Universidad Nacional Autónoma de México

Facultad de Ciencias

Algunos Aspectos Probabilísticos de Teoría Ergódica

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
MATEMÁTICO

PRESENTA:

FRANCISCO PÉREZ CARBAJAL

DIRECTOR DE TESIS:
DR. GUILLERMO GRABINSKY STEIDER

FACULTAD DE CIENCIAS
UNAM

2010

Hoja de Datos del Jurado

1. Datos del Alumno Apellido Paterno Apellido Materno Nombre(s) Teléfono Universidad Nacional Autónoma de México Facultad de Ciencias Carrera Número de Cuenta	1. Datos del Alumno Peréz Carbajal Francisco 57000683 Universidad Nacional Autónoma de México Facultad de Ciencias Matemáticas 09924849-7
2. Datos del Tutor Grado Apellido Paterno Apellido Materno Nombre(s)	2. Datos del Tutor Dr. Grabinsky Steider Guillermo
3. Datos del Sinodal 1 Grado Apellido Paterno Apellido Materno Nombre(s)	3. Datos del Sinodal 1 Dr. Barrera Sánchez Pablo
4. Datos del Sinodal 2 Grado Apellido Paterno Apellido Materno Nombre(s)	4. Datos del Sinodal 2 Dr. López García Francisco Marcos
5. Datos del Sinodal 3 Grado Apellido Paterno Apellido Materno Nombre(s)	5. Datos del Sinodal 3 Dr. Méndez Lango Héctor
6. Datos del Sinodal 4 Grado Apellido Paterno Apellido Materno Nombre(s)	6. Datos del Sinodal 4 Dr. Rincón Solís Luis Antonio
7. Datos del Trabajo Escrito Título Número de Páginas Año	7. Datos del Trabajo Escrito Algunos Aspectos Probabilísticos de Teoría Ergódica 141 p. 2010

Índice general

Índice de figuras

Introducción

La Teoría Ergódica es el estudio de sistemas dinámicos, para los cuales el espacio fase es un espacio de medida, y la transformación definida en este espacio **preserva la medida**. Su desarrollo fue inicialmente motivado por problemas en la Física Estadística. Un aspecto central de la teoría es el comportamiento de un sistema dinámico cuando se le permite evolucionar en el tiempo. Esto se expresa a través de teoremas ergódicos, que afirman que, bajo determinadas condiciones, el promedio temporal de una función en un punto existe casi dondequiera. Dos de los ejemplos más importantes son los **Teoremas Ergódicos de Birkhoff y von Neumann**. Para la clase especial de **sistemas ergódicos**, el promedio temporal de una función en un punto es el mismo para casi todos los puntos; estadísticamente hablando, cuando el sistema se desarrolla durante el tiempo "olvida" su estado inicial. Propiedades más fuertes como **fuertemente mezclante** y **débilmente mezclante**, también se han estudiado ampliamente. El problema de la clasificación de sistemas es otra parte importante de la Teoría Ergódica Abstracta. Un papel destacado en la teoría y sus aplicaciones es desempeñado por la noción de **entropía**.

Aplicaciones de la teoría, a otras partes de las matemáticas, por lo general implican el establecimiento de la propiedad de ergodicidad para el sistema dinámico que se forma a partir del problema. Esta teoría tiene conexiones con la Teoría de Números, la Teoría de la Información, la Teoría de la Probabilidad, entre otras.

El **Teorema de Recurrencia de Poincaré** afirma que ciertos sistemas, después de un tiempo suficientemente largo, vuelven a un estado muy cercano al estado inicial. Este resultado se aplica a los sistemas físicos que conservan la energía. El **Tiempo de recurrencia de Poincaré** es el tiempo transcurrido hasta la recurrencia. Este tiempo fue utilizado por S. Kakutani para construir una transformación inducida por una **transformación preservadora de medida**.

La tesis abordará principalmente sistemas dinámicos a tiempo discreto, excepto por el Teorema Ergódico para flujos. Se utilizarán conceptos y resultados de Teoría de la Medida y Análisis, la mayoría de los cuales vienen enunciados en los apéndices al final de la tesis. El contenido de cada capítulo es el siguiente:

En el Capítulo 1 se introducen las definiciones básicas, y ejemplos que usaremos a lo largo del trabajo, así como algunas propiedades de las transformaciones preservadoras de medida. Éstas son las siguientes: recurrencia, ergodicidad, débilmente mezclante y fuertemente mezclante. Se probará el Teorema de Recurrencia de Poincaré, el Teorema Ergódico de von Neumann, el Teorema Ergódico de Birkhoff-Khintchine, así como algunos teoremas y aplicaciones que son consecuencia del último teorema. Será mostrada la transformación inducida de Kakutani. También se introduce un operador en $L_p(\mu)$ inducido por una transformación preservadora de medida. Daremos equivalencias de las propiedades mencionadas utilizando este operador. Se muestra una transformación que es ergódica pero no es débilmente mezclante. Al final se relacionan las propiedades de ergodicidad y mezcla con las del operador inducido en $L_2(\mu)$.

El Capítulo 2 aborda los conceptos de isomorfismo espectral e isomorfismo de transformaciones preservadoras de medida. Los cuales se utilizan para clasificar estas transformaciones. Por ejemplo, el isomorfismo espectral nos ayudará a relacionar las transformaciones vía el espectro del operador en $L_2(\mu)$ introducido en el capítulo anterior. Estos conceptos serán retomados en el capítulo 4 con la finalidad de clasificar a las transformaciones vía la entropía de éstas. Al final se verán ejemplos de transformaciones que son isomorfas.

El objetivo del Capítulo 3 es mostrar consecuencias de lo realizado en el primer capítulo para cadenas de Markov. Se construye un espacio de probabilidad asociado a ciertas cadenas. Se da una transformación en este espacio que resulta ser invariante bajo la medida definida. De esta manera relacionamos los conceptos del primer capítulo con la cadena. Se probará un teorema de convergencia exponencial que será de vital importancia para mostrar una aplicación de la Teoría Ergódica para la búsqueda de información en internet.

En el Capítulo 4 se introduce el concepto de entropía y algunas de sus propiedades son desarrolladas. Se prueba que ésta es una propiedad invariante bajo isomorfismo de transformaciones y se introducen algunos métodos para calcularla. Se calcula la entropía para algunos de nuestros ejemplos, y daré un ejemplo con entropía infinita. Se enuncia el **Teorema de Ornstein**, el cual no será probado en esta tesis. Este teorema dice que la entropía es una propiedad invariante completa para una clase de transformaciones. En la penúltima sección de este capítulo se dice cuál es esta clase. Se introduce otra clase de transformaciones que contienen a la anterior. Sin embargo, no se profundiza en estas clases. Al final aparece una breve introducción a la Teoría de la Información y algunas aplicaciones hacia esta teoría.

En el desarrolo de la tesis se utilizan resultados así como conceptos de Teoría de la Medida y Análisis. Los cuales vienen incluidos en los Apéndices al final de la tesis. Denotaré como λ a la medida de Lebesgue o una restricción de ésta.

Capítulo 1

Transformaciones preservadoras de medida

Considérese el experimento aleatorio de lanzar una moneda. El espacio muestral o espacio de estados de este experimento es el conjunto formado por las dos caras de la moneda, para simplificar ideas sea $X = \{0, 1\}$ donde 0 representa un lado de la moneda y 1 el otro lado. Si las condiciones en que se lanza la moneda no cambian en el tiempo, entonces el resultado de este experimento varía de acuerdo con leyes de probabilidad constante, es decir, la probabilidad de obtener cualquier cara de la moneda es constante en el tiempo. La Teoría Ergódica es clave para entender estas variaciones.

Supongamos que el experimento anterior se realiza una vez cada día desde el día de hoy (por ejemplo). Podemos pensar al experimento anterior como uno más grande donde cada realización de éste es representada por una sucesión infinita $x = (x(1), x(2), \dots)$, donde $x(i)$ representa el resultado del experimento en el día i con $i \in \mathbb{N}$, es decir, $x(i) \in X$ para toda $i \in \mathbb{N}$, la medida de probabilidad μ asociada a este experimento, es la obtenida por el método descrito en Halmos [11, págs. 154-160]. Para reflejar matemáticamente la idea de que el paso del tiempo no afecta las leyes probabilistas que rigen este experimento, introduciremos una transformación en el espacio muestral del experimento, es decir, en $X' = \{0, 1\}^{\mathbb{N}}$. Dada por $(T(x))_i = x_{i+1}$ con $i \in \mathbb{N}$, a esta T la podemos pensar como borrar la primera entrada del vector infinito x y poner lo que resta en su imagen, como si recorriéramos un lugar a la izquierda el vector x, además podemos pensar a $T(x)$ como si dejáramos pasar un día en una realización del experimento. En otras palabras $T(x)$ representa la historia de una realización del experimento a partir de mañana. Como $m(\{x\}) = 0$, entonces debemos fijarnos en subconjuntos de X', si $A \subset X'$, entonces los elementos de X' que se encuentren en A después de un día es el conjunto $T^{-1}(A)$ y si queremos que las leyes de probabilidad que rigen nuestro experimento sean constantes en el tiempo debemos tener que $\mu(A) = \mu(T^{-1}(A))$, es decir que la probabilidad de que $x \in A$ y la probabilidad de que $x \in T^{-1}(A)$ es la misma. Lo anterior nos lleva a estudiar las transformaciones que preservan una medida de probabilidad o más generalmente las que preservan una medida.

La Teoría Ergódica estudia las cuartetas $(X, \mathscr{F}, \mu, T)$, donde $(X, \mathscr{F}, \mu)$ es un espacio de medida y $T : X \to X$ es una transformación medible y que preserva la medida μ. Así como las funciones continuas son transformaciones que preservan la estructura de los espacios topológicos, las transformaciones que preservan la medida preservan la estructura del espacio de medida. Los sistemas preservadores de medida son utilizados comúnmente como modelos en procesos que involucran el tiempo y para los cuales las leyes que rigen la evolución en el tiempo no cambian. Dentro de las aplicaciones de esta teoría se encuentran las áreas de estadística, mecánica estadística, teoría de números, mecánica clásica y teoría de la información entre otras.

1.1. Definiciones y resultados básicos

Definición 1.1. Sean $(X, \mathscr{F})$ y $(X', \mathscr{F}')$ dos espacios medibles y $T : X \to X'$ una transformación:

(a) Decimos que T es **medible** (con respecto a $\mathscr{F}$ y $\mathscr{F}'$) si: $T^{-1}(A') \in \mathscr{F}$, para toda $A' \in \mathscr{F}'$ denotado por: $T^{-1}(\mathscr{F}') \subset \mathscr{F}$.

(b) Si además μ y μ' son medidas en $\mathscr{F}$ y $\mathscr{F}'$ (respectivamente) diremos que T **preserva las medidas** si T es medible y $\mu(T^{-1}(A')) = \mu'(A')$ para toda $A' \in \mathscr{F}'$ denotado por: $\mu \circ T^{-1} = \mu'$.

(c) Si T es invertible y T^{-1} preserva las medidas, diremos que T es una **transfomación preservadora de medidas invertible**.

Observaciones 1.1.

(1) Estaremos interesados en el caso $(X, \mathscr{F}, \mu) = (X', \mathscr{F}', \mu')$, si $T : X \to X$ preserva las medidas, diremos que T es una **transformación preservadora de medida** o T **preserva a** μ.

(2) Si T es una transformación preservadora de medida, entonces $T^n = T \circ \cdots \circ T$ (n veces) es preservadora de medida también para toda $n \geq 2$.

(3) Si T preserva las medidas entre dos espacios de medida, T^{-1} existe y es medible, entonces preserva las medidas también.

(4) Si $(X, \mathscr{F}, \mu) = (X', \mathscr{F}', \mu')$, entonces $T : X \to X$ dada por $T(x) = x$ es una transformación preservadora de medida.

(5) Considerando el espacio de medida $((0,1), \mathscr{B}_{(0,1)}, \lambda|_{\mathscr{B}_{(0,1)}})$ donde $\mathscr{B}_{(0,1)}$ es la σ-álgebra de los Borelianos del $(0,1)$ (ver Definición B.11, pág 131) y λ es la medida de Lebesgue (ver Definición B.22, pág 134), entonces la transformación $T(x) = x^2$ no es una transformación preservadora de medida; ya que $T^{-1}((0,1/2)) = (0, 1/\sqrt{2})$ y $\lambda((0,1/2)) = 1/2$ la cual es distinta a $\lambda((0, 1/\sqrt{2})) = 1/\sqrt{2}$.

La Observación 1.1.(2) dice que si T es una transformación preservadora de medida, entonces el conjunto $\{T^n : n \in \mathbb{N} \cup \{0\}\}$ donde T^0 es la función identidad en X, está formado por transformaciones preservadoras de medida, y también es un semigrupo con elemento neutro bajo la operación de composición de funciones. La Observación 1.1.(3) junto con la 1.1.(2) dice que si T^{-1} existe y es medible, entonces el conjunto $\{T^k : k \in \mathbb{Z}\}$, donde nuevamente T^0 es la función identidad en X, sigue estando formado por transformaciones preservadoras de medidas, es un grupo bajo la misma operación y que basta con pedir que T^{-1} sea medible para garantizar que T sea una transformación preservadora de medida invertible.

Más adelante veremos que no basta con pedir que T sea invertible para garantizar que T^{-1} sea medible. Ahora procedemos a probar las Observaciones 1.1.(2) y 1.1.(3), necesitaremos los siguientes lemas:

Lema 1.1. *Sea $T : X \to X$ una transformación entonces:*

$$(T^n)^{-1}(A) = \underbrace{T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots))}_{n-veces} \quad \textit{para toda } n \in \mathbb{N}, A \subseteq X.$$

Demostración. Por inducción a partir de $n = 2$, probaré que $(T^2)^{-1}(A) = T^{-1}(T^{-1}(A))$. Pero los siguientes enunciados son equivalentes:

(i) $x \in T^{-1}(T^{-1}(A))$.

(ii) $T(x) \in T^{-1}(A)$.

(iii) $T(T(x)) \in A$, $T^2(x) \in A$.

(iv) $x \in (T^2)^{-1}(A)$.

Suponemos válido para n, probaremos válido para $n+1$, es decir,

$$(T^{n+1})^{-1}(A) = \underbrace{T^{-1}(T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots)))}_{n+1-veces}, \text{ pero;}$$

los siguientes enunciados son equivalentes:

(i) $x \in \underbrace{T^{-1}(T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots)))}_{n+1-veces}$.

(ii) $T(x) \in \underbrace{T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots))}_{n-veces}$.

(iii) $T(x) \in (T^n)^{-1}(A)$ por hipótesis de inducción.

(iv) $T^n(T(x)) \in A$.

(v) $T^{n+1}(x) \in A$.

(vi) $x \in (T^{n+1})^{-1}(A)$. □

Lema 1.2. *Sea $T : X \to Y$ una transformación invertible y T^{-1} la transformación inversa de T, entonces $(T^{-1})^{-1}(A) = T(A)$ para toda $A \subseteq X$.*

Demostración. Los siguientes enunciados son equivalentes:

(i) $y \in (T^{-1})^{-1}(A)$.

(ii) $T^{-1}(y) \in A$ $T^{-1}(y) = a$ para algún $a \in A$.

(iii) $y = T(a)$ para algún $a \in A$.

(iv) $y \in T(A)$. □

Ahora podemos probar las Observaciones 1.12 y 1.1.3.

Demostración.

1.1.(2). Por el Lema 1.1 $(T^n)^{-1}(A) = T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots))$ pero $T^{-1}(A) \in \mathscr{F}$ pues T es medible entonces $T^{-1}(T^{-1}(A)) \in \mathscr{F}$ ya que T es medible y $T^{-1}(A) \in \mathscr{F}$, continuando con el razonamiento anterior podemos concluir que $T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots))$ pertenece a $\mathscr{F}$ por lo tanto $(T^n)^{-1}(A) \in \mathscr{F}$ para toda $A \in \mathscr{F}$, por lo tanto T^n es medible. Como T preserva la medida entonces $\mu(T^{-1}(A)) = \mu(A)$, por demostrar $\mu((T^n)^{-1}(A)) = \mu(A)$ pero por el Lema 1.1 se tienen las siguientes igualdades:

$$\begin{aligned} \mu((T^n)^{-1}(A)) &= \mu(T^{-1}(T^{-1}(\cdots T^{-1}(A)\cdots))) = \mu(T^{-1}(\cdots T^{-1}(A)\cdots)) \\ &= \cdots = \mu(A), \quad \text{por lo tanto } T^n \text{ preserva la medida.} \end{aligned}$$

1.1(3). Sean $(X, \mathscr{F}, \mu)$ y $(X', \mathscr{F}', \mu')$ dos espacios de medida, $A \in \mathscr{F}$, $T : X \to X'$ una transformación que cumpla con las hipótesis requeridas, entonces:

$$\mu(A) = \mu(T^{-1}(T(A))) = \mu'(T(A)) = \mu'((T^{-1})^{-1}(A)),$$

donde la primera igualdad se sigue de que T^{-1} es medible y el Lema 1.2 por lo que $T(A) \in \mathscr{F}'$; la segunda se obtiene de que T preserva las medidas; finalmente la última es consecuencia del Lema 1.2. Por lo tanto T^{-1} preserva las medidas. □

Observación 1.2. Nótese que **no** basta con que T^{-1} exista para poder garantizar la medibilidad de ésta. Considere $T : (\mathbb{R}, \mathscr{A}^*_{\mathbb{R}}, \lambda) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}}, \lambda|_{\mathscr{B}_{\mathbb{R}}})$ dada por $T(x) = x$, donde $\mathscr{B}_{\mathbb{R}}$ denota la σ-álgebra de Borel, $\mathscr{A}^*_{\mathbb{R}}$ es la σ-álgebra de los conjuntos Lebesgue-medibles y λ es la medida de Lebesgue (ver B.10, pág. 131 y B.22, pág. 134); entonces T es medible, T^{-1} existe, pero T^{-1} no es medible pues $\mathscr{B}_{\mathbb{R}} \subsetneq \mathscr{A}^*_{\mathbb{R}}$ (ver Bartle [3, pág. 154]).

Para poder probar que una transformación es medible a partir de la definición tendríamos que conocer los elementos de $T^{-1}(\mathscr{F}')$, lo cual pocas veces sucede. A continuación veremos que podemos probar la medibilidad mediante una colección de subconjuntos de X', $\mathbb{E}'$ tal que $\sigma(\mathbb{E}') = \mathscr{F}'$ (ver Definición B.3, pág. 130), lo que en muchas ocasiones simplificará la prueba de la medibilidad de T.

Teorema 1.1. *Sean $(X, \mathscr{F})$ y $(X', \mathscr{F}')$ dos espacios medibles, $T : X \to X'$ una transformación y $\mathbb{E}' \subset \mathscr{F}'$ una familia de subconjuntos de X' tal que la σ-álgebra generada por $\mathbb{E}'$ coincide con $\mathscr{F}'$[1] (denotado por $\sigma(\mathbb{E}') = \mathscr{F}'$), entonces: T es medible (con respecto a $\mathscr{F}$ y $\mathscr{F}'$) si y sólo si $T^{-1}(E') \in \mathscr{F}$, para todo $E' \in \mathbb{E}'$.*

Demostración. Claramente basta probar la (Suficiencia).

Sea $\mathscr{C}' = \{A' \subset X' : T^{-1}(A') \in \mathscr{F}\}$, entonces por propiedades de la imagen inversa se comprueba fácilmente que $\mathscr{C}'$ es una σ-álgebra que contiene a $\mathbb{E}'$, por lo que $\mathbb{E}' \subset \mathscr{F}' \subset \mathscr{C}'$, de donde obtenemos $T^{-1}(\mathscr{F}') \subset T^{-1}(\mathscr{C}') \subset \mathscr{F}$. □

Sin embargo, la medibilidad de T no es la única que nos interesa que también preserve las medidas, para ello necesitaremos los siguientes conceptos (debidos a Dynkin):

Definición 1.2. Sea X un conjunto no vacío, $\mathcal{C}, \mathcal{L} \subset \mathcal{P}(X)$ familias no vacías de subconjuntos de X.

(a) Decimos que $\mathcal{C}$ es un π**-sistema** si: $A \cap B \in \mathcal{C}$ para todo A, B en $\mathcal{C}$ (es decir $\mathcal{C}$ es cerrado bajo intersecciones finitas).

(b) Decimos que $\mathcal{L}$ es un λ**-sistema** si:

  (1) $X \in \mathcal{L}$.

  (2) Si $A, B \in \mathcal{L}$ y $A \subseteq B$, entonces $B \setminus A \in \mathcal{L}$.

  (3) Si (A_n) es una sucesión creciente de elementos de $\mathcal{L}$ (es decir: $A_n \subset A_{n+1}$ para toda n), entonces: $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{L}$.

Observación 1.3. La intersección no vacía de π-sistemas es un π-sistema.

Demostración. Sean $X \neq \emptyset$, $\mathscr{C} = \{\mathcal{C} : \mathcal{C} \subset \mathcal{P}(X), \mathcal{C}$ es un π-sistema$\}$, tal que $\bigcap_{\mathcal{C} \in \mathscr{C}} \mathcal{C} \neq \emptyset$. Sean $\mathcal{C}' = \bigcap_{\mathcal{C} \in \mathscr{C}} \mathcal{C}$ y $A, B \in \mathcal{C}'$, entonces $A \in \mathcal{C}$ para todo $\mathcal{C} \in \mathscr{C}$ y $B \in \mathcal{C}$ para todo $\mathcal{C} \in \mathscr{C}$. Por lo que $A \cap B \in \mathcal{C}$ para todo $\mathcal{C} \in \mathscr{C}$ ya que para todo $\mathcal{C} \in \mathscr{C}$, $\mathcal{C}$ es un π-sistema. Por lo tanto $\mathcal{C}'$ es un π-sistema. □

¿Qué tienen que ver estos conceptos con el de σ-álgebra, que es el que nos interesa? La siguiente observación muestra su relación.

[1]$\mathbb{E}'$ se llama un **generador** de la σ-álgebra $\mathscr{F}'$.

Observación 1.4. Una familia $\mathscr{F} \subset \mathcal{P}(X)$ es una σ-álgebra si y sólo si $\mathscr{F}$ es un π-sistema y un λ-sistema (esto será de utilidad en el siguiente teorema).

Demostración.
(Necesidad). Como $\mathscr{F}$ es una σ-álgebra entonces es un λ-sistema por definición, por lo que basta probar que $\mathscr{F}$ es un π-sistema. Sean $A, B \in \mathscr{F}$, entonces $(A \setminus B), (B \setminus A) \in \mathscr{F}$ así que $(A \setminus B) \cup (B \setminus A) \in \mathscr{F}$ por lo que $A \bigtriangleup B \in \mathscr{F}$. Pero: $A \cap B = (A \cup B) \setminus (A \bigtriangleup B)$ por lo tanto $\mathscr{F}$ es un π-sistema.

(Suficiencia). Se prueban las tres partes de la Definición 1.2(b):

(1) $X \in \mathscr{F}$ por definición de λ-sistema.

(2) Sean $A, B \in \mathscr{F}$, entonces $A \cap B \in \mathscr{F}$ por ser π-sistema, así que $A \setminus (A \cap B) \in \mathscr{F}$ por ser λ-sistema, recordando que $A \setminus B = A \setminus (A \cap B)$ se tiene el resultado.

(3) Basta con probar que $\mathscr{F}$ es cerrada bajo uniones finitas. Pero:

$$A \cup B = X \setminus ((X \setminus A) \cap (X \setminus B)),$$

donde $X \setminus A$ y $X \setminus B$ están en $\mathscr{F}$ por ser λ-sistema la intersección de ambos está en $\mathscr{F}$ por ser π-sistema lo que nos da el resultado, observe que puede extenderse a una unión finita arbitraria de elementos de $\mathscr{F}$. Si (E_n) es una sucesión de elementos de $\mathscr{F}$ consideramos $G_1 = E_1$, $G_2 = E_1 \cup E_2, \dots G_k = \bigcup_{n=1}^{k} E_n, \dots$. Por lo ya probado $G_k \in \mathscr{F}$ para toda $k \in \mathbb{N}$ y (G_k) es una sucesión creciente tal que $\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} G_n \in \mathscr{F}$ por ser λ-sistema. Por lo tanto $\mathscr{F}$ es una σ-álgebra. $\square$

El siguiente resultado será de utilidad para esta sección y en el de la Teoría de la Probabilidad. Es conocido como el "Lema de las Clases Monótonas" (L.C.M.), y relaciona la σ-álgebra generada por un π-sistema y un λ-sistema que contiene a este π-sistema.

Teorema 1.2 (L.C.M.)**.** *Sea $\mathcal{C} \subset \mathcal{P}(X)$ un π-sistema y $\mathcal{L} \subset \mathcal{P}(X)$ un λ-sistema tal que $\mathcal{C} \subset \mathcal{L}$, entonces $\sigma(\mathcal{C}) \subset \mathcal{L}$.*

Demostración. Sea $\mathcal{T}$ la clase de todos los λ-sistemas que contienen a $\mathcal{C}$; consideramos $\mathcal{L}(\mathcal{C}) = \bigcap\{\mathcal{L} : \mathcal{L} \in \mathcal{T}\}$. Note que la intersección es sobre una familia diferente del vacío (ya que $\mathcal{P}(X) \in \mathcal{T}$), y que $\mathcal{L}(\mathcal{C})$ es un λ-sistema que contiene a $\mathcal{C}$.

Afirmación. $\mathcal{L}(\mathcal{C})$ es un π-sistema.

Para $E \in \mathcal{C}$ fijo definimos $\mathcal{L}_E = \{F \subset X : E \cap F \in \mathcal{L}(\mathcal{C})\}$. Como $\mathcal{C}$ es un π-sistema tenemos que $\mathcal{C} \subset \mathcal{L}_E$; además es fácil verificar que $\mathcal{L}_E$ es un λ-sistema, así pues $\mathcal{L}_E \in \mathcal{T}$, por lo que $\mathcal{L}(\mathcal{C}) \subset \mathcal{L}_E$. Ahora sea $F \in \mathcal{L}(\mathcal{C})$ fijo y definimos $\mathcal{L}'_F = \{E \subset X : E \cap F \in \mathcal{L}(\mathcal{C})\}$. Por lo ya probado: $E \cap F \in \mathcal{L}(\mathcal{C})$ para todo $E \in \mathcal{C}$ y para toda $F \in \mathcal{L}_E$, en particular $E \cap F \in \mathcal{L}(\mathcal{C})$ para toda $E \in \mathcal{C}$ y para toda $F \in \mathcal{L}(\mathcal{C})$, es decir, $\mathcal{C} \subset \mathcal{L}'_F$. Nuevamente es sencillo comprobar que $\mathcal{L}'_F$ es un λ-sistema para toda $F \in \mathcal{L}(\mathcal{C})$. Así que: $\mathcal{L}'_F \in \mathcal{T}$, por lo que $\mathcal{L}(\mathcal{C}) \subset \mathcal{L}'_F$ para toda $F \in \mathcal{L}(\mathcal{C})$. Lo anterior dice en particular que $E \cap F \in \mathcal{L}(\mathcal{C})$ para todo $E, F \in \mathcal{L}(\mathcal{C})$. Por lo tanto $\mathcal{L}(\mathcal{C})$ es un π-sistema y como además es un λ-sistema que contiene a $\mathcal{C}$, es una σ-álgebra que contiene a $\mathcal{C}$, de donde concluimos que $\sigma(\mathcal{C}) \subset \mathcal{L}(\mathcal{C}) \subset \mathcal{L}$. □

Utilizando los resultados anteriores podremos mostrar un criterio para probar que una transformación T preserva las medidas, utilizando un π-sistema que genere a $\mathscr{F}$. Este criterio será utilizado frecuentemente durante este trabajo.

Teorema 1.3. *Sean $(X, \mathscr{F}, \mu)$ y $(X', \mathscr{F}', \mu')$ espacios de medida con $\mu(X) = \mu'(X') < \infty$. Sea $T : X \to X'$ medible. Supongamos que $\mathcal{C}' \subset \mathscr{F}'$ es un π-sistema que genera a $\mathscr{F}'$ (o $\sigma(\mathcal{C}') \supset \mathscr{F}'$) y tal que $\mu(T^{-1}(E')) = \mu'(E')$ para toda $E' \in \mathcal{C}'$, entonces T preserva las medidas.*

Demostración. Sea $\mathcal{L}' = \{E' \subset \mathscr{F}' : \mu(T^{-1}(E')) = \mu'(E')\}$. Por hipótesis $\mathcal{C}' \subset \mathcal{L}' \subset \mathscr{F}'$ y en virtud del Teorema 1.2 será suficiente verificar que $\mathcal{L}'$ es un λ-sistema.

(1) $X' \in \mathcal{L}'$ por hipótesis.

(2) Sean $E', F' \in \mathcal{L}'$ con $E' \subset F'$, entonces como las medidas son **finitas** obtenemos:

$$\begin{aligned}
\mu\left(T^{-1}\left(F' \setminus E'\right)\right) = \mu\left(T^{-1}(F') \setminus T^{-1}(E')\right) &= \mu(T^{-1}(F')) - \mu(T^{-1}(E')) \\
&= \mu'(F') - \mu'(E') \\
&= \mu'\left(F' \setminus E'\right).
\end{aligned}$$

Así que $F' \setminus E' \in \mathcal{L}'$.

(3) Sea (E'_n) una sucesión creciente de elementos de $\mathcal{L}'$ entonces:

$$\mu\left(T^{-1}\left(\bigcup_{n=1}^{\infty} E'_n\right)\right) = \mu\left(\bigcup_{n=1}^{\infty} T^{-1}(E'_n)\right) = \lim_{n \nearrow \infty} \mu(T^{-1}(E'_n)) = \lim_{n \nearrow \infty} \mu'(E'_n) = \mu'\left(\bigcup_{n=1}^{\infty} E'_n\right),$$

por lo que $\bigcup_{n=1}^{\infty} E'_n \in \mathcal{L}'$. □

Observación 1.5. Para probar que T preserva las medidas tenemos que probar que T es medible y $\mu(T^{-1}(E')) = \mu'(E')$ para toda $E' \in \mathscr{F}'$. Notemos que con el Teorema 1.1 y el teorema anterior podemos probar las dos condiciones al mismo tiempo (basta con probar ambas condiciones en un π-sistema indicado).

Como podemos apreciar el teorema anterior nos es de utilidad sólo cuando los espacios de medida son finitos. El siguiente teorema nos dará una extensión en el caso de que es espacio de medida sea σ-finito, sin embargo necesitaremos hipótesis más fuertes.

Teorema 1.4. *Sean $(X, \mathscr{F}, \mu)$, $(Y, \mathscr{G}, \nu)$ dos espacios de medida σ-finitos, $T : X \to Y$ medible con respecto a las σ-álgebras correspondientes. Si existe $\mathcal{A} \subset \mathscr{G}$ un álgebra tal que $\sigma(\mathcal{A}) = \mathscr{G}$ y $\mu(T^{-1}(A)) = \nu(A)$ para todo $A \in \mathcal{A}$, entonces T preserva las medidas.*

Demostración. Definimos $\nu_0 : \mathscr{G} \to \mathbb{R} \cup \{\infty\}$ como sigue $\nu_0(G) = \mu(T^{-1}(G))$ para todo $G \in \mathscr{G}$, claramente ν_0 es una medida tal que $\nu_0|_{\mathcal{A}} = \nu|_{\mathcal{A}}$, así que ν_0 es una extensión a $\mathscr{G}$ de $\nu|_{\mathcal{A}}$. Como la extensión es única se tiene el resultado (ver Teorema B.8, pág. 134). $\square$

1.2. Ejemplos de Transformaciones Preservadoras de Medida

En esta sección introduciremos ejemplos de transformaciones preservadoras de medida que estaremos utilizando durante el desarrollo de este trabajo, por lo que estaremos refiriéndonos a estos constantemente. Algunos de ellos nos ayudarán para dar contra-ejemplos de teoremas que se verán más adelante.

Ejemplo 1.1. Para $m \in \mathbb{N}$ con $m > 1$ fijo, sea $X = [\,0\,,1\,)$, $\mathscr{F} = \mathscr{B}_{[\,0\,,1\,)}$ y μ la medida de Lebesgue. Definimos $T : X \to X$ como sigue:

$$T(x) = \begin{cases} mx & \text{si} \quad x \in \left[0\,, \dfrac{1}{m}\right) \\[2ex] mx - 1 & \text{si} \quad x \in \left[\dfrac{1}{m}\,, \dfrac{2}{m}\right) \\ \vdots & \vdots \quad \vdots \\ mx - m + 1 & \text{si} \quad x \in \left[\dfrac{m-1}{m}\,, 1\right) \end{cases}$$

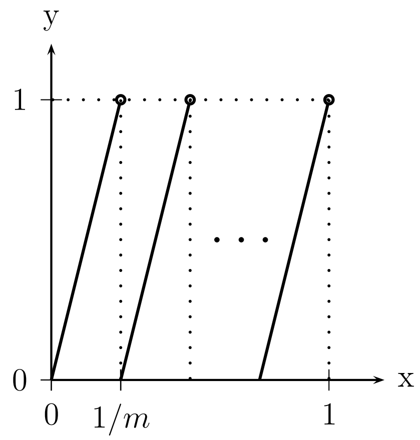

Figura 1.1: Gráfica de $T(x) = mx$ (mod 1).

Es decir, $T(x) = mx$ (mod 1), comprobaremos que T es medible y que preserva a μ:

Sea $\mathcal{C} = \big\{[\,a\,, b\,) \,:\, 0 \le a < b < 1\big\}$. Claramente $\mathcal{C}$ es un π-sistema que genera a $\mathscr{F}$, también $T^{-1}\left([\,a\,, b\,)\right) = \bigcup_{i=0}^{m-1} [\,(a+i)/m\,, (b+i)/m\,) \in \mathscr{F}$ y además $\mu(T^{-1}\left([\,a\,, b\,)\right)) = b - a$, así que por el Teorema 1.1 y el Teorema 1.3, T es medible y preserva a μ. Observe que T es esencialmente m a 1 (es decir todo punto de X posee m preimágenes) por lo tanto T no es invertible.

Ejemplo 1.2. Para $\alpha \in (0, 1)$ fijo, sea $X = [0, 1)$, $\mathscr{F} = \mathscr{B}_{[0,1)}$ y $\mu = \lambda$. Definimos $T : X \to X$ dada por $T(x) = x + \alpha$ (mod 1), veamos la gráfica de T;

$$T(x) = \begin{cases} x+\alpha & \text{si } x \in [\,0\,,1-\alpha\,) \\ \\ x+\alpha-1 & \text{si } x \in [\,1-\alpha\,,1\,) \end{cases}$$

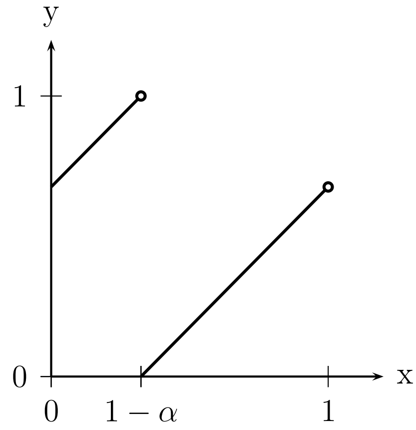

Figura 1.2: Gráfica de $T(x) = x + \alpha$ (mod 1).

