\right)
\end{aligned}
$$

Por el Teorema de Convergencia de Martingalas se tiene que $\lim_n E\left(\mathcal{X}_{A_i}|\Sigma\left(\alpha_n\right)\right)(x) = \mathcal{X}_{A_i}(x)$ casi en toda parte para todo $A_i \in \alpha$, y siendo

$$
I\left(\alpha|\Sigma\left(\alpha_n\right)\right)(x) = -\sum_{i\in\mathbb{N}} \mathcal{X}_{A_i}(x) \log \circ E\left(\mathcal{X}_{A_i}|\Sigma\left(\alpha_n\right)\right)(x) = -\log \circ E\left(\mathcal{X}_{A_x}|\Sigma\left(\alpha_n\right)\right)(x)
$$

con $x \in A_x \in \alpha$, de la continuidad de log se obtiene que $\lim_n I\left(\alpha|\Sigma\left(\alpha_n\right)\right)(x) = -\log \mathcal{X}_{A_x}$ casi en toda parte. Como $\log \mathcal{X}_{A_i} \equiv 0$ en cada $A_i \in \alpha$, entonces por teorema de convergencia monótona

$\lim_n \int I(\alpha|\Sigma(\alpha_n))\, d\mu = 0$, luego $\lim_n H(\alpha|\Sigma(\alpha_n)) = 0$. Así, $h(T,\alpha) \leq \lim_n h(T,\alpha_n) \leq h(T)$, y por definición de $h(T)$ también $h(T) \leq \lim_n h(T,\alpha_n)$.

Definición 3.6. En $(X, \mathcal{X}, \mu, T)$ sea $\alpha \subset \mathcal{X}$ una partición contable de X:

- Cuando T es invertible, se dice que α es un generador si $\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha$ genera $\mathcal{X}$.
- Se dice que α es un generador fuerte si $\bigcup_{i=0}^{+\infty} T^{-i}\alpha$ genera $\mathcal{X}$.

Teorema 3.3. De Sinai. En $(X, \mathcal{X}, \mu, T)$ sea $\alpha \subset \mathcal{X}$ una partición contable de X con $H(\alpha) < +\infty$. Si T es invertible y α es un generador, o si α es un generador fuerte, entonces $h(T) = h(T,\alpha)$.

Demostración:

Cuando T es invertible y α es un generador, para $n \in \mathbb{N}$, se tiene que

$$\begin{aligned}
h\left(T, \bigvee_{j=-n}^{n} T^{-j}\alpha\right) &= \lim_k \frac{1}{k} H\left(\bigvee_{i=0}^{k-1} T^{-i}\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right)\right) \\
&= \lim_k \frac{1}{k} H\left(\bigvee_{j=-n}^{n} T^{-j}\alpha \vee \bigvee_{j=-(n-1)}^{n-1} T^{-j}\alpha \vee \ldots \vee \bigvee_{j=-(n+k-1)}^{n+k-1} T^{-j}\alpha\right) \\
&= \lim_k \frac{1}{k} H\left(T^{n}\alpha \vee \ldots \vee T^{-n}\alpha \vee T^{-(n-1)}\alpha \vee \ldots \vee T^{-(n+k-1)}\alpha\right) \\
&= \lim_k \frac{1}{k} H\left(T^{n}\left(\alpha \vee \ldots \vee T^{-(2n+k-1)}\alpha\right)\right) \\
&= \lim_k \frac{1}{k} H\left(\alpha \vee \ldots \vee T^{-(2n+k-1)}\alpha\right) \\
&= \lim_k \frac{2n+k}{k} \frac{1}{2n+k} H\left(\bigvee_{i=0}^{2n+k-1} T^{-i}\alpha\right) \\
&= h(T,\alpha).
\end{aligned}$$

También $\bigvee_{j=-n}^{n} T^{-j}\alpha \leq \bigvee_{j=-(n+1)}^{n+1} T^{-j}\alpha$, y $\Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right) = \mathcal{X}$. Se puede concluir además que

$$\Sigma\left(\bigcup_{n\in\mathbb{N}} \Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right)\right) = \Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right),$$

lo que se obtiene de la siguiente cadena de implicaciones:

$$\bigvee_{j=-n}^{n} T^{-j}\alpha \subset \Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right)$$

implica

$$\Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right) \subset \Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right),$$

que implica

$$\bigcup_{n\in\mathbb{N}} \Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right) \subset \Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right),$$

y que implica

$$\Sigma\left(\bigcup_{n\in\mathbb{N}} \Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right)\right) \subset \Sigma\left(\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha\right).$$

Por otro lado, se tiene que $\bigcup_{i=-\infty}^{+\infty} T^{-i}\alpha \subset \bigcup_{n\in\mathbb{N}} \Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right)$, de donde

$$\Sigma\left(\bigcup_{n\in\mathbb{N}} \Sigma\left(\bigvee_{j=-n}^{n} T^{-j}\alpha\right)\right) = \mathcal{X}.$$

Por lo anterior y por el teorema de Abramov 3.2., se obtiene que $h(T,\alpha) = \lim_{n} h\left(T, \bigvee_{j=-n}^{n} T^{-j}\alpha\right) = h(T)$.

Haciendo un razonamiento completamente análogo se obtiene el mismo resultado en el caso en que α es un generador fuerte.

CAPÍTULO 4

La entropía y los corrimientos de Bernoulli

En esta sección se exhibe un caso en el que la entropía es un invariante completo de isomorfismos.

Se considera el conjunto $\Sigma_{\mathcal{N}}^{\mathbb{Z}} := \mathcal{N}^{\mathbb{Z}} = \{(x_i) : x_i \in \mathcal{N},\ i \in \mathbb{Z}\}$ de sucesiones bilaterales con elementos en el alfabeto $\mathcal{N}$ compuesto por los símbolos $1, 2, \ldots, N$. A continuación se establecen las estructuras necesarias para considerar un espacio métrico $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, d\right)$ y de medida $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, \mathcal{B}, \mu\right)$.

Los subconjuntos más importantes de $\Sigma_{\mathcal{N}}^{\mathbb{Z}}$ en este trabajo son los cilindros. Estos se definen como sigue:

- Los cilindros simétricos de la forma $\{\,(x_i) : x_i = j_i,\ i \in \{-m, \ldots, m\}\,\}$, $m \in \mathbb{N}$, determinado por $j_{-m}, j_{-m+1}, \ldots, j_{m-1}, j_m \in \mathcal{N}$.
- Los cilindros asimétricos, que tienen la forma $\{\,(x_i) : x_i = j_i,\ i \in \{-m, \ldots, n\}\,\}$, $n, m \in \mathbb{N}$, determinado por $j_{-m}, j_{-m+1}, \ldots, j_{n-1}, j_n \in \mathcal{N}$.

No es difícil ver que todo cilindro asimétrico es unión finita de cilindros simétricos, por lo que se considerarán solamente estos últimos, denotados en adelante por:

- $[\ j_{-m}\ \cdots\ j_m\ ]$, que prescribe los valores de (x_i) en las posiciones $-m, -m+1, \ldots, m-1, m$.
- Formas como $[\ i\ j\ k\ ]$ o $[\ i\ ]$ que prescriben los valores de (x_i) en $-1, 0, 1$ y en 0, respectivamente.

Los cilindros tendrán un papel central para establecer las estructuras nombradas y hacer seguimiento a algunas funciones que se definirán más adelante.

En $\Sigma_{\mathcal{N}}^{\mathbb{Z}}$ se pueden definir de manera independiente una métrica y una topología de la siguiente forma. La función $d : \Sigma_{\mathcal{N}}^{\mathbb{Z}} \times \Sigma_{\mathcal{N}}^{\mathbb{Z}} \to \mathbb{R}_0^+$ definida por $((x_i),(y_i)) \longmapsto 2^{-n}$, donde $n = \text{mín}\{|i| : x_i \neq y_i\}$ es una métrica (tomando $d((x_i),(x_i)) = 0$). Por otro lado, la topolpgía que se obtiene al considerar la colección de cilindros simétricos junto con $\emptyset$ como sub-base. Resulta que la topología inducida por la métrica d es equivalente a la topología de los cilindros, ya que:

- $x \in B_{2^{-m}}(x) \subset [\, j_{-m} \cdots j_m \,]$, para cada $x \in [\, j_{-m} \cdots j_m \,]$.
- $y \in [\, x_{-(m-1)} \cdots x_{m-1} \,] \subset B_\epsilon(x)$, para cada $y \in B_\epsilon(x)$, donde $m = \text{mín}\{|i| : x_i \neq y_i\}$ satisface $2^{-m} < \epsilon \leq 1$. Para $1 < \epsilon$, $B_\epsilon(x) = \Sigma_{\mathcal{N}}^{\mathbb{Z}}$.