Entonces:

$$T^{-1}([a,b)) = \begin{cases} [a-\alpha, b-\alpha) & \text{si} \quad \alpha \le a \\ [a-\alpha+1, 1) \cup [0, b-\alpha) & \text{si} \quad a < \alpha < b \\ [a-\alpha+1, b-\alpha+1) & \text{si} \quad b \le \alpha. \end{cases}$$

Por lo que T es medible y preserva a μ.

Ejemplo 1.3. Sea $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ la circunferencia unitaria, $\mathscr{F}$ la σ-álgebra generada por los arcos y $\mu : \mathscr{F} \to [\,0\,,1\,)$ definida en los arcos como: $\mu(\text{arco}) = 1/(2\pi)$(longitud de arco) y extendida a los elementos de $\mathscr{F}$ de acuerdo al procedimiento de extensión de Hahn-Caratheodory (ver Teorema B.7, pág. 134). Para $n \in \mathbb{N}$ fija, definimos $T : S^1 \to S^1$ poniendo $T(z) = z^n$, entonces T es medible y preserva a μ. Denotaremos $S^1 = \{e^{i\theta} : \theta \in [\,0\,,2\pi\,)\}$ y $\mathcal{C} = \Big\{ \{e^{ix} : \alpha \le x < \beta\} \,:\, 0 \le \alpha \le \beta \le 2\pi \Big\}$, entonces $\mathcal{C}$ es un π-sistema que genera a $\mathscr{F}$. Claramente:

$$T^{-1}\Big(\{e^{ix} : \alpha \le x < \beta\} \Big) = \bigcup_{k=0}^{n-1} \left\{ e^{ix} : \frac{\alpha + 2\pi k}{n} \le x < \frac{\beta + 2\pi k}{n} \right\} \quad \text{(disjunta)}.$$

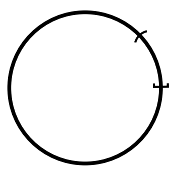

Figura 1.3: $\left\{e^{ix} : 0 \le x < \frac{\pi}{4}\right\}$.

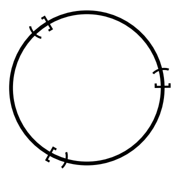

Figura 1.4: $T^{-1}\left(\left\{e^{ix} : 0 \le x < \frac{\pi}{4}\right\}\right)$ y $n = 3$.

De donde es claro que T es medible y preserva a μ. T es n a 1 por lo que es invertible si y sólo si $n = 1$.

Ejemplo 1.4 (Rotaciones en S^1)**.** Sea $(X, \mathscr{F}, \mu)$ el espacio de medida del ejemplo anterior. Para $(\,\rho \in [\,0\,,2\pi\,))$ fijo definimos una transformación $T : S^1 \to S^1$ poniendo $T(z) = az$ donde $a = e^{i\rho}$, entonces:

$$T^{-1}\left\{e^{ix}:\alpha\le x<\beta\right\}=\begin{cases}\left\{e^{it}:\alpha-\rho\le t<\beta-\rho\right\} & \text{si } \rho\le\alpha\\[2ex] \left\{e^{it}:0\le t<\beta-\rho\right\}\cup\left\{e^{it}:2\pi+\alpha-\rho\le t<2\pi\right\} & \text{si } \alpha<\rho<\beta\\[2ex] \left\{e^{it}:2\pi+\alpha-\rho\le t<2\pi+\beta-\rho\right\} & \text{si } \beta\le\rho<2\pi\end{cases}$$

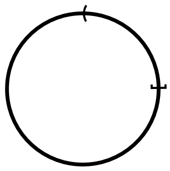

Figura 1.5: $\left\{e^{ix}:0\le x<\dfrac{\pi}{2}\right\}$

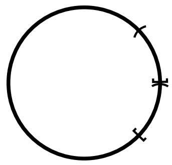

Figura 1.6: $T^{-1}\left(\left\{e^{ix}:0\le x<\dfrac{\pi}{2}\right\}\right)$ y $\rho=\dfrac{\pi}{4}$.

Por lo que T es medible y preservadora de medida (T se llama la rotación en S^1 a través del ángulo ρ). Claramente T es invertible con inversa medible y preservadora de medida.

Ejemplo 1.5. Sea $X=\mathbb{R}^n$, $\mathscr{F}=\mathscr{B}_{\mathbb{R}^n}$ y $\mu=$ medida de Lebesgue sobre $\mathscr{F}$. Sea $T:\mathbb{R}^n\to\mathbb{R}^n$ una transformación afín $T(\vec{x})=A(\vec{x})+\vec{b}$ ($\vec{b}\in\mathbb{R}^n$ fijo) con $A=(a_{ij})$ la matriz asociada a T (con respecto a la base canónica de $\mathbb{R}^n$). Así que, si $\vec{y}=T(\vec{x})$, entonces $y_i=\sum_{j=1}^n a_{ij}x_j+b_i$ para toda $i=1,\dots,n$.

Caso 1) A es una matriz invertible, entonces T puede ser escrita como composición de las siguientes transformaciones elementales invertibles (ver Friedberg [8, pág. 139]):

(1) $y_i=x_i+b_i$ para toda i (traslación).

(2) $y_i=cx_i$ para i fijo y $c\neq 0$.

(3) $y_i=x_i$ para toda $i\neq j$ y $y_j=x_j+cx_k$ ($j\neq k, c\in\mathbb{R}$).

(4) $y_i=x_i$ si $i\neq j$, $i\neq k$ y $y_j=x_k$, $y_k=x_j$.

Sea $E=\{(x_1,\dots,x_n)\in\mathbb{R}^n:\alpha_i\le x_i<\beta_i\}$, entonces $T^{-1}(E)$ es otra celda de igual volumen si T es del tipo (1) o (4). Si T es del tipo (2) entonces $T^{-1}(E)$ es otra celda cuyo volumen es igual a $|\,c\,|^{-1}$ veces el volumen de E. Si T es del tipo (3), entonces T^{-1} se define por: $x_i=y_i$ para toda $i\neq j$ y $x_j=y_j-cy_k$ ($j\neq k, c\in\mathbb{R}$), por lo que $T^{-1}(E)$ es un paralelogramo n-dimensional que tiene el mismo volumen que E (por la fórmula de cambio de variable en $\mathbb{R}^n$, en este caso el Jacobiano tiene valor absoluto igual a 1). Como T general es la composición de los casos considerados, y la colección de todas las celdas es un π-sistema que genera a $\mathscr{B}_{\mathbb{R}^n}$, obtenemos que T es medible y preserva a μ si y sólo si en la descomposición de T no aparecen transformaciones elementales del tipo (2) o bien $c=\pm 1$.

Caso 2) A no es invertible, entonces $T(\mathbb{R}^n)$ es un hiperplano de dimensión $< n$, por lo que es Borel medible y $\mu(T(\mathbb{R}^n)) = 0$. Sin embargo $\mathbb{R}^n = T^{-1}(T(\mathbb{R}^n))$ y $\mu(\mathbb{R}^n) = \infty$, por lo que T **no** puede preservar a μ. No obstante T es siempre medible lo cual puede establecerse como en el caso anterior.

Recordando la fórmula de cambio de variable para T no singular obtenemos:

$$\mu(T^{-1}(E)) = \frac{\mu(E)}{|\det A\,|}, \quad \text{así que: } T \text{ preserva a } \mu \text{ si y sólo si } \det A = \pm 1.$$

Ejemplo 1.6 (Transformación Producto)**.** Sean $(X, \mathscr{F}, \mu)$ y $(Y, \mathscr{G}, \nu)$ dos espacios de medida finita, y $U : X \to X$ y $T : Y \to Y$ dos transformaciones, entonces:

(1) Si U y T son medibles (en sus respectivos espacios), entonces $U \times T : X \times Y \to X \times Y$ dada por $U \times T\,(x, y) = (U(x), T(y))$ es medible con respecto a $\mathscr{F} \otimes \mathscr{G}$ (ver Definición B.27, pág. 135).

(2) Si U y T preservan sus respectivas medidas, entonces $U \times T$ preserva a $\mu \otimes \nu$ (ver Teorema B.11, pág. 136).

Demostración. Sea $\mathcal{C} = \{A \times B : A \in \mathscr{F} \text{ y } B \in \mathscr{G}\}$ la familia de rectángulos medibles contenidos en $X \times Y$. Claramente $\mathcal{C}$ es un π-sistema y $\sigma(\mathcal{C}) = \mathscr{F} \otimes \mathscr{G}$. Por definición de $U \times T$ es claro que:

$$(U \times T)^{-1}(A \times B) = U^{-1}(A) \times T^{-1}(B) \in \mathscr{F} \otimes \mathscr{G} \quad \text{por lo tanto } U \times T \text{ es medible,}$$

(si Uy T lo son). Si U y T preservan y son medibles entonces:

$$\begin{aligned} \mu \otimes \nu\left((U \times T)^{-1}(A \times B)\right) = \mu \otimes \nu\left(U^{-1}(A) \times T^{-1}(B)\right) &= \mu\left(U^{-1}(A)\right)\nu\left(T^{-1}(B)\right) \\ &= \mu \otimes \nu(A \times B). \end{aligned}$$

Por lo tanto $U \times T$ preserva a $\mu \otimes \nu$. □

Ejemplo 1.7. Para $(k > 1)$ fijo, sea $E = \{0, 1, \ldots, k-1\}$ definimos una medida sobre $\mathcal{P}(E)$ poniendo:

$$\pi(\{i\}) = p_i \ \text{ con } \ p_i \geq 0 \ (i = 0, \ldots, k-1), \ \sum_{i=0}^{k-1} p_i = 1\,; \ \text{ luego } \ \pi(A) = \sum_{i \in A} p_i \quad \text{para toda } A \subset E.$$

Sea $X = \prod_{-\infty}^{\infty} E = \{0, 1, \ldots, k-1\}^{\mathbb{Z}}$. X también puede verse como el espacio de todas la funciones $x : \mathbb{Z} \to \{0, 1, \ldots, k-1\}$ e identificamos $x \in X$ con $(\ldots, x_{-1}, x_0, x_1, \ldots)$. Definimos $T : X \to X$ poniendo $T(x) = y$, donde $y_i = x_{i+1}$, $i \in \mathbb{Z}$. Gráficamente:

A continuación definimos una σ-álgebra de subconjuntos de X y una medida. Sean $n_1 < n_2 < \ldots < n_r \in \mathbb{Z}$ y $a_i, \ldots, a_r \in E$ dados, denotamos:

$$Z\,(n_1, \ldots, n_r; a_1, \ldots, a_r) = \{x \in X : x_{n_1} = a_1, \ldots, x_{n_r} = a_r\} \quad (r \in \mathbb{N}),$$

$$\begin{array}{ccccccccccc} & & & & & \boxed{\text{lugar } 0} & & & & & \\ x & = & (\ldots & x_{-2}, & x_{-1}, & \boxed{x_0}, & x_1, & x_2, & \ldots) & & \\ & & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & & & \downarrow T \\ y & = & (\ldots & x_{-1}, & x_0, & x_1, & x_2, & x_3, & \ldots) & & \end{array}$$

Figura 1.7: Es decir, T recorre a x un lugar hacia la izquierda y claramente es invertible (con $T^{-1}(x) = y$ donde $y_i = x_{i-1}$, $i \in \mathbb{Z}$).

y lo llamamos un **cilindro** y definimos a $\mathscr{F}$ como la σ-álgebra generada por todos los cilindros. Definimos

$$\mu\left(Z\left(n_1, \ldots, n_r; a_1, \ldots, a_r\right)\right) = p_{a_1} p_{a_2} \cdots p_{a_r},$$

y extendemos a todo elemento de $\mathscr{F}$ para obtener el espacio de probabilidad $(X, \mathscr{F}, \mu)$ [la existencia de dicho espacio queda garantizada por el Teorema de consistencia de Kolmogorov (ver Halmos [11, págs. 154-160])] y el espacio en cuestión no es más que el producto infinito $\prod_{-\infty}^{\infty}(E, \mathcal{P}(E), \pi))$. Probaremos que T es medible y preserva a μ. Sea $\mathcal{C} = \{\emptyset\} \cup \{cilindros\}$ entonces por definición $\mathscr{F} = \sigma(\mathcal{C})$ y es claro que es un π-sistema, además:

$$T^{-1}\left(Z\left(n_1, \ldots, n_r; a_1, \ldots, a_r\right)\right) = Z\left(n_1 + 1, \ldots, n_r + 1; a_1, \ldots, a_r\right) \in \mathcal{C},$$

lo que prueba simultáneamente que T es medible y preservadora de medida.[2]

Definición 1.3. Llamamos a T el **corrimiento bilateral de Bernoulli** con parámetros $(p_0, p_1, \ldots, p_{k-1})$.

Observación 1.6 (Condiciones de Kolmogorov). La medida que dimos para los cilindros satisface:

(1) $\mu(Z(n_1, \ldots, n_m; a_1, \ldots, a_m)) \geq 0$.

(2) $\sum_{a \in E} \mu(Z(n_1, \ldots, n_m, n_{m+1}; a_1, \ldots, a_m, a)) = \mu(Z(n_1, \ldots, n_m; a_1, \ldots, a_m))$.

(3) $\sum_{a \in E} \mu(Z(n_1; a)) = 1$.

Las cuales corresponden a las condiciones de consistencia de Kolmogorov en nuestro contexto (ver Shiryaev [21, pág. 163]).

Demostración.

(1) Esto se debe a que $p_i \geq 0$ para toda $i \in E$.

(2) $\sum_{a \in E} \mu(Z(n_1, \ldots, n_m, n_{m+1}; a_1, \ldots, a_m, a)) = \sum_{a \in E} p_{a_1} \cdots p_{a_m} p_a$ debido a que $\sum_{a \in E} p_a = 1$, se obtiene el resultado.

(3) $\sum_{a \in E} \mu(Z(n_1; a)) = \sum_{a \in E} p_a = 1$ por definición. $\square$

[2]Si los enteros $n_1 < n_2 < \ldots < n_r$ son consecutivos, entonces $Z\left(n_1, \ldots, n_r; a_1, \ldots, a_r\right)$ se llama un **cilindro flaco**. Observe que todo cilindro es la unión finita y disjunta de cilindros flacos. Por lo que la familia de cilindros flacos también genera a $\mathscr{F}$.

Ejemplo 1.8. Modificando el espacio anterior definimos ahora: $X = \prod_{0}^{\infty} E = \{0, 1, \ldots, k-1\}^{\mathbb{N}}$ y $\mathscr{F}$ = σ-álgebra generada por los cilindros, y μ la medida producto (como antes se definió). Finalmente ponemos: $T : X \to X$ mediante la fórmula: $(T(x))_i = x_{i+1}$ $i = 0, 1, 2 \ldots$. Observe que T no es invertible (de hecho todo punto de X tiene exactamente k preimágenes). Es fácil ver que T es medible y preservadora de medida.

Definición 1.4. Llamamos a T el **corrimiento unilateral de Bernoulli** con parámetros $(p_0, p_1, \ldots, p_{k-1})$.

Ejemplo 1.9. Sea $(X\mathscr{F}, \mu)$ el espacio de medida correspondiente al corrimiento unilateral de Bernouilli con parámetros $(p_0, \ldots, p_{k-1})$ y $\tau : \mathbb{N} \to \mathbb{N}$ una permutación fija. Definimos $T_\tau : X \to X$ como sigue $(T_\tau(x))_i = x_{\tau(i)}$. Entonces T es preservadora de medida.

Demostración. Por el Teorema 1.3, basta mostrar el resultado para los elementos del π-sistema formado por los cilindros. Pero:

$$T^{-1}(Z(n_1, \ldots, n_r; a_1, \ldots, a_r)) = Z(\tau^{-1}(n_1), \ldots, \tau^{-1}(n_r); a_1, \ldots, a_r),$$

como estos dos cilindros tienen la misma medida se tiene el resultado. □

Ejemplo 1.10 (Rotaciones en grupos localmente compactos)**.** Sea $(G, \cdot)$ un grupo localmente compacto (ver Halmos [11, págs. 216-249]). El Teorema de Alfréd Haar garantiza la existencia de una medida regular m definida sobre los Borelianos de G (denotados $\mathscr{B}_G$), y que además satisface que $m(a \cdot \mathcal{U}) = m(\mathcal{U})$ para todo $\mathcal{U}$ abierto y para toda $a \in G$, donde $a \cdot \mathcal{U} = \{a \cdot x : x \in \mathcal{U}\}$ (es decir m es invariante bajo traslaciones por la izquierda) (ver Halmos [11, págs. 250-265]). Así que si definimos $T : G \to G$ por medio de la fórmula $T(x) = a \cdot x$, entonces T es medible y preserva a m. (Para verificar la medibilidad observe que $\mathcal{U} \subset G$ es abierto si y sólo si $a \cdot \mathcal{U}$ es abierto y se sigue para $E \in \mathscr{B}_G$ general. Para la preservación de medida, basta usar la regularidad de m). El ejemplo (1.3) es un caso especial de éste.

1.3. Operador Inducido en L_p

Todo espacio de medida $(X, \mathscr{F}, \mu)$ tiene asociado los espacios de Banach

$$L_p^{\mathbb{C}}(\mu) = \{f : X \to \mathbb{C} \,|\, f \text{ es } \mathscr{F}\text{-medible y } \int |f|^p d\,\mu < \infty\} \text{ con } (p \geq 1) \text{ o}$$

$$L_p^{\mathbb{R}}(\mu) = \{f : X \to \mathbb{R} \,|\, f \text{ es } \mathscr{F}\text{-medible y } \int |f|^p d\,\mu < \infty\} \text{ con } (p \geq 1),$$

(ver Definición B.24, pág. 135), estos espacios son de utilidad en Teoría de la Medida. En esta sección se consideran dos espacios de medida $(X_i, \mathscr{F}_i, \mu_i)$ $i = 1, 2$ y para todo $p \geq 1$ se define un operador lineal de $L_p^{\mathbb{C}}(\mu_2)$ en $L_p^{\mathbb{C}}(\mu_1)$ inducido por una transformación que preserve las medidas. Se sabe que $L_p^{\mathbb{C}}(\mu)$ es un espacio de Hilbert (ver Definición A.4, pág. 128) si y sólo si $p=2$. El operador lineal de $L_2^{\mathbb{C}}(\mu_2)$ en $L_p^{\mathbb{C}}(\mu_1)$ nos llevará al estudio de las transformaciones preservadoras de medida vía el "espectro" de este operador, además será de utilidad para definir uno de los conceptos de "isomorfismo" entre transformaciones preservadoras de medida. En esta sección utilizaré los siguientes conceptos de análisis funcional:

Definición 1.5. (Norma de un Operador lineal) Sean $(X, \|\cdot\|_1)$ y $(Y, \|\cdot\|_2)$ dos espacios vectoriales normados y $\mathcal{U} : X \to Y$ un operador lineal, la **norma** de $\mathcal{U}$, denotada como $\|\mathcal{U}\|$, está dada por:

$$\|\mathcal{U}\| = \sup_{\|f\|_1=1} \|\mathcal{U}(f)\|_2.$$

$\mathcal{U}$ se llama **acotado**, si $\|\mathcal{U}\| < \infty$; $\mathcal{U}$ se llama **contracción**, si $\|\mathcal{U}\| \leq 1$; y $\mathcal{U}$ se llama **isometría lineal** , si $\|\mathcal{U}(f)\|_2 = \|f\|_1$ para toda $f \in X$. Si X, Y son espacios vectoriales sobre $\mathbb{R}$. $\mathcal{U}$ se llama **operador positivo** , si $\mathcal{U}(f) \geq 0$ para toda $f \in X$ tal que $f \geq 0$.

Definición 1.6. Sean $(X_i, \mathscr{F}_i, \mu_i)$ espacios de medida con $i = 1, 2$, $T : X_1 \to X_2$ una trasformación que preserva las medidas. Para $p \geq 1$ fijo definimos $\mathcal{U}_T : L_p^{\mathbb{C}}(\mu_2) \to L_p^{\mathbb{C}}(\mu_1)$ dada por $\mathcal{U}_T(f) = f \circ T$, lo llamamos **el operador inducido por un transformación que preserva las medidas**. Del mismo modo se definen los operadores $\mathcal{U}_T : L_p^{\mathbb{R}}(\mu_2) \to L_p^{\mathbb{R}}(\mu_1)$.

Observación 1.7. Este operador es lineal. Si $f \in L_p^{\mathbb{R}}$ y cumple que $f \geq 0$, entonces $\mathcal{U}_T(f) \geq 0$, es decir, $\mathcal{U}_T$ es un operador positivo. Note que $\mathcal{U}(\chi_{F_2}) = \chi_{T^{-1}(F_2)}$, $F_2 \in \mathscr{F}_2$. Por último quiero ver que este operador está bien definido, es decir, que $\mathcal{U}_T(f) \in L_p^{\mathbb{C}}(\mu_1)$ para toda $f \in L_p^{\mathbb{C}}(\mu_2)$, lo anterior es consecuencia del siguiente lema.

Lema 1.3. *Sea* $T : (X, \mathscr{F}, \mu) \to (Y, \mathscr{G}, \nu)$ *preservadora de las medidas. Si* $f \in L_1^{\mathbb{C}}(\nu)$ *entonces* $f \circ T \in L_1^{\mathbb{C}}(\mu)$ *y* $\int\limits_{T^{-1}(G)} (f \circ T) d\mu = \int\limits_G f d\nu$ *para toda* $G \in \mathscr{G}$.

Demostración. Basta probar el resultado para funciones $f \in L_1^{\mathbb{R}}(\nu)$, y considerando la parte positiva y negativa de f (es decir f^+ y f^-), basta con probar el resultado para f no negativa. El resultado es válido para funciones s-simples ya que; si s es una función f-simple, entonces $s = \sum_{i=1}^n \alpha_i \chi_{A_i}$. Si $s \in L_1^{\mathbb{R}}(\nu)$ se tiene que $\sum_{i=1}^n \alpha_i \nu(A_i) < \infty$, como $s \circ T = \sum_{i=1}^n \alpha_i \chi_{A_i} \circ T = \sum_{i=1}^n \alpha_i \chi_{T^{-1}(A_i)}$, entonces $\int s \circ T = \sum_{i=1}^n \alpha_i \mu(T^{-1}(A_i)) = \sum_{i=1}^n \alpha_i \nu(A_i) < \infty$ ya que T preserva las medidas. Por lo tanto $s \circ T \in L_1^{\mathbb{R}}(\mu)$ y $\int s \circ T d\mu = \int s d\nu$ para toda f-simple. Ahora consideramos una f no negativa, elegimos una sucesión de funciónes f-simples $\{s_n\}$ tales que $s_n \nearrow f$, entonces $\{s_n \circ T\}$ es una sucesión de funciones f-simples tales que $s_n \circ T \nearrow f \circ T$. Por el T.C.M. (ver Teorema B.3. pág. 132), se tiene que:

$$\int\limits_B f d\nu = \lim_{n \nearrow \infty} \int\limits_B s_n d\nu = \lim_{n \nearrow \infty} \int\limits_{T^{-1}(B)} s_n \circ T d\mu = \int\limits_{T^{-1}(B)} f \circ T d\mu. \qquad \square$$

Teorema 1.5. *Utilizando la notación de la definición anterior, si* $p \geq 1$, *entonces*

$$\mathcal{U}_T(L_p^{\mathbb{C}}(\mu_2)) \subset L_p^{\mathbb{C}}(\mu_1), \text{ más aun } \|\mathcal{U}_T(f)\|_p = \|f\|_p \text{ para toda } f \in L_p^{\mathbb{C}}(\mu_2).$$

Demostración. Considérese la función $|f|^p$ y utilice el lema anterior. □

Observación 1.8. Por lo anterior una transformación que preserva las medidas induce una isometría lineal de $L_p^{\mathbb{C}}(\mu_2)$ en $L_p^{\mathbb{C}}(\mu_1)$ para toda $p \geq 1$. El operador $\mathcal{U}_T$ será de utilidad en los conceptos de ergodicidad y en las propiedades mezclantes.

1.4. Recurrencia

Recordemos que estamos interesados en el comportamiento del conjunto $\{T^n(x_0)|$ para algún $x_0 \in X\}$, es decir, la órbita de x_0 donde T es una transformación preservadora de medida. Preguntarse por este comportamiento intuitivamente es preguntarse por la evolución del sistema en el tiempo. En otras palabras si $x_0 \in X$, entonces se puede pensar que $T^n(x_0)$ es el estado en el que se encuentra x_0 después de transcurridos n días. Henri Poincaré demostró uno de los primeros resultados sobre este comportamiento en espacios de medida finita. En esta parte demostraremos su teorema y veremos propiedades para extender su resultado a espacios de medida infinita. Necesitaremos las siguientes definiciones:

Definición 1.7. Sean $(X, \mathscr{F}, \mu)$ un espacio de medida y $T : X \to X$ una transformación.

(a) T se llama **recurrente** si para todo $E \in \mathscr{F}$ con $\mu(E) > 0$ existe $E_0 \subset E$ $(E_0 \in \mathscr{F})$ con $\mu(E_0) = 0$ tal que para toda $x \in E \setminus E_0$ se tiene que existe $n = n(E, x) \in \mathbb{N}$ tal que $T^n(x) \in E$.

(b) T se llama **infinitamente recurrente** si para todo $E \in \mathscr{F}$ con $\mu(E) > 0$ existe $E_0 \subset E$ $(E_0 \in \mathscr{F})$ con $\mu(E_0) = 0$ tal que para toda $x \in E \setminus E_0$, se tiene que $T^{n_i}(x) \in E$ para alguna subsucesión $n_1 < n_2 < \cdots$ (que depende de E y x).

Teorema 1.6 (De Recurrencia de Poincaré 1912)**.** *Sea T una transformación medible definida en un espacio de medida finita $(X, \mathscr{F}, \mu)$ en sí mismo. Si T preserva a μ entonces T es infinitamente recurrente.*

Demostración. Se probará una afirmación un poco más general, a saber: *Para todo $E \in \mathscr{F}$ con $\mu(E) > 0$, existe $F \subset E$ $(F \in \mathscr{F})$ tal que $\mu(E \setminus F) = 0$ y tal que para todo $x \in F$, $T^n(x) \in F$ para una infinidad de naturales.* Para $E \in \mathscr{F}$ con $\mu(E) > 0$ y $N \geq 0$ definimos $E_N = \bigcup_{n=N}^{\infty} T^{-n}(E)$ y $E^* = \limsup_{n\to\infty} T^{-n}(E)$ (el límite superior de la sucesión $\{T^{-n}(E)\}$ ver Definición A.2, pág. 128). Así que también: $E^* = \bigcap_{N=0}^{\infty} E_N$ y es el conjunto de puntos en X que entran a E una infinidad de veces bajo iteraciones positivas de T. Sea $F = E \cap E^*$, probaremos que $\mu(E \setminus F) = 0$.

Claramente: $E_{N+1} = T^{-1}(E_N) \subset E_N$ para toda $N \geq 0$ y como T preserva a μ se sigue de la sustractividad que: $\mu(E_N \setminus T^{-1}(E_N)) = 0$ para toda $N \geq 0$ en consecuencia $\mu(E_0 \setminus E_k) = 0$ para toda $k \geq 0$ y

$$\mu(E_0 \setminus E^*) = \mu\left(\bigcup_{k=0}^{\infty} E_0 \setminus E_k\right) \leq \sum_{k=0}^{\infty} \mu(E_0 \setminus E_k) = 0.$$

Por otro lado $\mu(E \setminus F) = \mu(E \setminus E^*) \leq \mu(E_0 \setminus E^*) = 0$. Sea $x \in F$, entonces existen $n_1 < n_2 < \cdots$ tal que $T^{n_i}(x) \in E$ (ver Proposición A.1, pág. 128), probaremos de hecho $T^{n_i}(x) \in F$ para toda i. Por construcción $T^{n_i}(x) \in E$ y $T^{n_i - n_1}(T^{n_1}(x)) \in E$ para toda $i \geq 2$, entonces $T^{n_1}(x) \in E \cap E^* = F$. De manera análoga se prueba que $T^{n_i}(x) \in F$ para toda $i \geq 2$. □

Este resultado nos asegura que si E cumple que $\mu(E) > 0$, entonces casi todos sus puntos regresan a él una infinidad de veces durante la evolución del sistema conforme transcurre el tiempo. Regresando al Ejemplo 1.4, supongamos que éste describe el movimiento de una partícula, es decir, $T^n(z_0)$ donde $z_0 \in S^1$ proporciona la posición de la partícula después de n días, si ésta comenzó a moverse a partir de la posición z_0. Si tomamos un arco con longitud positiva, el teorema anterior nos asegura que el número de veces que la partícula toma alguna de las posiciones dadas por el arco es infinito para casi todo punto que empieza a moverse en el arco. Note que en la demostración anterior $F \neq \emptyset$ porque $\mu(F) \geq \mu(E \cap F) = \mu(E) > 0$.

Observación 1.9. Siguiendo la notación del teorema de Poincaré, podemos mostrar que $E \cap E^* = F \cap F^*$ por lo que todo punto de F regresa a F una infinidad de veces bajo iteraciones positivas de T lo que prueba de otra manera al teorema anterior.

Demostración. Sea $x \in E \cap E^*$ entonces existe una sucesión $n_1 < n_2 < \cdots$ tal que $T^{n_i}(x) \in E$ para toda $i \geq 1$ y $x \in E$. Por lo que $x \in F$. Para ver que $x \in F^*$, probaré que $T^{n_i}(x) \in F$ para toda $i \geq 1$. Para $i = 1$, considero la sucesión $n_i' = n_i - n_1$ para toda $i \geq 2$, notando que $T^{n_i'}(T^{n_1}(x)) = T^{n_i}(x) \in E$ para toda $i \geq 2$, se tiene que $T^{n_1}(x) \in F$. Para $i = 2$, considero la sucesión $n_i'' = n_i - n_2$ para toda $i \geq 3$, entonces $T^{n_i''}(T^{n_2}(x)) = T^{n_i}(x) \in E$ para toda $i \geq 3$, por lo que $T^{n_2}(x) \in F$, continuando de esta manera tenemos que $T^{n_i} \in F^*$ para toda $i \geq 1$. Por lo que $E \cap E^* \subset F \cap F^*$, la otra contención es clara debido a que $F = E \cap E^*$. Por lo tanto $E \cap E^* = F \cap F^*$. □

Podemos preguntarnos si existe un resultado análogo al anterior para espacios de medida infinita. En general no existe tal resultado y la siguiente observación muestra la afirmación anterior.

Observación 1.10. Sea $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}), \mu)$ donde μ es la medida del conteo (ver Ejemplo B.2). Definimos $T : \mathbb{Z} \to \mathbb{Z}$ poniendo $T(k) = k + 1$. Sea $E = \{0\}$, E^* y F como en el teorema anterior. Note que $\mu(E) > 0$ y $F = \emptyset$ por lo que es necesario que el espacio sea de medida finita en el teorema anterior.

Sin embargo existe una condición con la que podemos extender el resultado del Teorema 1.6 para espacios de medida infinita.

Definición 1.8. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y $T : X \to X$ una transformación medible. T se llama **incompresible**. Si $H \in \mathscr{F}$ y $T^{-1}(H) \subset H$, entonces $\mu(H \setminus T^{-1}(H)) = 0$.

Teorema 1.7. *El resultado anterior permanece válido en el caso finito o infinito si en lugar de suponer que T es preservadora de medida, suponemos que T es incompresible.*

Demostración. Definimos F como en el Teorema 1.6, pero $T^{-1}(E_N) = E_{N+1} \subset E_N$ entonces:

$$\mu(E_N \setminus T^{-1}(E_N)) = 0 \quad \text{para toda} \quad N \geq 0 \ \text{ por lo que } \ \mu(E_0 \setminus E^*) = 0 \ \text{ y } \ \mu(E \setminus F) = 0,$$

el resto es igual. $\square$

Observación 1.11. Si $(X, \mathscr{F}, \mu)$ es un espacio de medida finita. Note que si T preserva a μ, entonces T es incomprensible. Sin embargo , si T es incomprensible, entonces se tiene que un subconjunto de $\mathscr{F}$ cumple que $\mu(T^{-1}(H)) = \mu(H)$, este subconjunto puede coincidir o no con $\mathscr{F}$. Por lo tanto incomprensibilidad no implica que preserve la medida.

Definición 1.9. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y $T : X \to X$ medible. T se llama **disipativa** si existe $W \in \mathscr{F}$ con $\mu(W) > 0$ tal que $W, T^{-1}(W), T^{-2}(W), \ldots$ son disjuntos. Llamamos a W un conjunto **errante** para T. Finalmente si T no es disipativa, entonces T se llama **conservativa**.

Hemos visto en la Observación 1.1(5) que la transformación $T : (0,1) \to (0,1)$ dada por $T(x) = x^2$ no preserva a λ, análogamente se puede ver que las transformaciones $T_k : (0,1) \to (0,1)$ dadas por $T_k(x) = x^k$ no preservan a λ para toda $k \geq 2$. Uno de los problemas de la Teoría Ergódica, es encontrar una medida μ en $\mathscr{B}_{(0,1)}$, tal que $\mu \circ T^{-1} = \mu$. De tal manera, que tenga los mismos conjuntos nulos que la medida de Lebesgue, es decir, $\mu \ll \lambda$ y $\lambda \ll \mu$ (ver Definición B.25, pág. 135). Como podremos ver en el siguiente ejemplo existe una medida $m_f : \mathscr{B}_{(0,1)} \to \mathbb{R} \cup \{\infty\}$ que es preservada por estas transformaciones, además $m_f \ll \lambda$ y $\lambda \ll m_f$.

Ejemplo 1.11. Para $k \geq 2$ fijo, sea $T : (0,1) \to (0,1)$ dada por $T(x) = x^k$, $\mathscr{F} = \mathscr{B}_{(0,1)}$, $f : (0,1) \to \mathbb{R}$ dada por $f(x) = -1/(x\log(x))$, y $m_f : \mathscr{B}_{(0,1)} \to \mathbb{R} \cup \{\infty\}$ dada por $m_f(F) = \int\limits_F f d\,\lambda$. Entonces T preserva a m_f y $W = [a^k, a)$ $(a \in (0,1))$ es errante bajo T.

Demostración. Consideramos el siguiente π-sistema $\mathcal{C} = \{[a,b) : 0 < a < 1,\, 0 < b < 1\}$, como $T^{-1}([a,b)) = [\sqrt[k]{a}, \sqrt[k]{b})$ (lo anterior puesto que $x \in T^{-1}([a,b))$ si y sólo si $a \leq x^k < b$ si y sólo si $\sqrt[k]{a} \leq x < \sqrt[k]{b}$), por otro lado

$$\begin{aligned}
m_f([\sqrt[k]{a}, \sqrt[k]{b})) &= \int\limits_{(\sqrt[k]{a}, \sqrt[k]{b})} -1/(x\log(x))d\,\lambda = -\int\limits_{\sqrt[k]{a}}^{\sqrt[k]{b}} 1/(x\log(x))d\,x = -\log(|\log(x)|)\Big|_{\sqrt[k]{a}}^{\sqrt[k]{b}} \\
&= -\log(|\log(\sqrt[k]{b})|) + \log(|\log(\sqrt[k]{a})|) = \log(|\log(a)|) - \log(|\log(b)|) \\
&= \int\limits_a^b -\frac{1}{x\log(x)} = m_f([a,b)),
\end{aligned}$$

por lo tanto T preserva a m_f.

Sea $W = [a^k, a)$, entonces $T^{-1}(W) = [a, \sqrt[k]{a})$, $T^{-n}(W) = [\sqrt[k^{n-1}]{a}, \sqrt[k^n]{a})$, como $[a^k, a)$, $[a, \sqrt[k]{a}), \ldots,$ son disjuntos entonces $W, T^{-1}(W), T^{-2}(W), \ldots$ son disjuntos y

$$m_f([a^k, a)) = \log(\log(a^k)) - \log(\log(a)) = \log(k) > 0,$$

se tiene que W es errante para T. □

Observación 1.12. Si T es preservadora de medida en un espacio de medida finito, entonces T es conservativa, ya que si existiera un conjunto errante para T, entonces

$$\mu\left(\bigcup_{i=0}^{\infty} T^{-i}(W)\right) = \sum_{i=0}^{\infty} \mu(W) = \infty,$$

lo que nos da una contradición con el supuesto de que el espacio es de medida finita.

Teorema 1.8. *T es conservativa si y sólo si T es incompresible.*

Demostración.

(Necesidad). Si T no es incompresible, entonces existe $H \in \mathscr{F}$ tal que $T^{-1}(H) \subset H$ pero $\mu(H \setminus T^{-1}(H)) > 0$. Sea $W = H \setminus T^{-1}(H)$, veamos que W es errante para T; como

$$T^{-n}(W) = T^{-n}(H) \setminus T^{-(n+1)}(H) \text{ y } T^{-(n+1)}(H) \subset T^{-n}(H),$$

para toda $n \geq 0$ por propiedades de la imagen inversa. Ahora si $n \neq m$, sin perdida de generalidad podemos suponer que $n < m$, entonces $n+1 \leq m$ por lo anterior se tiene que $T^{-m}(H) \subseteq T^{-(n+1)}(H)$ así que $T^{-m}(W) \cap T^{-n}(W) \subset T^{-n}(W) \cap (T^{-(n+1)}(H) \setminus T^{-(m-1)}(H)) = \emptyset$; por lo tanto T no es conservativa.

(Suficiencia). Supongamos que T es incompresible, sea $W \in \mathscr{F}$ tal que: $W, T^{-1}(W), T^{-2}(W), \ldots$ son disjuntos. Sea $H = \bigcup_{i=0}^{\infty} T^{-i}(W)$, entonces $T^{-1}(H) \subset H$ por lo tanto $\mu(H \setminus T^{-1}(H)) = 0$, pero:

$$H \setminus T^{-1}(H) = \bigcup_{i=0}^{\infty} T^{-i}(W) \setminus \bigcup_{i=1}^{\infty} T^{-i}(W) = W^{[3]}, \quad \text{entonces} \quad \mu(W) = 0,$$

por lo tanto T es conservativa. □

Observación 1.13. Note que la transformación del ejemplo anterior es disipativa y por el teorema anterior no es incompresible, pero preserva a m_f; es decir, en el caso $\mu(X) = \infty$, si T preserva la medida no implica que T sea incompresible.

Finalmente de los Teoremas 1.7 y 1.8 obtenemos:

[3] Esta igualdad es valida porque los conjuntos W, $T^{-1}(W), \ldots$ son disjuntos.

Corolario 1.8.1. *Sea $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ medible y conservativa, entonces T es infinitamente recurrente.*

Demostración. Como T es conservativa por el teorema anterior T es incomprensible, así por el Teorema 1.7 T es infinitamente recurrente. $\square$

Observación 1.14. El enunciado del corolario anterior es más común, sin embargo la prueba es mas complicada, por ello es que introducimos el concepto de incomprensibilidad.

Podemos preguntarnos por la primera vez que x regresa a $F \in \mathscr{F}$. Se puede pensar en asignarle a cada $x \in F$ su primer retorno. Más aún, podremos inducir una transformación medible, que presevará una medida, si la transformación original preserva la medida en su espacio. Esta transformación es la siguiente:

Proposición 1.1 (Transformación inducida de S. Kakutani 1943)**.** *Sea $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ medible, conservativa e **invertible**. Para $F \in \mathscr{F}$ con $\mu(F) > 0$ fijo definimos $n_F : F \to \mathbb{N}$ como sigue:*

$$n_F(x) = \begin{cases} \min\{k \geq 1 : T^k(x) \in F\} \\ \\ 0 \quad si \quad T^k(x) \notin F \quad para\ toda\ k \geq 1. \end{cases}$$

Definimos $T_F : F \to F$ como sigue $T_F(x) = T^{n_F(x)}(x)$.

*Entonces $T_F : (F, \mathscr{F} \cap F, \mu_F) \to (F, \mathscr{F} \cap F, \mu_F)$ es medible e **invertible**, si T preserva a μ entonces T_F preserva a μ_F.*

Demostración. Para mostrar la medibilidad consideramos los conjuntos

$$F_n = (F \cap T^{-n}(F)) \setminus (\bigcap_{j=1}^{n-1} T^{-j}(F)) \text{ con } n \geq 2,\ F_1 = F \cap T^{-1}(F) \text{ y } G_n = T^{-n}(F) \setminus (\bigcup_{j=0}^{n-1} T^{-j}(F)),$$

observemos que $F_n \cap F_m = \emptyset$ y $G_n \cap G_m = \emptyset$ si $n \neq m$, como T es medible, entonces $F_n, G_n \in \mathscr{F}$ para toda n, $F_n \subset F$ y para toda $E \in \mathscr{F}$ $T_F^{-1}(E) = \bigcup_{n=1}^{\infty}(F_n \cap T^{-1}(E))$, por lo tanto T_F es medible.

Para probar que T_F preserva a μ_F, observamos que $T^{-1}(F) = E_1 \cup F_1$, $T^{-1}(G_n) = F_{n+1} \cup G_{n+1}$ para $n \geq 1$, si $E \in \mathscr{F} \cap F$, entonces por lo observado anteriormente

$$T^{-1}(E) = (F_1 \cap T^{-1}(E)) \cup (G_1 \cap T^{-1}(E)),\ T^{-1}(G_n \cap T^{-n}(E)) = (F_{n+1} \cap T^{-n-1}(E)),$$

como T preserva a μ, se tiene que $\mu(F) = \sum_{j=1}^{n} \mu(F_j \cap T^{-j}(E)) + \mu(G_n \cap T^{-n}(E))$, pero la sucesión $\{G_n\}$ es disjunta, por lo que $\mu(G_n \cap T^{-n}(E)) \to 0$. Por lo tanto

$$\mu(F) = \sum_{n=1}^{\infty} \mu(F_n \cap T^{-n}(E)) = \mu(T_F^{-1}(E))).$$

Para probar que T_F^{-1} existe basta notar que $T_F^{-1} = (T^{-1})_F$. $\square$

Hemos visto propiedades para garantizar que T sea infinitamente recurrente para un espacio de media, el siguiente lema será de utilidad para la siguiente sección.

Lema 1.4. *Sean* $T : (X, \mathscr{F}, \mu) \to (Y, \mathscr{G}, \nu)$ *medible, conservativa y* $f : X \to \mathbb{R}$ *medible no negativa, entonces* $\sum\limits_{n=1}^{\infty} f(T^n(y))$ *diverge casi dondequiera para* $y \in \{x \in X : f(x) > 0\}$.

Demostración. Sean $A = \{x \in X : f(x) > 0\}$ y $B_k = \{x \in A : f(x) > 1/k\}$ con $k \in \mathbb{N}$, entonces $A = \bigcup_{n=1}^{\infty} B_k$, por el corolario anterior T es infinitamente recurrente, entonces existen $F_k \subset B_k$ tal que $\mu(B_k \setminus F_k) = 0$ y para todo $x \in F_k$, $T^n(x) \in F_k$ para una infinidad de naturales. Sean $N_k = B_k \setminus F_k$, $N = \bigcup_{k=1}^{\infty} N_k$ (notemos que $\mu(N) = 0$), ahora consideramos $x \in A \setminus N$, entonces $x \in B_k$ para algún k y $x \notin N_k$ para toda $k \geq 1$, se tiene que $T^n(x) \in B_k$ para una infinidad de naturales, por lo que $\sum_{n=1}^{\infty} f(T^n(y)) > \sum_{n \in I} 1/k$, donde I es una sucesión creciente de naturales, por lo tanto $\sum_{n=1}^{\infty} f(T^n(y))$ diverge casi dondequiera en A. $\square$

Observación 1.15. El lema anterior nos asegura que si $\mu(F) > 0$, entonces la siguiente serie $\sum_{i=0}^{\infty} \chi_F(T^i(x))$ diverge casi dondequiera en F , es decir, para casi todo punto en F, el número de veces que x regresa a F bajo iteraciones positivas de T es una infinidad.

1.5. Ergodicidad

Sean $(X, \mathscr{F}, \mu)$ un espacio de medida y $T : X \to X$ una transformación preservadora de medida. Si $T^{-1}(F) = F$ para algún $F \in \mathscr{F}$, entonces $T^{-1}(X \setminus F) = X \setminus F$, por lo que podríamos estudiar a T a partir de las transformaciones $T|_F$ y $T|_{X \setminus F}$, lo que simplifica el estudio de T si $0 < \mu(F) < \mu(X)$. Si $\mu(F) = 0$ ó $\mu(X \setminus F) = 0$, entonces podemos ignorar a F ó $X \setminus F$ recordando que en Teoría de la Medida los conjuntos de medida cero usualmente no son tomados en cuenta, por lo tanto no tendría caso simplificar el estudio de T. Será de nuestro interés el estudio de las transformaciónes que no se pueden descomponer, a éstas las llamaremos ergódicas. A continuación definiremos estas transformaciones.

Definición 1.10. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y $T : X \to X$ una transformación preservadora de medida. Decimos que T es **ergódica** si siempre que $F \in \mathscr{F}$ es tal que $T^{-1}(F) = F$ entonces $\mu(F) = 0$ ó $\mu(X \setminus F) = 0$, es decir, **no** existe una descomposición $X = F \cup G$ (ajena) tal que $\mu(F) > 0$, $\mu(G) > 0$ y $T^{-1}(F) = F$. Conjuntos $F \in \mathscr{F}$ tal que $T^{-1}(F) = F$ los llamamos **estrictamente invariantes** (bajo T). Si $F \in \mathscr{F}$ es tal que $\mu(F \triangle T^{-1}(F)) = 0$ lo llamamos **invariante** (bajo T).

Observación 1.16. Considerando:

$$\mathscr{I} = \{F \in \mathscr{F} : \mu(F \triangle T^{-1}(E)) = 0\}, \ \mathscr{I}_e = \{F \in \mathscr{F} : T^{-1}(F) = F\},$$

entonces $\mathscr{I}$ e $\mathscr{I}_e$ son sub-σ-álgebras de $\mathscr{F}$ y $\mathscr{I}$ es la μ-completación[4] de $\mathscr{I}_e$ si T es invertible.

Demostración. $N = \bigcup_{n=0}^{\infty} T^{-n}(F) \triangle T^{-n-1}(F)$ $F = F \setminus N \cup F \cap N$. $\square$

Más adelante daremos otros criterios de ergodicidad. A continuación veremos un ejemplo de una transformación ergódica. Sea $a \in S^1$ fijo y $T : S^1 \to S^1$ dada por $T(z) = az$ con la σ-álgebra y medida definida como en el Ejemplo 1.4 (pág. 9). Tenemos 2 casos:

Caso 1) $a = e^{i\rho}$ con ρ **conmensurable** con 2π, es decir, existe $r \in \mathbb{Q} \cap [\,0\,,1\,)$ tal que $\rho = 2\pi r$. Si $r = p/q$ $(p, q \in \mathbb{N})$ y $(p, q) = 1$, entonces a es una raíz de la unidad (de hecho es una q-raíz primitiva de la unidad).

$$\text{Sea} \quad F = \bigcup_{j=0}^{q-1} \left\{ e^{ix} : \frac{2j\pi}{q} \leq x < \frac{2\left(j+\frac{1}{2}\right)\pi}{q} \pmod{2\pi} \right\} \quad \text{entonces} \quad \mu(F) = \frac{1}{2} \;\; \text{y} \;\; T^{-1}(F) = F$$

por lo tanto T no es ergódica, gráficamente:

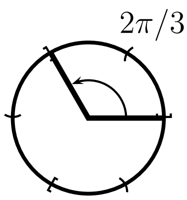

Figura 1.8: F y $\rho = \dfrac{2\pi}{3}$

Observación 1.17. En este caso la órbita de x_0 es periódica para todo x_0, es decir, $\{T^n(x_0) | n \in \mathbb{N}\} = \{x_0, T^1(x_0), \ldots, T^{q-1}(x_0)\}$.

Caso 2) ρ es **inconmensurable** con 2π, así pues $a^n \neq 1$ para toda $n \in \mathbb{N}$ $(n \neq 0)$. Probaremos que T es ergódica pero antes necesitaremos el siguiente resultado.

Lema 1.5 (Teorema de Jacobi)**.** *Si $a^n \neq 1$ para toda $n \in \mathbb{N}$ $(n \neq 0)$ entonces $\{T^n(z)\}_{n=0}^{\infty}$ es denso en S^1 para toda $z \in S^1$.* (*La métrica en S^1 es la longitud circular*).

Demostración. Como $a^n \neq 1$, entonces $T^n(z) \neq T^m(z)$ para toda n, $m \in \mathbb{N}$ $(n \neq m)$, así pues la semiórbita de z bajo T consiste de un número infinito de puntos de S^1, todos ellos distintos. Como S^1 es compacto, $\{T^n(z)\}$ posee un punto límite, entonces para toda $\varepsilon > 0$ existe $n \neq m$ tal que $d(T^n(z), T^m(z)) < \varepsilon$ (longitud circular). Sea $p = |n - m| \in \mathbb{N}$, como T es una rotación $d(T^p(z), z) < \varepsilon$ y también $d(T^{jp}(z), T^{(j-1)p}(z)) < \varepsilon$ para toda $j \geq 1$; así pues $z, T^p(z), \ldots, T^{kp}(z), \ldots$

[4] $(X, \mathscr{F}, \mu)$ es un espacio de medida completo si para todo $F \in \mathscr{F}$ tal que $\mu(F) = 0$ y $E \subset F$ se tiene que $E \in \mathscr{F}$. Si $(X, \mathscr{F}, \mu)$ no es completo, es posible construir un espacio de medida completo $(X, \mathscr{F}', \mu')$ tal que $\mathscr{F} \subset \mathscr{F}'$ y $\mu'|_{\mathscr{F}} = \mu$ llamado la μ-completación de $(X, \mathscr{F}, \mu)$.

"parten" a S^1 en arcos de **igual** longitud positiva menor que ε. Por lo tanto $\{T^n(z)\}_{n=0}^{\infty}$ es denso en S^1. $\square$

Teorema 1.9. *Si $a^n \neq 1$ $(n \neq 0)$, entonces $T(z) = az$ es ergódica.*

Demostración. Sea $A \subset S^1$ medible, estrictamente invariante y tal que $\mu(A) > 0$. Debemos probar que $\mu(A) = 1$. Primero notamos que dada $\varepsilon \in (0, 1)$ es posible hallar un arco I tal que $\mu(I) \leq \varepsilon$ y $\mu(A \cap I) > (1-\varepsilon)\mu(I)$. [*Demostración.* Por definición de la medida circular de Lebesgue es posible hallar una sucesión **disjunta** de arcos (I_n) con $\mu(I_n) < \varepsilon$ tal que

$$A \subset \bigcup_{n=1}^{\infty} I_n \quad \text{y} \quad \frac{\mu(A)}{1-\varepsilon} \geq \sum_n \mu(I_n),$$

esto pues recordando la definición de medida exterior (ver Definición B.21, pág. 134), como $\mu(A)/(1-\varepsilon) > \mu(A) > 0$. Por definición de ínfimo existe $\{I_n\}_{n=1}^{\infty}$ tal que $A \subset \bigcup_{i=1}^{\infty} I_n$ y $\mu(A) \leq \sum_{i=1}^{\infty} \mu(I_n) < \mu(A)/(1-\varepsilon)$, entonces $\mu(A) \geq (1-\varepsilon)\sum_{i=1}^{\infty} \mu(I_n)$. Si $\{I_n\}$ no es ajena podemos hallar $\{I'_n\}$ ajenos tales que $\bigcup_{i=1}^{\infty} I_n = \bigcup_{i=1}^{\infty} I'_n$, de la siguiente manera elegimos $I'_1 = I_1$ y $I'_n = I_n - \bigcup_{i=1}^{n-1} I_i$ si $n \geq 2$; ahora para que $\mu(I'_n) < \varepsilon$ para toda n, como $\sum_{i=1}^{\infty} \mu(I'_n) < \infty$ entonces $\lim \mu(I'_n) \to 0$, por lo que existe una $N \in \mathbb{N}$ tal que $\mu(I'_n) < \varepsilon$ para toda $n' \geq N$. Como la sucesión $\mu(I'_n)$ es convergente está acotada, sea $M > 0$ tal que $\mu(I'_n) < M$, considerando a los arcos $I'_1, \ldots, I'_{N-1}$, elegimos una m talque $M/m < \varepsilon$ partiendo cada arco anterior en m arcos disjuntos obtenemos que : $\sum_{n=1}^{\infty} \mu(A \cap I'_n) = \mu(A) \geq (1-\varepsilon)\sum_{n=1}^{\infty} \mu(I'_n)$. Lo anterior prueba que existe **al menos** un arco **no degenerado** I'_k tal que $\mu(A \cap I'_k) \geq (1-\varepsilon)\mu(I'_k)$. Ponemos $I = I'_k$ y **lo fijamos**]. Ya que T es una rotación y preserva a μ, obtenemos que:

$$\mu(A \cap T^{-n}(I)) \geq (1-\varepsilon)\mu(T^{-n}(I)) \quad \text{para toda } n \in \mathbb{N}.$$

Como la semiórbita del extremo inicial de I es densa en S^1 es posible hallar $n_1, \ldots, n_k$ naturales tales que $T^{-n_1}(I), T^{-n_2}(I), \ldots, T^{-n_k}(I)$ disjuntos por lo que:

$$\mu(A) \geq \sum_{j=1}^{k} \mu(A \cap T^{-n_j}(I)) \geq (1-\varepsilon)\mu(\bigcup_{j=1}^{k} T^{-n_j}(I)),$$

y **además** suficientes en número para que $T^{-n_1}(I), T^{-n_2}(I), \ldots, T^{-n_k}(I)$ cubran a S^1, salvo por un conjunto de medida menor que 2ε, es decir, $\mu(\bigcup_{j=1}^{k} T^{-n_j}(I)) > 1 - 2\varepsilon$[5]. En este caso obtendremos que $\mu(A) > (1-\varepsilon)(1-2\varepsilon)$, y como $\varepsilon > 0$ es arbitrario tenemos que $\mu(A) = 1$. $\square$

Observación 1.18. En este caso la órbita de z_0 es densa en S^1 para todo z_0, lo que nos lleva a preguntarnos: ¿Los sistemas ergódico tienen la propiedad de que la órbita de $x_0 \in X$ es "densa" para casi todo punto? Más adelante aclararemos lo anterior.

[5]Sea $k = \min\{j \in \mathbb{N} : 1 - 2\varepsilon < k\mu(I)\}$ entonces $1 - 2\varepsilon < k\mu(I) < 1 - \varepsilon$, ya que si $1 - \varepsilon \leq k\mu(I)$ entonces $1 - 2\varepsilon < 1 - \varepsilon - \mu(I) \leq (k-1)\mu(I)$, contradiciendo la elección de k. Elegimos n_1 de tal modo que el extremo inicial de $T^{-n_1}(I)$ diste del extremo final de I en menos de ε/k, luego elegimos n_2 de tal modo que el extremo inicial de $T^{-n_2}(I)$ diste del extremo final de $T^{-n_1}(I)$ es menos de ε/k. Continuando de este modo se obtienen k arcos **ajenos** $I, T^{-n_1}(I), T^{-n_2}(I), \ldots, T^{-n_k}(I)$ de longitud total $> 1 - 2\varepsilon$.

Otro ejemplo de transformación ergódica es el corrimiento bilateral (unilateral) de Bernoulli con parámetros $(p_0, \ldots, p_{k-1})$, haremos uso de un lema de aproximación que se utiliza frecuentemente en Teoría de la Medida, el cual consiste en aproximar a los elementos de $\mathscr{F}$ mediante elementos de $\mathcal{A}$ donde $\mathcal{A}$ es un álgebra tal que $\sigma(\mathcal{A}) = \mathscr{F}$. Nos aproximaremos a los elementos de $\mathscr{F}$ mediante elementos de $\mathcal{A}$ que "dependen de solo un número finito de coordenadas". Note que la demostración es válida para los dos tipos de corrimientos.

Teorema 1.10. *El corrimiento bilateral (unilateral) de Bernoulli con parámetros $(p_0, \ldots, p_{k-1})$ es ergódico.*

Demostración. Sea $\mathcal{A}$ =álgebra generada por los cilindros medibles (ver Ejemplo 1.7, pág. 11), entonces $\sigma(\mathcal{A}) = \mathscr{F}$. Es fácil comprobar que $A \in \mathcal{A}$ si y sólo si A es la unión finita y disjunta de conjuntos de la forma:

$$\{x \in X : x_{n_1} \in E_1, \ldots, x_{n_m} \in E_m\} \quad \text{con} \quad \emptyset \subsetneq E_j \subseteq E \quad \text{y con} \quad n_1 < n_2 \cdots < n_m \in \mathbb{Z} \quad (\text{o } \in \mathbb{N}),$$

(en el caso del corrimiento unilateral). Sea $F \in \mathscr{F}$ tal que $T^{-1}(F) = F$, dada $\varepsilon > 0$ existe $A \in \mathcal{A}$ tal que $\mu(F \triangle A) < \varepsilon$ (ver Lema B.1, pág. 131). Como $A \in \mathcal{A}$ depende sólo de un número finito de coordenadas, por lo que si j es suficientemente grande, entonces $B = T^{-j}(A)$ $(\in \mathcal{A})$ depende del mismo número finito de coordenadas pero **distintas** de las de A entonces:

$$\mu(B \cap A) = \mu(B)\mu(A) = \mu(A)^2, \qquad \text{además:}$$

$$\mu(F \triangle B) = \mu(T^{-j}(F) \triangle T^{-j}(A)) = \mu(T^{-j}(F \triangle A)) = \mu(F \triangle A) < \varepsilon,$$

y como $F \triangle (A \cap B) \subset (F \triangle A) \cup (F \triangle B)$ tenemos que:

$$\begin{aligned}
|\mu(F) - \mu(A \cap B)| &= |\mu(F \cap A \cap B) + \mu(F \setminus A \cap B) - \mu(A \cap B \cap F) - \mu(A \cap B \setminus F)| \\
&= |\mu(F \setminus A \cap B) - \mu(A \cap B \setminus F)| \\
&\leq \mu(F \setminus A \cap B) + \mu(A \cap B \setminus F) \\
&\leq \mu(F \triangle (A \cap B)) < 2\varepsilon,
\end{aligned}$$

de donde:

$$\begin{aligned}
\left|\mu(F) - \mu(F)^2\right| &\leq |\mu(F) - \mu(A \cap B)| + \left|\mu(A \cap B) - \mu(F)^2\right| \\
&< 2\varepsilon + (\mu(A) + \mu(F))\,|\mu(A) - \mu(F)| < 4\varepsilon.
\end{aligned}$$

Como $\varepsilon > 0$ es arbitrario se tiene que $\mu(F) = \mu(F)^2$, por lo tanto $\mu(F) = 0$ ó 1, mostrando con esto que T es ergódica. Recordemos que estamos trabajando en un espacio de probabilidad entonces $(\mu(A) + \mu(F)) \leq 2$. $\square$

Ejemplo 1.12. Sean $X = \{0, \ldots, k-1\}$, $\mathscr{F} = \mathcal{P}(X)$ y $\mu(\{j\}) = 1/k$, $j \in X$. Definimos $T : X \to X$ dada por

$$T(n) = \begin{cases} n+1 & \text{si } n \neq k-1 \\ \\ 0 & \text{si } n = k-1. \end{cases}$$

Entonces:

(1) T es ergódica.

(2) T^k no es ergódica.

Demostración.

(1) Sea $F \in \mathcal{P}(E)$, $F \neq \emptyset$ tal que $T^{-1}(F) = F$, sea $m = \max\{F\}$, si $m = 0$ entonces $k-1 \in F$, por lo que $k-2 \in F, \ldots, 1 \in F$, por lo tanto $F = X$. En caso contrario análogamente llegaríamos a que $0 \in F$, y utilizando el mismo razonamiento llegaríamos a que $X = F$, por lo tanto T es ergódica.

(2) Se puede ver mediante cálculos directos que T^k es la función identidad, por lo tanto T no es ergódica. ☐

Observación 1.19. El ejemplo anterior nos dice que las potencias de una transformación ergódica no tienen porque ser ergódicas.

Como hemos mostrado anteriormente, probar que una transformación es ergódica a partir de la definición resulta bastante complicado. A continuación mostraremos condiciones equivalentes a la ergodicidad. El siguiente lema nos será de utilidad. Recordemos que

Lema 1.6. *Sean $T : X \to X$ una transformación, $E \subset X$ y $E^* = \limsup_{n\to\infty} T^{-n}(E)$, entonces:*

$$E \bigtriangleup E^* \subset \bigcup_{n=0}^{\infty} \left(T^{-n}(E) \bigtriangleup T^{-n-1}(E)\right).$$

Demostración. Basta probar que: $X \setminus \bigcup_{n=0}^{\infty} (T^{-n}(E) \bigtriangleup T^{-n-1}(E)) \subset X \setminus (E \bigtriangleup E^*)$. Observe que $x \notin T^{-n}(E) \bigtriangleup T^{-n-1}(E)$ si y sólo si $\chi_E(T^n(x)) = \chi_E(T^{n+1}(x))$. Así que los siguientes enunciados son equivalentes:

(i) $x \in X \setminus \bigcup_{n=0}^{\infty} (T^{-n}(E) \bigtriangleup T^{-n-1}(E))$

(ii) $x \notin (T^{-n}(E) \bigtriangleup T^{-n-1}(E))$ para toda $n \geq 0$

(iii) $\chi_E(T^{-n}(x)) = \chi_E(T^{-n-1}(x))$ para toda $n \geq 0$

(iv) $\chi_E(T^n(x)) = 0$ para toda $n \geq 0$ ó $\chi_E(T^n(x)) = 1$ para toda $n \geq 0$

(v) $T^n(x) \notin E$ para toda $n \geq 0$ ó $T^n(x) \in E$ para toda $n \geq 0$,

por lo tanto $x \in X \setminus (E \triangle E^*)$. □

Observación 1.20. El conjunto E^* tiene la propiedad de ser estrictamente invariante, es decir, $T^{-1}(E^*) = E^*$.

Demostración.

$\subseteq$) $x \in T^{-1}(E^*)$, así pues $T(x) \in E^*$, entonces existen $n_1 < n_2 < \cdots$ tal que $T^{n_1}(T(x)) \in E$, $T^{n_2}(T(x)) \in E, \ldots$; así que $T^{n_1+1}(x) \in E$, $T^{n_2+1}(x) \in E, \ldots$, por lo tanto $x \in E^*$.

$\supseteq$) $x \in E^*$ entonces existen $n_1 < n_2 < \cdots$ tal que $T^{n_1}(x) \in E$, $T^{n_2}(x) \in E, \ldots$, así pues $T^{n_1-1}(T(x)) \in E$, $T^{n_2-1}(T(x)) \in E, \ldots$, por lo tanto $T(x) \in E^*$. □

Teorema 1.11. *Las siguientes condiciones son equivalentes para una transformación conservativa $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ que preserve a μ:*

(i) T es ergódica.

(ii) Si $E \in \mathscr{F}$ tal que $\mu(E \triangle T^{-1}(E)) = 0$, entonces $\mu(E) = 0$ o $\mu(X \setminus E) = 0$.

(iii) Para todo $F, G \in \mathscr{F}$ con $\mu(F)\,\mu(G) > 0$ existe $n \in \mathbb{N}$ tal que $\mu(T^{-n}(F) \cap G) > 0$.

(iv) Para todo $F, G \in \mathscr{F}$ con $\mu(F)\mu(G) > 0$ existe una sucesión $n_1 < n_2 < \cdots$ tal que $\mu(T^{-n_i}(F) \cap G) > 0$ para toda $i \geq 1$.

Demostración.

(i) implica (ii). Supongamos que T es ergódica y que $\mu(E \triangle T^{-1}(E)) = 0$. Sea $E^* = \limsup_{n\to\infty} T^{-n}(E)$, entonces $T^{-1}(E^*) = E^*$ así que $\mu(E^*) = 0$ *ó* $\mu(X \setminus E^*) = 0$. Demostraremos que $\mu(E \triangle E^*) = 0$ (por lo que también se tendrá $\mu(X \setminus E \triangle X \setminus E^*) = 0$) lo que será suficiente. Como T preserva a μ, $\mu(T^{-n}(E) \triangle T^{-n-1}(E)) = 0$ para toda $n \geq 0$ pero $E \triangle E^* \subset \bigcup_{n=0}^{\infty} T^{-n}(E) \triangle T^{-n-1}(E)$, por lo tanto $\mu(E \triangle E^*) = 0$.

(ii) implica (iii). Supongamos que (iii) es falso, entonces existen $F, G \in \mathscr{F}$ con $\mu(F)\,\mu(G) > 0$ tal que $\mu(T^{-n}(F) \cap G) = 0$ para toda $n \in \mathbb{N}$, entonces $\mu(\bigcup_{n=1}^{\infty} T^{-n}(F) \cap G) = 0$. Sea $\widetilde{F} = \bigcup_{n=1}^{\infty} T^{-n}(F)$ entonces $T^{-1}(\widetilde{F}) \subset \widetilde{F}$, y como T es incompresible (conservativa) entonces $\mu(\widetilde{F} \setminus T^{-1}(\widetilde{F})) = 0$ así pues $\mu(\widetilde{F} \triangle T^{-1}(\widetilde{F})) = 0$. Por hipótesis $\mu(\widetilde{F}) = 0$ ó $\mu(X \setminus \widetilde{F}) = 0$ pero como $\mu(\widetilde{F}) \geq \mu(F) > 0$, se debe tener $\mu(X \setminus \widetilde{F}) = 0$ entonces,

$$0 < \mu(G) = \mu(G \cap \widetilde{F}) + \mu(G \setminus \widetilde{F}) \leq \mu(G \cap \widetilde{F}) + \mu(X \setminus \widetilde{F}) = 0,$$

lo cual es imposible.

(iii) implica (iv). Sean $F, G \in \mathscr{F}$ con $\mu(F)\mu(G) > 0$, entonces por (iii) existe una $n \in \mathbb{N}$ tal que $\mu(T^{-n}(F) \cap G) > 0$. Sea $n_1 = \text{mín}\{n \in \mathbb{N} : \mu(T^{-n}(F) \cap G) > 0\}$, entonces $\mu(T^{-n_1}(F))\mu(G) > 0$, nuevamente por (iii) existe una $n \in \mathbb{N}$ tal que $\mu(T^{-n}(T^{-n_1}(F)) \cap G) > 0$, consideramos $n_2 = n_1 + \text{mín}\{n \in \mathbb{N} : \mu(T^{-n}(T^{-n_1}(F)) \cap G) > 0\}$. Continuando de esta manera obtenemos $n_1 < n_2 < \cdots$ tales que $\mu(T^{-n_i}(F) \cap G) > 0$ para toda $i \geq 1$ lo que termina la prueba.

(iv) implica (i). Supongamos que (i) es falso, entonces existe $F \in \mathscr{F}$ tal que $T^{-1}(F) = F$ y $\mu(F)\mu(X \setminus F) > 0$, entonces $\mu(\bigcup_{n=1}^{\infty} T^{-n}(F) \cap (X \setminus F)) = 0$, lo que prueba que (iv) es falso. ☐

Observación 1.21. Examinando la condición (iii) del teorema anterior, podemos aclarar la pregunta realizada en la Observación 1.18 en el siguiente sentido: en un sistema ergódico se cumple que para cualesquiera dos regiones del sistema A y B con medida positiva, entonces existe una $n > 0$ y $a \in A$ tal que $T^n(a) \in B$, en otras palabras se puede acceder a cualquier región desde cualquier otra con iteraciones positivas de T.

A continuación daremos otros criterios relacionados con el operador inducido por una transformación preservadora de medida para verificar la ergodicidad de una transformación.

Teorema 1.12. *Sea $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ una tranformación conservativa y que preserva a μ. Las siguientes condiciones son equivalentes:*

(i) T es ergódica.

(ii) Si $f : X \to \mathbb{R}$ es una función $\mathscr{F}$-medible y $f(T(x)) \geq f(x)$ casi dondequiera, entonces f es constante casi dondequiera.

*(iii) Si $f : X \to \mathbb{R}$ es una función $\mathscr{F}$-medible y **T-invariante** (es decir $f(T(x)) = f(x)$ casi dondequiera $x \in X$), entonces f es constante casi dondequiera.*

*(iv) Si $f : X \to \mathbb{R}$ es una función $\mathscr{F}$-medible y **estríctamente T-invariante** (es decir $f(T(x)) = f(x)$ (para toda $x \in X$), entonces f es constante casi dondequiera.*

Además si $\mu(X) < \infty$.

(v) Si $f \in L_p^{\mathbb{R}}(\mu)$ es T-invariante, entonces f es constante casi dondequiera ($p \geq 1$).

(vi) Si $f \in L_p^{\mathbb{R}}(\mu)$ es estrictamente T-invariante, entonces f es constante casi dondequiera ($p \geq 1$).

Demostración. Es claro que se tiene (ii) implica (iii), (iii) implica (iv), (iii) implica (v), (iv) implica (vi), por lo que sólo basta probar (i) implica (iii), (iv) implica (i), (i) implica (ii), finalmente (vi) implica (i).