Como la compacidad y la compacidad secuencial son equivalentes en los espacios métricos, podemos comprobar que $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, d\right)$ es compacto. En efecto,

- Sea (x^i) una sucesión de rango infinito. Como $\Sigma_{\mathcal{N}}^{\mathbb{Z}} = \biguplus_{j \in \mathcal{N}} [\, j \,]$, existe $[\, j_1 \,]$ que contiene infinitos elementos de la sucesión. Luego $[\, j_1 \,] = \biguplus_{j,j' \in \mathcal{N}} [\, j \; j_1 \; j' \,]$ por lo que existe $[\, j_2 \; j_1 \; j_2' \,]$ que contiene infinitos elementos de la sucesión. Continuando de esta manera se obtiene una sucesión de cilindros encajados: $\ldots \supset [\, j_{-m_{m+1}} \cdots j_2 \; j_1 \; j_2' \cdots j_{m_{m+1}}' \,] \supset \ldots \supset [\, j_3 \; j_2 \; j_1 \; j_2' \; j_3' \,] \supset [\, j_2 \; j_1 \; j_2' \,] \supset [\, j_1 \,]$, cada uno con infinitos elementos de la sucesión. Sea $(y_i) = \left(\ldots, j_3, j_2, j_1, j_2', j_3', \ldots\right)$. Para $k = 1$ sea $(x^{i_1}) \in [\, j_2 \; j_1 \; j_2' \,] \subset B_{2^{-1}}(y_i)$, y, sucesivamente, para $k = m$ sea $(x^{i_k}) \in [\, j_{-m_{m+1}} \cdots j_2 \; j_1 \; j_2' \cdots j_{m_{m+1}}' \,] \subset B_{2^{-m}}(y_i)$. La sucesión (x^{i_k}) es una sub-sucesión de (x^i), que converge a (y_i).

El espacio $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, d\right)$ es entonces un espacio polaco (completo y separable). Se establecen ahora las estructuras de medida necesarias para obtener un espacio de probabilidad estándar.

Proposición 4.1. En $\Sigma_{\mathcal{N}}^{\mathbb{Z}}$, la colección de cilindros simétricos $C \cup \{\emptyset\}$ es una semi-álgebra.

Demostración:

- $[\, j_{-m_1} \cdots j_{m_1} \,] \bigcap [\, j_{-m_2} \cdots j_{m_2} \,] \in \left\{ \emptyset, [\, j_{-m} \cdots j_m \,] \right\}$, donde $m = \text{máx}\{m_1, m_2\}$ para cualquier par de cilindros.
- $\left([\, j_{-m} \cdots j_m \,]\right)^c = \biguplus [\, k_{-m} \ldots k_m \,]$,donde $k_{-m}, \ldots, k_m \in \mathcal{N}$ y $\exists i : k_i \neq j_i$. Además, $\Sigma_{\mathcal{N}}^{\mathbb{Z}} = \biguplus_{i=1}^{N} [\, i \,]$.

Por la proposición anterior, con la topología inducida por la métrica d, los cilindros resultan ser subconjuntos a la vez abiertos y cerrados y, de esto último, son también compactos.
Para establecer una medida en $\Sigma_{\mathcal{N}}^{\mathbb{Z}}$ se parte de una pre-medida en $C \cup \{\emptyset\}$.

Nota 4.1. La definición de una pre-medida sobre una colección de conjuntos requiere que cada vez que la unión disjunta de elementos de la colección sea un elemento de la colección se pueda verificar la σ-aditividad, y esta condición no puede relajarse incluso si la colección que se considera es una semi-álgebra, como lo muestra el ejemplo a continuación.

Ejemplo 4.1. En $\Sigma_{\{1,2\}}^{\mathbb{Z}}$ la unión disjunta de los cilindros obtenidos por el rango de la sucesión definida en $\mathbb{Z}^+$ como $n \longmapsto [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ]$ no es un cilindro en C ni se reduce a una unión finita de elementos de la unión.

- Suponga que
$$\biguplus_{n=1}^{+\infty} [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ] = [\ j_{-m}\ \cdots\ j_m\ ].$$
Sea $(x_i) \in [\ j_{-m}\ \cdots\ j_m\ ]$. Entonces $(x_i) \in [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ]$ para algún $n \geq 1$.

  Si $m \leq n-1$, $[\ j_{-m}\ \cdots\ j_m\ ] = [\ 2_{-m}\ \cdots\ 2_m\ ]$ y entonces, (y_i) dado por $y_i = 2$, para $-m \leq i \leq m$, $y_i = 1$ para $i = m + (2k+1)$ y $y_i = 2$ para $i = m + 2k$ es tal que $(y_i) \in [\ j_{-m}\ \cdots\ j_m\ ]$. Pero $(y_i) \notin \biguplus_{n=1}^{+\infty} [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ]$, de donde $[\ j_{-m}\ \cdots\ j_m\ ] \not\subset \biguplus_{n=1}^{+\infty} [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ]$, lo que es una contradicción.

  Por otro lado, si $n-1 < m$ y $n \leq m$,
$$[\ j_{-m}\ \cdots\ j_m\ ] = [\ j_{-m}\ \cdots\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ \cdots\ j_m\ ]$$
y entonces (y_i) dado por $y_i = 2$, para $-n \leq i \leq n$ y $y_{\pm(n+1)}, y_{\pm(n+2)} = 1$ es tal que $(y_i) \in [\ 1_{-(n+2)}\ 1_{-(n+1)}\ 2_{-n}\ \cdots\ 2_n\ 1_{n+1}\ 1_{n+2}\ ]$. Pero $(y_i) \notin [\ j_{-m}\ \cdots\ j_m\ ]$ por la posición $\pm n$, de donde
$$\biguplus_{n=1}^{+\infty} [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ] \not\subset [\ j_{-m}\ \cdots\ j_m\ ]$$
lo que es una contradicción.

- Suponga que

$$\begin{aligned} & \biguplus_{n=1}^{+\infty} [\ 1_{-(n+1)}\ 1_{-n}\ 2_{-(n-1)}\ \cdots\ 2_{n-1}\ 1_n\ 1_{n+1}\ ] \\ = & \biguplus_{i=1}^{m} [\ 1_{-(n_i+1)}\ 1_{-n_i}\ 2_{-(n_i-1)}\ \cdots\ 2_{n_i-1}\ 1_{n_i}\ 1_{n_i+1}\ ], \end{aligned}$$

con $n_1 < n_2 < \ldots < n_m$. Entonces (y_i) dado por $y_i = 2$, para $-n_m \leq i \leq n_m$ y $y_i = 1$ para $i \notin \{-n_m, \ldots, n_m\}$ es tal que

$$(y_i) \in [\ 1_{-(n_m+2)}\ 1_{-(n_m+1)}\ 2_{-n_m}\ \cdots\ 2_{n_m}\ 1_{n_m+1}\ 1_{n_m+2}\ ].$$

Pero $(y_i) \notin \biguplus_{i=1}^{m} [\ 1_{-(n_i+1)}\ 1_{-n_i}\ 2_{-(n_i-1)}\ \cdots\ 2_{n_i-1}\ 1_{n_i}\ 1_{n_i+1}\ ]$.

Para obtener la estructura de medida se introduce el siguiente objeto. Se dice que $p = (p_1, \ldots, p_N)$ es un vector de probabilidad si satisface $0 \leq p_i \leq 1$ y $\sum p_i = 1$. Con un vector de probabilidad p se define π_p en $C \cup \{\emptyset\}$ por

$$\pi_p \left([\ j_{-m}\ \cdots\ j_m\ ] \right) = p_{j_{-m}} p_{j_{-(m-1)}} \ldots p_{j_{m-1}} p_{j_m}, \text{ y } \pi_p(\emptyset) = 0,$$

la cual resulta ser una pre-medida:

Proposición 4.2. π_p es una pre-medida en $C \cup \{\emptyset\}$.

Demostración:

π_p es no negativa. Si $\left([\ j_{-n_i}\ \cdots\ j_{n_i}\ ] \right)$ es una sucesión de cilindros disjuntos tal que $\biguplus_{i=1}^{+\infty} [\ j_{-n_i}\ \cdots\ j_{n_i}\ ] = [\ j_{-n}\ \cdots\ j_n\ ]$ entonces, de la compacidad de los cilindros,

$$\biguplus_{i=1}^{+\infty} [\ j_{-n_i}\ \cdots\ j_{n_i}\ ] = \biguplus_{k=1}^{m} [\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ ],$$

se reduce a una unión finita, con $n_i < n_{i_1} < n_{i_2} < \ldots < n_{i_m}$. Cada cilindro de longitud menor que n_{i_m} puede descomponerse en cilindros de esta longitud sin alterar la medida:

- $[\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ ] = \biguplus_{m', m \in \mathcal{N}} [\ m'\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ m\ ]$.
-

$$\begin{aligned} \sum_{m', m \in \mathcal{N}} \pi_p \left([\ m'\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ m\ ] \right) &= \sum_{m=1}^{N} \left(\sum_{m'=1}^{N} \left(p_{m'} p_{j_{-n_{i_k}}} \ldots p_{j_{n_{i_k}}} p_m \right) \right) \\ &= \sum_{m=1}^{N} \left(p_{j_{-n_{i_k}}} \ldots p_{j_{n_{i_k}}} p_m \right) \\ &= p_{j_{-n_{i_k}}} \ldots p_{j_{n_{i_k}}} \\ &= \pi_p \left([\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ ] \right). \end{aligned}$$

- $\left[\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \right] = \biguplus \left[\ a_{-n_{i_m}}\ \cdots\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \cdots\ a_{n_{i_m}}\ \right]$

  con $a_{\pm(n_{ik}+1)}, \ldots, a_{\pm(n_{i_m})} \in \mathcal{N}$.