(i) implica (iii). Supongamos que $f \circ T = f$ c.d. y que T es ergódica para cada $k \in \mathbb{Z}$ y $n \in \mathbb{N}$. Definimos:

$$X(k,n) = \left\{x \in X : \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\right\} = f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right) \ (\in \mathscr{F}) \ k \in \mathbb{Z}, \ n \in \mathbb{N}.$$

Consideremos $X(k,n) \bigtriangleup T^{-1}(X(k,n))$, veamos que el conjunto anterior está contenido en $\{x \in X : f(T(x)) \ne f(x)\}$. Sea $x \in X(k,n) \bigtriangleup T^{-1}(X(k,n))$, entonces $x \in X(k,n) \setminus T^{-1}(X(k,n))$ ó $x \in T^{-1}(X(k,n)) \setminus X(k,n)$, así que:

$$\frac{k}{n} \le f(x) < \frac{k+1}{n} \quad \text{y} \quad f(T(x)) \notin \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) \quad \text{ó}$$

$$\frac{k}{n} \le f(T(x)) < \frac{k+1}{n} \quad \text{y} \quad f(x) \notin \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

entonces $f(x) \ne f(T(x))$, por hipótesis $\mu(X(k,n) \bigtriangleup T^{-1}(X(k,n))) = 0$. Por el Teorema 1.11(ii) se tiene que $\mu(X(k,n)) = 0$ ó $\mu(X \backslash X(k,n)) = 0$. Pero para toda $n \in \mathbb{N}$ fija $X = \bigcup_{k\in\mathbb{Z}} X(k,n)$ (disjunta), (ya que $\mathbb{R} = \bigcup_{k\in\mathbb{Z}}[k/2^n, (k+1)/2^n))$, entonces existe una única k_n tal que $\mu(X \backslash X(k_n,n)) = 0$ (sea $n \in \mathbb{N}$, si para toda $k \in \mathbb{Z}\mu(X \backslash X(k,n)) \ne 0$, entonces $\mu(X) = 0$, lo cual no es posible, por lo que para cada $n \in \mathbb{N}$ tenemos que existe una $k(n)$ tal que $\mu(X \backslash X(k(n),n)) = 0$, sin embargo estas $k(n)$ deben ser la misma k ya que si existen $k_1(n) \ne k_2(n)$, tales que $\mu(X \backslash X(k_1(n),n)) = \mu(X \backslash X(k_2(n),n)) = 0$; como $X(k_1(n),n) \subset X \backslash X(k_2(n),n)$, entonces $0 = \mu(X(k_1(n),n)) = \mu(X)$ lo cual no es posible, por lo tanto para cada $n \in \mathbb{N}$ existe una única $k(n) = k_n$ tal que $\mu(X \setminus X(k_n,n)) = 0$). Sea $Y = \bigcap_{n=1}^{\infty} X(k_n,n)$, entonces también $\mu(X \setminus Y) = 0$ puesto que $\mu(X \setminus Y) \le \sum_{n=0}^{\infty} \mu(X \setminus X(k_n\,,n)) = 0$, por lo tanto $f(x) = \lim_{n\to\infty} k_n/2^n$ para todo $x \in Y$[6], es decir, f es constante casi dondequiera.

(iv) implica (i). Supongamos que $T^{-1}(E) = E$ $(E \in \mathscr{F})$ entonces $f = \chi_E$ es $\mathscr{F}$-medible y $f \circ T = f$, entonces f es constante c.d., por lo que $f = 0$ c.d. ó $f = 1$ c.d., es decir, $\mu(E) = 0$ ó $\mu(X \setminus E) = 0$ (respectivamente), por lo tanto T es ergódica.

(vi) implica (i). Supongamos que $T^{-1}(E) = E$ $(E \in \mathscr{F})$, entonces $f = \chi_E \in L_p(\mu)$ (ya que $\mu(X) < \infty$) y $f \circ T = f$, así que f es constante c.d., por lo tanto $\mu(E) = 0$ ó $\mu(X \setminus E) = 0$.

(i) implica (ii). Supongamos que T es ergódica y que $f(T(x)) \ge f(x)$ c. d. Si f no es constante c. d., entonces existe $y \in \mathbb{R}$ tal que $A = \{x \in X | f(x) \ge y\} = f^{-1}([y,\infty))$ tal que $0 < \mu(A) < \mu(X)$. Pero $A \subset T^{-1}(A)$ c.d. por lo que $\mu(A \bigtriangleup T^{-1}(A)) = 0$, así que $\mu(A) = 0$ ó $\mu(A) = \mu(X)$, lo cual contradice que $0 < \mu(A) < \mu(X)$. ☐

Observación 1.22. En el resultado anterior los incisos (iii) y (iv) las funciones pueden ser cambiadas por funciones $f : X \to \mathbb{C}$, análogamente (v) y (vi) por funciones en $L_p^{\mathbb{C}}(\mu)$.

[6]Note que la sucesión $\{[\,k_n/2^n\,, k_{n+1}/2^{n+1}\,]\}_{n=1}^{\infty}$ está anidada (ya que, existe una única k_1 tal que $\mu(X \backslash X(k_1,1)) = 0$; como existe una única k_2, entonces $\mu(X \backslash X(2k_1,2)) = 0$ ó $\mu(X \backslash X(2k_1+1,2)) = 0$ continuando así se tiene la sucesión mencionada está anidada) y la longitud de los intervalos tiende a cero por el Principio de Intervalos Anidados (PIA) (ver Bartle [1, pág. 46]), la intersección consiste de un solo punto $l = \lim_{n\to\infty} \frac{k_n}{2^n}$.

A continuación veremos dos pruebas de ergodicidad, basadas en el teorema anterior.

Teorema 1.13. *$T : S^1 \to S^1$ dada por $T(z) = az$ es ergódica con $a = e^{i\rho}$ donde $\frac{\rho}{2\pi} \notin \mathbb{Q}$.*

Demostración. Sea $f \in L_2(S^1, \mathscr{B}_{S^1}, \mu)$ estrictamente invariante. Por Teoría de Fourier $\sum_{-\infty}^{\infty} \hat{f}(n)e^{inx}$ (la serie de Fourier de f) converge en L_2 a f (ver Observación A.3, pág. 128). Por otro lado la serie de Fourier de $f \circ T$ está dada por $\sum_{-\infty}^{\infty} \hat{f}(n)e^{in\rho}e^{inx}$, pero $f \circ T = f$ y por la unicidad de los coeficientes de Fourier debe tenerse:

$$\hat{f}(n) = \hat{f}(n)e^{in\rho} \quad \text{para toda } n \in \mathbb{Z}, \text{entonces } \hat{f}(n) = 0 \quad \text{para toda } n \neq 0,$$

por lo tanto $f = \hat{f}(0)$ casi dondequiera. □

Teorema 1.14. *$T : S^1 \to S^1$ dada por $T(z) = z^k$ es ergódica si y sólo si $k \neq \pm 1$.*

Demostración. Razonando como en el teorema anterior si $f \in L_2(\mu)$ es tal que $f \circ T = f$, entonces $f \sim \sum_{-\infty}^{\infty} \hat{f}(n)e^{inx}$ y también $f \sim \sum_{-\infty}^{\infty} \hat{f}(n)e^{iknx}$ y por unicidad de los coeficientes de Fourier obtenemos: $\hat{f}(n) = \hat{f}(kn) = \cdots = \hat{f}(k^j n)$ para toda $j \geq 0$. Supongamos que $|k| > 1$, entonces por desigualdad de Bessel (ver Teorema A.2, pág. 128):

$$\sum_{j \in \mathbb{Z}} \left| \hat{f}(k^j n) \right|^2 \leq \|f\|_2^2 < \infty \text{ por lo tanto } \hat{f}(l) = 0 \quad \text{para toda } l \neq 0 \text{ y } f = \hat{f}(0) \text{ casi dondequiera.}$$

Inversamente si $k = \pm 1$ basta tomar $f : S^1 \to \mathbb{R}$ definida por $f(z) = \Re(z)$ entonces f **no** es constante c.d. (donde $\Re(z)$ es la parte real de z, observemos que $\Re(z) = \Re(\overline{z}) = \Re(z^{-1})$ y $\Re(z) = \Re(z)$), entonces $f \circ T = f$ por lo que T no es ergódica. □

Ejemplo 1.13. Si T^n es ergódica para alguna $n \in \mathbb{N}$, entonces T es ergódica.

Demostración. Supongamos que $f \circ T = f$, entonces $f \circ T^2 = f \circ T = f, \ldots, f \circ T^n = f$ por el Teorema 1.12(iv) f es constante c.d., por lo tanto T es ergódica. □

Observación 1.23. El ejemplo anterior nos asegura la ergodicidad de T cuando alguna de sus potencias lo sea.

Ejemplo 1.14. Sea $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ preservadora de medida invertible. Entonces T es ergódica si y sólo si T^{-1} es ergódica.

Demostración.

(Necesidad). Sea $F \in \mathscr{F}$ tal que $(T^{-1})^{-1}(F) = F$, como T es invertible se tiene que $(T^{-1})^{-1}(F) = T(F) = F$, por lo que $F = T^{-1}(F)$, entonces $\mu(F) = 0$ ó $\mu(X \setminus F) = 0$.

(Suficiencia). Sea $F \in \mathscr{F}$ tal que $T^{-1}(F) = F$, entonces $F = T(F) = (T^{-1})^{-1}(F)$, como T^{-1} es ergódica se tiene que $\mu(F) = 0$ ó $\mu(X \setminus F) = 0$. □

1.6. Teoremas Ergódicos

Sea $(X, \mathscr{F}, \mu)$ un espacio de medida, por el Teorema de Poincaré y la Observación 1.15, sabemos que la serie $\sum_{i=0}^{\infty} \chi_F(T^i(x))$ diverge si $0 < \mu(F)$ y T es medible y conservativa. Pero ¿qué tan rápido diverge?, de manera exponencial, polinomial o lineal. Podremos decir que diverge de manera lineal salvo un conjunto de medida cero si T es una transfromación preservadora de medida y el espacio de medida es σ-finito. Además si T es ergódica, entonces no sólo se puede decir qué tan rápido converge si no a dónde converge exactamente. Más aún considerando funciones $f \in L_1(X, \mathscr{F}, \mu)$ más complejas, se probará que existe una función $f^* \in L_1(X, \mathscr{F}, \mu)$ T-invariante y tal que: $\{\sum_{j=0}^{n-1} f(T^j(x))\} \sim \{nf^*(x)\}$ casi dondequiera $x \in X$. Si T es además ergódica, entonces f^* es constante casi dondequiera y su valor es igual a $1/\mu(X) \int f d\mu$. Lo anterior es consecuencia del teorema más importante de este trabajo, el Teorema Ergódico Individual.

Antes de comenzar veamos la importancia de este teorema en Teoría de la Probabilidad. Si $f \in L_1^{\mathbb{R}}(\mu)$ y $\mu(X) = 1$, diremos que f es una variable aleatoria con esperanza finita (aqui la esperanza de f es igual al número $\int f d\,\mu$). Un resultado que se utiliza frecuentemente en la Teoría de la Probabilidad es la Ley Fuerte de los Grandes Números, la cual nos dice que: si $\{f_n\}_{n=1}^{\infty} \subset L_1^{\mathbb{R}}(\mu)$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, entonces $\{(1/n)\sum_{j=1}^{n} f_j(x)\}$ converge en probabilidad a la esperanza de la variable aleatoria f_1 (en este caso convergencia en probabilidad es la misma que convergencia casi dondequiera). El Teorema Ergódico Individual es una generalización de esta ley. En Física la importancia de este teorema radica en la hipótesis ergódica de Boltzmann, ésta se pregunta sobre la existencia de límites de la forma $\{\text{lím}(1/n)\sum_{j=0}^{n-1} f(T^j(x))\}$ y si existen se pregunta si $\{\text{lím}(1/n)\sum_{j=0}^{n-1} f(T^j(x)) = \int f d\,\mu\}$. Este teorema responde estas preguntas y dice que en los sistemas ergódicos es válida la igualdad anterior casi dondequiera. Para probar este teorema necesitaremos los siguientes resultados.

Teorema 1.15 (Teorema Ergódico Maximal)**.** *Sea* $(X, \mathscr{F}, \mu)$ *un espacio de medida y* $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ *una transformación preservadora de medida. Sea* $f \in L_1^{\mathbb{R}}(X, \mathscr{F}, \mu)$*, si* $\mathcal{U} : L_1^{\mathbb{R}}(\mu) \to L_1^{\mathbb{R}}(\mu)$ *es un operador lineal* ***positivo*** *y una* ***contracción*** $(\|\mathcal{U}(f)\|_1 \leq \|f\|_1$ *para toda* $f \in L_1^{\mathbb{R}}(\mu))$*, entonces: si* $f_0 = 0$, $f_n = f + \mathcal{U}(f) + \cdots + \mathcal{U}^{n-1}(f)$ $(n \geq 1)$ *y* $F_N = \text{máx}_{0 \leq n \leq N}\{f_n\}$*, entonces:*

$$\int_{\{x:F_N(x)>0\}} f\, d\mu \geq 0.$$

Demostración. (Adriano Garsia, 1961). Claramente $F_N \in L_1(\mu)$. Para $0 \leq n \leq N$, como $F_N \geq f_n$ y $\mathcal{U}$ es un operador lineal positivo, entonces $\mathcal{U}(F_N) \geq \mathcal{U}(f_n)$, por lo que:

$$\mathcal{U}(F_N) + f \geq \mathcal{U}(f_n) + f = f_{n+1}.$$

De las desigualdades anteriores obtenemos: $\mathcal{U}(F_N) + f \geq \text{máx}_{1 \leq n \leq N}\{f_n\}$. Así que:

$$(\mathcal{U}(F_N) + f)(x) \geq \max_{0 \leq n \leq N}\{f_n(x)\} \quad \text{si} \quad F_n(x) > 0,$$

es decir, $\mathcal{U}(F_N) + f \geq F_N$ en $\{x : F_N(x) > 0\}$, por lo tanto:

$$\begin{aligned}
\int\limits_{\{x:F_N(x)>0\}} f\,d\mu &\geq \int\limits_{\{x:F_N(x)>0\}} F_N d\mu - \int\limits_{\{x:F_N(x)>0\}} \mathcal{U}(F_N) d\mu \\
&= \int\limits_X F_N\,d\mu - \int\limits_{\{x:F_N(x)>0\}} \mathcal{U}(F_N)\,d\mu \quad \text{pues } F_N = 0 \text{ en } X \setminus \{x : F_N(x) > 0\} \\
&\geq \int\limits_X F_N\,d\mu - \int\limits_X \mathcal{U}(F_N)\,d\mu \;\; \text{ya que } F_N \geq 0 \text{ por lo tanto } \mathcal{U}(F_N) \geq 0 \\
&\geq 0 \quad \text{ya que } \|\mathcal{U}\| \leq 1.
\end{aligned}$$

□

Corolario 1.15.1. *Sea $T : (X, \mathscr{F}, \mu) \to (X, \mathscr{F}, \mu)$ preservadora de medida. Para cada $g \in L^1_{\mathbb{R}}(\mu)$ y $\alpha \in \mathbb{R}$ definimos:*

$$B_\alpha = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i(x)) > \alpha \right\} \quad \textit{entonces :}$$

$$\int\limits_{B_\alpha \cap A} g\,d\mu \geq \alpha\mu(B_\alpha \cap A) \quad \textit{para todo} \quad A \in \mathscr{F} \quad \textit{tal que} \quad T^{-1}(A) = A \quad \textit{y} \quad \mu(A) < \infty.$$

Demostración.

Caso 1) $A = X$ y $\mu(X) < \infty$. Sea $f = g - \alpha$, como $\mu(X) < \infty$, entonces $f \in L_1^{\mathbb{R}}(\mu)$. Observamos que $B_\alpha = \bigcup_{N=0}^{\infty} \{x : F_N(x) > 0\}$, ya que:

$$\begin{aligned}
B_\alpha = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} \sum_{i=0}^{n-1} \big(g(T^i(x)) - \alpha\big) > 0 \right\} &= \left\{ x : \sup_{n\geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) > 0 \right\} \\
&= \left\{ x : \sup_{n\geq 1} \{f_n(x)\} > 0 \right\} \\
&= \left\{ x : \sup_{N\geq 0} \{F_N(x)\} > 0 \right\} \\
&= \bigcup_{N=0}^{\infty} \{x : F_N(x) > 0\}.
\end{aligned}$$

Por el Teorema Ergódico Maximal aplicado al operador $\mathcal{U}_T$ de $L_1(\mu)$ en sí mismo obtenemos:

$$\int\limits_{B_\alpha} f\,d\mu \geq 0 \;\; \text{es decir} \;\; \int\limits_{B_\alpha} g\,d\mu \geq \alpha\mu(B_\alpha).$$

(Observemos que $0 = F_0 \leq F_1 \leq F_2 \ldots$ por lo que $\{x : F_N(x) > 0\} \subset \{x : F_{N+1}(x) > 0\}$ así que de hecho: $\int_{B_\alpha} f\,d\mu = \lim_{N\to\infty} \int_{\{x:F_N(x)>0\}} f\,d\mu \geq 0$ por el T.C.D. (ver Teorema B.4, pág. 133).

Caso 2) Sea $A \in \mathscr{F}$ con $T^{-1}(A) = A$ y $\mu(A) < \infty$. Definimos una transformación, $T|_A : (A, \mathscr{F} \cap A, \mu_A) \to (A, A \cap \mathscr{F}, \mu_A)$ poniendo $T|_A(x) = T(x)$ para toda $x \in A$, en donde

$\mathscr{F} \cap A = \{E \cap A : E \in \mathscr{F}\}$ y $\mu_A(E \cap A) = \mu(E \cap A)$, entonces $T|_A$ es una transformación preservadora de medida. Repitiendo exactamente los mismos argumentos anteriores con $T|_A$ en lugar de T, A en lugar de X y μ_A en lugar de μ obtenemos:

$$\int\limits_{B_\alpha} g \, d\mu_A \geq \alpha \mu_A(B_\alpha) \quad \text{o bien} \quad \int\limits_{B_\alpha \cap A} g \, d\mu \geq \alpha \mu(B_\alpha \cap A). \qquad \square$$

Ahora se prueba el teorema principal de este trabajo con la ayuda del lema y corolario anterior. Cabe señalar que esta prueba no es la prueba original dada por Birkhoff.

Teorema 1.16 (Teorema Ergódico Individual Birkhoff-Khintchine 1931)**.** *Sean $(X, \mathscr{F}, \mu)$ un espacio de medida σ-finito, $T : X \to X$ una transformación preservadora de medida y $f \in L_1(\mu)$ dados. Entonces la sucesión de funciones $\left\{(1/n)\sum_{i=0}^{n-1} f(T^i(x))\right\}_{n=1}^{\infty}$ converge casi dondequiera a una función $f^* \in L_1(\mu)$ tal que:*

(i) $f^* \circ T = f^*$ casi dondequiera.

(ii) $\int_F f^* \, d\mu = \int_F f \, d\mu$ para todo $F \in \mathscr{F}$ con $F = T^{-1}(F)$ y $\mu(F) < \infty$. (En particular, si $\mu(X) < \infty$ entonces $\int f^* \, d\mu = \int f \, d\mu$).

(iii) Si T es además ergódica, entonces: $f^* = (1/\mu(X)) \int f \, d\mu$ casi dondequiera si $\mu(X) < \infty$, y $f^* = 0$ casi dondequiera si $\mu(X) = \infty$.

Demostración.

(i). Sea $f \in L_1(\mu)$ dada, definimos:

$$f_*(x) = \liminf_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \quad \text{y} \quad f^*(x) = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)).$$

Claramente $f_* \leq f^*$. Por otro lado:

$$\frac{1}{n+1} \sum_{i=0}^{n} f(T^i(x)) = \left(\frac{n}{n+1}\right) \frac{1}{n} \sum_{i=0}^{n-1} f(T^{i+1}(x)) + \frac{f(x)}{n+1}.$$

Observando que $\lim_{n\to\infty} f(x)/(n+1) = 0$ y $\lim_{n\to\infty} n/(n+1) = 1$. Tomando $\limsup$ y $\liminf$ de ambos lados obtenemos que $f^*(T(x)) = f^*(x)$ y $f_*(T(x)) = f_*(x)$ (ver Teorema A, pág. 126). Consideramos $E = \left\{x \in X : (1/n)\sum_{i=0}^{n-1} f(T^i(x)) \text{ no converge}\right\}$, entonces $E = \bigcup_{\alpha, \beta \in \mathbb{Q}} E_{\alpha, \beta}$ donde $E_{\alpha,\beta} = \{x \in X : f_*(x) < \beta, \alpha < f^*(x)\}$. Como f_* y f^* son estrictamente T-invariantes, entonces: $T^{-1}(E_{\alpha,\beta}) = E_{\alpha,\beta}$. Probaremos que $\mu(E_{\alpha,\beta}) = 0$ para toda α, β pero primero estableceremos $\mu(E_{\alpha,\beta}) < \infty$. **Supongamos que** $\alpha > 0$. Note que:

$$E_{\alpha,\beta} = E_{\alpha,\beta} \cap \left\{x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) > \alpha\right\},$$

ya que si $x \in E_{\alpha,\beta}$ (con $\alpha > 0$) se tiene que $\limsup \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) > \alpha$, es decir, $\inf\{\sup_{n\geq m}\{(1/n)\sum_{i=0}^{n-1} f(T^i(x))\} : m \in \mathbb{N}\} > \alpha$, por lo que $\sup_{n\geq 1}(1/n)\sum_{i=0}^{n-1} f(T^i(x)) > \alpha$, así que $x \in B_\alpha$ por lo tanto $x \in E_{\alpha,\beta} \cap B_\alpha$. Para cada $C \in \mathscr{F}$, $C \subset E_{\alpha,\beta}$ de medida **finita**, sea $h = f - \alpha\chi_C \in L_1(\mu)$ (el cual existe ya que $X = \bigcup_{n=1}^{\infty} X_n$ con $\mu(X_n) < \infty$ para toda $n \geq 1$ como $E_{\alpha,\beta} = \bigcup_{n=1}^{\infty} X_n \cap E_{\alpha,\beta}$, se tiene que $\mu(X_n \cap E_{\alpha,\beta}) < \infty$), entonces por el Teorema Ergódico Maximal:

$$\int\limits_{\bigcup\limits_{N=0}^{\infty}\{x:H_N(x)>0\}} \left(f - \alpha\chi_C\right) d\mu \geq 0 \quad \text{con} \quad H_N = \max_{1\leq n\leq N}\left\{0, \sum_{i=0}^{n-1} h \circ T^i\right\}.$$

Como $C \subset E_{\alpha,\beta} \subset \bigcup_{N=0}^{\infty}\{x \in X : H_N(x) > 0\}$ tenemos:

$$\int\limits_X |f|\, d\mu \geq \int\limits_{\{\sup\limits_N H_N(x)>0\}} |f|\, d\mu \geq \left|\int\limits_{\{\sup\limits_N H_N(x)>0\}} f\, d\mu\right| \geq \alpha\mu(C),$$

por lo que $\mu(C) < (1/\alpha)\int_X |f|\, d\mu < \infty$, por lo tanto $\mu(E_{\alpha,\beta}) < \infty$ (ver Proposición B.2, pág. 131), (si $\alpha \leq 0$, entonces $\beta < 0$. Se puede aplicar el argumento anterior a $-f$ y $-\beta$ en vez de f y α para obtener $\mu(E_{\alpha,\beta}) < \infty$. Observemos que en este caso $E_{\alpha,\beta}(f) = E_{-\beta,-\alpha}(-f)$ ya que $(-f)_* = -f^*$ y $(-f)^* = -f_*$ (ver Propiedades A, pág. 126) y obtenemos $\mu(E_{\alpha,\beta}(f)) \leq (1/(-\beta))\int_X |f|\, d\mu < \infty$).

(i)$'$ $\mu(E_{\alpha,\beta}) = 0$. Sea $B_\alpha = \left\{x \in X : \sup_{n\geq 1}(1/n)\sum_{i=0}^{n-1} f(T^i(x)) > \alpha\right\}$, entonces por el corolario del Teorema Maximal obtenemos:

$$\int\limits_{E_{\alpha,\beta}} f\, d\mu = \int\limits_{E_{\alpha,\beta}\cap B_\alpha} f\, d\mu \geq \alpha\mu(E_{\alpha,\beta} \cap B_\alpha) = \alpha\mu(E_{\alpha,\beta}).$$

$$\text{Es decir,} \quad \alpha\mu(E_{\alpha,\beta}) \geq \int\limits_{E_{\alpha,\beta}} f\, d\mu. \tag{*}$$

Reemplazando f, α, β por $-f, -\beta, -\alpha$ (respectivamente) obtenemos:

$$\int\limits_{E_{\alpha,\beta}} f\, d\mu \leq \beta\mu(E_{\alpha,\beta}). \tag{**}$$

Como $\beta < \alpha$ debe tenerse que $\mu(E_{\alpha,\beta}) = 0$ para toda α, β, por lo tanto $\mu(E) = 0$, entonces $\left\{(1/n)\sum_{i=0}^{n-1} f(T^i(x))\right\}_{n=1}^{\infty}$ converge a $f^*(x)$ $(= f_*(x))$ para toda $x \notin E$.

(ii) Probaremos primero que $f^* \in L_1(\mu)$. Sea $g_n(x) = \left|(1/n)\sum_{i=0}^{n-1} f(T^i(x))\right|$ entonces $g_n \geq 0$ y $\int g_n\, d\mu \leq (1/n)\sum_{i=0}^{n-1}\int |f(T^i(x))|\, d\mu = \int |f|\, d\mu < \infty$, por lo que $g_n \in L_1(\mu)$ para toda n, además $g_n \to |f^*|$ c.d. Así que usando el lema de Fatou (ver Corolario B.3.1, pág. 132), obtenemos:

$$\int |f^*|\, d\mu \leq \liminf \int g_n\, d\mu \leq \int |f|\, d\mu < \infty \quad \text{por lo tanto} \quad f^* \in L_1(\mu).$$

Sea $F \in \mathscr{F}$ con $F = T^{-1}(F)$ y $\mu(F) < \infty$. Para cada $k \in \mathbb{Z}$ y $n \in \mathbb{N}$ sea $D_k^n = \{x \in F : k/n \leq f^*(x) < (k+1)/n\}$, entonces si $\varepsilon > 0$ es suficientemente pequeña obtenemos:

$$D_k^n \subset B_{(\,k/n-\varepsilon)} = \left\{ x \in X : \sup_{n\geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) > \frac{k}{n} - \varepsilon \right\},$$

(pues si $x \in D_k^n$, se tiene que $f^*(x) \geq k/n$, por lo que dada $\varepsilon > 0$ existe m tal que $(1/m)\sum_{i=0}^{m-1} f(T^i(x)) > k/n - \varepsilon$, por lo tanto $x \in B_{(\,k/n-\varepsilon)}$). Así que: $D_k^n \cap B_{(\,k/n-\varepsilon)} = D_k^n$ y como D_k^n es T-invariante y $\mu(D_k^n) < \infty$ concluimos que: $\int_{D_k^n} f\, d\mu \geq (\,k/n - \varepsilon)\mu(D_k^n)$ (por el Corolario 1.15.1). Como $\varepsilon > 0$ es arbitrario obtenemos: $\int_{D_k^n} f\, d\mu \geq (k/n)\mu(D_k^n)$, lo que implica que:

$$\int\limits_{D_k^n} f^*\, d\mu \leq \frac{k+1}{n}\mu(D_k^n) \leq \int\limits_{D_k^n} f\, d\mu + \frac{1}{n}\mu(D_k^n) \quad \text{para toda } k \in \mathbb{Z},$$

sumando sobre k, obtenemos que:

$$\int_F f^*\, d\mu \leq \int_F f\, d\mu + \frac{1}{n}\mu(F) \ \ \text{(Note: } F = \bigcup_{k=-\infty}^{\infty} D_k^n \text{ para toda } n \in \mathbb{N}).$$

Haciendo tender $n \to \infty$ obtenemos finalmente que: $\int_F f^*\, d\mu \leq \int_F f\, d\mu$. Aplicando lo anterior a $-f$ en lugar de f, hallamos:

$$\int_F (-f)^*\, d\mu \leq \int_F (-f)\, d\mu \ \text{ es decir } \ \int_F f\, d\mu \leq \int_F f_*\, d\mu,$$

pero $f_* = f^*$ c.d. así que $\int_F f\, d\mu \leq \int_F f^*\, d\mu$, por lo que tenemos la igualdad.

(iii) Supongamos que T es ergódica. Como f^* es T-invariante, se sigue del Teorema 1.12 que $f^* = k$ c.d. (k constante). Si $\mu(X) < \infty$, entonces como $\int f^*\, d\mu = \int f\, d\mu$ debemos tener $k = (1/\mu(X))\int_X f\, d\mu$. Si $\mu(X) = \infty$, entonces como $f^* \in L_1(\mu)$ y es constante c. d. se debe tener que $f^* = 0$ c. d. □

Observaciones 1.24.

(1) Sean $(X, \mathscr{F}, \mu)$ un espacio de medida finita, $T : X \to X$ una transformación preservadora de medida ergódica y $F \in \mathscr{F}$, entonces el Teorema Ergódico Individual nos asegura que $(1/n)\sum_{i=0}^{n-1} \chi_F(T^i(x)) \to \mu(F)/\mu(X)$ casi dondequiera, es decir, para casi todo $x \in X$ el número de visitas que hace x a F en promedio es igual a la proporción de medida que tiene F con respecto a $\mu(X)$. En sistemas dinámicos es de interés este hecho ya que relaciona el comportamiento asintótico de las órbitas con la medida del conjunto. Otra manera de ver lo anterior es que si $\mu(F_1) < \mu(F_2)$, intuitivamente el número de veces que x entra a F_1 es menor que el número de veces que entra a F_2.

(2) Utilizando las hipótesis del inciso anterior, definimos el promedio temporal de f en x como sigue:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)), \qquad \text{si este límite existe,}$$

y el promedio espacial de f como sigue:

$$\int_X f d\mu.$$

El Teorema Ergódico Individual nos asegura que los promedios temporales son iguales casi dondequiera al promedio espacial para toda $f \in L_1(\mu)$ si y sólo si T es ergódica (más adelante veremos el reciproco de esta afirmación). Considerando un flujo $\{T_t : t \in \mathbb{R}\}$ de transformaciones preservadoras de medida, la afirmación anterior sigue siendo válida, es decir, $\lim_{T\to\infty}(1/T)\int_0^T f(T_t(x))dt \to \int_X f d\mu$ si el flujo es ergódico (más adelante veremos este resultado).

Ahora veremos algunas consecuencias de este importante teorema.

Corolario 1.16.1. *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida **finita** y $T : X \to X$ una transformación preservadora de medida, entonces T es ergódica si y sólo si para todo $A, B \in \mathscr{F}$ se tiene que:*

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu\left(T^{-i}(A)\cap B\right) \longrightarrow \frac{\mu(A)\mu(B)}{\mu(X)}.$$

Demostración.

(Necesidad). Supongamos que T es ergódica y sea $f = \chi_A$. Entonces por el Teorema 1.16(iii):

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(T^i(x)) \longrightarrow \frac{\mu(A)}{\mu(X)} \ \text{ c. d. o bien } \ \frac{1}{n}\sum_{i=0}^{n-1} \chi_{T^{-i}(A)}(x) \longrightarrow \frac{\mu(A)}{\mu(X)} \ \text{ c. d.}$$

Multiplicando por χ_B obtenemos:

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_{T^{-i}(A)\cap B}(x) \longrightarrow \frac{\mu(A)}{\mu(X)}\chi_B(x) \ \text{ c.d.}$$

Pero la sucesión está dominada por la función integrable χ_B. Por lo que aplicando el Teorema de Convergencia Dominada de Lebesgue (ver Teorema B.5, pág. 133), obtenemos:

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A)\cap B) \longrightarrow \frac{\mu(A)\mu(B)}{\mu(X)}.$$

(Suficiencia). Sea $A \in \mathscr{F}$ tal que $\mu(A \triangle T^{-1}(A)) = 0$, como

$$(A \triangle T^{-i}(A)) \subset \bigcup_{j=0}^{i}(T^{-j}(A) \triangle T^{-j-1}(A)) = \bigcup_{j=0}^{i} T^{-j}(A \triangle T^{-1}(A)),$$

entonces $\mu(A \triangle T^{-i}(A)) = 0$ para toda $i \geq 1$, así que $\mu(T^{-i}(A) \cap A) = \mu(A)$ para toda $i \geq 1$. Por hipótesis tenemos que:

$$\mu(A) = \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap A) \longrightarrow \frac{\mu(A)^2}{\mu(X)} \text{ es decir } \mu^2(A) = \mu(A)\mu(X),$$

por lo que $\mu(A) = 0$ ó $\mu(A) = \mu(X)$, por lo tanto T es ergódica. □

Observación 1.25. Si $\mu(X) = \infty$ se tiene que si T es ergódica, entonces

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) \longrightarrow 0 \quad \text{para toda } A, B \in \mathscr{F} \text{ con } \mu(A)\mu(B) < \infty.$$

Sin embargo el recíproco correspondiente es falso.

Combinando el Teorema Ergódico Individual y el Teorema de Egorov (ver Teorema B.10, pág. 135), obtenemos la siguiente mejora sobre el Teorema 1.16.

Teorema 1.17. *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida finita y $T : X \to X$ una transformación ergódica, entonces para toda $\delta > 0$ y para toda $f \in L_1(\mu)$, existe $F = F(\delta, f) \in \mathscr{F}$ tal que $\mu(F) < \delta$ y $(1/n)\sum_{i=0}^{n-1} f(T^i(x)) \to (1/\mu(X))\int f\, d\mu$* ***uniformemente*** *en $X \setminus F$.*

Demostración. Sea $\delta > 0$ y $f \in L_1(\mu)$ por el Teorema Ergódico Individual:

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \to \frac{1}{\mu(X)}\int f d\,\mu \qquad \text{c.d.},$$

como $\mu(X) < \infty$ por el Teorema de Egorov

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)) \to \frac{1}{\mu(X)}\int f d\,\mu \qquad \text{c.u.},$$

es decir, existe $F = F(\delta, f) \in \mathscr{F}$ tal que $\mu(F) < \delta$ y $(1/n)\sum_{i=0}^{n-1} f(T^i(x)) \to (1/\mu(X))\int f d\,\mu$ uniformemente en $X \setminus F$. □

Ahora examinamos la convergencia en media $p \in [\,1\,, \infty\,)$ de los promedios $\{(1/n)\sum_{i=0}^{n} f \circ T^i\}$.

Teorema 1.18 (Teorema Ergódico L_p (J.Von Neumann))**.** *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida* ***finita****, $T : X \to X$ preservadora de medida, $p \in [\,0\,, \infty\,)$ fija y $f \in L_p(\mu)$ dada, entonces existe $f^* \in L_p(\mu)$ tal que $f^* \circ T = f^*$ casi dondequiera y:*

$$\left\| \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - f^* \right\|_p \longrightarrow 0, \quad \textit{cuando} \quad n \longrightarrow \infty.$$

Demostración.