- $\sum \pi_p \left(\left[\ a_{-n_{i_m}}\ \cdots\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \cdots\ a_{n_{i_m}}\ \right]\right) = \pi_p \left(\left[\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \right]\right)$,

  con $a_{\pm(n_{ik}+1)}, \ldots, a_{\pm(n_{i_m})} \in \mathcal{N}$.

Nótese ahora que, si se descomponen todos los cilindros desde 1 hasta m de esta forma, deberán tener todas las posibles combinaciones de símbolos en $\mathcal{N}$ de longitud $2n_{i_m} + 1$, con el bloque de $-n$ a n fijo y dado por $[\ j_{-n}\ \cdots\ j_n\ ]$ (pues $\left[\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \right] \subset [\ j_{-n}\ \cdots\ j_n\ ]$), por lo cual estos cilindros conforman de hecho la colección

$$C' = \left\{\left[\ a_{-n_{i_m}}\ \cdots\ j_{-n_i}\ \cdots\ j_{n_i}\ \cdots\ a_{n_{i_m}}\ \right] : a_{-n_{i_m}}, \ldots, a_{-n_i-1}, \ldots, a_{n_i+1}, \ldots, a_{n_{i_m}} \in \mathcal{N}\right\}.$$

Entonces,

$$\begin{aligned}
& \sum_{i=1}^{+\infty} \pi_p \left(\left[\ j_{-n_i}\ \cdots\ j_{n_i}\ \right]\right) = \\
= \quad & \sum_{k=1}^{m} \pi_p \left(\left[\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \right]\right) \\
= \quad & \sum_{k=1}^{m} \left(\sum \pi_p \left(\left[\ a_{-n_{i_m}}\ \cdots\ j_{-n_{i_k}}\ \cdots\ j_{n_{i_k}}\ \cdots\ a_{n_{i_m}}\ \right]\right)\right) \\
= \quad & \sum_{C'} \pi_p \left(\left[\ a_{-n_{i_m}}\ \cdots\ j_{-n_i}\ \cdots\ j_{n_i}\ \cdots\ a_{n_{i_m}}\ \right]\right) \\
= \quad & \sum_{a_{\pm(n+1)} \in \mathcal{N}} \left(\cdots \left(\sum_{a_{\pm(n_{i_m})} \in \mathcal{N}} p_{a_{-(n_{i_m})}} \cdots p_{a_{-(n+1)}} p_{j_{-n}} \cdots p_{j_n} p_{a_{n+1}} \cdots p_{a_{n_{i_m}}}\right) \cdots\right) \\
= \quad & \sum_{a_{\pm(n+1)} \in \mathcal{N}} \left(p_{a_{-(n+1)}} p_{j_{-n}} \cdots p_{j_n} p_{a_{n+1}}\right) \\
= \quad & p_{j_{-n}} \cdots p_{j_n} \\
= \quad & \pi_p \left(\biguplus_{i=1}^{+\infty} \left[\ j_{-n_i}\ \cdots\ j_{n_i}\ \right]\right).
\end{aligned}$$

Mediante el Teorema de Extensión de Carathéodory (Teorema 1.1.) podemos ahora obtener la medida en la σ-álgebra $\mathcal{B}$ generada por los cilindros (la σ-álgebra de Borel[1]). Así, $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, \mathcal{B}, \pi_p\right)$ es un espacio de medida, donde $\mathcal{B}$ es la σ-álgebra de Borel generada por la topología.

[1]Pues la topología inducida por $C \cup \{\emptyset\}$ es la misma que la generada por d.

$\left(\Sigma_{N}^{\mathbb{Z}}, \mathcal{B}, \pi_p\right)$ es un espacio de probabilidad estándar. La aproximación de conjuntos medibles arbitrarios mediante conjuntos abiertos y cerrados (regularidad) tiene como una consecuencia la siguiente proposición, la cual nos permitirá verificar solamente sobre cilindros algunas de las propiedades que siguen más adelante.

Proposición 4.3. Si $E \in \mathcal{B}$ y $\epsilon > 0$ entonces existen cilindros simétricos[2]

$$C_1, C_2, \ldots, C_n$$

tales que $\pi_p\left(E \triangle \bigcup_{i=1}^{n} C_i\right) < \epsilon$.

Demostración:

Sea $E \in \mathcal{B}$ y $\epsilon > 0$. Por regularidad, existe un subconjunto abierto U y un subconjunto cerrado F tales que $F \subset E \subset U$ y $\pi_p(U \setminus F) < \epsilon/2$. Para cada $x \in U$ sea $\epsilon_x > 0$ tal que $B_{\epsilon_x}(x) \subset U$, y a su vez, el cilindro simétrico C_x tal que $x \in C_x \subset B_{\epsilon_x}(x) \subset U$. Entonces $F \subset \bigcup_{x \in U} C_x$, y dado que F es compacto, esto implica que existen $C_1, \ldots, C_n$ tales que $F \subset \bigcup_{i=1}^{n} C_i$. De las relaciones:

$$E \cap \left(\bigcup_{i=1}^{n} C_i\right)^c \subset E \cap F^c \subset U \cap F^c,$$

$$E^c \cap \left(\bigcup_{i=1}^{n} C_i\right) \subset F^c \cap \left(\bigcup_{i=1}^{n} C_i\right) \subset F^c \cap U$$

se obtiene la estimación requerida.

Se introduce ahora una transformación de interés en $\Sigma_{N}^{\mathbb{Z}}$ una vez se tiene el contexto de espacio de medida. La siguiente transformación de $\Sigma_{N}^{\mathbb{Z}}$ es de una simplicidad tan enorme como su importancia en la Teoría Ergódica.

Definición 4.1. El corrimiento de Bernoulli es la función $\mathcal{T}_p : \Sigma_{N}^{\mathbb{Z}} \to \Sigma_{N}^{\mathbb{Z}}$ definida por $\mathcal{T}_p(x_i) = (x_{i+1})$, esto es, al calcular $\mathcal{T}_p$ en la sucesión $i \mapsto x_i$ se obtiene la sucesión $i \mapsto x_{i+1}$.

Para verificar que el corrimiento es medible se parte de una propiedad más fuerte, en el contexto de σ-álgebras de Borel:

Proposición 4.4. El corrimiento $\mathcal{T}_p$ es continuo respecto a la métrica d definida en $\Sigma_{N}^{\mathbb{Z}}$.

Demostración:

[2]Por simplicidad de la notación el cilindro simétrico $[\ j_{-m_i} \ \cdots \ j_{m_i}\ ]$ se denota por C_i

Para $[\, j_{-m}\ \cdots\ j_m \,] \in \mathcal{C}$, se tiene que

$$\mathcal{T}_p^{-1}[\, j_{-m}\ \cdots\ j_m \,] = \biguplus_{i,k \in \mathcal{N}} [\, k_{-(m+1)}\ i_{-m}\ \cdots\ j_{m-1_m}\ j_{m_{m+1}} \,].$$

Como la topología generada por $\mathcal{C} \cup \{\emptyset\}$ coincide con la inducida por la métrica d, se obtiene que $\mathcal{T}_p$ es continuo.

Proposición 4.5. En los espacios de medida $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$, con $\mathcal{X}$ y $\mathcal{Y}$ σ-álgebras de Borel generadas por las topologías $\mathbb{A}_X$ y $\mathbb{A}_Y$ respectivamente, si $T : X \to Y$ es continua ($T^{-1}\mathbb{A}_Y \subset \mathbb{A}_X$) entonces es medible.

Demostración:

Dado que $T^{-1}\mathcal{Y}$ es una σ-álgebra de X y por continuidad $T^{-1}\mathbb{A}_Y \subset \mathbb{A}_X$, entonces $T^{-1}\mathcal{Y} = T^{-1}\Sigma(\mathbb{A}_Y) = \Sigma\left(T^{-1}\mathbb{A}_Y\right) \subset \Sigma\left(\mathbb{A}_X\right) = \mathcal{X}$.

Para verificar ahora que $\mathcal{T}_p$ preserva la medida se hace uso del Teorema de Extensión de Carathéodory[3]:

Proposición 4.6. $\mathcal{T}_p$ preserva la medida π_p.

Demostración:

Se define una medida $\tilde{\pi}$ en $\mathcal{B}$ y se muestra que coincide con π_p en $\mathcal{C}$:

- Para $A \in \mathcal{B}$ sea $\tilde{\pi}(A) := \pi_p\left(\mathcal{T}_p^{-1}A\right)$. De las propiedades de ímagen recíproca y la σ-aditividad de π_p se obtiene que $\tilde{\pi}$ es una medida en $\mathcal{B}$.
- $\pi_p\big|_{\mathcal{C}} \equiv \tilde{\pi}|_{\mathcal{C}}$ pues

$$\tilde{\pi}\left([\, j_{-m}\ \cdots\ j_m \,]\right) = \sum_{i,k \in \mathcal{N}} p_i p_k \pi_p\left([\, j_{-m}\ \cdots\ j_m \,]\right) = \pi_p\left([\, j_{-m}\ \cdots\ j_m \,]\right)$$

La unicidad en la extensión de Carathéodory asegura que $\tilde{\pi} \equiv \pi_p$ en $\mathcal{B}$, por lo cual para todo $A \in \mathcal{B}$ se tiene que $\pi_p\left(\mathcal{T}_p^{-1}A\right) = \tilde{\pi}(A) = \pi_p(A)$.

El corrimiento de Bernoulli es una transformación ergódica, como resulta de demostrar que es de hecho una mezcla fuerte respecto a la medida π_p.

[3]En este caso permite extender una propiedad si ésta se formula en términos de una medida.

Proposición 4.7. $\mathcal{T}_p$ es una mezcla fuerte respecto a π_p.

Demostración:

La relación $\lim_k \pi_p\left(\mathcal{T}_p^{-k} A \cap B\right) = \pi_p(A)\,\pi_p(B)$ se verifica primero cuando A y B son cilindros simétricos. Sean entonces $[\, j_{-m} \cdots j_m \,]$ y $[\, i_{-n} \cdots i_n \,]$ dos cilindros simétricos. Obsérvese que la preimagen mediante $\mathcal{T}_p$ de $[\, j_{-m} \cdots j_m \,]$ puede expresarse como una union finita y disjunta de cilindros simétricos de mayor longitud que tienen el bloque $j_{-m} \ldots j_m$ fijo en las últimas posiciones. Más aún, pueden tomarse preimagenes hasta que ese bloque quede prescrito en posiciones disjuntas a las del cilindro $[\, i_{-n} \cdots i_n \,]$, que se encuentra fijo. En efecto, si $k \geq m+n+1$ entonces $\mathcal{T}_p^{-k}[\, j_{-m} \cdots j_m \,]$ tiene la forma

$$\biguplus_{a^1,\ldots,a^{2k}\in\mathcal{N}} \left[\, a^1_{-(m+k)} \cdots a^{2m+1}_{-(-m+k)}\ a^{2(m+1)}_{-(-(m+1)+k)} \cdots a^{2k}_{-(m+1)+k}\ j_{-m_{-m+k}} \cdots j_{m_{m+k}} \,\right],$$

y entonces al calcular la medida de $\mathcal{T}_p^{-k}[\, j_{-m} \cdots j_m \,] \cap [\, i_{-n} \cdots i_n \,]$, dado que los únicos valores prescritos son $i_{-n}, \ldots, i_n, j_{-m}, \ldots, j_m$, se obtiene como factor común $p_{i_{-n}} \ldots p_{i_n} p_{j_{-m}} \ldots p_{j_m}$, mientras los demás valores recorren todos los símbolos del afabeto $\mathcal{N}$. Así

$$\begin{aligned} & \pi_p\left(\mathcal{T}_p^{-k}[\, j_{-m} \cdots j_m \,] \cap [\, i_{-n} \cdots i_n \,]\right) \\ = \; & p_{i_{-n}} \ldots p_{i_n} p_{j_{-m}} \ldots p_{j_m} \\ = \; & \pi_p[\, j_{-m} \cdots j_m \,]\pi_p[\, i_{-n} \cdots i_n \,] \end{aligned}$$

Sean ahora $A, B \in \mathcal{B}$ subconjuntos medibles arbitrarios y $\epsilon > 0$. Si $\pi_p(A) = 0$ o $\pi_p(B) = 0$ el resultado es inmediato. Supóngase entonces que $\pi_p(A), \pi_p(B) > 0$. Por la proposición 1.3 sean $C_1^A, \ldots, C_{n_A}^A$ y $C_1^B, \ldots, C_{n_B}^B$ cilindros simétricos tales que $\pi_p\left(A \triangle \bigcup_{i=1}^{n_A} C_i^A\right) < \epsilon/4$ y $\pi_p\left(B \triangle \bigcup_{i=1}^{n_B} C_i^B\right) < \epsilon/4$. Denotemos a $\bigcup_{i=1}^{n_A} C_i^A$ por C_A y a $\bigcup_{i=1}^{n_B} C_i^B$ por C_B. Nótese que, análogamente a la parte anterior, puede escogerse $k' \in \mathbb{N}$ a partir del cual, si $k \geq k'$, los elementos en $\mathcal{T}_p^{-k} C_A \cap C_B$ tienen valores prescritos por $\mathcal{T}_p^{-k} C_A$ y C_B en posiciones distintas, con lo cual $\pi_p\left(\mathcal{T}_p^{-k} C_A \cap C_B\right) = \pi_p(C_B)\,\pi_p\left(\mathcal{T}_p^{-k} C_A\right)$. Por otro lado, $\pi_p\left(\mathcal{T}_p^{-k} C_A\right) = \pi_p(C_A)$ y $\pi_p\left(\mathcal{T}_p^{-k} C_B\right) = \pi_p(C_B)$. Ahora, una vez tenemos estimado $\pi_p(B \triangle C_B)$ podemos estimar $\left|\pi_p(B) - \pi_p(C_B)\right|$, pues de las identidades

$$\pi_p(B) = \pi_p(B \cap C_B) + \pi_p\left(B \cap C_B^c\right),$$

$$\pi_p(C_B) = \pi_p(B \cap C_B) + \pi_p(B^c \cap C_B)$$

se obtiene que $\left|\pi_p(B) - \pi_p(C_B)\right| \leq \pi_p\left(B \cap C_B^c\right) + \pi_p(B^c \cap C_B) = \pi_p(B \triangle C_B)$. Consideremos la siguiente desigualdad:

$$\begin{aligned}
\left|\pi_p(A)\pi_p(B) - \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right)\right| &\leq \left|\pi_p(A)\left(\pi_p(B) - \pi_p(C_B)\right)\right| \\
&+ \left|\pi_p(C_B)\left(\pi_p(A) - \pi_p(C_A)\right)\right| \\
&+ \left|\pi_p(C_A)\left(\pi_p(C_B) - \pi_p\left(\mathcal{T}_p^{-k}C_B\right)\right)\right| \\
&+ \left|\pi_p\left(\mathcal{T}_p^{-k}C_B\right)\left(\pi_p(C_A) - \pi_p\left(\mathcal{T}_p^{-k}C_A\right)\right)\right| \\
&+ \left|\pi_p\left(\mathcal{T}_p^{-k}C_B\right)\pi_p\left(\mathcal{T}_p^{-k}C_A\right) - \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right)\right|.
\end{aligned}$$

Por la escogencia de $k^{'}$ se tiene que $\pi_p\left(\mathcal{T}_p^{-k}C_B\right)\pi_p\left(\mathcal{T}_p^{-k}C_A\right) = \pi_p\left(\mathcal{T}_p^{-k}C_A \cap C_B\right)$, por lo que el último término en la suma anterior es igual a

$$\left|\pi_p\left(\mathcal{T}_p^{-k}C_A \cap C_B\right) - \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right)\right|.$$

Estimemos ahora la medida de $\left(\mathcal{T}_p^{-k}C_A \cap C_B\right) \triangle \left(\mathcal{T}_p^{-k}A \cap B\right)$. Obsérvese que, en general, para cualquier selección de $A, B, C, D \subset X$ se tiene que $(A \cap B) \triangle (C \cap D) \subset (A \triangle C) \cup (B \triangle D)$, de donde

$$\left(\mathcal{T}_p^{-k}C_A \cap C_B\right) \triangle \left(\mathcal{T}_p^{-k}A \cap B\right) \subset \left(\mathcal{T}_p^{-k}C_A \triangle \mathcal{T}_p^{-k}A\right) \cup (C_B \triangle B),$$

y entonces

$$\left|\pi_p\left(\mathcal{T}_p^{-k}C_A \cap C_B\right) - \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right)\right| \leq \pi_p(C_A \triangle A) + \pi_p(C_B \triangle B).$$

Entonces $k \geq k^{'}$ implica que $\left|\pi_p(A)\pi_p(B) - \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right)\right| < \epsilon$, por lo cual $\lim_k \pi_p\left(\mathcal{T}_p^{-k}A \cap B\right) = \pi_p(A)\pi_p(B)$.

Una vez se obtiene $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}}, \mathcal{B}, \pi_p, \mathcal{T}_p\right)$, se usa el teorema de Sinai 3.3. para calcular $h\left(\mathcal{T}_p\right)$.

Proposición 4.8. La entropía de $\mathcal{T}_p$ respecto a π_p es $-\sum_{i=1}^{N} p_i \log p_i$.