Caso 1) f es acotada c.d., como $\mu(X) < \infty$, entonces $f \in L_1(\mu)$ y por el Teorema 1.16 se tiene que $(1/n)\sum_{i=0}^{n-1} f \circ T^i(x) \longrightarrow f^*(x)$ c.d. Es claro que f^* es acotada casi dondequiera también y $\left|(1/n)\sum_{i=0}^{n-1} f \circ T^i(x) - f^*(x)\right|^p \longrightarrow 0$ c.d. La sucesión anterior está acotada casi dondequiera para toda $n \in \mathbb{N}$ (De hecho:$\left|(1/n)\sum_{i=0}^{n-1} f \circ T^i(x) - f^*(x)\right|^p \leq 2^p\|f\|_\infty^p$ c.d.), aplicando el Teorema de Convergencia Acotada (ver Teorema B.4, pág. 133), obtenemos que:

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x) - f^*(x)\right\|_p \longrightarrow 0.$$

Caso 2) f general. Sea $\varepsilon > 0$, hallamos $g \in L_\infty(\mu)$ tal que $\|f-g\|_p < \varepsilon$ entonces para toda $n \geq 1$ y $k \geq 1$:

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} f \circ T^i\right\|_p &\leq \left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i\right\|_p + \\
&\quad \left\|\frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i\right\|_p + \\
&\quad \left\|\frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} f \circ T^i\right\|_p \\
&< \varepsilon + \left\|\frac{1}{n}\sum_{i=0}^{n-1} g \circ T^i - \frac{1}{n+k}\sum_{i=0}^{n+k-1} g \circ T^i\right\|_p + \varepsilon \\
&< 3\varepsilon,
\end{aligned}$$

si n se elige suficientemente grande para toda $k \geq 1$ (esto por el caso 1) aplicado a g). Así pues $\left\{(1/n)\sum_{i=0}^{n-1} f \circ T^i\right\}_{n\geq 1}$ es $\|\|_p$-Cauchy y por completez existe $f^* \in L_p(\mu)$ tal que $\left\|(1/n)\sum_{i=0}^{n-1} f \circ T^i - f^*\right\|_p \longrightarrow 0$. Para probar que $f^* \circ T = f^*$ c.d. dada $\varepsilon > 0$ hallamos $N \in \mathbb{N}$ tal que $\left\|f^* - (1/n)\sum_{i=0}^{n-1} f \circ T^i\right\|_p < \varepsilon$ si $n \geq N$, entonces:

$$\begin{aligned}
\|f^* - f^* \circ T\|_p &\leq \left\|f^* - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i\right\| + \left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i - \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^{i+1}\right\|_p + \\
&\quad \left\|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^{i+1} - f^* \circ T\right\|_p \\
&< \varepsilon + \frac{1}{n}\|f - f \circ T^n\|_p + \varepsilon \\
&\leq 2\varepsilon + \frac{2}{n}\|f\|_p \\
&< 3\varepsilon,
\end{aligned}$$

si n es grande por lo tanto $f^* = f^* \circ T$ casi dondequiera. □

Corolario 1.18.1. *Sean $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y T es una transformación preservadora de medida, entonces para toda $f \in L_1(\mu)$, f^* es igual casi dondequiera a la esperanza* *condicional de f con respecto a $\mathscr{I}$ (ver Observación 1.16, pág. 20); es decir, $f^* = E(f|\mathscr{I})$ c.d. (ver Definición B.29, pág. 137). Si además T es ergódica, entonces $\mathscr{I}$ es trivial por lo que: $f^* = E(f)$ casi dondequiera.*

Demostración. Por el Teorema Ergódico Individual, tenemos que f^* es $\mathscr{I}$-medible ya que $f^* \circ T = f^*$, es decir, $T^{-1}(f^{*-1}(B)) = f^{*-1}(B)$; como $f^* \in L_1(\mu)$ y para toda $F \in \mathscr{I}$ se tiene que $\int_F f^*\, d\mu = \int_F f\, d\mu$, entonces f^* cumple con las tres propiedades de la esperanza condicional y por unicidad $f^* = E(f|\mathscr{I})$ casi donde quiera. Si T es ergódica entonces $f^* = (1/\mu(X)) \int f\, d\mu = E(f)$ casi dondequiera. □

Observación 1.26. El corolario anterior es como normalmente se presenta el Teorema Ergódico Individual en Teoría de la Probabilidad. Recordemos que la esperanza condicional de f es una variable aleatoria que contiene la información de f sobre la σ-álgebra con la que estamos condicionando. En otras palabras la esperanza condicional recoge la información sobre una σ-álgebra que sea de nuestro interés.

Recordemos que el operador $\mathcal{U}_T$ es una isometría lineal en el espacio de Hilbert $L_2(\mu)$ (ver Observación 1.8, pág. 15), existirá un análogo al Teorema Ergódico Individual, para esta transformación. Más general si H es un espacio de Hilbert y $\mathcal{U} : H \to H$ es una isometría lineal, ¿Podemos decir algo sobre los promedios $(1/n)\sum_{i=0}^{n-1} \mathcal{U}^i(f)$ para $f \in H$?, la respuesta es afirmativa y es conocido como el Teorema Ergódico en Promedio. Denotaré al producto interior de dos elementos f, g de un espacio de Hilbert H, como $\langle f, g\rangle$. Para probar este teorema necesitaremos los siguientes resultados y conceptos de espacios de Hilbert:

Definición 1.11. Sea H un espacio de Hilbert. El **operador adjunto** $\mathcal{U}^*$ de un operador lineal acotado $\mathcal{U}$ en H, está caracterizado por $\langle \mathcal{U}(f), h\rangle = \langle f, \mathcal{U}^*(h)\rangle$, válido para toda f, h.

Lema 1.7. *Si $\mathcal{U}$ es una isometría lineal sobre un espacio de Hilbert H, entonces $\mathcal{U}^* \circ \mathcal{U} = I_H$, donde $\mathcal{U}^*$ es el operador adjunto de $\mathcal{U}$, y I_H es la función identidad en H.*

Demostración. Si $\|\mathcal{U}(f)\| = \|f\|$ se tiene que $\langle \mathcal{U}(f), \mathcal{U}(f)\rangle = \langle f, f\rangle$, así que:

$$\begin{aligned} \|\mathcal{U}^*(\mathcal{U}(f)) - f\|^2 &= \langle \mathcal{U}^*(\mathcal{U}(f)), \mathcal{U}^*(\mathcal{U}(f))\rangle - \langle \mathcal{U}^*(\mathcal{U}(f)), f\rangle - \langle f, \mathcal{U}^*(\mathcal{U}(f))\rangle + \langle f, f\rangle \\ &= \langle \mathcal{U}(f), \mathcal{U}(f)\rangle - \langle \mathcal{U}(f), \mathcal{U}(f)\rangle - \langle \mathcal{U}(f), \mathcal{U}(f)\rangle + \langle \mathcal{U}(f), \mathcal{U}(f)\rangle \\ &= 0, \end{aligned}$$

por lo que $\mathcal{U}^*(\mathcal{U}(f)) = f$ para toda f por lo tanto $\mathcal{U}^* \circ \mathcal{U} = I_H$. □

Lema 1.8. *Si $\mathcal{U}^*$ es el operador adjunto de $\mathcal{U}$, entonces*

$$I_{\mathcal{U}} = \{h \in H : \mathcal{U}(h) = h\} = \{h \in H : \mathcal{U}^*(h) = h\} = I_{\mathcal{U}^*}.$$

Demostración. Supongamos que $h \in I_{\mathcal{U}}$, es decir, $\mathcal{U}(f) = f$; aplicando de ambos lados el operador adjunto $\mathcal{U}^*$ y utilizando el hecho de que $\mathcal{U}^* \circ \mathcal{U} = I_H$, donde I_H es la función identidad, por el Lema 1.7 se tiene que $f = \mathcal{U}^*(f)$, es decir, $f \in I_{\mathcal{U}^*}$. Para el recíproco si $f \in I_{\mathcal{U}}$, se tiene que:

$$\|\mathcal{U}(f) - f\|^2 = \|\mathcal{U}(f)\|^2 - \langle f\,,\mathcal{U}(f)\rangle - \langle \mathcal{U}(f)\,,f\rangle + \|f\|^2,$$

pero:

$$\|\mathcal{U}(f)\|^2 = \|f\|^2, \quad \langle f\,,\mathcal{U}(f)\rangle = \langle \mathcal{U}^*(f)\,,f\rangle = \langle f\,,f\rangle = \|f\|^2,$$
$$\langle \mathcal{U}(f)\,,f\rangle = \langle f\,,\mathcal{U}^*(f)\rangle = \langle f\,,f\rangle = \|f\|^2,$$

por lo tanto $\|\mathcal{U}(f) - f\|^2 = 0$, es decir, $f \in I_{\mathcal{U}}$. □

Ahora podemos pasar a la demostración del Teorema J. Von Neumann.

Teorema 1.19 (J. Von Neumann 1932). *Sea H un espacio de Hilbert sobre $\mathbb{C}$ o $\mathbb{R}$, y $\mathcal{U} : H \to H$ una isometría lineal. Denotamos por $I_{\mathcal{U}} = \{h \in H : \mathcal{U}(h) = h\}$, entonces para toda $f \in H$, la sucesión de promedios $(1/n)\sum_{i=0}^{n-1}\mathcal{U}^i(f)$ converge en H a la* ***proyección ortogonal*** *de f sobre $I_{\mathcal{U}}$ (que denotaremos por $proj(f|I_{\mathcal{U}})$).*[7]

Demostración. Si $f \in I_{\mathcal{U}}$ claramente $(1/n)\sum_{i=0}^{n-1}\mathcal{U}^i(f) \to f = proj(f|I_{\mathcal{U}})$, así pues el teorema es válido para $f \in I_{\mathcal{U}}$; supongamos que f es de la forma $f = \mathcal{U}(g) - g$ para alguna $g \in H \cdots (*)$. Entonces:

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=0}^{n-1}\mathcal{U}^i(f)\right\| &= \left\|\frac{1}{n}\sum_{i=0}^{n-1}\left(\mathcal{U}^{i+1}(g) - \mathcal{U}^i(g)\right)\right\| \quad \text{la cual es una suma telescópica}\\
&= \left\|\frac{1}{n}\left(\mathcal{U}^n(g) - g\right)\right\|\\
&\leq \frac{2}{n}\|g\| \longrightarrow 0 \quad \text{si} \quad n \longrightarrow \infty.
\end{aligned}$$

Por otro lado, si f es de la forma $(*)$, entonces $proj(f|I_{\mathcal{U}}) = 0$. Lo anterior ya que si $f \in I_{\mathcal{U}}$ se tiene que:

$$\begin{aligned}
\langle f\,,h\rangle = \langle \mathcal{U}(g) - g\,,h\rangle &= \langle \mathcal{U}(g)\,,h\rangle - \langle g,h\rangle\\
&= \langle \mathcal{U}(g)\,,\mathcal{U}(h)\rangle - \langle g\,,h\rangle\\
&= \langle g\,,\mathcal{U}^*(\mathcal{U}(h))\rangle - \langle g\,,h\rangle\\
&= \langle g\,,h\rangle - \langle g\,,h\rangle \quad \text{por Lema 1.7}\\
&= 0.
\end{aligned}$$

Por lo tanto $f \perp I_{\mathcal{U}}$, es decir, $proj(f|I_{\mathcal{U}}) = 0$; así pues el teorema es válido para toda f de la forma $(*)$.

[7]Note que $I_{\mathcal{U}}$ es un subespacio **cerrado** de H.

Consideremos $E_{\mathcal{U}} = \{f \in H : f = \mathcal{U}(g) - g$ para alguna $g \in H\}$, entonces $E_{\mathcal{U}}$ es un subespacio vectorial de H, el cual no necesariamente es cerrado, sea $\overline{E_{\mathcal{U}}}$ la cerradura de este espacio. Sea $f \in \overline{E_{\mathcal{U}}}$ y $\varepsilon > 0$, entonces existe $f_\varepsilon \in E_{\mathcal{U}}$ tal que $\|f - f_\varepsilon\| < \varepsilon$, así que:

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f) \right\| &\leq \left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f - f_\varepsilon) \right\| + \left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f_\varepsilon) \right\| \\
&\leq \frac{1}{n}\sum_{i=0}^{n-1} \|\mathcal{U}^i(f - f_\varepsilon)\| + \left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f_\varepsilon) \right\| \\
&< \varepsilon + \left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f_\varepsilon) \right\| \quad \text{pues } \mathcal{U} \text{ es una isometría.}
\end{aligned}$$

Por lo anterior, sabemos que existe una $N \in \mathbb{N}$ tal que si $n \geq N$ entonces:

$$\left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f_\varepsilon) \right\| < \varepsilon, \quad \text{por lo tanto} \quad \lim_{n\to\infty} \left\| \frac{1}{n}\sum_{i=0}^{n-1} \mathcal{U}^i(f) \right\| = 0.$$

Como $E_{\mathcal{U}}^{\perp} = \overline{E_{\mathcal{U}}}^{\perp}$ entonces $proj(f|I_{\mathcal{U}}) = 0$ para toda $f \in \overline{E_{\mathcal{U}}}^{\perp}$ (donde $E_{\mathcal{U}}^{\perp}$ y $\overline{E_{\mathcal{U}}}^{\perp}$ denotan el complemento ortogonal de $E_{\mathcal{U}}$ y $\overline{E_{\mathcal{U}}}$ respectivamente), es decir, el teorema es válido para toda $f \in \overline{E_{\mathcal{U}}}^{\perp}$. Solo nos falta probar que $H = I_{\mathcal{U}} \oplus \overline{E_{\mathcal{U}}}^{\perp}$, para lo cual bastará probar que $\overline{E_{\mathcal{U}}}^{\perp} = I_{\mathcal{U}}$. De lo anterior obtenemos que $I_{\mathcal{U}} \subset \overline{E_{\mathcal{U}}}^{\perp}$. Sea $f \in \overline{E_{\mathcal{U}}}^{\perp}$, en particular se tiene que $\langle f, \mathcal{U}(g) - g\rangle = 0$ para toda $g \in H$, así que:

$$\langle f, g\rangle = \langle f, \mathcal{U}(g)\rangle = \langle \mathcal{U}^*(f), g\rangle \quad \text{para toda } g \in H,$$

entonces $\langle f - \mathcal{U}^*(f), g\rangle = 0$ para toda $g \in H$, por lo tanto $f - \mathcal{U}^*(f) = 0$, así que $f \in I_{\mathcal{U}}^*$, por el Lema 1.8, se tiene que $f \in I_{\mathcal{U}}$, es decir $I_{\mathcal{U}} \subset \overline{E_{\mathcal{U}}}^{\perp}$. □

El operador $\mathcal{U}_T$ realiza una conexión entre la Teoría Ergódica y Análisis Funcional. Aplicando el teorma anterior, al operador $\mathcal{U}_T$ en $L_2^{\mathbb{R}}$, tenemos el siguiente corolario:

Corolario 1.19.1. *Sea $(X, \mathscr{F}, \mu)$ un espacio de probabilidad, y T una transformación ergódica. Entonces para toda $f \in L_2^{\mathbb{R}}(X, \mathscr{F}, \mu)$ se tiene que:*

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \underset{L_2}{\longrightarrow} \int_X f\, d\mu.$$

Demostración. Como $L_2^{\mathbb{R}}(\mu)$ es un espacio de Hilbert y $\mathcal{U}_T$ es una isometría lineal de este espacio en sí mismo. Además como T es ergódica, entonces $I_{\mathcal{U}_T} = \{\text{constantes}\}$ [ver Teorema 1.12(vi)], se tiene que $proj(f|I_{\mathcal{U}_T}) = \int_X f d\mu$ .para toda $f \in L_2^{\mathbb{R}}$. En vista de lo aterior, y aplicando el teorema anterior a $L_2(\mu)$ y $\mathcal{U}_T$ se tiene el resulado. □

Observación 1.27. Retomando la hipótesis ergódica de Boltzmann. El Teorema de Von Neumann junto con el corolario anterior aseguran: la existencia de los promedios temporales en el sentido de convergencia $L_2^{\mathbb{R}}$, y éstos son iguales al promedio espacial en el caso ergódico.

1.7. Algunas Aplicaciones

Como hemos mencionado entre las aplicaciones de la Teoría Ergódica se encuentran áreas como: Física, Teoría de la Información, Estadística, Probabilidad y Teoría de Números (entre otras). El propósito de esta sección es mostrar su relación en algunas de estas áreas (ver Pollicott [18, pág. 103], Parry [17, págs. 9, 27 y 28], Silva [22, pág. 72]) y Walters [23, cap. 6].

Para empezar veremos una aplicación del Teorema Ergódico Individual a Teoría de Números. Sea $m \in \mathbb{N}$ tal que $m \geq 2$ para $x \in [\,0\,,1\,)$ podemos expandir a x de la siguiente forma:

$$x = \frac{a_1}{m} + \frac{a_2}{m^2} + \frac{a_3}{m^3} + \cdots$$

donde $a_1, a_2, a_3, \dots \in \{0, 1, \dots, m-1\}$ (esta expansión no necesariamente es única). Cuando $m = 10$ esta expansión es la expansión decimal usual. Diremos que x es un número normal en base m si para cada $j \in \{0, 1, \dots, m-1\}$ el término j aparece u ocurre en la expansión de x con densidad $1/m$, es decir, $\lim_{N\to\infty}(1/N)\text{card}(\{1 \leq n \leq N : a_n = j\}) = 1/m$. Diremos que x es un **número normal** si x es normal en cualquier base m. E. Borel probó que casi todo número es normal. A continuación enunciaremos este resultado y lo probaremos.

Teorema 1.20 (De los números normales de Borel)**.** *Para casi toda $x \in [\,0\,,1\,)$ (relativo a la medida de Lebesgue) x es un número normal.*

Demostración. Sea $Y = \{x \in [\,0\,,1\,) : x$ admite dos expansiones en base $m\,\}$, entonces Y es numerable por lo que $\mu(Y) = 0$. Sea $X = [\,0\,,1\,) \setminus Y$, definimos $T : X \to X$ como en el Ejemplo 1.1. En páginas previas probamos que T preserva a λ. Modificando el argumento empleado para probar que $T(z) = z^k$ en S^1 es ergódica, obtenemos que T lo es también. Sea $f = \chi_{[\,j/m\,,(j+1)/m\,)}$ entonces:

$$f(T^k(x)) = f\left(\sum_{i=1}^{\infty} \frac{a_{i+k}}{m^i}\right) = \begin{cases} 1 & \text{si} \quad a_{k+1} = j \\ \\ 0 & \text{si} \quad a_{k+1} \neq j. \end{cases}$$

Así que: $1/n \sum_{k=0}^{n-1} f(T^k(x)) =$(número de jotas en la expansión en base m de x en los primeros n dígitos)$/n$ y por el Teorema Ergódico Individual:

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) \longrightarrow \frac{1}{k} \text{ casi dondequiera,}$$

lo que prueba el Teorema. □

Uno de los primeros teoremas ergódicos que se demostró es el siguiente, éste no sólo asegura la convergencia casi segura si no también convergencia uniforme, en el caso f continua.

Teorema 1.21 (Equipartición módulo 1, Bohl, Sierpinskii y Weyl)**.** *Sea $T : S^1 \to S^1$ la rotación irracional y $f : S^1 \to \mathbb{R}$ una función Riemann integrable entonces:*

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T^j(x)) \to \int_{S^1} f\, d\mu \quad \textit{para toda } z \in S^1,$$

en donde la convergencia es ***uniforme*** *si f es continua.*

Demostración. Sea $T(z) = az$ con $a = e^{i\rho}$ y $\rho/2\pi \notin \mathbb{Q}$.

Caso 1) $f(z) = z^p$ $p \in \mathbb{Z}$. Entonces:

$$\frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) = \frac{1}{N}\sum_{j=0}^{N-1} (a^j z)^p = \begin{cases} 1 & \text{si } \ p = 0 \\ \\ \frac{1}{N} z^p \left(\frac{a^{Np}-1}{a^p-1}\right) & \text{si } \ p \neq 0. \end{cases}$$

Por otro lado

$$\int_{S^1} f\, d\mu = \begin{cases} 1 & \text{si } \ p = 0 \\ \\ 0 & \text{si } \ p \neq 0 \end{cases} \qquad \text{esto pues: } \ \frac{1}{2\pi i}\int z^n\, dz = \begin{cases} 1 & \text{si } \ n = -1 \\ \\ 0 & \text{si } \ n \neq -1. \end{cases}$$

Ahora bien como $\rho/2\pi \notin \mathbb{Q}$, entonces $a^p - 1 \neq 0$ para toda $p \in \mathbb{Z}$ y como se tiene que $\left|a^{Np} - 1\right| < 2$ obtenemos que:

$$\frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) \to \begin{cases} 1 & \text{si } \ p = 0 \\ \\ 0 & \text{si } \ p \neq 0, \end{cases}$$

por lo tanto $f^* = \int_{S^1} f\, d\mu$.[8]

Caso 2) f es un polinomio trigonométrico $\sum_{p\in\mathbb{Z}} a_p z^p$ $z \in S^1$, $a_p = 0$ para una infinidad de $p \in \mathbb{Z}$. De 1) obtenemos que $f^* \equiv a_0 = \int_{S^1} f\, d\mu$ (la convergencia es nuevamente uniforme).

Caso 3) Caso general (f Riemann-Integrable). Dada $\varepsilon > 0$ existen polinomios trigonométricos P_ε^- y P_ε^+ tales que $P_\varepsilon^- < f(z) < P_\varepsilon^+$ para toda $z \in S^1$ y $\int_{S^1}(P_\varepsilon^+ - P_\varepsilon^-)\, d\mu < \varepsilon$. Por el caso 2) se obtiene que:

$$\int_{S^1} P_\varepsilon^-\, d\mu \leq \liminf_{N\to\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) \leq \limsup_{N\to\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) \leq \int_{S^1} P_\varepsilon^+\, d\mu \quad \text{para toda } z \in S^1,$$

por lo que

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) - \limsup_{N\to\infty} \frac{1}{N}\sum_{j=0}^{N-1} f(T^j(z)) \leq \varepsilon.$$

[8]Note que en este caso la convergencia es uniforme.

Como $\varepsilon > 0$ es arbitrario se tiene que $f^*(z)$ existe para toda $z \in S^1$ y es igual a $\int_{S^1} f\,d\mu$; esto ya que: $\int_{S^1} P_\varepsilon^-\,d\mu \leq f^*(z) \leq \int_{S^1} P_\varepsilon^+\,d\mu$ para toda $z \in S^1$ y $\int_{S^1} P_\varepsilon^-\,d\mu \leq \int_{S^1} f\,d\mu \leq \int_{S^1} P_\varepsilon^+\,d\mu$ para toda $z \in S^1$, por lo tanto $\left|f^*(z) - \int_{S^1} f\,d\mu\right| \leq \int (P_\varepsilon^+ - P_\varepsilon^-)\,d\mu < \varepsilon$.

b) Si f es continua entonces para toda $\varepsilon > 0$ existe P un polinomio trigonométrico tal que $\|f - P\|_\infty < \varepsilon$ por lo que:

$$\left|\int_{S^1} f\,d\mu - \int_{S^1} P\,d\mu\right| < \varepsilon \quad \text{y} \quad \left\|\frac{1}{N}\sum_{j=0}^{N-1} f\circ T^j - \frac{1}{N}\sum_{j=0}^{N-1} P\circ T^j\right\|_\infty < \varepsilon.$$

Así que:

$$\begin{aligned}
\left\|\frac{1}{N}\sum_{j=0}^{N-1} f\circ T^j - \int\limits_{S^1} f\,d\mu\right\|_\infty &\leq \left\|\frac{1}{N}\sum_{j=0}^{N-1} f\circ T^j - \frac{1}{N}\sum_{j=0}^{N-1} P\circ T^j\right\|_\infty + \\
&\quad \left\|\frac{1}{N}\sum_{j=0}^{N-1} P\circ T^j - \int\limits_{S^1} P\,d\mu\right\|_\infty + \left|\int\limits_{S^1} P\,d\mu - \int\limits_{S^1} f\,d\mu\right| \\
&< 2\varepsilon + \left\|\frac{1}{N}\sum_{j=0}^{N-1} P\circ T^j - \int\limits_{S^1} P\,d\mu\right\|_\infty \\
&< 3\varepsilon.
\end{aligned}$$

si N es suficientemente grande. Por lo tanto

$$\frac{1}{N}\sum_{j=0}^{N-1} f\circ T^j \longrightarrow \int\limits_{S^1} f\,d\mu \quad \text{uniformemente.}$$ □

El teorema anterior será de utilidad para lo siguiente: considerando la sucesión formada por los primeros dígitos de potencias de dos. Las primeras 25 potencias de 2 son:

$$2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536$$

$$131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432.$$

Los primeros 40 términos de la sucesión considerada son:

2	4	8	1	3	6	1	2	5	1
2	4	8	1	3	6	1	2	5	1
2	4	8	1	3	6	1	2	5	1
2	4	8	1	3	6	1	2	5	1

Como podemos apreciar en estos térmimos no aparece el dígito 7 ni el 9, cabe preguntarse: ¿En la sucesión de los primeros dígitos de potencias de 2 aparece el dígito 7 o 9?, y si aparecen estos dígitos ¿con qué frecuencia aparecen? La respuesta es afirmativa , de hecho hay una infinidad de potencias de 2 cuyo primer término es 7 o 9, además se puede calcular su frecuencia. Esto está contenido en el siguiente resultado.

Corolario 1.21.1. *(Cálculo de la frecuencia relativa del primer dígito de 2^n)*

Obtención: El primer dígito de 2^n es igual a $k \in \{1,2,\dots,9\}$ si y sólo si $k10^r \leq 2^n < (k+1)10^r$ para algún $r \in \mathbb{N} \cup \{0\}$, es decir, $r + \log_{10}(k) \leq n\log_{10}(2) < r + \log_{10}(k+1)$. Sea $\rho = \log_{10}(2)$, $T : S^1 \to S^1$ la rotación irracional $T(z) = az$ con $a = e^{i2\pi\rho}$,

$$A_k = \{z \in S^1 : z = e^{i\theta} \quad \text{y} \quad 2\pi\log_{10}(k) \leq \theta < 2\pi\log_{10}(k+1)\},$$

es decir, A_k es un arco y χ_{A_k} es Riemann-integrable. Por el teorema anterior se obtiene que para toda $z \in S^1$:

$$\frac{1}{n}\sum_{j=0}^{n-1}\chi_{A_k}(T^j(z)) \longrightarrow \mu(A_k) = \log_{10}(k+1) - \log_{10}(k) = \log_{10}\left(1+\frac{1}{k}\right).$$

Por otro lado, si $z = 1$, entonces $\chi_{A_k}(T^j(1)) = \chi_{A_k}(e^{i2\pi\rho j})$ es uno o cero de acuerdo a que $2\pi\rho j \in [\,2\pi(r+\log_{10}(k))\,, 2\pi(r+\log_{10}(k+1))\,)$, o no (para alguna $r \in \mathbb{N}$) (respectivamente), equivalentemente uno o cero de acuerdo a que, $j\log_{10}(2) \in [\,r+\log_{10}(k)\,, r+\log_{10}(k+1)\,)$ para alguna $r \in \mathbb{N}\cup\{0\}$ o no; así pues $(1/n)\sum_{j=0}^{n-1}\chi_{A_k}(T^j(1)) = (1/n)$(número de veces que el primer dígito de la sucesión $1, 2, \dots, 2^{n-1}$ es igual a k). Así que la frecuencia relativa de apariciones de k como primer dígito de la sucesión 2^n está dada por $f(k) = \log_{10}(1+1/k)$. Observamos que si $k < k'$ entonces $f(k) > f(k')$ $(k, k' \in \{1,\dots,9\})$, por lo que $f(1) > f(2) > \cdots > f(9)$, en particular $f(7) > f(8)$, hecho que resulta sorpresivo si se examinan los primeros 47 términos (¡que es el primer momento en que se obtiene $k = 7$, mientras que ya han salido cinco ochos!).

Teorema 1.22 (Ley Débil de los Grandes Números). *Sean $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $\{f_n\}$ una sucesión de variables aleatorias integrables independientes e idénticamente distribuidas (es decir, $\{f_n\} \subset L_1^{\mathbb{R}}(\mu)$, $\mu(f_n^{-1}(B))$, no depende de n para toda $B \in \mathscr{B}_{\mathbb{R}}$ y*

$$\mu(f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \cap \cdots \cap f_n^{-1}(B_n)) = \mu(f_1^{-1}(B_1))\mu(f_2^{-1}(B_2))\cdots\mu(f_n^{-1}(B_n))$$

para todo n, para toda $B_1, B_2, \cdots B_n \in \mathscr{B}_{\mathbb{R}}$), entonces

$$\frac{1}{N}\sum_{n=1}^{N} f_n(x) \to \int f_1 d\,\mu \qquad \textit{en medida.}$$

Demostración. Por hipótesis existe una única medida de probabilidad ν sobre $(\prod_{n=1}^{\infty}\mathbb{R}, \bigotimes_{n=1}^{\infty}\mathscr{B}_{\mathbb{R}})$ (ver Ejemplo B.3, pág. 136). Considero la transformación $T : \prod_{n=1}^{\infty}\mathbb{R} \to \prod_{n=1}^{\infty}\mathbb{R}$ dada por el corrimiento, es decir, $(T(x))_n = x_{n+1}$ para toda n. Como tenemos variables aleatorias independientes e idénticamente distribuidas, se tiene que T preserva a ν en $\prod_{n=1}^{\infty}\mathbb{R}$. Más aún tenemos que T es ergódica y que $f(x) = x_1$ para toda $x \in \prod_{n=1}^{\infty}\mathbb{R}$ tiene la misma distribución que f_1 (es decir, $\nu(f^{-1}(B)) = \mu(f_1^{-1}(B))$ para toda $B \in \mathscr{B}_{\mathbb{R}}$).

Por el Corolario 1.16.1 tenemos que si

$$B_\lambda = \{x \in X || \frac{1}{n}\sum_{k=1}^{n} f_k| > \lambda\},$$

entonces

$$\int_{B_\lambda} f_1 d\mu \geq \lambda\mu(B_\lambda),$$

es decir, $\mu(B_\lambda) \leq (\int f_1 d\,\mu)/\lambda$, por lo que se tiene el resultado observando que $\int \sum_{k=1}^{n} |f_n| = \int |f_1| d\mu$, y la desigualdad de Markov la cual nos asegura que $\mu(|f_1| \geq \lambda) \leq (\int |f| d\mu)/\lambda$. □

Teorema 1.23 (Ley Fuerte de los Grandes Números)**.** *Sean $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $\{f_n\}$ una sucesión de variables aleatorias integrables independientes e idénticamente distribuidas, entonces*

$$\frac{1}{N}\sum_{n=1}^{N} f_n(x) \to \int f_1 d\,\mu \qquad \textit{casi dondequiera.}$$

Demostración. Utilizando la notación del teorema anterior y sus observaciones, el Teorema Ergódico Individual nos asegura que:

$$\frac{1}{n}\sum_{k=1}^{n} f(T^n(x)) \to \int f d\,\nu \qquad \text{casi dondequiera,}$$

por lo anterior tenemos que:

$$\frac{1}{n}\sum_{k=1}^{n} f_n(\omega) \to \int f_1 d\,\mu \qquad \text{casi dondequiera.}$$ □

Un caso donde podremos asegurar la existencia de medidas de probabilidad μ tales que $\mu \circ T^{-1} = \mu$, es cuando el espacio X es un espacio métrico compacto y $T : X \to X$ es una transformación continua, daré un bosquejo de la prueba de este hecho. Para mayor referencia ver Walters [23, págs. 150-153]. Para esto necesitaremos lo siguiente:

Teorema 1.24. *Sean $(X, \mathscr{F})$ un espacio de medida, $T : X \to X$ una transformación medible, y $\mathcal{M}(X,T) = \{\mu : \mathscr{F} \to [0,1] | \mu$ es de probabilidad, $\mu \circ T^{-1} = \mu\}$. Entonces:*

(i) $\mathcal{M}(X,T)$ es convexo (es decir, si $\mu, \nu \in \mathcal{M}(X,T)$, entonces $\alpha\mu + (1-\alpha)\nu \in \mathcal{M}(X,T)$ para toda $\alpha \in (0,1)$).

(ii) Si μ y ν hacen ergódica a T, entonces $\mu = \nu$ ó $\mu \perp \nu$ (ver Definición B.25, pág. 135).

(iii) Si μ hace ergódica a T, entonces es un punto extremo de $\mathcal{M}(X,T)$ (ver Definición A.3, pág. 128).

Demostración. Primero establezcamos sobre qué espacio vectorial estamos trabajando. Pensando en el conjunto de las medidas con signo finitas en $(X, \mathscr{F})$ (ver Definición B.26, pág. 135), este conjunto forma un espacio vectorial sobre $\mathbb{R}$ definiendo las operaciones como sigue

$(\mu+\nu)(F) = \mu(F) + \nu(F)$ y $(\beta \cdot \mu)(F) = \beta \cdot \mu(F)$ para toda $F \in \mathscr{F}$ y $\beta \in \mathbb{R}$, entonces $\mathcal{M}(X,T)$ es un subconjunto de este espacio vectorial.

(i) Sean $\mu, \nu \in \mathcal{M}(X,T)$, entonces $(\alpha\mu+(1-\alpha)\nu)(\emptyset) = 0$, como $\alpha \in (0,1)$ $(\alpha\mu+(1-\alpha)\nu)(F) \geq 0$ para toda $F \in \mathscr{F}$ y $(\alpha\mu+(1-\alpha)\nu)(X) = \alpha+(1-\alpha) = 1$. Sólo basta probar la σ-aditividad, pero si $\{F_n\}$ es una sucesión de elementos de $\mathscr{F}$ disjunta dos a dos se tiene que las series $\sum_{i=1}^{\infty} \mu(F_n)$, $\sum_{i=1}^{\infty} \nu(F_n)$ son convergentes y $(\alpha\mu+(1-\alpha)\nu)(\bigcup_{i=1}^{\infty} F_n) = \sum_{i=1}^{\infty}(\alpha\mu+(1-\alpha)\nu)(F_n)$, finalmente

$$\begin{aligned}(\alpha\mu+(1-\alpha)\nu)(T^{-1}(F)) &= \alpha\mu(T^{-1}(F)) + (1-\alpha)\nu(T^{-1}(F)) \\ &= \alpha\mu(F) + (1-\alpha)\nu(F) \\ &= (\alpha\mu+(1-\alpha)\nu)(F),\end{aligned}$$

por lo tanto $\mathcal{M}(X,T)$ es convexo.