Demostración:

Consideremos $\alpha = \left\{[\, i \,] : i \in \mathcal{N}\right\} \subset \mathcal{B}$, que es una partición contable de $\Sigma_{\mathcal{N}}^{\mathbb{Z}}$ con $H(\alpha) = -\sum_{i=1}^{N} p_i \log p_i < +\infty$. Para verificar que es un generador fuerte, nótese que

- $\mathcal{T}_p^{-1}[\, i \,] = \bigcup_{k_{-1}, k_0 \in \mathcal{N}} [\, k_{-1} \; k_0 \; i \,]$,
- $\mathcal{T}_p[\, i \,] = \bigcup_{k_0, k_1 \in \mathcal{N}} [\, i \; k_0 \; k_1 \,]$,

- $[\, i_{-1}\ i_0\ i_1\,] = [\, i_0\,] \cap \mathcal{T}_p^{-1}[\, i_1\,] \cap \mathcal{T}_p[\, i_{-1}\,]$.

Inductivamente se puede demostar que cada cilindro $[\, j_{-m}\ \cdots\ j_m\,]$ es la intersección de elementos en $\bigcup_{k=-\infty}^{+\infty} \mathcal{T}_p^{-k}\alpha$. Entonces $\mathcal{C} \subset \Sigma\left(\bigcup_{k=-\infty}^{+\infty} \mathcal{T}_p^{-k}\alpha\right)$ y así

$$\mathcal{B} \subset \Sigma\left(\bigcup_{k=-\infty}^{+\infty} \mathcal{T}_p^{-k}\alpha\right).$$

Ahora, dado que cada $\mathcal{T}_p^{-k}[\, i\,]$ puede escribirse como unión de cilindros en $\mathcal{C}$, se obtiene que $\bigcup_{k=-\infty}^{+\infty} \mathcal{T}_p^{-k}\alpha \subset \Sigma(\mathcal{C})$, con lo cual $\Sigma\left(\bigcup_{k=-\infty}^{+\infty} \mathcal{T}_p^{-k}\alpha\right) = \mathcal{B}$. Por el teorema de Sinai 3.3. $h\left(\mathcal{T}_p\right) = h\left(\mathcal{T}_p, \alpha\right)$. Para calcular $H\left(\bigvee_{j=0}^{n-1} \mathcal{T}_p^{-j}\alpha\right)$ se debe tener en cuenta que $\{(x_i) : x_i = j_i\, , i \in \{0, 1, \ldots, n-1\}\}$ es un subconjunto medible, pues es la unión de cilindros simétricos $[\, i_{-(n-1)}\ \cdots\ j_{-1}\ j_0\ j_1\ \cdots\ j_{n-1}\,]$ para los cuales $p_{j_0} p_{j_1} \ldots p_{j_{n-1}}$ es un factor común en sus medidas. Por la σ-aditividad se obtiene entonces que $\pi_p\left(\{(x_i) : x_i = j_i\, , i \in \{0, 1, \ldots, n-1\}\}\right) = p_{j_0} p_{j_1} \ldots p_{j_{n-1}}$.

Extendiendo la notación a este elemento medible escribimos

$$[\, j_0\ j_1\ \cdots\ j_m\,] := \{(x_i) : x_i = j_i\, , i \in \{0, 1, \ldots, m\}\}\, .$$

Entonces, si $\eta = \sum_{i=1}^{N} p_i \log p_i$

$$\begin{aligned}
& H\left(\bigvee_{j=0}^{n-1} \mathcal{T}_p^{-j}\alpha\right) = \\
= \quad & -\sum_{j_0=1}^{N}\left(\ldots\left(\sum_{j_{n-1}=1}^{N} \pi_p\left([\, j_0\ \cdots\ j_{n-1}\,]\right) \log \pi_p\left([\, j_0\ \cdots\ j_{n-1}\,]\right)\right)\ldots\right) \\
= \quad & -\sum_{j_0=1}^{N}\left(\ldots\left(\sum_{j_{n-1}=1}^{N} p_{j_0} \ldots p_{j_{n-1}} \log p_{j_0} \ldots p_{j_{n-1}}\right)\ldots\right) \\
= \quad & -\sum_{j_0=1}^{N}\left(\ldots\left(\sum_{j_{n-2}=1}^{N}\left(\sum_{j_{n-1}=1}^{N} p_{j_0} \ldots p_{j_{n-2}} p_{j_{n-1}} \left(\log p_{j_0} \ldots p_{j_{n-2}} + \log p_{j_{n-1}}\right)\right)\right)\ldots\right) \\
= \quad & -\sum_{j_0=1}^{N}\left(\ldots\left(\sum_{j_{n-2}=1}^{N}\left(p_{j_0} \ldots p_{j_{n-2}}\left(\sum_{j_{n-1}=1}^{N} p_{j_{n-1}} \log p_{j_0} \ldots p_{j_{n-2}}\right.\right.\right.\right. \\
+ \quad & \left.\left.\left.\left.\sum_{j_{n-1}=1}^{N} p_{j_{n-1}} \log p_{j_{n-1}}\right)\right)\right)\ldots\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{j_0=1}^{N}\left(\dots\left(\sum_{j_{n-2}=1}^{N}\left(p_{j_0}\dots p_{j_{n-2}}\left(\log p_{j_0}\dots p_{j_{n-2}}+\eta\right)\right)\right)\dots\right)\\
&= -\sum_{j_0=1}^{N}\left(\dots\left(\sum_{j_{n-2}=1}^{N} p_{j_0}\dots p_{j_{n-2}}\left(\log p_{j_0}\dots p_{j_{n-3}}+\log p_{j_{n-2}}\right)\right.\right.\\
&+ \left.\left.\sum_{j_{n-2}=1}^{N} p_{j_0}\dots p_{j_{n-2}}\eta\right)\dots\right)\\
&\vdots\\
&= -\sum_{j_0=1}^{N}\left(p_{j_0}\log p_{j_0}+(n-1)p_{j_0}\eta\right)\\
&= -\left(\eta+(n-1)\eta\sum_{j_0=1}^{N}p_{j_0}\right)\\
&= -n\eta,
\end{aligned}
$$

Por lo cual,

$$
h\left(\mathcal{T}_p\right)=h\left(\mathcal{T}_p,\alpha\right)=\lim_n \frac{1}{n}H\left(\bigvee_{j=0}^{n-1}\mathcal{T}_p^{-j}\alpha\right)=-\sum_{i=1}^{N}p_i\log p_i.
$$

Se pueden considerar ahora dos alfabetos $\mathcal{N}$ y $\mathcal{K}$ y sus respectivos espacios de medida junto con el corrimiento de Bernoulli: $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}},\mathcal{B}_{\mathcal{N}},\pi_p,\mathcal{T}_p\right)$ y $\left(\Sigma_{\mathcal{K}}^{\mathbb{Z}},\mathcal{B}_{\mathcal{K}},\pi_q,\mathcal{T}_q\right)$. Aunque el corrimiento actúa igual en cada espacio de sucesiones bilaterales, el sub-índice permite discriminar bajo cual estructura de medida está operando. El resultado que establece que la entropía es un invariante completo para los corrimientos de Bernoulli es el Teorema de Ornstein:

Teorema 4.1. De Ornstein [16]. En $\left(\Sigma_{\mathcal{N}}^{\mathbb{Z}},\mathcal{B}_{\mathcal{N}},\pi_p,\mathcal{T}_p\right)$ y $\left(\Sigma_{\mathcal{K}}^{\mathbb{Z}},\mathcal{B}_{\mathcal{K}},\pi_q,\mathcal{T}_q\right)$, $\mathcal{T}_p$ y $\mathcal{T}_q$ son isomorfos si $h\left(\mathcal{T}_p\right)=h\left(\mathcal{T}_q\right)$.

La demostración de este teorema es extensa y requiere la introducción de nuevos conceptos. Su demostración se omite, pero puede consultarse en [15] con más detalles que la demostración original de Ornstein [16].

CAPÍTULO 5

El conjunto de valores propios de una transformación ergódica y las transformaciones con espectro discreto

5.1. El conjunto de valores propios de una transformación ergódica

En esta sección se muestra que el operador de Koopman U_T asociado a una transformación ergódica T tiene un comportamiento particular en $\mathcal{L}^2$ y permite formular propiedades para T en términos de propiedades conocidas de U_T en el contexto de transformaciones lineales sobre espacios vectoriales.

Consideremos el operador de Koopman $U_T : \mathcal{L}^2(X, \mathcal{X}, \mu) \to \mathcal{L}^2(X, \mathcal{X}, \mu)$ asociado a T en $(X, \mathcal{X}, \mu, T)$ y el sub-espacio vectorial $\left\{[f] \in \mathcal{L}^2(X, \mathcal{X}, \mu) : \ U_T[f] = [f]\right\}$ de $\mathcal{L}^2(X, \mathcal{X}, \mu)$ asociado al valor propio 1 de U_T. Por la proposición 1.2. y teniendo en cuenta que $U_T[1] = [1]$ se obtiene que

- T es ergódica si y sólo si $\dim\left\{[f] \in \mathcal{L}^2(X, \mathcal{X}, \mu) : \ U_T[f] = [f]\right\} = 1$.

Con esto y la caracterización de mezclas fuertes de la proposición 1.3., la ergodicidad y la mezcla fuerte son invariantes espectrales:

Proposición 5.1. Si T y S en $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$ son espectralmente isomorfas, entonces

- Si T es ergódica entonces S es ergódica.

- Si T es una mezcla fuerte entonces S es una mezcla fuerte.