(ii) Supongamos que $\mu \neq \nu$, entonces existe $F \neq \emptyset \in \mathscr{F}$ tal que $\mu(F) \neq \nu(F)$. Sin pérdida de generalidad supongamos que $\mu(F) > 0$, considerando el conjunto:

$$A = \left\{ x \in X : \frac{1}{n}\sum_{i=0}^{n-1} \chi_F(T^i(x)) \to \mu(F) \right\},$$

como $\mu(F) \neq \nu(F)$ y por el Teorema Ergódico Individual tenemos que $\nu(A) = \mu(X \setminus A) = 0$.

(iii) Supongamos que $\mu = \alpha\nu_1 + (1-\alpha)\nu_2$ con $\nu_1, \nu_2 \in \mathcal{M}(X,T)$ y $\mu \perp \nu_1$, entonces existen A, B tales que $X = A \cup B$ y $\mu(A) = \nu_1(B) = 0$, así que $\mu(B) > 0$; por el Teorema de Poincaré $\mu(B^*) > 0$ y $T^{-1}(B^*) = B^*$, ahora $0 < \mu(B^*) = (\alpha\nu_1 + (1-\alpha)\nu_2)(B^*) < 1$ lo que contradice que T sea ergódica. □

Cuando tenemos un espacio métrico compacto (X, d), se puede considerar a la mínima σ-álgebra generada por los conjuntos abiertos de X, llamada **σ-álgebra de Borel** y denotada como $\mathscr{B}_X$. Note que cualquier transformación continua $T : X \to X$ es $(X, \mathscr{B}_X)$-medible.

Sea $\mathcal{M}(X) = \{\mu : \mathscr{B}_X \to [0,1] \mid \mu \text{ es una medida de probabilidad}\}$. Sea $(C(X), \|\cdot\|)$ el espacio de las funciones continuas de X a $\mathbb{R}$, donde $\|f\| = \sup_{x\in X} |f(x)|$. La **topología débil*** en $M(X)$ es la mínima topología que hace a cada función $\mu \to \int_X f d\,\mu$ ($f \in C(X)$) continua. En nuestro caso resulta que el espacio topólogico $(\mathcal{M}(X), \tau)$, donde τ es la topología débil*, es metrizable. Si $\{f_n : n \in \mathbb{N}\}$ es un conjunto denso en $C(X)$, entonces $d(\mu,\nu) = \sum_{i=1}^{\infty} \left| \int\limits_X f_i d\,\mu - \int\limits_X f_i d\,\nu \right|$ es una métrica en $\mathcal{M}(X)$ (ver Walters [23, pág. 148]).

Hemos dotado a $\mathcal{M}(X)$ de una topología y una métrica. Además resulta ser que bajo estas condiciones este espacio es un espacio métrico compacto (ver Walters [23, pág. 150]). Así que $\mathcal{M}(X,T)$ es un subconjunto convexo de este espacio, por el Teorema de Krein-Milman (ver Royden [19, pág. 242]), se tiene que $\mathcal{M}(X,T)$ es la cerradura de la envoltura convexa de sus puntos extremos. Finalmente por el Teorema anterior tenemos que $\mathcal{M}(X,T)$ es la cerradura de la envoltura convexa de las $\mu \in \mathcal{M}(X,T)$ que hacen ergódica a T. Con lo que no sólo se prueba la existencia de una medida de probabilidad μ, tal que $\mu \circ T^{-1} = \mu$, si no que la podemos aproximar tanto como queramos con combinaciones lineales convexas de medidas ν que hacen ergódica a T.

Teorema 1.25 (El Teorema ergódico para flujos)**.** *Sea $(T_t)_{t\geq 0}$ una familia de transformaciones preservadoras de medida de un espacio de medida σ-finito $(X, \mathscr{F}, \mu)$ en sí mismo tal que:*

(i) $T_0 = I_X$.

(ii) $T_{s+t} = T_s \circ T_t$ *para toda* $s, t \geq 0$.

(iii) $T : X \times [0,\infty) \to X$ *dada por* $T(x,t) = T_t(x)$ *es* $\mathscr{F} \times \mathscr{B}_{[0,\infty)}$*-medible.*

Entonces para toda $f \in L_1(\mu)$ existe $f^* \in L_1(\mu)$ tal que $f^* \circ T_s = f^*$ casi dondequiera tal que:

(i) $\lim\limits_{t\to\infty} \frac{1}{t}\int_0^t f \circ T_s d\,s = f^*$ *casi dondequiera.*

(ii) Si $F \in \mathscr{F}$ *tal que* $T_1^{-1}(F) = F$, $\mu(F) < \infty$, *entonces* $\int_F f d\,\mu = \int_F f^* d\,\mu$.

Demostración. Sea $g : X \to \mathbb{R}$ dada por $g(x) = \int_0^1 f(T_s(x))d\,s$, como $(X, \mathscr{F}, \mu)$ y $([0,\infty), \mathscr{B}_{\mathbb{R}}, \lambda)$ son dos espacios de medida σ-finitos entonces por el Teorema B.11, existe la medida $\mu \otimes \lambda$. Consideremos la función $g : X \times [0,1] \to \mathbb{R}$ dada por $g(x,s) = f(T_s(x))$, como

$$g^{-1}(B) = T_s^{-1}(f^{-1}(B)) \in \mathscr{F} \times \mathscr{B}_{[0,1]}$$

para toda $B \in \mathbb{R}$, se tiene que g es $\mathscr{F} \times \mathscr{B}_{[0,1]}$-medible, además

$$\int\limits_{X\times[0,1]} |g| d\,\mu \otimes \lambda = \int\limits_{[0,1]} \left(\int\limits_{X} |g| d\mu \right) d\lambda = \int_X |f| d\mu < \infty,$$

entonces $g \in L_1(\mu \times \lambda)$. Por el Teorema de Fubini (ver Teorema B.12, pág. 136), se tiene que $h(x) = \int_{[0,1]} f(T_s(x))d\,s \in L_1(\mu)$. Por el Teorema 1.16 existe una $f^* \in L_1(\mu)$ tal que $f^* \circ T_1 = f^*$ y $\int_F h = \int_F f^*$ para todo F tal que $T_1^{-1}(F) = F$ y $\mu(F) < \infty$, afirmamos que f^* cumple las propiedades del teorema, pero antes notemos que:

$$\begin{aligned}
\frac{1}{n}\int\limits_0^n f \circ T_s d\,s &= \frac{1}{n}\sum_{i=0}^{n-1}\int_i^{i+1} f \circ T_s(x) d\,s \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_0^1 f \circ T_{s+i}(x) d\,s \\
&= \frac{1}{n} h \circ T_1^i(x) \to f^*(x) \text{ c.d. si } n \to \infty,
\end{aligned}$$

ahora sea $t \geq 0$, consideramos $[t]$ (la parte entera de t), entonces:

$$\frac{1}{t}\int_0^t f \circ T_s(x) d\,s = \frac{[t]}{t}\left(\frac{1}{[t]}\int_0^{[t]} f \circ T_s d\,s\right) + \frac{1}{t}\int_{[t]}^t f \circ T_s(x) d\,s,$$

como

$$\lim_{t\to\infty}([t]/t) = 1,\ \lim_{t\to\infty}(1/[t])\int_0^{[t]} f \circ T_s d\,s \to f^* \quad \text{y} \quad \lim_{t\to\infty}(1/t)\left|\int_{[t]}^t f \circ T_s(x) d\,s\right| \leq \lim_{t\to\infty}(1/t)\|h\|_1,$$

se tiene que $\lim_{t\to\infty} \frac{1}{t}\int_0^t f \circ T_s(x) d\,s = f^*(x)$ casi dondequiera, el segundo inciso se sigue del Teorema Ergódico Individual. ☐

1.8. Propiedades Mezclantes

En lo que resta de este capítulo supondremos que $(X, \mathscr{F}, \mu)$ es un espacio de probabilidad y $T : X \to X$ preserva a μ. Utilizando la notación anterior, por el Corolario 1.16.1, tenemos que T es ergódica si y sólo si para toda $A, B \in \mathscr{F}$ se tiene que

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

Supongamos que tenemos una cuba con 10 % de ron y 90 % de Coca-Cola. Si A es la región original ocupada por el ron, entonces para cualquier región B del líquido, la cantidad de ron en B después de "mezclar" n veces, está dado por $\mu(T^{-i}(B) \cap A)/\mu(A)$. Si T es ergódica, entonces en promedio la cantidad de ron que hay en B es 10 %. Sin embargo, intuitivamente después de mezclar suficientemente la cantidad de ron en cada parte de B debería ser aproximadamente del 10 %. En otras palabras, si T "mezcla bien" entonces podríamos cambiar la convergencia en promedio por convergencia, es decir, $\lim_{n\to\infty} \mu(T^{-n}(B) \cap A) = \mu(B)\mu(A)$ para todo $A, B \in \mathscr{F}$. A una transformación T con esta propiedad la llamamos fuertemente mezclante, claramente toda transformación fuertemente mezclante es ergódica, sin embargo el recíproco no es cierto. También mostraremos una condición intermedia entre ergodicidad y fuertemente mezclante: la condición de ser débilmente mezclante.

Definición 1.12. Sea $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $T : X \to X$ una transformación preservadora de medida. Entonces:

(a) T se llama **débilmente mezclante** si para todo $A, B \in \mathscr{F}$ se tiene:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| = 0.$$

(b) T se llama **fuertemente mezclante** si para todo $A, B \in \mathscr{F}$ se tiene:

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

Los siguientes lemas justifican lo discutido anteriormente.

Lema 1.9. *Si T es fuertemente mezclante, entonces T es débilmente mezclante.*

Lema 1.10. *Si T es débilmente mezclante, entonces T es ergódica.*

La razón de lo anterior es púramente analítica, a saber:

Lema 1.11. *Sea $(a_i)_{i=0}^{\infty}$ una sucesión numérica entonces:*

(i) Si $a_n \to 0$, entonces $(1/n)\sum_{i=0}^{n-1} |\, a_i| \to 0$.

(ii) Si $(1/n)\sum_{i=0}^{n-1} |\, a_i| \to 0$, entonces $(1/n)\sum_{i=0}^{n-1} a_i \to 0$.

Demostración.
(i). Como $a_n \to 0$, entonces existe $N > 0$ tal que para toda $n \geq N$ $|\, a_n| < \varepsilon$ por lo tanto

$$\begin{aligned} \frac{1}{n}\sum_{i=0}^{n-1} |\, a_i| = \frac{1}{n}\left(\sum_{i=0}^{N} |\, a_i| + \sum_{m=N}^{n} |\, a_m|\right) &< \frac{1}{n}\sum_{i=0}^{N} |\, a_i| + \left(\frac{n-N}{n}\right)\varepsilon \\ &< 2\varepsilon, \end{aligned}$$

si n es suficientemente grande.

(ii). $\left|(1/n)\sum_{i=0}^{n-1} a_i\right| \leq (1/n)\sum_{i=0}^{n-1} |\, a_i| < \varepsilon$ si n es suficientemente grande. □

Aplicando el Lema 1.11 a $a_i = \mu(T^{-i}(A)\cap B) - \mu(A)\mu(B)$ se obtienen los Lemas 1.10 y 1.9.

Observación 1.28. Un ejemplo de una transformación que es ergódica pero no débilmente mezclante es la rotación irracional, esto se probará mas adelante. También existen ejemplos de transformaciones que son débilmente mezclantes pero no fuertemente mezclantes (ver Silva [22, pág. 218]). En el lenguaje de probabilidad una transformación T es ergódica si el evento $T^{-n}(A)$ es en promedio independiente de cualquier evento B bajo iteraciones de T. Si T es fuertemente mezclante entonces $T^{-n}(A)$ y B son asintóticamente independientes para todo $A, B \in \mathscr{F}$. En ocasiones es más fácil verificar que una transformación es fuertemente o débilmente mezclante que probar que es ergódica.

Nuevamente probar que una transformación es fuertemente o débilmente mezclante a partir de la definición es complicado. El siguiente resultado simplifica la tarea de verificar las propiedades mezclantes de una transformación, a partir de una álgebra que genera a la σ-álgebra.

Teorema 1.26. *$T : X \to X$ una transformación preservadora de medida y $\mathcal{A}$ un álgebra de conjuntos tal que $\sigma(\mathcal{A}) = \mathscr{F}$ entonces:*

(i) T es ergódica si y sólo si para toda $A, B \in \mathcal{A}$ se tiene que

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

(ii) T es débilmente mezclante si y sólo si para toda $A, B \in \mathcal{A}$ se tiene que

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left|\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| = 0.$$

(iii) T es fuertemente mezclante si y sólo si para toda $A, B \in \mathcal{A}$ se tiene que

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

Demostración. Se probará solamente (i), los otros incisos son similares. Claramente basta probar la (Suficiencia).

Sean $A, B \in \mathscr{F}$ y $\varepsilon > 0$ arbitrarios, hallamos $A_0, B_0 \in \mathcal{A}$ tal que $\mu(A \triangle A_0) < \varepsilon$ y $\mu(B \triangle B_0) < \varepsilon$, entonces $\mu((T^{-n}(A) \cap B) \triangle (T^{-n}(A_0 \cap B_0)) < 2\varepsilon$, en consecuencia

$$\left|\mu(T^{-n}(A) \cap B) - \mu(T^{-n}(A_0) \cap B_0)\right| < 2\varepsilon \quad \text{por lo que}$$

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)\right| &\leq \left|\frac{1}{n}\sum_{i=0}^{n-1} \left(\mu(T^{-i}(A) \cap B) - \mu(T^{-i}(A_0) \cap B_0)\right)\right| + \\
&\quad \left|\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A_0) \cap B_0) - \mu(A_0)\mu(B_0)\right| + \\
&\quad |\,\mu(A_0)\mu(B_0) - \mu(A_0)\mu(B)| + \\
&\quad |\,\mu(A_0)\mu(B) - \mu(A)\mu(B)| \\
&< 2\varepsilon + \left|\frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(A_0) \cap B_0) - \mu(A_0)\mu(B_0)\right| + 2\varepsilon \\
&< 5\varepsilon.
\end{aligned}$$

Si n se elige suficientemente grande. □

Como hemos mencionado anteriormente, la condición de ergodicidad es importante, además las propiedades mezclantes implican ergodicidad. Ahora voy a encontrar condiciones equivalentes a débilmente mezclante, para ello será necesario la siguiente definición:

Definición 1.13. Un subconjunto $J \subset \{0, 1, 2, \dots\}$ se dice que tiene **densidad cero** si

$$\lim_{n\to\infty} \frac{\operatorname{card}(J \cap \{0, 1, \cdots, n-1\})}{n} = 0,$$

y se denota por $d(J) = 0$.

Teorema 1.27. *Sean $(a_n)_{n=0}^{\infty}$ una sucesión **acotada** de números reales (o complejos), entonces las siguientes condiciones son equivalentes:*

(i) $\frac{1}{n}\sum_{i=0}^{n-1} |\, a_i| \longrightarrow 0.$

(ii) Existe $J \subset \{0, 1, 2 \dots\}$ de densidad cero tal que $\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0.$

(iii) $\frac{1}{n}\sum_{i=0}^{n-1} |\, a_i|^2 \longrightarrow 0.$

Demostración. Por simplificar la notación se denotará $\alpha_J(n) = \operatorname{card}(J \cap \{0, 1, \dots, n-1\})$.

(i) implica (ii). Para toda $k \geq 1$, sea $J_k = \{n \in \{0, 1, 2, \dots\} : |\, a_n| \geq (1/k)\}$. Claramente $J_1 \subset J_2 \subset \dots$ además $d(J_k) = 0$ para toda k (esto pues:$(1/n)(1/k)\alpha_{J_k}(n) \leq (1/n)\sum_{i=0}^{n-1} |\, a_i| \to 0$), luego entonces existen enteros $0 = l_0 < l_1 < l_2 < \cdots$ tal que si $n \geq l_k$: $(1/n)\alpha_{J_{k+1}}(n) < 1/(k+1)$. Sea $J = \bigcup_{k=0}^{\infty}\{J_{k+1} \cap [\, l_k\,, l_{k+1}\,)\}$ probaremos que $d(J) = 0$ y que $\lim_{\substack{n\to\infty \\ n\notin J}} a_n = 0$.

Ahora bien por la definición de J tenemos que si $n \in [\, l_k\,, l_{k+1}\,)$, entonces $J \cap [\,0\,, l_k\,) \subset J_k \cap [\,0\,, l_k\,)$ y $J \cap [\, l_k\,, n\,) \subset J_{k+1} \cap [\,0\,, n\,)$ de donde: $J \cap [\,0\,, n\,) \subset (J_k \cap [\,0\,, l_k\,)) \cup (J_{k+1} \cap [\,0\,, n\,))$, por lo que:

$$\frac{1}{n}\alpha_J(n) \leq \frac{1}{n}\left[\alpha_{J_k}(l_k) + \alpha_{J_{k+1}}(n)\right] < \frac{1}{k} + \frac{1}{k+1} \quad \text{pues} \quad n \in [\, l_k\,, l_{k+1}\,),$$

pero si $n \to \infty$ entonces $l_{k+1} \to \infty$, por lo que $k \to \infty$ también, por lo tanto $d(J) = 0$. Por otro lado si $n > l_k$ y $n \notin J$ se tiene que $n \notin J_{k+1}$ (pues si $n \in J_{k+1}$, entonces $n \in J_r$ para toda $r \geq k+1$, por lo que pertenecería a J). Así pues $|\, a_n| < 1/(k+1)$ por lo tanto $\lim_{\substack{n\to\infty \\ n\notin J}} |\, a_n| = 0$.

(ii) implica (i). Supongamos que $|\, a_n| \leq K$ para toda n. Sea $\varepsilon > 0$, entonces existe $N_\varepsilon \in \mathbb{N}$ tal que si $n \geq N_\varepsilon$ y $n \notin J$ se tiene que $|\, a_n| < \varepsilon$. Además existe $M_\varepsilon \in \mathbb{N}$ tal que si $n \geq M_\varepsilon$, entonces $\alpha_J(n)/n < \varepsilon$, por lo que si $n \geq \max\{N_\varepsilon, M_\varepsilon\}$ obtenemos:

$$\frac{1}{n}\sum_{i=0}^{n-1} |\, a_i| = \left\{\frac{1}{n}\sum_{j\in\{0,\dots,n-1\}\cap J} |\, a_j| + \frac{1}{n}\sum_{j\in\{0,\dots,n-1\}\setminus J} |\, a_j|\right\} < \frac{K}{n}\alpha_J(n) + \varepsilon + \frac{K\max\{N_\varepsilon, M_\varepsilon\}}{n},$$

por lo tanto $(1/n)\sum_{i=0}^{n-1} |\, a_i| < 3\varepsilon$ si n es suficientemente grande.

(i) si y sólo si (iii). Esto resulta obvio pues por lo ya probado, basta notar que:

$$\lim_{\substack{n\to\infty \\ n\notin J}} |\, a_n| = 0 \quad \text{si y sólo si} \quad \lim_{\substack{n\to\infty \\ n\notin J}} |\, a_n|^2 = 0 \quad \text{para toda } J \quad \text{con} \quad d(J) = 0. \qquad \square$$

El teorema anterior da una manera de vefificar la propiedad de ser débilmente mezclante a partir de un subconjunto de los naturales con densidad cero. Utilice la expresión $\lim_{J\not\ni n\to\infty}$, la cual significa tomar el límite sobre los naturales que no están en J.

Corolario 1.27.1. *T es débilmente mezclante si y sólo si para todo $A, B \in \mathscr{F}$ existe $J = J(A,B) \subset \{0,1,\ldots\}$ tal que:*

$$d(J) = 0 \quad y \quad \lim_{J\not\ni n\to\infty} \mu(T^{-n}(A)\cap B) = \mu(A)\mu(B),$$

si y sólo si para todo $A, B \in \mathscr{F}$ se tiene que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\mu(T^{-i}(A)\cap B) - \mu(A)\mu(B)\right|^2 = 0.$$

Observación 1.29. ¿Es posible encontrar $J \subset \{0,1,2\ldots\}$ tal que

$$d(J) = 0 \quad \text{y} \quad \lim_{J\not\ni n\to\infty} \mu(T^{-n}(A)\cap B) = \mu(A)\mu(B) \quad \text{para toda } A, B \in \mathscr{F}?$$

Un caso en el que esto ocurre es cuando existe una familia numerable $\{E_k\}_{k=1}^{\infty}$ tal que para toda $\varepsilon > 0$ y para toda $A \in \mathscr{F}$ con $\mu(A) < \infty$, entonces existe E_k con $\mu(A \triangle E_k) < \varepsilon$. (Por ejemplo, la medida de Lebesgue sobre $\mathscr{B}_{[0,1)}$ tiene esta propiedad). La familia $\{E_k\}_{k=1}^{\infty}$ se llama una **base numerable** de $(X, \mathscr{F}, \mu)$.

Teorema 1.28. *Sea $(X, \mathscr{F}, \mu)$ un espacio de probabilidad con base numerable $\{E_i\}_{i=1}^{\infty}$ y $T : X \to X$ preservadora de medida. Entonces T es débilmente mezclante si y sólo si existe $J \subset \{0,1,2,\ldots\}$ con $d(J) = 0$ tal que para toda $A, B \in \mathscr{F}$*

$$\lim_{J\not\ni n\to\infty} \mu(T^{-n}(A)\cap B) = \mu(A)\mu(B).$$

Demostración. Basta probar la (Necesidad).

Sea $a_n = \sum_{i,j=1}^{\infty} |\mu(T^{-n}(E_i)\cap E_j) - \mu(E_i)\mu(E_j)| \cdot 2^{-i-j}$ como T es débilmente mezclante, se tiene que $(1/n)\sum_{k=0}^{n-1} a_k \to 0$. (Esto requiere de un argumento: sea $f_n : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ dada por $f_n(i,j) = (1/n)\sum_{k=0}^{n-1}\left|\mu(T^{-k}(E_i)\cap E_j - \mu(E_i)\mu(E_j)\right|$. Definimos una medida ν sobre $\mathcal{P}(\mathbb{N}\times\mathbb{N})$ extendiendo σ-aditivamente los valores $\nu(\{(i,j)\}) = 2^{-i-j}$. Notamos que $\nu(\mathbb{N}\times\mathbb{N}) = 1$. Como T es débilmente mezclante entonces $f_n(i,j) \to 0$ si $n \to \infty$ para toda i, j, además es claro que $|f_n(i,j)| \leq 1$ para toda i, j, n. Aplicando el teorema de convergencia acotada a $\{f_n\}$ obtenemos que:

$$\int_{\mathbb{N}\times\mathbb{N}} f_n\,d\nu \longrightarrow 0 \quad \text{es decir} \quad \frac{1}{n}\sum_{k=0}^{n-1} a_k \longrightarrow 0).$$

Por el Teorema 1.27 existe $J \subset \{0,1,2,\ldots\}$ con $d(J) = 0$ tal que: $\lim_{J\not\ni n\to\infty} a_n = 0$ por lo que:

$$\lim_{J\not\ni n\to\infty} \mu(T^{-n}(E_i)\cap E_j) = \mu(E_i)\mu(E_j) \quad \text{para toda } i, j.$$

Utilizando un razonamiento análogo a la prueba del Teorema 1.26 junto con la hipótesis de que el espacio en cuestión tiene base numerable se obtiene el resultado. Por lo tanto

$$\lim_{J\not\ni n\to\infty} \mu(T^{-n}(A)\cap B) = \mu(A)\mu(B). \qquad \square$$

Es posible expresar las nociones de ergodicidad, débilmente mezclante y fuertemente mezclante en términos del operador unitario inducido $\mathcal{U}_T : L_2^{\mathbb{R}}(\mu) \to L_2^{\mathbb{R}}(\mu)$, será de utilidad para mostrar propiedades mezclantes para nuestros ejemplos.

Teorema 1.29. *Sean $(X, \mathscr{F}, \mu)$ un espacio de probabilidad, $T : X \to X$ preservadora de medida y $\mathcal{U}_T : L_2^{\mathbb{R}}(\mu) \to L_2^{\mathbb{R}}(\mu)$ dada por $\mathcal{U}_T(f) = f \circ T$ para toda $f \in L_2^{\mathbb{R}}(\mu)$, entonces:*

(i) Las siguientes condiciones son equivalentes:

(1) *T es ergódica.*

(2) *Para toda $f, g \in L_2^{\mathbb{R}}(\mu)$ se tiene que:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \int \mathcal{U}_T^i(f) g\, d\mu = \int f\, d\mu \cdot \int g\, d\mu.$$

(3) *Para toda $f \in L_2^{\mathbb{R}}(\mu)$ se tiene que:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \int \mathcal{U}_T^i(f) f\, d\mu = \int f\, d\mu \cdot \int f\, d\mu.$$

(ii) Las siguientes condiciones son equivalentes:

(1) *T es débilmente mezclante.*

(2) *Para toda $f, g \in L_2^{\mathbb{R}}(\mu)$ se tiene que*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left| \int \mathcal{U}_T^i(f) g\, d\mu - \int f\, d\mu \cdot \int g\, d\mu \right| = 0.$$

(3) *Para toda $f \in L_2^{\mathbb{R}}(\mu)$ se tiene que*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left| \int \mathcal{U}_T^i(f) f\, d\mu - \int f\, d\mu \cdot \int f\, d\mu \right| = 0.$$

(iii) Las siguientes condiciones son equivalentes:

(1) *T es fuertemente mezclante.*

(2) *Para toda $f, g \in L_2^{\mathbb{R}}(\mu)$ se tiene que*

$$\lim_{n\to\infty} \int \mathcal{U}_T^n(f) g\, d\mu = \int f\, d\mu \cdot \int g\, d\mu.$$

(3) *Para toda* $f \in L_2^{\mathbb{R}}(\mu)$ *se tiene que*

$$\lim_{n\to\infty} \int \mathcal{U}_T^n(f) f \, d\mu = \int f \, d\mu \cdot \int f \, d\mu.$$

Demostración. (i), (ii) y (iii) se demuestran de manera similar, siendo la diferencia sólo en lo elaborado (más no díficil) del manipuleo algebraico, por lo que únicamente probaremos (iii). Para la prueba utilizo resultados conocidos de espacios de Hilbert, recordando que el producto interior en $L_2^{\mathbb{R}}(\mu)$ está dado por $\langle f, g\rangle = \int fg \, d\mu$.

(2) implica (1). Note que reescribiendo (2) en términos del producto interior, tenemos que para toda $f, g \in L_2^{\mathbb{R}}(\mu)$, $\langle \mathcal{U}_T^i(f), g\rangle \to \langle f, 1\rangle\langle 1, g\rangle$. Si $f = \chi_A$, $g = \chi_B$ $(A, B \in \mathscr{F})$, entonces obtenemos $\mu(T^{-i}(A) \cap B) \to \mu(A)\mu(B)$, por lo tanto T es fuertemente mezclante.

(1) implica (3). Supongamos que T es fuertemente mezclante, entonces:

$$\langle \mathcal{U}_T^n(\chi_A), \chi_B\rangle \to \langle \chi_A, 1\rangle\langle 1, \chi_B\rangle \quad \text{para toda } A, B \in \mathscr{F}.$$

Fijemos $B \in \mathscr{F}$, entonces por linealidad del producto interior obtenemos:

$$\langle \mathcal{U}_T^n(s), \chi_B\rangle \to \langle s, 1\rangle\langle 1, \chi_B\rangle \quad \text{para toda } s,\ S\text{-simple}.$$

Fijando ahora s y empleando el argumento anterior obtenemos:

$$\langle \mathcal{U}_T^n(s), s\rangle \to \langle s, 1\rangle\langle 1, s\rangle \quad \text{para toda } s.$$

Para $f \in L_2(\mu)$ fija, dada $\varepsilon > 0$, hallamos $s = s_\varepsilon$ S-simple tal que $\|f - s_\varepsilon\|_2 < \varepsilon$ y a continuación hallamos $N(\varepsilon) \in \mathbb{N}$ tal que

$$|\langle \mathcal{U}_T^n(s), s\rangle - \langle s, 1\rangle\langle 1, s\rangle| < \varepsilon \quad \text{si} \quad n \geq N(\varepsilon) \quad \text{entonces:}$$

$$\begin{aligned}
|\langle \mathcal{U}_T^n(f), f\rangle - \langle f, 1\rangle\langle 1, f\rangle| &\leq |\langle \mathcal{U}_T^n(f), f\rangle - \langle \mathcal{U}_T^n(s), f\rangle| \\
&\quad + |\langle \mathcal{U}_T^n(s), f\rangle - \langle \mathcal{U}_T^n(s), s\rangle| \\
&\quad + |\langle \mathcal{U}_T^n(s), s\rangle - \langle s, 1\rangle\langle 1, s\rangle| \\
&\quad + |\langle s, 1\rangle\langle 1, s\rangle - \langle f, 1\rangle\langle 1, s\rangle| \\
&\quad + |\langle f, 1\rangle\langle 1, s\rangle - \langle f, 1\rangle\langle 1, f\rangle| \\
&= I_1(n) + I_2(n) + I_3(n) + I_4 + I_5.
\end{aligned}$$

Estimamos usando la desigualdad de Schwarz:

$$\begin{aligned}
I_1(n) &= |\langle \mathcal{U}_T^n(f - s), f\rangle| \leq \|\mathcal{U}_T^n(f - s)\|_2 \|f\|_2 < \varepsilon \|f\|_2 \\
I_2(n) &= |\langle \mathcal{U}_T^n(s), f - s\rangle| \leq \|\mathcal{U}_T^n(s)\|_2 \|f - s\|_2 < \varepsilon \|s\|_2 \\
I_3(n) &< \varepsilon \quad \text{si } n \geq N(\varepsilon) \\
I_4 &= |\langle s, 1\rangle - \langle f, 1\rangle| \, |\langle 1, s\rangle| \leq \|f - s\|_2 \|s\|_2 < \varepsilon \|s\|_2 \\
I_5 &= |\langle f, 1\rangle| \, |\langle 1, s\rangle - \langle 1, f\rangle| \leq \|f\|_2 \|s - f\|_2 < \varepsilon \|f\|_2.
\end{aligned}$$

Así que $|\langle \mathcal{U}_T^n(f), f\rangle - \langle f, 1\rangle\langle 1, f\rangle| \leq \varepsilon(2\varepsilon + 1 + 4\|f\|_2)$ si $n \geq N(\varepsilon)$.

(3) implica (2). Supongamos que $\langle \mathcal{U}_T^n(f), f\rangle \to \langle f, 1\rangle\langle 1, f\rangle$ para toda $f \in L_2(\mu)$. Fijamos $f \in L_2(\mu)$ y sea $\mathscr{H}_f = \{$menor subespacio **cerrado** contenido en $L_2(\mu)$ tal que contiene a f, las constantes y $\mathcal{U}_T(\mathscr{H}_f) \subset \mathscr{H}_f\}$ y sea $\mathscr{G}_f = \{g \in L_2(\mu) : \langle \mathcal{U}_T^n(f), g\rangle \to \langle f, 1\rangle\langle 1, g\rangle\}$. Probaremos que $\mathscr{G}_f = L_2(\mu)$.

Claramente $\mathscr{G}_f$ contiene a las constantes, contiene a f por hipótesis, probaremos que es un subespacio cerrado de $L_2(\mu)$. Sea $h \in L_2(\mu)$ y $(h_n) \subset \mathscr{G}_f$, tal que $h_n \to h$, entonces

$$\begin{aligned}
\left|\langle \mathcal{U}_T^i(f), h\rangle - \langle f, 1\rangle\langle 1, h\rangle\right| &\leq \left|\langle \mathcal{U}_T^i(f), h\rangle - \langle \mathcal{U}_T^i(f), h_n\rangle\right| + \\
&\quad \left|\langle \mathcal{U}_T^i(f), h_n\rangle - \langle f, 1\rangle\langle 1, h_n\rangle\right| + \\
&\quad |\langle f, 1\rangle\langle 1, h_n\rangle - \langle f, 1\rangle\langle 1, h\rangle| \\
&\leq 2\|f\|_2\|h_n - h\|_2 + \\
&\quad |\langle \mathcal{U}^i(f), h_n\rangle - \langle f, 1\rangle\langle 1, h_n\rangle| \\
&< \varepsilon,
\end{aligned}$$

si i, n son suficientemente grandes, por lo que $h \in \mathscr{G}_f$, por lo tanto $\mathscr{G}_f$ es cerrado. Así que $\mathscr{H}_f \subset \mathscr{G}_f$. Ahora supongamos que $g \perp \mathscr{H}_f$, así pues $g \perp \mathcal{U}_T^n(\mathscr{H}_f)$ para toda n, por lo que $\langle \mathcal{U}_T^n(f), g\rangle = 0$ para toda n y además $\langle 1, g\rangle = 0$ (pues $1 \in \mathscr{H}_f$), se tiene que $\mathscr{H}_f^\perp \subset \mathscr{G}_f$ por lo tanto

$$L_2(\mu) = \mathscr{H}_f \oplus \mathscr{H}_f^\perp \subset \mathscr{G}_f. \qquad \square$$

El siguiente resultado muestra que la noción de ser débilmente mezclante no es un concepto artificial que se halla "entre" ergodicidad y el ser fuertemente mezclante, sino que se relaciona con la ergodicidad del producto cartesiano.

Teorema 1.30. *Sean $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $T : X \to X$ preservadora de medida. Las siguientes condiciones son equivalentes:*

(i) T es débilmente mezclante.

(ii) $T \times T$ es ergódica.

(iii) $T \times T$ es débilmente mezclante.

Demostración.

(i) implica (iii). Sean A, B, C y $D \in \mathscr{F}$ entonces existen $J_1, J_2 \subset \mathbb{N}$ ($J_1(A,B)$ y $J_2(C,D)$) de densidad cero tal que

$$\lim_{J_1 \not\ni n \to \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \quad \text{y} \quad \lim_{J_2 \not\ni n \to \infty} \mu(T^{-n}(C) \cap D) = \mu(C)\mu(D).$$

Sea $J = J_1 \cup J_2$, entonces $d(J) = 0$ y

$$\begin{aligned}
\lim_{J \not\ni n \to \infty} \mu \otimes \mu((T \times T)^{-n}(A \times C) \cap (B \times D)) &= \lim_{J \not\ni n \to \infty} \mu(T^{-n}(A) \cap B)\mu(T^{-n}(C) \cap D) \\
&= \mu(A)\mu(B)\mu(C)\mu(D) \\
&= \mu \otimes \mu(A \times C)\mu \otimes \mu(B \times D).
\end{aligned}$$

Así que la condición adecuada es válida para rectángulos medibles y por ende para el álgebra generada por dichos rectángulos (consistente de uniones finitas disjuntas de rectángulos medibles). Como dicha álgebra genera a $\mathscr{F}$ obtenemos que $T \times T$ es débilmente mezclante (por Teorema 1.26(ii)).

(iii) implica (ii). Es claro puesto que si T es débilmente mezclante entonces T es ergódica.