Demostración:

Sea $\mathcal{W} : \mathcal{L}^2(X, \mathcal{X}, \mu) \to \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ el operador lineal mediante el cual T y S son espectralmente isomorfas.

- Dado que $\mathcal{W}[1] \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ y que $[\mathcal{W}[1]] = [\mathcal{W}U_T[1]] = U_S[\mathcal{W}[1]]$, se obtiene que $[\mathcal{W}[1]] \in \left\{[g] \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda) : U_S[g] = [g]\right\}$.

  Si $[g] \in \left\{[g] \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda) : U_S[g] = [g]\right\}$, entonces $\mathcal{W}^! [g] \in \mathcal{L}^2(X, \mathcal{X}, \mu)$, y como

  $$U_T\left[\mathcal{W}^![g]\right] = \left[\mathcal{W}^![U_S[g]]\right] = \mathcal{W}^![g]$$

  se obtiene que $\left[\mathcal{W}^![g]\right] \in \left\{[f] \in \mathcal{L}^2(X, \mathcal{X}, \mu) : U_T[f] = [f]\right\}$. Por lo tanto $\left[\mathcal{W}^![g]\right] = c[1]$ y así $[g] = c[\mathcal{W}[1]]$, lo cual implica que

  $$\dim\left\{[g] \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda) : U_S[g] = [g]\right\} = 1.$$

- Se usa la caracterización de mezcla fuerte de la proposición 1.3.[1] Sean $g_1, g_2 \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$, y $f_1, f_2 \in \mathcal{L}^2(X, \mathcal{X}, \mu)$ tales que $\mathcal{W}f_1 = g_1$ y $\mathcal{W}f_2 = g_2$. Entonces

  $$\left\langle g_1, U_S^n g_2\right\rangle = \left\langle \mathcal{W}f_1, U_S^n \mathcal{W}f_2\right\rangle = \left\langle \mathcal{W}f_1, \mathcal{W}U_T^n f_2\right\rangle = \left\langle f_1, U_T^n f_2\right\rangle,$$

  y entonces $\lim_n \left\langle g_1, U_S^n g_2\right\rangle = \lim_n \left\langle f_1, U_T^n f_2\right\rangle = \langle f_1, 1\rangle \overline{\langle f_2, 1\rangle}$. Ahora, como $U_T 1 = 1$ y $U_S\mathcal{W}1 = \mathcal{W}U_T 1 = \mathcal{W}1$, entonces, siendo $\mathcal{W}1 = c$ se obtiene $|c|^2 = \|c\|_2^2 = \langle c, c\rangle = \langle \mathcal{W}1, \mathcal{W}1\rangle = \langle 1, 1\rangle = \|1\|_2^2 = 1$. Así $\langle f_1, 1\rangle = \langle g_1, \mathcal{W}1\rangle = \langle g_1, c\rangle = \overline{c}\langle g_1, 1\rangle$ y también

  $$\overline{\langle f_2, 1\rangle} = \overline{\langle \mathcal{W}^! g_2, 1\rangle} = \overline{\langle g_2, \mathcal{W}1\rangle} = \overline{\langle g_2, c\rangle} = c\overline{\langle g_2, 1\rangle},$$

  con lo cual $\lim_n \left\langle g_1, U_S^n g_2\right\rangle = \overline{c}\langle g_1, 1\rangle c\overline{\langle g_2, 1\rangle} = \langle g_1, 1\rangle \overline{\langle g_2, 1\rangle}$.

Los valores propios de U_T motivan la definición de un invariante espectral de T.

Definición 5.1. El espectro puntual de T en $(X, \mathcal{X}, \mu, T)$, denotado por $H(T)$, se define por

$$H(T) := \left\{\alpha \in \mathbb{C} : \text{ existe } f \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\} \text{ tal que } f \circ T = U_T f = \alpha f\right\}$$

[1]Por simplicidad se prescindirá de la notación con clases de equivalencia.

Proposición 5.2. Si T y S en $(X, \mathcal{X}, \mu, T)$ y $(Y, \mathcal{Y}, \lambda, S)$ son espectralmente isomorfas entonces $H(T) = H(S)$.

Demostración:

Sea $\mathcal{W} : \mathcal{L}^2(X, \mathcal{X}, \mu) \to \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ el operador lineal mediante el cual T y S son espectralmente isomorfas.

- Si $\alpha \in H(T)$, sea $f \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ tal que $U_T = \alpha f$. Dado que $\mathcal{W}f \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ y es tal que $U_S(\mathcal{W}f) = \mathcal{W}(U_T f) = \alpha \mathcal{W} f$, entonces $\alpha \in H(S)$.

- Si $\beta \in H(S)$, sea $g \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda) \setminus \{0\}$ tal que $U_S = \beta g$. Dado que $\mathcal{W}^! g \in \mathcal{L}^2(X, \mathcal{X}, \mu)$ y es tal que $U_T\left(\mathcal{W}^! g\right) = \mathcal{W}^!(U_S g) = \beta \mathcal{W}^! g$, entonces $\beta \in H(T)$.

Así $H(T) = H(S)$.

Cuando T es ergódica, $H(T)$ resulta ser un sub-grupo contable del círculo unitario $\{c \in \mathbb{C} : |c| = 1\}$ y los sub-espacios de funciones asociados a cada valor propio son ortogonales entre sí, y de dimensión 1. Para verificar esto se deben tener en cuenta las siguientes afirmaciones.

Proposición 5.3. Si T en $(X, \mathcal{X}, \mu, T)$ es ergódica entonces:

- $H(T) \subset \{c \in \mathbb{C} : |c| = 1\}$.

- Si $f \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ satisface $U_T f = \alpha f$ para algún $\alpha \in \mathbb{C}$ entonces $|f| \overset{\mu}{=} c$ para algún $c \in \mathbb{C}$.

- $H(T)$ es sub-grupo del círculo unitario.

- Si $\alpha, \beta \in H(T)$ con $\alpha \neq \beta$, si f y g son vectores propios de U_T asociados a α y β respectivamente, entonces $\langle f, g \rangle = 0$.

- Si $f, g \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ son vectores propios de U_T asociados al valor propio α entonces $f \overset{\mu}{=} cg$ para algún $c \in \mathbb{C}$.

Demostración:

- Sea $\alpha \in H(T)$ y $f \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ tal que $U_T f = \alpha f$. Dado que $\|f\|_2 = \|U_T f\|_2 = |\alpha| \|f\|_2$ entonces $|\alpha| = 1$.

- Si $f \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ satisface $U_T f = \alpha f$ para algún $\alpha \in \mathbb{C}$, dado que $U_T |f| = |f| \circ T = |f \circ T| = |\alpha f| = |\alpha| |f| = |f|$, entonces, de la caracterización de ergodicidad, se obtiene que $|f| \overset{\mu}{=} c$ para algún $c \in \mathbb{C}$.

- Sean $\alpha, \beta \in H(T)$, y f, g vectores propios de α y β, respectivamente. Dado que $|f| \overset{\mu}{=} c$ y $|g| \overset{\mu}{=} d$, y siendo $f\overline{g} \in \mathcal{M}(X, \mathcal{X})$ con $\int |f\overline{g}|^2 \, d\mu = \int |c|^2 |d|^2 \, d\mu < +\infty$, entonces $f\overline{g} \in \mathcal{L}^2(X, \mathcal{X}, \mu)$ es tal que $U_T f\overline{g} = \alpha\overline{\beta} f\overline{g} = \alpha\beta^{-1} f\overline{g}$, esto es $\alpha\beta^{-1} \in H(T)$.

- Sean $\alpha, \beta \in H(T)$ con $\alpha \neq \beta$, y f, g vectores propios de α y β, respectivamente. Por la caracterización de transformaciones que preservan la medida, se tiene que

$$
\begin{aligned}
\langle f, g \rangle &= \int f\,\overline{g}\, d\mu \\
&= \int (f\,\overline{g}) \circ T \, d\mu \\
&= \int (f \circ T)(\overline{g} \circ T) \, d\mu \\
&= \int \alpha f\, \overline{\beta g}\, d\mu \\
&= \alpha\overline{\beta} \langle f, g \rangle .
\end{aligned}
$$

  Esto implica que $\langle f, g \rangle \left(1 - \alpha\overline{\beta}\right) = 0$, pero dado que $1 - \alpha\overline{\beta} = 0$ si y sólo si $\alpha\overline{\beta} = 1$ si y sólo si $\alpha |\beta|^2 = \beta$, lo cual no se tiene pues $\alpha \neq \beta$, entonces se concluye que $\langle f, g \rangle = 0$.

- Sean $f, g \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ vectores propios de U_T asociados al valor propio α. Nótese que $f/g \in \mathcal{M}(X, \mathcal{X})$ y por la caracterización de T ergódica, $f/g \circ T = f \circ T/g \circ T = f/g$ implica $f/g \overset{\mu}{=} c$ para algún $c \in \mathbb{C}$. Entonces $f \overset{\mu}{=} cg$.