(ii) implica (i). Sean $A, B \in \mathscr{F}$ arbitrarios entonces:

$$\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu \otimes \mu((T \times T)^{-i}(A \times X) \cap (B \times X)) \\
&= \mu(A \times X)\mu(B \times X) \\
&= \mu(A)\mu(B).
\end{aligned}$$

Por otro lado

$$\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} [\mu(T^{-i}(A) \cap B)]^2 &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu \otimes \mu((T \times T)^{-i}(A \times A) \cap (B \times B)) \\
&= \mu \otimes \mu(A \times A)\mu \otimes \mu(B \times B) \\
&= (\mu(A))^2(\mu(B))^2.
\end{aligned}$$

De donde:

$$\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} [\mu(T^{-i}(A) \cap B) - \mu(A)\mu(B)]^2 &= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} [\mu(T^{-i}(A) \cap B) - \\
&\quad -2\mu(T^{-i}(A) \cap B)\mu(A)\mu(B) + \mu(A)^2\mu(B)^2] \\
&= 2\mu(A)^2\mu(B)^2 - 2\mu(A)^2\mu(B)^2 \\
&= 0.
\end{aligned}$$

Por lo tanto T es débilmente mezclante por el Corolario 1.27.1. □

Corolario 1.30.1. *La rotación irracional en S^1 es ergódica pero no débilmente mezclante.*

Demostración. Consideramos el siguiente conjunto $A = \left\{(e^{2\pi i\theta}, e^{2\pi i\gamma}) : |\theta - \gamma| \leq \frac{1}{4}\right\}$. La representación de A en el cuadrado unitario sería la siguiente:

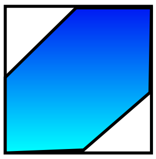

Figura 1.9: $A = \left\{(e^{2\pi i\theta}, e^{2\pi i\gamma}) : |\theta - \gamma| \leq \frac{1}{4}\right\}$.

Observamos que $T^{-1}(A) = A$ ya que los siguientes enunciados son equivalentes:

(i) $(x, y) \in T^{-1}(A)$,

(ii) $(T(x), T(y)) \in A$,

(iii) $(e^{i(\alpha+\theta)}, e^{i(\alpha+\gamma)}) \in A$,

(iv) $|\alpha + \theta - (\alpha + \gamma)| \leq \frac{1}{4}$,

(v) $|\theta - \gamma)| \leq \frac{1}{4}$,

(vi) $(x, y) \in A$.

Como $0 < \mu(A) < 1$ se tiene que $T \times T$ no es ergódica, por el Teorema 1.28 se sigue que T no es débilmente mezclante. $\square$

Observación 1.30. Argumentos similares (aún más simples) prueban que T es fuertemente mezclante si y sólo si $T \times T$ es fuertemente mezclante, por lo que también tiene tal propiedad multiplicativa.

1.9. Propiedades del Operador Inducido

En esta sección se relacionan los conceptos de ergodicidad, débilmente y fuertemente mezclante con las propiedades espectrales del operador inducido por una transformación preservadora de medida. Recordando que $\mathcal{U}_T(f) = f \circ T$, se da la siguiente definición:

Definición 1.14. Sea $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $T : X \to X$ una transformación preservadora de medida. Decimos que $\lambda \in \mathbb{C}$ es un **valor propio** de T, si existe $f \in L_2(\mu)$ $(f \neq 0)$ tal que $\mathcal{U}_T(f) = \lambda f$. Llamamos a f un **vector propio** correspondiente al valor propio λ.

Teorema 1.31. *Sea $(X, \mathscr{F}, \mu)$ un espacio de probabilidad y $T : X \to X$ preservadora de medida. Si además T es ergódica entonces:*

(i) Los vectores propios de T tienen módulo constante (casi dondequiera).

(ii) Los valores propios de T forman un subgrupo de S^1.

(iii) Los vectores propios correspondientes a valores propios diferentes son ortogonales.

(iv) Los espacios propios correspondientes son unidimensionales.

Demostración.

(i) Si λ es un valor propio de T y $f \neq 0$ un vector propio asociado a λ, entonces como $|\lambda| = 1$ y $|f| \circ T = |f \circ T| = |\lambda f| = |\lambda|\,|f| = |f|$ entonces $|f|$ es constante casi dondequiera (T ergódica se uso aquí).

(ii) Es suficiente probar que si λ y τ son valores propios de T, entonces $\lambda\tau^{-1} = \lambda\overline{\tau}$ es un valor propio de T. Pero si f y g son vectores propios correspondientes a λ y τ (respectivamente) entonces $\overline{g}$ es un vector propio asociado a $\overline{\tau}$. Además $\mathcal{U}_T(f\overline{g}) = \lambda\overline{\tau}(f\overline{g})$ y como $f\overline{g} \in L_2(\mu)$ (pues tienen módulo constante c. d.) entonces $\lambda\overline{\tau}$ es un valor propio de T y $f\overline{g}$ es un vector propio asociado a $\lambda\overline{\tau}$. (T preservadora de medida).

(iii) Sean λ y τ valores propios **diferentes** de T. Si f y g son vectores propios asociados a λ y τ entonces como $\mathcal{U}_T$ preserva el producto interior se sigue:

$$\langle f\,, g\rangle = \langle \mathcal{U}_T(f)\,, \mathcal{U}_T(g)\rangle = \lambda\overline{\tau}\langle f\,, g\rangle \quad \text{así pues} \quad \langle f\,, g\rangle = 0.$$

(T preservadora de medida).

(iv) Si λ es un valor propio y g un vector propio entonces $|g|$ es constante c. d. (constante distinta de cero). Si f es otro vector propio, entonces f/g es un vector propio asociado a 1 por lo tanto f/g es constante casi dondequiera. □

Proposición 1.2. *T es ergódica si y sólo si 1 es un valor propio* ***simple*** *(es decir el subespacio propio correspondiente a 1, $\{g \in L_2(\mu) : \mathcal{U}_T(g) = g\}$ es unidimensional).*

Demostración.

(Necesidad). Si g es un vector propio correspondiente a 1, entonces g es constante casi dondequiera pues T es ergódica, entonces $\{g \in L_2(\mu) : \mathcal{U}_T(g) = g\}$ =subespacio generado por la función constante $= 1$. Por lo tanto 1 es un valor propio simple.

(Suficiencia). Claramente cualquier función constante $\neq 0$ es un vector propio correspondiente a 1 entonces $\{g \in L_2(\mu) : \mathcal{U}_T(g) = g\} = \{\text{constantes}\}$, por lo que T es ergódica. □

Observación 1.31. En cualquier caso $E_\lambda = \{g \in L_2(\mu) : \mathcal{U}_T(g) = \lambda g\}$ es siempre un subespacio **cerrado** de $L_2(\mu)$.

Demostración. Sea $h \in L_2(\mu)$ y $\{h_n\} \subset E_\lambda$ tal que $\|h_n - h\|_2 \to 0$, entonces

$$\|\mathcal{U}_T(h) - \lambda h\|_2 = \|\mathcal{U}_T(h) - \mathcal{U}_T(h_n) + \mathcal{U}_T(h_n) - \lambda h_n + \lambda h_n - \lambda h\|_2 \leq \|h - h_n\|_2 + |\lambda|\|h - h_n\|_2 < \varepsilon$$

si n es suficientemente grande, por lo tanto E_λ es cerrado. □

Definición 1.15. Decimos que T (como en la Definición 1.14, pág. 56), tiene **espectro continuo** si 1 es el único valor propio y los únicos vectores propios son las funciones constantes.

Observación 1.32. T tiene espectro continuo si y sólo si 1 es el único valor propio y T es ergódica.

Teorema 1.32. *T es dédilmente mezclante si y sólo si T tiene espectro continuo. (T invertible).*

Demostración.

(Necesidad). Supongamos que λ es un valor propio de T y f un vector propio correspondiente a λ (es decir $\mathcal{U}_T(f) = \lambda f$ $(f \neq 0)$). Supongamos que $\lambda \neq 1$, entonces como $\langle \mathcal{U}_T(f), 1\rangle = \langle f, 1\rangle$ (pues T es preservadora de medida), entonces $\lambda\langle f, 1\rangle = \langle f, 1\rangle$ por lo tanto $\langle f, 1\rangle = 0$. Como T es débilmente mezclante se tiene que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\langle \mathcal{U}_T^i(f), f\rangle\right| = 0, \quad \text{así que,} \quad \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\lambda|^i \,|\langle f, f\rangle| = 0, \quad \text{o bien,}$$

$$\lim_{n\to\infty} \|f\|_2^2 \left(\frac{1}{n}\sum_{i=0}^{n-1} |\lambda|^i\right) = 0,$$

lo cual sólo es posible si $f = 0$ (pues $|\lambda| = 1$, dado que $\mathcal{U}_T$ es isometría) por lo tanto f no es un vector propio, lo cual contradice lo supuesto. Concluimos que si $\lambda \neq 1$, entonces no puede ser valor propio de T, ciertamente 1 sí lo es y su subespacio propio asociado es unidimensional pues T es ergódica, por lo tanto T tiene espectro continuo. (En esta parte **no se usó** el hecho de que T sea invertible).

(Suficiencia). Esta parte requiere del uso de la descomposición espectral $\{E\}$ del operador unitario $\mathcal{U}_T$. Denotemos por $\mu_f : \mathscr{B}_{S^1} \to \mathbb{R}$ la medida de Borel definida por:

$$\mu_f(B) = \langle E(B)f, f\rangle \quad \text{para toda } f \in L_2(\mu), \quad \text{entonces} \quad \langle \mathcal{U}_T^n(f), f\rangle = \int_{S^1} z^n d\mu_f(z) \quad n \geq 0,$$

(ver Halmos [10, págs. 71-73]). Es un hecho conocido que si $\mathcal{U}_T$ tiene espectro continuo y $\langle 1, f\rangle = 0$ entonces μ_f es una medida no atómica (es decir $\mu_f(\{w\}) = 0$ para todo $w \in \mathbb{C}$). Debemos probar que

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\langle \mathcal{U}_T^i(f), f\rangle - \langle f, 1\rangle\langle 1, f\rangle\right|^2 = 0 \quad \text{para toda } f \in L_2(\mu).$$

Como lo anterior es evidente para f constante, es suficiente verificarlo sólo para aquellas $f \in L_2(\mu)$ tal que $\langle 1, f\rangle = 0$ (ya que si $h \in L_2(\mu)$, entonces $h = h_1 + \mathit{cte}$ con $h_1 \perp 1$ $h_1 \in L_2(\mu)$). En cuyo caso lo anterior se reduce a verificar que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\langle \mathcal{U}_T^i(f), f\rangle\right|^2 = 0 \quad \text{para toda } f \in L_2(\mu) \quad \text{con} \quad \langle 1, f\rangle = 0, \quad \text{es decir,}$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \left|\int_{S^1} \lambda^i d\mu_f(\lambda)\right|^2 = 0 \quad \text{para toda } f \in L_2(\mu) \quad \text{con} \quad \langle 1, f\rangle = 0.$$

Pero:

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1}\left|\int_{S^1}\lambda^i d\mu_f(\lambda)\right|^2 &= \frac{1}{n}\sum_{i=0}^{n-1}\int_{S^1}\lambda^i d\mu_f(\lambda)\cdot\int_{S^1}\overline{\tau}^{-i}d\mu_f(\tau)\\
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_{S^1\times S^1}\lambda^i\overline{\tau}^i d(\mu_f\otimes\mu_f)(\lambda,\tau)\\
&= \int_{S^1\times S^1}\left(\frac{1}{n}\sum_{i=0}^{n-1}(\lambda\overline{\tau})^i d(\mu_f\otimes\mu_f)(\lambda,\tau)\right).
\end{aligned}$$

Sea D =diagonal de $S^1\times S^1$. Si $(\lambda,\tau)\notin D$ (es decir $\lambda\overline{\tau}\neq 1$) entonces:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}(\lambda\overline{\tau})^i = \lim_{n\to\infty}\frac{1}{n}\left[\frac{1-(\lambda\overline{\tau})^n}{1-(\lambda\overline{\tau})}\right]=0.$$

Por otro lado, como μ_f no posee átomos entonces $\mu_f\otimes\mu_f(D)=0$, así que:

$$\lim_{n\to\infty}\int_{S^1\times S^1}\left(\frac{1}{n}\sum_{i=0}^{n-1}(\lambda\overline{\tau})^i d(\mu_f\otimes\mu_f)\right)=\lim_{n\to\infty}\frac{1}{n}\int_{S^1\times S^1\setminus D}\left(\frac{1-(\lambda\overline{\tau})^n}{1-(\lambda\overline{\tau})}\right)d(\mu_f\otimes\mu_f)=0,$$

por el T.C.D.L.[9] (ver Teorema B.5, pág. 133). □

Recapitulando: hemos caracterizado las transformaciones ergódicas y las débilmente mezclantes, en términos espectrales. Al final del capítulo anterior probamos que la rotación irracional no es débilmente mezclante, ahora probaremos este hecho de una manera más sencilla como sigue:

Corolario 1.32.1. *La rotación irracional $T: S^1\to S^1$ dada por $T(z)=az$ ($a=e^{i\rho}$ con $\rho/2\pi\notin\mathbb{Q}$) no es débilmente mezclante.*

Demostración. Es suficiente comprobar que $\mathcal{U}_T$ no tiene espectro continuo. Sea $n\in\mathbb{Z}\setminus\{0\}$, considérese $f_n(z)=z^n$. Note que $f_n\circ T=f_n(az)=a^nz^n=a^nf_n$, por lo tanto f_n es un vector propio asociado al valor propio a^n, el cual es distinto de 1 para toda $n\neq 0$. □

Aún más, $\{z^n\}_{n\in\mathbb{Z}}$ construye un sistema ortonormal **completo** de $L_2(\mu)$. A transformaciones con esta propiedad (es decir en la que existe una base de vectores propios para $L_2(\mu)$) las llamaremos transformaciones con **espectro discreto**.

Observación 1.33. Otra prueba del corolario anterior es como sigue: $F: S^1\times S^1 :\to\mathbb{C}$ definida por $F(x,y)=x^ny^{-n}$ no es constante (c. d. relativo a $\mu\otimes\mu$) y es $T\times T$-invariante, así que $T\times T$ no es ergódica, por lo tanto T no es débilmente mezclante.

[9]Notése que $\left|(1/n)\sum_{i=0}^{n-1}(\lambda\overline{\tau})^i\right|\leq(1/n)\sum_{i=0}^{n-1}|(\lambda\overline{\tau})^i|=1\in L(\mu_f\otimes\mu_f)$ y que $\mu_f(X)=\|f\|_2^2<\infty$.

Capítulo 2

Isomorfismo Espectral e Isomorfismo

Uno de los problemas de la Teoría Ergódica consiste en clasificar a las transformaciones preservadoras de medida, es decir, ¿cuándo se considera que dos de estas transformaciones son "isomorfas"? En este capítulo se abordará este problema con el fin de dar una mejor clasificación de estas transformaciones. Sin embargo, no se profundizará en este tema (ver Walters [23, cap. 2,3]).

En este parte se definen los conceptos de isomorfismo e isomorfismo espectral. Para el último se utiliza el operador lineal unitario inducido por una transformación preservadora de la medida T; considerando dos transformaciones T_1 y T_2 con sus respectivos espacios $L_2(\mu)$ y $L_2(\mu')$; éstos en particular son espacios de Hilbert; basandose en la definición de isomorfo para espacios de Hilbert se define el isomorfismo espectral. Para la definición de isomorfismo se debe recordar que en Teoría de la Medida un conjunto de medida cero puede ser ignorado; dos espacios de medida pueden ser considerados iguales salvo por un conjunto de medida cero; basándose en lo anterior se define isomorfismo.

2.1. Isomorfismo Espectral

Ahora introduciremos uno de los conceptos de "isomorfismo" entre transformaciones preservadoras de medida. Para esto utilizo el operador unitario inducido por estas transformaciones. Recordando que estos operadores están definidos en espacios de Hilbert, intuitivamente bajo las condiciones mencionadas lo menos que deberíamos pedir es que los correspondientes espacios de Hilbert de las transformaciones sean isomorfos, sin embargo necesitaremos pedir una condición más para considerar que dos transformaciónes preservadoras de medida sean "isomorfas" en este sentido. Más adelante daremos otra noción de isomorfismo.

Definición 2.1 (Isomorfismo Espectral)**.** Sean $(X_i, \mathscr{F}_i, \mu_i)$ $(i = 1, 2)$ espacios de probabilidad y $T_i : X_i \to X_i$ transformaciones preservadoras de medida $(i = 1, 2)$. Decimos que T_1 y T_2 son

espectralmente isomorfas (denotado $T_1 \overset{e}{\simeq} T_2$) si existe $W : L_2(\mu_2) \to L_2(\mu_1)$ un operador lineal entre espacios de Hilbert tal que:

(a) W es invertible.

(b) $\langle W(f), W(h) \rangle = \langle f, h \rangle$ para toda $f, h \in L_2(\mu_2)$.[1]

(c) $\mathcal{U}_{T_1} \circ W = W \circ \mathcal{U}_{T_2}$.

(Las condiciones (a) y (b) dicen que W es un isomorfismo de espacios de Hilbert). Varios de los

$$\begin{tikzcd} L_2(\mu_2) \arrow[r, "\mathcal{U}_{T_2}"] \arrow[d, "W"'] & L_2(\mu_2) \arrow[d, "W"] \\ L_2(\mu_1) \arrow[r, "\mathcal{U}_{T_1}"] & L_2(\mu_1) \end{tikzcd}$$

Figura 2.1: Diagrama de $T_1 \overset{e}{\simeq} T_2$.

problemas de Teoría Ergódica se resuelven a partir de este concepto. Note que $\overset{e}{\simeq}$ es una relación de equivalencia. Era de esperarse que este concepto preserve las propiedades de ergodicidad y mezlantes, lo anterior es consecuencia del siguiente resultado.

Proposición 2.1. *Sean $T_i : X_i \to X_i$ $(i = 1, 2)$ transformaciones preservadoras de medida (en sus respectivos espacios) espectralmente isomorfas, entonces:*

(i) T_1 es ergódica si y sólo si T_2 lo es.

(ii) T_1 es débilmente mezclante si y sólo si T_2 lo es.

(iii) T_1 es fuertemente mezclante si y sólo si T_2 lo es.

Demostración.

(i) Sabemos que T es ergódica si y sólo si 1 es valor propio simple. Ahora supongamos que $T_1 \overset{e}{\simeq} T_2$ y que T_1 es ergódica, por lo anterior basta con probar que T_2 tiene a 1 como valor propio simple. Sean $E_1^{T_1} = \{f \in L_2(\mu_1) : \mathcal{U}_{T_1}(f) = f\}$, $E_1^{T_2} = \{g \in L_2(\mu_2) : \mathcal{U}_{T_1}(g) = g\}$, como las constantes pertenecen a $E_1^{T_2}$, basta con probar que $\dim(E_1^{T_2}) = 1$. Sea $g \in E_1^{T_2}$, entonces $\mathcal{U}_{T_1}(W(g)) = W(\mathcal{U}_{T_2}(g)) = W(g)$, como T_1 es ergódica entonces $W(g)$ es constante c.d., por lo tanto $W(E_1^{T_2}) \subset E_1^{T_1}$, como W es inyectivo se tiene que $\dim(E_1^{T_2}) = 1$.

[1]Note que son productos interiores en espacios de Hilbert probablemente diferentes.

(ii) Por el Teorema 1.27, basta con probar que:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} |\langle \mathcal{U}_{T_2}^i(g)\,,g\rangle - \langle g\,,1\rangle\langle 1\,,g\rangle| = 0 \quad \text{para toda } g \in L_2(\mu_2),$$

como es válido para las constantes y $E_1^{T_2}$ es cerrado basta con probarlo para las $g \in L_2(\mu_2)$ tales que $\langle g, 1\rangle = 0$, además por el inciso anterior W manda el espacio de funciones contantes de $L_2(\mu_2)$ al espacio de funciones contantes de $L_2(\mu_1)$ por lo que $\langle W(g), 1\rangle = 0$, por lo tanto:

$$\frac{1}{n}\sum_{i=0}^{n-1} |\langle \mathcal{U}_{T_2}^i(g), g\rangle| = \frac{1}{n}\sum_{i=0}^{n-1} |\langle W(\mathcal{U}_{T_2}^i(g)), W(g)\rangle| = \frac{1}{n}\sum_{i=0}^{n-1} |\langle \mathcal{U}_{T_1}^i(W(g)), W(g)\rangle| \to 0.$$

(iii) Análogo al anterior. □

Observación 2.1. Si $T_1 \overset{e}{\simeq} T_2$ entonces $\mathcal{U}_{T_1}$ y $\mathcal{U}_{T_2}$ tienen los mismos valores propios. Si f es un vector propio para T_1 correspondiente a λ, entonces $W(f)$ es un vector propio para T_2 correspondiente a λ. Inversamente si g es un vector propio para T_2 correspondiente a λ, entonces $W^{-1}(g)$ es un vector propio para T_1 correspondiente a λ.

Otro modo de enunciar la Proposición 2.1 es la siguiente: Las nociones de ergodicidad, débilmente mezclante y fuertemente mezclante son **invariantes** bajo isomorfismo espectral. Hemos visto que si dos transformaciones son isomorfas entonces tienen las mismas propiedades que hemos mencionado, que es lo mínimo que se podía esperar de una relación de isomorfismo. Notemos que una rotación irracional no puede ser isomorfa a una transformación que sea débilmente o fuertemente mezclante.

Hemos relacionado las propiedades de ergodicidad y débilmente mezclante, con las propiedades espectrales del operador $\mathcal{U}_T$, cabe preguntarnos: ¿Existirá una condición sobre las propiedades espectrales de los operador $\mathcal{U}_T$ que nos diga cuando dos transformaciones preservadoras de medida son espectralmente isomorfas? La respuesta es afirmativa, para ello necesitaremos la siguiente definición:

Definición 2.2. Sean $(X, \mathscr{F}, \mu)$ un espacio de medida con base numerable (ver Teorema 1.28, pág. 51), y $T : X \to X$ una transformación preservadora de medida invertible, diremos que T tiene **espectro de Lebesgue numerable** si existe una sucesión $\{f_i\}_{i=0}^{\infty} \subset L_2(\mu)$ con $f_0 \equiv 1$, tal que $\{f_0\} \cup \{\mathcal{U}_T^k(f_i) | i \geq 1, k \in \mathbb{Z}\}$ es una base ortonormal de $L_2(\mu)$.

Gráficamente la base tiene la forma:

$$\begin{array}{ccccccc}
 & & & f_0 & & & \\
\ldots & \mathcal{U}_T^{-2}(f_1) & \mathcal{U}_T^{-1}(f_1) & f_1 & \mathcal{U}_T(f_1) & \mathcal{U}_T^2(f_1) & \ldots \\
\ldots & \mathcal{U}_T^{-2}(f_2) & \mathcal{U}_T^{-1}(f_2) & f_2 & \mathcal{U}_T(f_2) & \mathcal{U}_T^2(f_2) & \ldots \\
 & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}$$

Teorema 2.1. *Cualesquiera dos transformaciones preservadoras de medida invertibles con espectro de Lebesgue numerable son espectralmente isomorfas.*

Demostración. Supongamos que $L_2(\mu_1)$ tiene base $A = \{f_0\} \cup \{\mathcal{U}_{T_1}^k(f_i) | i \geq 1, k \in \mathbb{Z}\}$ y $L_2(\mu_2)$ tiene base $B = \{g_0\} \cup \{\mathcal{U}_{T_2}^k(g_i) | i \geq 1, k \in \mathbb{Z}\}$. Definimos $W : L_2(\mu_2) \to L_2(\mu_1)$ como sigue $W(g_0) = f_0$, $W(\mathcal{U}_{T_2}^k(g_i)) = \mathcal{U}_{T_1}^k(f_i)$ y la extendemos por linealidad. Como A, B son bases ortonormales se tiene que $\langle W(f'), W(g')\rangle = \langle f', g'\rangle$ para toda $f', g' \in B$, además

$$\mathcal{U}_{T_1}(W(\mathcal{U}_{T_2}^k(g_i))) = \mathcal{U}_{T_1}(\mathcal{U}_{T_1}^k(f_i)) = \mathcal{U}_{T_1}^{k+1}(f_i) \quad \text{y} \quad W(\mathcal{U}_{T_2}(\mathcal{U}_{T_2}^k(g_i))) = W(\mathcal{U}_{T_2}^{k+1}(g_i)) = \mathcal{U}_{T_1}^{k+1}(f_i).$$

Así que es válido para combinaciones lineales finitas de elementos de B. Como el conjunto de combinaciones lineales finitas de B es denso en $L_2(\mu_2)$, se tiene que $\langle W(f), W(g)\rangle = \langle f, g\rangle$ para toda $f, g \in L_2(\mu_2)$ y $\mathcal{U}_{T_1} \circ W = W \circ \mathcal{U}_{T_2}$. Por lo tanto T_1 y T_2 son espectralmente isomorfas. $\square$

Lo anterior será de utilidad más adelante, específicamente en el capítulo cuatro.

2.2. Isomorfismo

Hemos visto una noción de isomorfismo utilizando el espacio de Hilbert $L_2(\mu)$ asociado a un espacio de medida $(X, \mathscr{F}, \mu)$. Sin embargo existe otra noción de isomorfismo. Recordando que en Teoría de la Medida los espacios de medida cero pueden ser ignorados, utilizando este hecho daremos una noción más fuerte de isomorfismo (lo anterior será probado en este capítulo). Considerese el siguiente ejemplo:

Ejemplo 2.1. Para $\alpha \in [0, 1)$ fijo. Sea $T_1(x) = x + \alpha \pmod 1$ con el espacio y la medida definidas en el Ejemplo 1.2 (pág. 8). Sea $T_2(z) = az$, donde $a = e^{i\alpha}$, con el espacio y la medida definidas en el Ejemplo 1.4 (pág. 9). Considero la transformación $\phi : [0, 1) \to S^1$, dada por $\phi(x) = e^{2\pi i x}$. A partir del Teorema 1.3 (pág. 7) y considerando el π-sistema C utilizado en el Ejemplo 1.4, se tiene que ϕ es una transformación preservadora de medidas invertible. Además $\phi(T_1(x)) = T_2(\phi(x))$ para todo $x \in [0, 1)$. Se puede considerar a T_1 y T_2 como "las mismas" ya que existe un isomorfismo ϕ entre espacios de medida que manda T_2 a $\phi \circ T_2 \circ \phi^{-1}$ que es T_1. Note que en este ejemplo no hubo que considerar conjuntos de medida cero. Más adelante, daré un ejemplo, donde considero los espacios de medida dados después de remover subconjuntos de medida cero de éstos.

Definición 2.3. Sean T_i $(i = 1, 2)$ transformaciones preservadoras de medida actuando sobre espacios de probabilidad $(X_i, \mathscr{F}_i, \mu_i)$ $(i = i, 2)$. Decimos que T_1 y T_2 son **isomorfas** (denotado $T_1 \simeq T_2$) si existen $M_i \subset X_i$ $(M_i \in \mathscr{F}_i)$ tales que:

(a) $\mu_i(M_i) = 1$ y $T_i(M_i) \subset M_i$ $(i = 1, 2)$.

(b) Existe $\phi : M_1 \to M_2$ una transformación **invertible y preservadora de medidas** (es decir, $\mu_2 = \mu_1 \circ \phi^{-1}$ o bien $\mu_1 = \mu_2 \circ \phi$), tal que $\phi \circ T_1(x) = T_2 \circ \phi(x)$ para toda $x \in M_1$.

A M_i lo equipamos con la σ-álgebra relativa $\mathscr{F}_i \cap M_i = \{E_i \cap M_i : E_i \in \mathscr{F}\}$.

$$\begin{tikzcd}
X_1 & X_2 \\
\bigcup & \bigcup \\
M_1 \arrow[r, "\phi"] \arrow[d, "T_1"'] & M_2 \arrow[d, "T_2"] \\
M_1 \arrow[r, "\phi"] & M_2 \\
\bigcap & \bigcap \\
X_1 & X_2
\end{tikzcd}$$

Figura 2.2: Diagrama de $T_1 \simeq T_2$.

Observaciones 2.2.

(1) $\simeq$ es una relación de equivalencia.

(2) Se puede probar que si $T_1 \simeq T_2$ entonces $T_1^n \simeq T_2^n$ para cada $n \in \mathbb{N}$.

(3) Si T_1 y T_2 son invertibles y $T_1 \simeq T_2$, existen M_1' M_2' tales que $\mu_i(M_i') = 1$ y $T_i(M_i') = M_i'$ sólo debemos tomar $M_i' = \bigcap_{k=-\infty}^{\infty} T^k(M_i)$.

(4) Dos espacios de medida son isomorfos si las transformaciones identidad son isomorfas.

Teorema 2.2. *Si T_1 y T_2 son isomorfas entonces son espectralmente isomorfas.*

Demostración. Como $\mu_i(X_i \setminus M_i) = 0$ $(i = 1, 2)$, entonces los espacios $L_2(X_i, \mathscr{F}_i, \mu_i)$ y $L_2(M_i, \mathscr{F}_i \cap M_i, \mu_i)$ $(i = 1, 2)$ pueden considerarse como idénticos (pues funciones iguales casi dondequiera se consideran idénticas), y así lo haremos. Definimos $W : L_2(M_1) \to L_2(M_2)$ poniendo $W(f) = f \circ \phi^{-1}$. Claramente $W(f) \in L_2(M_2)$. Debemos verificar que: W es invertible, preserva productos interiores y $\mathcal{U}_{T_2} \circ W = W \circ \mathcal{U}_{T_1}$.

Claramente $W^{-1} : L_2(M_2) \to L_2(M_1)$ definido por $W^{-1}(g) = g \circ \phi$ es la inversa de W y ambas son transformaciones lineales. Además:

$$\begin{aligned}
\langle W(f)\,, W(h)\rangle = \langle f \circ \phi^{-1}\,, h \circ \phi^{-1}\rangle &= \int (f \circ \phi^{-1})(\overline{h \circ \phi^{-1}}) d\mu_2 \\
&= \int (f\overline{h}) \circ \phi^{-1} d\mu_2 \\
&= \int f\overline{h}\mu_1 \quad \text{por Lema 1.3} \\
&= \langle f, h\rangle.
\end{aligned}$$

Finalmente si $f \in L_2(\mu_1)$, entonces:

$$\mathcal{U}_{T_2} \circ W(f) = \mathcal{U}_{T_2}(f \circ \phi^{-1}) = f \circ \phi^{-1} \circ T_2 = f \circ T_1 \circ \phi^{-1} = (\mathcal{U}_{T_1}(f)) \circ \phi^{-1} = W \circ \mathcal{U}_{T_1}(f). \qquad \square$$

El teorema anterior nos dice que la noción de isomorfismo espectral es más débil que la de isomorfismo. El recíproco del teorema anterior es falso, sin embargo para mostrarlo tendremos que esperar hasta el cuarto capítulo. A continuación daremos ejemplos:

Ejemplo 2.2. Sea $T_2 : [\,0\,,1\,) \to [\,0\,,1\,)$ dada por $T_2(x) = 2x \pmod 1$ ($X = [\,0\,,1\,)$, $\mathscr{F}_2 = \mathscr{B}_{[\,0\,,1\,)}$) y μ_2 =Lebesgue en $[\,0\,,1\,)$) y $T_1 : \{0,1\}^{\mathbb{N}} \to \{0,1\}^{\mathbb{N}}$ el corrimiento **unilateral** de Bernoulli con parámetros $(1/2, 1/2)$, con su espacio de medida definido en el Ejemplo 1.8. Probaré que $T_1 \simeq T_2$.

Demostración. Definimos $\phi : \{0,1\}^{\mathbb{N}} \to [\,0\,,1\,)$ como sigue: $\phi(x) = \sum_{i=1}^{\infty} x(i)/2^i$. Observe que ϕ no es 1 a 1 solamente en aquellos x tal que $x(i)$ es constante a partir de cierto momento. Sea

$$D_1 = \left\{x \in \{0,1\}^{\mathbb{N}} : \text{existe } n \in \mathbb{N} \quad \text{tal que} \quad x(m) = 0 \quad \text{para toda } m \geq n\right\} \cup$$
$$\left\{x \in \{0,1\}^{\mathbb{N}} : \text{existe } n \in \mathbb{N} \quad \text{tal que} \quad x(m) = 1 \quad \text{para toda } m \geq n\right\}.$$

Claramente D_1 es numerable, por lo que $\mu_1(D_1) = 0$. Sea $D_2 = \{k/2^n : k = 0, 1 \ldots, 2^n - 1\}$, entonces D_2 es numerable y además $\mu_2(D_2) = 0$. Es claro que $\phi : \{0,1\}^{\mathbb{N}} \setminus D_1 \to [0,1) \setminus D_2$ es una biyección. Ahora bien para todo $x \in \{0,1\}^{\mathbb{N}} \setminus D_1$ se tiene que:

$$\phi \circ T_1(x) = \phi((x_2, x_3, \ldots)) = \sum_{i=1}^{\infty} \frac{x(i+1)}{2^i} \quad \text{y}$$

$$T_2 \circ \phi(x) = T_2\left(\sum_{i=1}^{\infty} \frac{x(i)}{2^i}\right) = 2\left(\sum_{i=1}^{\infty} \frac{x(i)}{2^i}\right) \quad (\text{mod}1) = \sum_{i=1}^{\infty} \frac{x(i+1)}{2^i}.$$

Por lo tanto $\phi \circ T_1 = T_2 \circ \phi$ en $\{0,1\}^{\mathbb{N}} \setminus D_1$. Probemos que ϕ es medible y que preserva medidas; verificamos ambas condiciones en el álgebra $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}(\mathcal{D}_n)$ diádica (ver Ejemplo B.1, pág. 130).