Para verificar que $H(T)$ es contable se requieren dos resultados. El primero es que en todo espacio de Hilbert separable un conjunto ortonormal es a lo más numerable. El segundo, siempre que $(X, \mathcal{X}, \mu)$ sea un espacio de probabilidad estándar, como se ha restringido a lo largo de este documento, es que si μ es σ-finita y $\mathcal{X}$ es contablemente generada entonces $\mathcal{L}^p(X, \mathcal{X}, \mu)$ es separable.

Las siguientes proposiciones se toman de [2].

Proposición 5.4. Lema 3.4.6. [2]. Sea $(X, \mathcal{Y}, \mu)$ un espacio de medida finito y $\mathcal{A}$ un álgebra de subconjuntos de X tal que $\Sigma(\mathcal{A}) = \mathcal{Y}$. Entonces $\mathcal{A}$ es densa en $\mathcal{Y}$ en el sentido en que, para cada $E \in \mathcal{Y}$ y $\epsilon > 0$, existe $A_0 \in \mathcal{A}$ tal que $\mu(E \triangle A_0) < \epsilon$.

Proposición 5.5. Lema 3.4.7. [2]. Sea $(X, \mathcal{Y}, \mu)$ un espacio de medida. Suponga que $\mathcal{A}$ es un álgebra de subconjuntos de X tal que $\Sigma(\mathcal{A}) = \mathcal{Y}$ y que $X = \bigcup_{i\in\mathbb{N}} A_i$ con $A_i \in \mathcal{A}$ y $\mu(A_i) < +\infty$ para todo $i \in \mathbb{N}$. Entonces, para cada $E \in \mathcal{Y}$ con $\mu(E) < +\infty$ y $\epsilon > 0$, existe $A_0 \in \mathcal{A}$ que satisface $\mu(E \triangle A_0) < \epsilon$.

Proposición 5.6. Proposición 3.4.5. [2]. En $(X, \mathcal{X}, \mu, T)$, si μ es σ-finita y $\mathcal{X}$ es contablemente generada entonces $\mathcal{L}^p(X, \mathcal{X}, \mu)$ es separable para $1 \leq p < +\infty$.

Demostración:

Sea $C \subset \mathcal{X}$ contable y tal que $\Sigma(C) = \mathcal{X}$. Como μ es σ-finita sea $\{B_n\} \subset \mathcal{X}$ tal que $X = \bigcup_{n\in\mathbb{N}} B_n$ y $\mu(B_n) < +\infty$, para todo $n \in \mathbb{N}$. Consideremos $C' = C \cup \{B_n\}$, que es una colección contable de $\mathcal{X}$. Nótese que $\Sigma(C) \subset \Sigma\left(C'\right) \subset \mathcal{X} = \Sigma(C)$, de donde $\Sigma\left(C'\right) = \mathcal{X}$. Sea $\tilde{C} = \left\{A^c : A \in C'\right\} \cup C'$, y considérese $\mathcal{A} = \mathcal{A}\left(C'\right)$. $\mathcal{A}$ resulta ser la intersección finita de elementos en $\tilde{C}$, por lo cual $\mathcal{A}$ es contable, y además $C' \subset \mathcal{A}$, por lo que $\Sigma\left(C'\right) = \mathcal{X} \subset \Sigma(\mathcal{A})$, y siendo $\mathcal{A} \subset \mathcal{X}$ se obtiene que $\Sigma(\mathcal{A}) = \mathcal{X}$, y $\{B_n\} \subset C' \subset \mathcal{A}$, luego $\mathcal{A}$ satisface las condiciones de la proposición 5.5. Sea

$$\mathcal{G} = \left\{\sum_{j=1}^{N} d_j \mathcal{X}_{D_j} : N \in \mathbb{N},\, d_j \in \mathbb{Q},\, D_j \in \mathcal{A} \text{ con } \mu\left(D_j\right) < +\infty\right\}.$$

$\mathcal{G}$ es contable y $\mathcal{G} \subset \mathcal{L}^p(X, \mathcal{X}, \mu)$. Sea $f \in \mathcal{L}^p(X, \mathcal{X}, \mu)$ y $\epsilon > 0$. Sea $\varphi \in \mathcal{L}^p(X, \mathcal{X}, \mu)$ una función simple tal que $\|f - \varphi\|_p < \epsilon$. Dado que $\varphi = \sum_{j=1}^{N} A_j \mathcal{X}_{A_j}$ con $a_j \in \mathbb{R}$ y $A_j \in \mathcal{X}$ con $\mu\left(A_j\right) < +\infty$, sean:

- $d_j \in \mathbb{Q}$ tales que $\left\|\varphi - \sum_{j=1}^{N} d_j \mathcal{X}_{A_j}\right\|_p \leq \sum_{j=1}^{N} \left|a_j - d_j\right| \left\|\mathcal{X}_{A_j}\right\|_p < \epsilon$.
- $D_j \in \mathcal{A}$ tales que $\mu\left(A_j \triangle D_j\right) < \left(\frac{\epsilon}{N|d_j|}\right)^p$.

Entonces, teniendo en cuenta que

$$\left|\mathcal{X}_{A_j} - \mathcal{X}_{D_j}\right| = \left|\mathcal{X}_{A_j \cup D_j}\mathcal{X}_{A_j \cap D_j^c} - \mathcal{X}_{A_j \cup D_j}\mathcal{X}_{A_j^c \cap D_j}\right| \leq \mathcal{X}_{A_j \cap D_j^c} + \mathcal{X}_{A_j^c \cap D_j},$$

$$\begin{aligned}
\left\|\sum_{j=1}^{N} d_j \mathcal{X}_{A_j} - \sum_{j=1}^{N} d_j \mathcal{X}_{D_j}\right\|_p &= \left\|\sum_{j=1}^{N} d_j \left(\mathcal{X}_{A_j} - \mathcal{X}_{D_j}\right)\right\|_p \\
&\leq \sum_{j=1}^{N} \left|d_j\right| \left\|\mathcal{X}_{A_j} - \mathcal{X}_{D_j}\right\|_p
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{N} \left|d_j\right| \left(\int \left|\mathcal{X}_{A_j} - \mathcal{X}_{D_j}\right| d\mu \right)^{\frac{1}{p}} \\
&\leq \sum_{j=1}^{N} \left|d_j\right| \mu\left(A_j \triangle D_j\right)^{\frac{1}{p}} \\
&< \epsilon
\end{aligned}
$$

Así, $\sum_{j=1}^{N} d_j \mathcal{X}_{D_j} \in \mathcal{G}$ es tal que

$$\left\| f - \sum_{j=1}^{N} d_j \mathcal{X}_{D_j} \right\|_p \leq \|f - \varphi\|_p + \left\| \varphi - \sum_{j=1}^{N} d_j \mathcal{X}_{D_j} \right\|_p + \left\| \sum_{j=1}^{N} d_j \mathcal{X}_{A_j} - \sum_{j=1}^{N} d_j \mathcal{X}_{D_j} \right\|_p < 3\epsilon$$

Entonces $\mathcal{G}$ es denso en $\mathcal{L}^p(X, \mathcal{X}, \mu)$ y por lo tanto $\mathcal{L}^p(X, \mathcal{X}, \mu)$ es separable.

Dado que los espacios de medida en los que se ha trabajado son espacios de probabilidad estándar, la proposición 2.3. garantiza la existencia de una base contable $\mathcal{U}$ para la topología que genera la σ-álgebra $\mathcal{B}$, como se ha definido. Si denotamos por $\mathbb{A}$ a esta topología, se obtiene que $\Sigma(\mathcal{U}) \subset \Sigma(\mathbb{A}) = \mathcal{B}$, y dado que cada elemento en $\mathbb{A}$ es la unión contable de una sub-colección de $\mathcal{U}$, por definición de σ-álgebra se obtiene que $\mathbb{A} \subset \Sigma(\mathcal{U})$. Entonces la σ-álgebra de todo espacio de probabilidad estándar es contablemente generada, y en consecuencia, por la proposición 5.6., $\mathcal{L}^p(X, \mathcal{B}, \mu)$ es separable para $1 \leq p < +\infty$, siempre que $(X, \mathcal{B}, \mu)$ sea un espacio de probabilidad estándar.

Se concluye finalmente que si T es ergódica entonces $H(T)$ es contable, pues por cada $\alpha \in H(T)$ se considera $f_\alpha \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\}$ vector propio de U_T asociado a α, con $\|f_\alpha\|_2 = 1$, y entonces $\left\{f_\alpha \in \mathcal{L}^2(X, \mathcal{X}, \mu) \setminus \{0\} : \alpha \in H(T)\right\}$ es a lo más contable (pues es un conjunto ortonormal), mediante la correspondencia entre α y f_α.

5.2. Las transformaciones con espectro discreto

Se considera ahora el caso en que los vectores propios asociados a los valores propios de U_T, con T ergódica, son una base del espacio vectorial $\mathcal{L}^2$.

Definición 5.2. Si T en $(X, \mathcal{X}, \mu, T)$ es ergódica, se dice que T tiene espectro discreto si existe una base ortonormal de $\mathcal{L}^2(X, \mathcal{X}, \mu)$ compuesta de funciones propias (vectores propios) de U_T.

Obsérvese que por lo expuesto en la sección anterior, la base tiene exactamente un vector propio asociado a cada valor propio, pues los sub-espacios de cada valor propio son ortogonales y de dimensión 1. Con esta definición, para el caso de las transformaciones con espectro discreto, el espectro puntual es un invariante espectral completo:

Teorema 5.1. Del espectro discreto. Sean T y S transformaciones ergódicas con espectro discreto en espacios de probabilidad $(X, \mathcal{X}, \mu)$ y $(Y, \mathcal{Y}, \lambda)$, respectivamente. Si $H(T) = H(S)$ entonces T y S son espectralmente isomorfas.

Demostración: Sean

$$B_X = \left\{ f_\alpha \in \mathcal{L}^2(X, \mathcal{X}, \mu) : U_T f_\alpha = \alpha f_\alpha \right\},$$
$$B_Y = \left\{ g_\alpha \in \mathcal{L}^2(Y, \mathcal{Y}, \lambda) : U_S g_\alpha = \alpha g_\alpha \right\}$$

las bases ortonormales correspondientes, esto es $\left\langle f_\alpha, f_\beta \right\rangle = 1$ si $\alpha = \beta$ y $\left\langle f_\alpha, f_\beta \right\rangle = 0$ si $\alpha \neq \beta$. Se define $\mathcal{W} : \mathcal{L}^2(Y, \mathcal{Y}, \lambda) \to \mathcal{L}^2(X, \mathcal{X}, \mu)$ por $g_\alpha \in B_Y \mapsto f_\alpha \in B_X$. La extensión a los demás elementos de $\mathcal{L}^2(Y, \mathcal{Y}, \lambda)$ resulta de que B_Y es base del espacio vectorial $\mathcal{L}^2(Y, \mathcal{Y}, \lambda)$. $\mathcal{W}$ es invertible mediante la correspondencia $f_\alpha \in B_X \mapsto g_\alpha \in B_Y$. Además:

- $\left\langle \mathcal{W} g_\alpha, \mathcal{W} g_\beta \right\rangle = \left\langle f_\alpha, f_\beta \right\rangle = 0$ si $\alpha \neq \beta$, caso en el que también $\left\langle g_\alpha, g_\beta \right\rangle = 0$.
- $\left\langle \mathcal{W} g_\alpha, \mathcal{W} g_\beta \right\rangle = \left\langle f_\alpha, f_\beta \right\rangle = 1$ si $\alpha = \beta$, caso en el que también $\left\langle g_\alpha, g_\beta \right\rangle = 1$.
- $\mathcal{W} \circ U_S g_\alpha = \mathcal{W}(\alpha g_\alpha) = \alpha f_\alpha = U_T f_\alpha = U_T \circ \mathcal{W} g_\alpha$.

Más allá

Hemos visto que para la clasificación mediante isomorfismos e isomorfismos espectrales se pueden definir ciertos objetos (entropía y el conjunto de valores propios de la transformación) que, una vez se reduce el conjunto de transformaciones objeto de clasificación, permiten obtener una clasificación completa, es decir, dan una caracterización de cuándo, según la relación de equivalencia inducida por la clasificación, las transformaciones están o no en la misma clase. Lo que se expone al comienzo de [17], y que llevó a la escogencia de los contenidos para desarrollar este documento, es la clasificación sobre las relaciones de equivalencia que permiten identificar transformaciones, es decir, el objeto de clasificación ya no son las transformaciones sino las relaciones de equivalencia con que se buscan clasificar las transformaciones, y una forma de clasificar esas relaciones de equivalencia es mediante la respuesta al interrogante: ¿Cúando puede asignarse un conjunto de objetos a esta relación, de forma tal que las clases de equivalencia por ella inducidas puedan reconstruirse a partir del espacio en el cual se encuentran esos objetos?. Los problemas de clasificación se describen entonces de la siguiente forma [17]:

Un problema de clasificación esta dado por:

- *Una colección de objetos X.*
- *Una relación de equivalencia $\mathcal{E}$ en X.*

Y una clasificación completa de X mediante $\mathcal{E}$ consiste de:

- *Un conjunto de invariantes $\mathcal{I}$.*

- *Una función $c : X \to \mathcal{I}$ tal que $x\mathcal{E}y$ si y sólo si $c(x) = c(y)$.*

Como se expuso en los capítulos 3 y 4, la clasificación de los corrimientos de Bernoulli es completa. En esta, el conjunto de invariantes $\mathcal{I}$ son los reales (los valores que toma la entropía) y la función c es $h(\cdot)$. Igualmente la clasificación de transformaciones ergódicas con espectro discreto, en la cual $\mathcal{I}$ son los sub-grupos contables del círculo unitario en $\mathbb{C}$ y la función c de define por $T \mapsto H(T)$. Obsérvese que en $\mathbb{R}$ y en el círculo unitario en $\mathbb{C}$ se pueden considerar, en lugar de los objetos $x \in \mathbb{R}$ y $H(T) \subset \mathbb{C}$, las clases de equivalencia definidas por las relaciones de igualdad en $\mathbb{R}$ y de igualdad de conjuntos en $\mathbb{C}$. Entonces la clasificación se formularía como: $x\mathcal{E}y$ si y sólo si $c(x)\mathcal{F}c(y)$. El estudio de formulaciones como ésta se hace en el contexto de la Teoría Descriptiva de Conjuntos, como lo comenta el resumen presentado en [17].

Bibliografía

[1] BARTLE, ROBERT G.*The Elements of Integration and Lebesgue Measure*. JOHN WILEY AND SONS, INC, 1995.

[2] COHN, DONALD L. *Measure Theory*. BIRKHÄUSER, 1980.

[3] HALMOS, PAUL R. *Measure Theory*. SPRINGER-VARLAG NEW YORK INC, 1974.

[4] RESNICK, SIDNEY L. *A Probability Path*. BIRKHÄUSER, 1998.

[5] MUKHERJEE, SAYAN. Extension of measure. [en línea] 2009. [fecha de consulta: 1 de julio de 2013]

Disponible en: <http://www.stat.duke.edu/courses/Fall09/sta205/lec/extension.pdf>.

[6] SARIG, OMRI. Lecture Notes on Ergodic Theory. [en línea] Diciembre 10, 2009. [fecha de consulta: 1 de julio de 2013]

Disponible en: <http://www.wisdom.weizmann.ac.il/s̃arigo/506/ErgodicNotes.pdf>.

[7] WALKDEN, CHARLES. Ergodic Theory lectures notes. [en línea] Abril 29, 2013. [fecha de consulta: 28 de agosto de 2013]

Disponible en: <http://www.maths.manchester.ac.uk/c̃walkden/ergodic-theory/>.

[8] HOCHMAN, MICHAEL. Notes on Ergodic Theory. Noviembre 17, 2012. [fecha de consulta: 28 de agosto de 2013]

Disponible en: <http://math.huji.ac.il/m̃hochman/courses/ergodic-theory-2012/notes.final.pdf>.

[9] DAJANI, KARMA, DIRKSIN, SJOERD. A Simple Introduction to Ergodic Theory. [en línea] Diciembre 18, 2008. [fecha de consulta: 28 de agosto de 2013]

Disponible en: <http://www.staff.science.uu.nl/k̃raai101/lecturenotes2009.pdf>.

[10] WALTERS, PETER. *An Introduction to Ergodic Theory*. SPRINGER-VARLAG NEW YORK INC, 1982.

[11] FETTER NATHANSKY, HELGA, GAMBOA DE BUEN, BERTA. *Introducción al Análisis Funcional y a la Geometría de Espacios de Banach*. GRUPO EDITORIAL IBEROAMÉRICA, S.A. DE C.V., 2002.

[12] BRIN, MICHAEL, STUCK, GARRET. *Introduction to Dynamical Systems*. CAMBRIDGE UNIVERSITY PRESS, 2004.

[13] ULCIGRAI, CORINNA. Dynamical Systems and Ergodic Theory. [en línea] 2012/2013. [fecha de consulta: 1 de julio de 2013]

Disponible en: <http://www.maths.bristol.ac.uk/m̃axcu/Teaching.html>.

[14] WARD, THOMAS. Entropy of Compact Group Automorphisms. [en línea] 1994. [fecha de consulta: 28 de agosto de 2013]

Disponible en: <http://www.maths.dur.ac.uk/t̃pcc68/lecturenotes/#entropy>.

[15] SHIELDS, PAULC. The theory of Bernoulli Shifts. [en línea] Octubre 3 de 2003. [fecha de consulta: 28 agosto de 2013]

Disponible en:

<http://www.impan.pl/ gutman/The %20Theory %20of %20Bernoulli %20Shifts.pdf>.

[16] D. ORNSTEIN. *Bernoulli shifts with the same entropy are isomorphic*. Adv. in Math. 4 (1970), 337-352.

[17] KECHRIS, ALEXANDER S., RUCKER-DROB, ROBIN D. The Complexity of Classification Problems in Ergodic Theory. [en línea] 2009. [fecha de consulta: 1 de mayo de 2013]

Disponible en: <http://www.math.cmu.edu/ẽschimme/Appalachian/KechrisNotes.pdf>.