Sea $n \in \mathbb{N}$ y $k \in \{0, \ldots, 2^n - 1\}$, entonces $\mu_2(D_n(k)) = 1/2^n$. Por otro lado:

$$\phi^{-1}(D_n(k)) = \left\{x \in \{0,1\}^{\mathbb{N}} : \phi(x) \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right\}.$$

Si escribimos $k = a_0 2^0 + a_1 2^1 + \cdots + a_{n-1} 2^{n-1}$ donde $a_i = 0$ ó 1, entonces

$$\frac{k}{2^n} = \frac{a_{n-1}}{2} + \frac{a_{n-2}}{2^2} + \cdots + \frac{a_0}{2^n} \quad \text{por lo que:}$$

$$\phi^{-1}(D_n(k)) = \left\{x \in \{0,1\}^{\mathbb{N}} : x(1) = a_{n-1}, \ldots, x(n) = a_0\right\} \in \mathscr{F}_1 \quad \text{y}$$

$$\mu_1(\phi^{-1}(D_n(k))) = \prod_{i=0}^{n} \frac{1}{2} = \frac{1}{2^n} = \mu_2(D_n(k)).$$

Lo anterior basta para probar que: $\phi^{-1}(\mathcal{A}) \subset \mathscr{F}_1$ y $\mu_1 \circ \phi^{-1} = \mu_2$ sobre $\mathcal{A}$ por lo tanto $\phi^{-1}(\mathscr{F}_2) \subset \mathscr{F}_1$ y $\mu_1 \circ \phi^{-1} = \mu_2$ y así $T_1 \simeq T_2$. $\square$

Ejemplo 2.3. Definimos $P : [\,0\,,1\,) \times [\,0\,,1\,) \to [\,0\,,1\,) \times [\,0\,,1\,)$ (la transformación "del panadero") mediante la fórmula:

$$P((x,y)) = \begin{cases} (2x, \frac{1}{2}y) & \text{si} \quad 0 \leq x < \frac{1}{2}, \\ \\ (2x-1, \frac{1}{2}(y+1)) & \text{si} \quad \frac{1}{2} \leq x < 1. \end{cases}$$

El efecto de P sobre $[\,0\,,1\,) \times [\,0\,,1\,)$ es como sigue:

Dividimos a $[\,0\,,1\,) \times [\,0\,,1\,)$ en dos rectángulos $[\,0\,,1/2\,) \times [\,0\,,1\,) = R_1$ y $[\,1/2\,,1\,) \times [\,0\,,1\,) = R_2$ y enviamos R_1 en $R_1' = [\,0\,,1\,) \times [\,0\,,1/2\,)$ multiplicando la primera coordenada por 2 y la segunda por 1/2. Análogamente enviamos R_2 en $R_2' = [\,0\,,1\,) \times [\,1/2\,,1\,)$. Es fácil comprobar que P es una

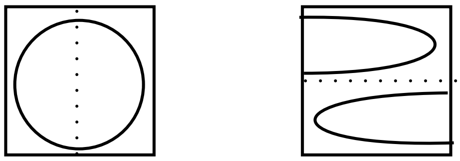

Figura 2.3: Acción de T sobre el cuadrado unitario.

transformación invertible. Si $X = [\,0\,,1\,) \times [\,0\,,1\,)$, $\mathscr{F}_1 = \mathscr{B}_{[\,0\,,1\,)\times[\,0\,,1\,)}$ y $\mu_1 = \lambda \otimes \lambda$ entonces se comprueba que P preserva a $\lambda \otimes \lambda$. Probaré que P es isomorfa al inverso de corrimiento bilateral de Bernoulli con parámetros (1/2,1/2).

Demostración. Sea $x \in \{0,1\}^{\mathbb{Z}}$, definimos $\phi(x) = (\alpha(x), \beta(x))$ como sigue:

$$\alpha(x) = \sum_{i=0}^{\infty} \frac{x(-i)}{2^{i+1}} \qquad \beta(x) = \sum_{i=1}^{\infty} \frac{x(i)}{2^{i}},$$

$\phi : \{0,1\}^{\mathbb{Z}} \to [\,0\,,1\,) \times [\,0\,,1\,)$ es una biyección, excepto en los puntos $(\alpha, \beta) \in [\,0\,,1\,) \times [\,0\,,1\,)$ tales que α o β es diádico. Sea D_2 el conjunto formado por dichas parejas entonces $\mu_2(D_2) = 0$ Sea $D_1 = \{x \in \{0,1\}^{\mathbb{Z}} : \text{existe } m \in \mathbb{N} \text{ tal que } x(j) \text{ es constante para toda } j \geq m \ 0\ j \leq m\}$. Claramente $\phi(D_1) = D_2$ y $\mu_1(D_1) = 0$. Además $D_1 \in \mathscr{F}_1$ y $\mu_1(D_1) = 0$. Sea $M_1 = \{0,1\}^{\mathbb{Z}} \setminus D_1$ y $M_2 = [\,0\,,1\,) \times [\,0\,,1\,) \setminus D_2$ entonces $\phi(M_1) = \phi(M_2)$ y es biyección y $\phi(M_i) \subset M_i$ $(i = 1, 2)$. Probaremos que ϕ^{-1} es medible y preserva las medidas.

Sea $A_i^j = \{x : x(i) = j\}$, $i \in \mathbb{Z}$ y $j = 0$ ó 1, entonces:

$$\phi\left(A_i^j\right) = \left\{ \left(\sum_{k=0}^{\infty} \frac{x(-k)}{2^{k+1}}, \sum_{k=1}^{\infty} \frac{x(k)}{2^{k}} \right) : x(i) = j \right\}.$$

Así que, si $i \geq 1$, entonces $\phi(A_i^j)$ consiste de 2^i "rectángulos" con base de longitud $1/2^{i+1}$ y altura de longitud $1/2^{i+1}$. Si $i \leq 0$, entonces $\phi(A_i^j)$ es la unión de $2^{|i|}$ "rectángulos" con base de longitud $1/2^{|i|+1}$ y altura de longitud 1.

En cualquier caso $\phi(A_i^j) \in \mathscr{F}_2$ y $\mu_1(A_i^j) = 1/2 = \mu_2(\phi(A_i^j))$, lo mismo ocurre para cilindros flacos (que son intersecciones finitas de conjuntos del tipo A_i^j) así como a cada cilindro, por lo tanto ϕ^{-1} es medible y preserva las medidas. Finalmente comprobemos que $\phi \circ T^{-1} \circ \phi^{-1} = P$. Sean $(\alpha, \beta) \in M_1$ dada por sus expansiones binarias:

$$\alpha = \sum_{i=0}^{\infty} \frac{x(-i)}{2^{i+1}} \quad \beta = \sum_{i=1}^{\infty} \frac{x(i)}{2^i}, \quad \text{es decir,} \quad (\alpha, \beta) = \phi(x), \text{ entonces:}$$

$$\phi^{-1}(\alpha, \beta) = (\dots, x(-1), x(0), x(1), \dots) \quad \text{y}$$
$$T^{-1}(\phi^{-1}((\alpha, \beta))) = (\dots, x(-2), x(-1), x(0), \dots),$$

por lo que

$$\begin{aligned}
\phi \circ T^{-1} \circ \phi^{-1}((\alpha, \beta)) &= \left(\sum_{k=1}^{\infty} \frac{x(-k)}{2^k}, \sum_{k=0}^{\infty} \frac{x(k)}{2^{k+1}} \right) \\
&= \begin{cases} \left(2\alpha, \dfrac{1}{2}\beta\right) & \text{si} \quad x(0) = 0 \quad \text{es decir} \quad 0 \le \alpha < \dfrac{1}{2} \\ \\ (2\alpha - 1, \dfrac{1}{2}(\beta + 1)) & \text{si} \quad x(0) = 1 \quad \text{es decir} \quad \dfrac{1}{2} \le \alpha < 1. \end{cases}
\end{aligned}$$

Por lo tanto el inverso del corrimiento bilateral de Bernouilli es isomorfo a $P(x, y)$. $\square$

A esta transformación se la llama "del panadero" porque el proceso de homogeneizar la masa consiste también en estirar (para homogeneizar) y plegar (para tener unas dimensiones manejables) la masa repetidas veces.

Capítulo 3

Cadenas de Markov Estacionarias

Las cadenas de Markov se utilizan para modelar fenómenos aleatorios en los cuales la probabilidad de que ocurra un evento depende solamente del evento inmediato anterior. Estas cadenas han tenido un gran desarrollo y tienen una gran variedad de aplicaciones (ver Caballero [6]). En este capítulo se considera únicamente cadenas de Markov con espacio de estados finito y estacionarias u homogéneas, es decir, que el conjunto de estados tiene cardinalidad finita y que la probabilidad de ir de un estado a otro en una unidad de tiempo no depende del tiempo. A partir de este momento, y en lo que resta del presente trabajo, denotaremos $(\Omega, \mathscr{F}, \mu)$ a un espacio de probabilidad.

3.1. Definiciones básicas

En nuestro estudio de las cadenas de Markov necesitaremos la siguiente definición:

Definición 3.1. Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad, sean $A, B \in \mathscr{F}$ tal que $\mu(A) > 0$, al siguiente número

$$\mu(B|A) = \frac{\mu(B \cap A)}{\mu(A)},$$

lo llamaremos **la probabilidad de B dado A**. En general en un espacio de probabilidad $\mu(F)$ la llamaremos probabilidad de F, para toda $F \in \mathscr{F}$.

Esta probabilidad condicional puede ser interpretada como sigue: la probabilidad de que ocurra el evento B (recordemos que en Teoría de la Probabilidad a los elementos de la σ-álgebra se les llama eventos) ya que ha ocurrido el evento A es $\mu(B|A)$, por ello esta probabilidad involucra la intersección de los eventos y la división de la probabilidad de A. Otra interpretación de esta probabilidad es como sigue: cuando un evento A ocurre, el espacio muestral Ω se reduce y queda conformado por los posibles resultados que forman al evento A. La probabilidad condicional $\mu(B|A)$

es la relación que existe entre la probabilidad del evento B, calculada con respecto a todo el espacio muestral, y la probabilidad del evento B, calculada en el espacio muestral reducido. Procedemos a dar la definición de cadena de Markov sobre un conjunto finito de estados a tiempo discreto.

Definición 3.2. Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad. Una **Cadena de Markov** es una colección de funciones $\mathscr{F}$-medibles $\{f_n : n \in \mathbb{N} \cup \{0\} \text{ ó } n \in \mathbb{Z}\}$ que toman valores en un conjunto E con $|E| < \infty$, que llamaremos conjunto de estados, y que satisface la propiedad de Markov; es decir, que para todo $n \geq 0$ y para cualesquiera estados $x_0, \ldots, x_{n+1} \in E$ se cumple que

$$\mu(f_{n+1} = x_{n+1} | f_n = x_n, \ldots, f_0 = x_0) = \mu(f_{n+1} = x_{n+1} | f_n = x_n),$$

siempre que $\mu(f_n = x_n, \ldots, f_0 = x_0) > 0$ y $\mu(f_n = x_n) > 0$.

Observación 3.1. La igualdad anterior dice que la probabilidad de pasar al estado x_{n+1}, dado que los estados anteriores fueron $x_0, \ldots, x_n$, sólo depende del estado inmediato anterior, es decir, el futuro sólo depende del presente. Recordemos que $\mu(f_n = x_{n+1}) = \mu(f_n^{-1}(\{x_{n+1}\}))$. A partir de este momento tomaremos como espacio de estados de una cadena de Markov al conjunto $E = \{0, \ldots, k-1\}$ para algún $k \in \mathbb{N}$.

Definición 3.3. Se dice que una cadena de Markov es **estacionaria u homogénea** si la probabilidad $\mu(f_{n+1} = b | f_n = a)$ no depende de n para todo $a, b \in E$, estos números los denotaremos como p_{ab} y diremos que p_{ab} es la probabilidad transición del estado a al estado b en un paso.

Definición 3.4. Sea $E = \{0, \ldots, k-1\}$ $(k > 1)$ el conjunto de estados de una Cadena de Markov. La **matriz cuadrada de probabilidades de transición** dada por $P(a, b) = p_{ab}$ con $a, b \in E$ cumple las siguientes dos propiedades:

(a) $p_{ab} \geq 0$ para todo $a, b \in E$.

(b) $\sum_{b \in E} p_{ab} = 1$ para toda $a \in E$.

Llamamos a los elementos p_{ab} de P, **probabilidades de transición**. En general cualquier matriz cuadrada que cumpla estas dos propiedades la llamaremos **matriz estocástica**.

Observación 3.2. Debido a la propiedad de Markov la matriz anterior contiene la información de la cadena y determina el comportamiento de ésta en el tiempo.

Definición 3.5. Sea P una matriz estocástica como en la definición anterior. Por una **distribución inicial estacionaria** para P entenderemos un vector $\pi = (\pi_0, \ldots, \pi_{k-1})$ tal que:

(a) $\pi_a \geq 0$ para toda $a \in E$.

(b) $\pi P = \pi$ es decir $\pi_b = \sum_{a \in E} \pi_a p_{ab}$ para toda $b \in E$.

Observaciones 3.3.

(1) Si P es una matriz estocástica y π es una distribución inicial estacionaria, entonces P^n es una matriz estocástica para toda $n = 0, 1, 2, \ldots$ y $\pi P^n = \pi$.

(2) Podemos suponer que la cadena parte de un estado inicial $a \in E$, es decir, $f_0(\omega) = a$. La distribución inicial estacionaria generaliza lo anterior, dando una distribución de probabilidad sobre el conjunto de estados, esto es, la cadena parte del estado $a \in E$ con probabilidad π_a.

(3) No es evidente que para toda cadena de Markov exista una distribución inicial estacionaria, menos aún que sea única, este es uno de los problemas de interés de estos procesos. Más adelante veremos condiciones para que esta distribución exista y sea única.

Definición 3.6 (Probabilidades de transición en n pasos)**.** Sean $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad y $\{f_n : n = 0, 1, \ldots\}$ una cadena de Markov, la probabilidad $\mu(f_{n+m} = b | f_m = a)$ es la probabilidad de pasar al estado a el tiempo m, al estado b en el tiempo $m + n$. Como estamos considerando Cadenas de Markov estacionarias, esta probabilidad no depende de m, es decir, $\mu(f_{n+m} = b | f_m = a) = \mu(f_n = b | f_0 = a)$. A esta probabilidad la denotaremos como $p_{ab}^{(n)}$ y la llamaremos **probabilidad de transición en n pasos**.

Observación 3.4. Debido a que la cadena de Markov es homogénea y a la **Ecuación de Chapman-Kolmogorov** (ver Shiryaev [21, pág. 116]), estas probabilidades coinciden con la componente correspondiente de los estados involucrados de la potencia n-ésima de la matriz de probabilidades de transición, es decir, $p_{ab}^{(n)} = P^n(a, b)$ para toda $a, b \in E$, para toda n, donde P^n es la matriz que se obtiene de multiplicar n veces P consigo misma.

3.2. Corrimientos de Markov

Estamos en posición de definir lo que llamaremos corrimientos de Markov, veremos que los corrimiento de Bernoulli son un caso particular de éstos. La idea central de este capítulo será construir un espacio de probabilidad y una transformación T que preserve la medida del espacio construido, de tal manera que podamos inferir condiciones acerca de la cadena de Markov; utilizando por supuesto la teoría de los capítulos anteriores. Ahora procedemos a la construcción de dicho espacio de probabilidad.

Definición 3.7. Sea $\Omega = E^{\mathbb{N}}$ (ó $E^{\mathbb{Z}}$), $\mathscr{F} = \sigma(\text{cilindros})$, $P = (p_{ab})$ con $a, b \in E$ una matriz estocástica, y $\pi = (\pi_0, \ldots, \pi_{k-1})$ una distribución inicial estacionaria para P. Si $n_1 < n_2 < \cdots < n_k \in \mathbb{N}$ (ó $\in \mathbb{Z}$) y $a_1, \ldots, a_k \in E$ definimos:

$$\mu(Z(n_1, n_2, \ldots, n_k; a_1, a_2, \ldots, a_k)) = \pi_{a_1} p_{a_1 a_2}^{(n_2 - n_1)} \cdots p_{a_{k-1} a_k}^{(n_k - n_{k-1})}, \tag{*}$$

donde $p_{ab}^{(r)}$ denota la componente (a, b) de la matriz P^r $(r \geq 1)$ y $\mu(Z(n; a)) = \pi_a$ para toda n y $a \in E$.

Es fácil comprobar que μ satisface las condiciones de consistencia de Kolmogorov (ver Observación B.1, pág. 137), por lo que existe una **única** medida de probabilidad $\mu : \mathscr{F} \to [\,0\,,1\,]$ (que **denotaremos con la misma letra**) y que coincide con μ en los cilindros. Ahora procedemos a dar la transformación mencionada anteriormente, recordemos que ya conocemos una transformación que preserva la media en los corrimientos de Bernoulli.

Definición 3.8. Sea $T : \Omega \to \Omega$ poniendo $(T(\omega))_i = \omega_{i+1}$ para toda $i \in \mathbb{N}$ (ó $i \in \mathbb{Z}$) de acuerdo a que $\Omega = E^{\mathbb{N}}$ (ó $\Omega = E^{\mathbb{Z}}$) y llamamos a las cuarteta $(\Omega, \mathscr{F}, \mu, T)$ el **corrimiento unilateral (o bilateral) de Markov con matriz de transición** P **y distribución inicial estacionaria** π y lo denotaremos: $M_{u\ \text{ó}\ b}(P, \pi)$.

Ejemplo 3.1. El corrimiento unilateral (o bilateral) de Bernoulli con parámetros $(p_0, \ldots, p_{k-1})$ es un caso especial del corrimiento de Markov unilateral (o bilateral).

Demostración. Definimos $P = (p_{ab})$ mediante la fórmula: $p_{ab} = p_b$ para toda $a, b \in E$, es decir, las componentes de cada columna son idénticas. Como $\sum_{i=0}^{k-1} p_i = 1$ y $p_1 \geq 0$, P es claramente una matriz estocástica y

$$P = \begin{pmatrix} p_0 & \cdots & p_{k-1} \\ p_0 & \cdots & p_{k-1} \\ \vdots & \vdots & \vdots \\ p_0 & \cdots & p_{k-1} \end{pmatrix},$$

es inmediato verificar que $\pi = (p_0, \ldots, p_{k-1})$ es una distribución inicial estacionaria para P. Note que $P^r = P$ para toda $r = 1, 2 \ldots$ por lo que la medida definida por la pareja (P, π) satisface: $\mu(Z(n_1, n_2, \ldots, n_k; a_1, \ldots, a_k)) = p_{a_1} p_{a_2} \cdots p_{a_k}$ que es precisamente el valor de la medida del corrimiento de Bernoulli en dicho cilindro. $\square$

Observación 3.5. El corrimiento T es una transformación preservadora de medida.

Nota: A partir de este momento se supondrá que $\pi = (\pi_0, \ldots, \pi_{k-1})$ es tal que $\pi_j > 0$ para toda $j = 0, \ldots, k-1$.

Proposición 3.1. *Sea $P = (p_{ab})$ con $a, b \in E$ una matriz estocástica y $\pi = (p_0, \ldots, p_{k-1})$ una distribución inicial estacionaria para P, entonces para todo $a, b \in E$ y $r = 1, 2, \ldots$ se tiene que:*

$$p_{ab}^{(r)} = \mu(\{\omega : \omega(t + r) = b \,|\, \omega(t) = a\}) \quad \textit{para toda } t \in \mathbb{N} \text{ (ó } t \in \mathbb{Z}) \quad \textit{si} \quad \Omega = E^{\mathbb{N}} \text{ (ó } \Omega = E^{\mathbb{Z}}).$$

Demostración. Para toda t se tiene que:

$$\begin{aligned} \mu(\{\omega:\omega(t)=a\})\mu(\{\omega:\omega(t+r)=b\,|\,\omega(t)=a\}) &= \mu(\{\omega:\omega(t)=a,\omega(t+r)=b\}) \\ &= \mu(Z(t,t+r;a,b)) \\ &= \pi_a p_{ab}^{(t+r-t)} \\ &= \pi_a p_{ab}^{(r)}, \end{aligned}$$

dividiendo entre π_a obtenemos el resultado. □

Teorema 3.1 (Propiedad de Markov)**.** *Sea $P=(p_{ab})$ con $a,b\in E$ una matriz estocástica y $\pi=(\pi_0,\dots,\pi_{k-1})$ una distribución inicial estacionaria para P. Sean*

$$n_1<n_2<\cdots<n_{m+k}\in\mathbb{N}\ (\text{ó }\in\mathbb{Z})\quad y\quad a_1,a_2,\dots,a_{m+k}\in E,$$

entonces:

$$\mu(\{\omega:\omega(n_{m+k})=a_{m+k},\dots,\omega(n_{m+1})=a_{m+1}\,|\,\omega(n_m)=a_m,\dots,\omega(n_1)=a_1\})=$$

$$\mu(\{\omega:\omega(n_{m+k})=a_{m+k},\dots,\omega(n_{m+1})=a_{m+1}|\omega(n_m)=a_m\}).$$

Demostración. El lado izquierdo es igual a

$$\frac{\mu(\{\omega:\omega(n_1)=a_1,\dots,\omega(n_m)=a_m,\dots,\omega(n_{m+k})=a_{m+k}\})}{\mu(\{\omega:\omega(n_1)=a_1,\dots,\omega(n_m)=a_m\})}=$$

$$\frac{\pi_{a_1}p_{a_1a_2}^{(n_2-n_1)}\cdots p_{a_{m-1}a_m}^{(n_m-n_{m-1})}p_{a_ma_{m+1}}^{(n_{m+1}-n_m)}\cdots p_{a_{m+k-1}a_{m+k}}^{(n_{m+k}-n_{m+k-1})}}{\pi_{a_1}p_{a_1a_2}^{(n_2-n_1)}\cdots p_{a_{m-1}a_m}^{(n_m-n_{m-1})}}=$$

$$p_{a_ma_{m+1}}^{(n_{m+1}-n_m)}\cdots p_{a_{m+k-1}a_m}^{(n_{m+k}-n_{m+k-1})}=\frac{\pi_{a_m}p_{a_ma_{m+1}}^{(n_{m+1}-n_m)}\cdots p_{a_{m+k-1}a_m}^{(n_{m+k}-n_{m+k-1})}}{\pi_{a_m}},$$

que es precisamente el lado derecho. □

Observación 3.6. Note que el teorema anterior afirma que la colección de funciones $f_n:\Omega\to E$, definidas como $f_n(\omega)=\omega(n)$ para toda $n\geq 0$ es una cadena de Markov en el espacio de probabilidad $(\Omega,\mathscr{F},\mu)$.

Observación 3.7. Por el teorema ergódico (con $T=M(P,\pi)$ ergódico o no); tenemos que $(1/n)\sum_{k=0}^{n-1}p_{ab}^{(k)}$ converge para toda $a,b\in E$ a un valor que denotaremos q_{ab}, y que definiremos como la componente (a,b) de una matriz cuadrada Q (de tamaño $k\times k$) y escribiremos:

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

Demostración. Sean $A=\{\omega\in\Omega|\omega(0)=a\}$, $B=\{\omega\in\Omega|\omega(0)=b\}$. Como $\chi_B\in L_1(\mu)$, por el Teorema Ergódico Individual se tiene que $(1/n)\sum_{i=0}^{n-1}\chi_{T^{-i}(B)}(\omega)\to f^*$ c.d. Multiplicando por χ_A, tenemos que $(1/n)\sum_{i=0}^{n-1}\chi_{T^{-i}(B)\cap A}(\omega) \to f^*\chi_A$. Pero la sucesión $\{(1/n)\sum_{i=0}^{n-1}\chi_{T^{-i}(B)\cap A}\}$ está acotada, entonces por el Teorema de Convergencia Dominada, se tiene que

$$\lim_{n\to\infty}\int \frac{1}{n}\sum_{i=0}^{n-1}\chi_{T^{-i}(B)\cap A}d\,\mu = \int f^*\chi_A, \quad \text{es decir,} \quad \frac{1}{n}\sum_{k=0}^{n-1}p_{ab}^{(k)} \to q_{a,b}.$$

$\square$

Lema 3.1. *La matriz Q anterior tiene las siguientes propiedades:*

(i) *Q es estocástica (pues $(1/n)\sum_{k=0}^{n-1}P^k$ lo es para toda $n \geq 1$).*

(ii) *$QP = Q = PQ$.*

(iii) *$Q^2 = Q$.*

(iv) *$\pi Q = \pi$.*

Demostración. Las propiedades anteriores son consecuencia de considerar la sucesión $\{(1/n)\sum_{k=0}^{n-1}P^k\}$, entonces para toda n cada una de estas matrices cumplen las propiedades anteriores, por lo que el límite también tendrá esa propiedad. Por ejemplo sea $H_n = (1/n)\sum_{k=0}^{n-1}P^k$, entonces $H_nP = ((n+1)/n)H_{n+1} - (1/n)I$ para toda $n \geq 1$, por lo que al tomar el límite de ambos lados obtenemos que $QP = Q$. $\square$

Observaciones 3.8.

(1) Hemos visto que para toda cadena de Markov estacionaria su comportamiento en el tiempo está determinado por la matriz de probabilidad de transición. La observación anterior dice que su comportamiento **en promedio** se estabiliza, sin embargo, cabe preguntar si existen cadenas de Markov que su comportamiento se estabilice en el tiempo. No todas las cadenas tienen este comportamiento. Más adelante veremos condiciones necesarias para que la cadena se estabilice, es decir, cambiar la convergencia en Césaro por convergencia.

(2) La existencia y unicidad de una distribución inicial estacionaria y la del $\lim_{n\to\infty} p_{a,b}^{(n)}$ son algunos de los problemas que se presentan en el estudio de las cadenas de Markov. Se sabe que si la matriz cumple ciertas condiciones entonces ambos o alguno de los problemas tiene solución. Relacionaremos estas condiciones con las propiedad de ergodicidad y las mezclantes del corrimiento de Markov, para ello necesitaremos examinar cuando el corrimiento de Markov tiene estas propiedades.

3.3. Ergodicidad de Cadenas de Markov

Veremos que se pueden obtener condiciones que hacen que la transformación de corrimiento de Markov sea ergódica, débilmente y fuertemente mezclante. Estas condiciones estarán relacionadas con la matriz de probabilidades de transición. Para ello necesitaremos las siguientes definiciones:

Definición 3.9. Un subconjunto de estados $E_0 \subset E$ es **cerrado** si $\sum_{b \in E_0} p_{ab} = 1$ para toda $a \in E_0$. (Alternativamente E_0 es cerrado si reenumerando los elementos de E es posible extraer de P una submatriz estocástica P_0 con índices en E_0).

Definición 3.10. Decimos que P es **irreducible** si E no posee subconjuntos cerrados propios.

Teorema 3.2. *Las siguientes condiciones son equivalentes para el corrimiento de Markov $T = M(P, \pi)$:*

(i) T *es ergódico.*

(ii) q_{ab} *es independiente de* a (*es decir* Q *posee columnas con elementos iguales*).

(iii) P *es irreducible.*

(iv) $q_{ab} > 0$ *para todo* $a, b \in E$.

(v) La ecuación $Q\vec{y} = \vec{y}$ $(\vec{y} \neq 0)$ *tiene solución única* (*salvo por factores constantes*).

(vi) La ecuación $\vec{y}Q = \vec{y}$ $(\vec{y} \neq 0)$ *tiene solución única* (*salvo por factores constantes*).

(vii) $P\vec{y} = \vec{y}$ $(\vec{y} \neq 0)$ *tiene solución única* (*salvo por factores constantes*).

(viii) $\vec{y}P = \vec{y}$ $(\vec{y} \neq 0)$ *tiene solución única* (*salvo por factores constantes*).

Demostración.

(i) implica (ii). Sean $B = \{\omega : \omega(0) = a\}$ y $C = \{\omega : \omega(0) = b\}$ con a y b elementos de E, por el Corolario 1.16.1 se tiene que:

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu(\{\omega : \omega(0) = a, \omega(k) = b\}) \longrightarrow \mu(\{\omega : \omega(0) = a\})\mu(\{\omega : \omega(0) = b\}), \quad \text{es decir,}$$

$$\frac{1}{n}\sum_{k=0}^{n-1} \pi_a p_{ab}^{(k)} \longrightarrow \pi_a \pi_b.$$

Dividiendo entre π_a, tenemos que $q_{a,b} = \pi_b$ para toda a, b, por lo tanto q_{ab} no depende de a.

(ii) implica (i). Como $\pi Q = \pi$, debemos tener $q_{ab} = \pi_b$ para toda $a \in E$. Sean $B = \{\omega : \omega(0) = b_0, \ldots, \omega(s) = b_{s-1}\}$ y $C = \{\omega : \omega(0) = c_0, \ldots, \omega(t) = c_{t-1}\}$ dos cilindros **flacos**, si $n > l \geq s$ entonces:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(B\cap T^{-k}(C)) &= \lim_{n\to\infty} \frac{1}{n}\sum_{l=s}^{n-1} \mu(B\cap T^{-l}(C)) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{l=s}^{n-1} (\pi_{b_0}p_{b_0b_1}\cdots p_{b_{s-2}b_{s-1}}p^{(l+1-s)}_{b_{s-1}c_0}p_{c_0c_1}\cdots p_{c_{t-2}c_{t-1}}) \\
&= \mu(B)\frac{\mu(C)}{\pi_{c_0}} \lim_{n\to\infty} \frac{1}{n}\sum_{l=s}^{n-1} p^{(l+1-s)}_{b_{s-1}c_0} \\
&= \mu(B)\frac{\mu(C)}{\pi_{c_0}} q_{b_{s-1}c_0} = \mu(B)\mu(C).
\end{aligned}$$

Por lo que

$$\frac{1}{n}\sum_{k=0}^{n-1} \mu(B\cap T^{-k}(C)) \longrightarrow \mu(B)\mu(C),$$

si B y C son cilindros flacos, por lo que esta misma relación es válida para $\mathcal{A} = \{$ uniones finitas y disjuntas de cilindros flacos$\}$, es fácil comprobar que $\mathcal{A}$ es un álgebra que genera a $\mathscr{F}$, por el Teorema 1.26 $T = M(P,\pi)$ es ergódico.

(i) implica (iii). Supongamos que P **no** es irreducible. Sean $E_0 \subsetneq E$ un subconjunto cerrado propio y $F = \{\omega : \omega(0) \in E_0\}$. Como E_0 es propio $0 < \mu(F) < 1$, además $\mu(F \setminus T^{-1}(F)) = 0$, ya que:

$$F \setminus T^{-1}(F) = \{\omega : \omega(0) \in E_0, \omega(1) \in E \setminus E_0\} = \bigcup_{\substack{a\in E_0 \\ b\in E\setminus E_0}} \{\omega : \omega(0) = a, \omega(1) = b\} \quad \text{(disjunta)},$$

pero

$$p_{ab} = \mu(\{\omega : \omega(1) = b \,|\, \omega(0) = a\}) = 0,$$

pues E_0 es cerrado por lo tanto $\mu(F \setminus T^{-1}(F)) = 0$.

Como $\mu(F) = \mu(T^{-1}(F))$ y $\mu(F \setminus T^{-1}(F)) = 0$, entonces $\mu(T^{-1}(F) \setminus F) = 0$ por lo tanto F es T-invariante por lo tanto T no es ergódica.

(iii) implica (iv). Como $Q = QP$ tenemos que $q_{ab} = \sum_c q_{ac}p_{cb} \geq q_{ac}p_{cb}$ para toda $c \in E \cdots (1)$. Sea $E_a = \{b \in E : q_{ab} > 0\}$ debemos probar que $E_a = E$ para toda $a \in E$. Claramente $E_0 \neq \emptyset$ pues Q es estocástica, además E_a es cerrado, ya que si $c \in E_a$, entonces $p_{cb} = 0$ para toda $b \notin E_a$ por (1), así pues $\sum_{d\in E_a} p_{cd} = 1$ para toda $c \in E_a$ y $E_a = E$ por hipótesis.

(iv) implica (ii). Supongamos que $q_{ab} > 0$ para todo $a, b \in E$. Consideremos la ecuación vectorial $\sum_{b\in E} q_{ab}y_b = y_a$ es decir $Q\vec{y} = \vec{y}$. Sea $m = \text{máx}_{b\in E}\{y_b\}$ y supongamos que $y_a < m$ para alguna $a \in E$, entonces

$$y_c = \sum_{b\in E} q_{cb}y_b < \sum_{b\in E} q_{cb}m = m, \quad \text{entonces} \quad y_c < m \quad \text{para toda } c \in E,$$

lo cual no es posible. Luego entonces toda solución de $Q\vec{y} = \vec{y}$ debe tener componentes iguales. Por otro lado $Q^2 = Q$, por lo que cada vector columna de Q es solución de la ecuación vectorial, así que cada columna de Q tiene las mismas componentes. Por lo tanto q_{ab} es independiente de a para toda $a \in E$.

(i) si y sólo si (v). En la demostración de (iv) implica (ii) obtuvimos que si $q_{ab} > 0$ para toda $a, b \in E$ entonces la ecuación vectorial $Q\vec{y} = \vec{y} \cdots (*)$ $(\vec{y} \neq 0)$ tiene solución única (salvo por multiplicación por constantes), **es decir,** si T es ergódica entonces (*) tiene solución única (salvo por multiplicación por constantes, de hecho $\vec{y} = \lambda Q^{(a)}$, con $Q^{(a)} = a$ columna de Q). Inversamente supongamos que (*) tiene solución única (salvo por multiplicación por constantes), entonces como Q es estocástica **cualquier** vector con componentes iguales resuelve a (*). Por otro lado como $Q^2 = Q$, cada columna de Q es solución de (*), y por unicidad tenemos que concluir que Q posee columnas con componentes iguales, es decir, (ii) se cumple. Por lo tanto T es ergódica.

Por otra parte a partir de (ii) y del hecho de que $Q^2 = Q$ entonces se prueba como arriba que: (i) si y sólo si (vi).

(i) si y sólo si (vii), y (i) si y sólo si (viii). Como $Q = \lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1} P^k$ y $PQ = Q = QP$, así que cualquier solución de $P\vec{y} = \vec{y}$ es solución de $Q\vec{y} = \vec{y}$ y viceversa. Análogamente cualquier solución de $\vec{y}P = \vec{y}$ es solución de $\vec{y}Q = \vec{y}$ y viceversa. [1] $\square$

Observación 3.9. Examinemos con detenimiento la condición (iii), es decir, P es irreducible. En el teorema anterior probamos que si P es irreducible, entonces $q_{ab} > 0$ para toda $a, b \in E$ (condición (iv)) pero $q_{ab} = \lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1} p_{ab}^{(k)}$, así pues obtenemos:

$$\text{para toda} \quad a, b \in E \quad \text{existe} \quad k = k(a,b) \in \{0, 1, \dots\} \quad \text{tal que} \quad p_{ab}^k > 0. \tag{3.2}$$

Ahora se prueba que 3.2 implica (iii).

Supongamos que (iii) es falso, es decir, existe $E_0 \subset E$ un subconjunto propio cerrado. Se probará que: $p_{ab}^n = 0$ para toda $a \in E_0$, para toda $b \in E \setminus E_0$ y para toda $n = 0, 1, 2, \dots$ claramente la afirmación es cierta para $n = 0$ y $n = 1$ supongamos que $n \geq 2$ y $a \in E_0$ $b \in E \setminus E_0$, entonces:

$$p_{ab}^n = \sum_{\gamma,\delta} p_{a\gamma} p_{\gamma\delta}^{n-2} p_{\delta b} = \Big(\sum_{\substack{\gamma\in E_0\\ \delta\in E_0}} + \sum_{\substack{\gamma\in E_0\\ \delta\in E\setminus E_0}} + \sum_{\substack{\gamma\in E\setminus E_0\\ \delta\in E\setminus E_0}} + \sum_{\substack{\gamma\in E\setminus E_0\\ \delta\in E_0}}\Big)\{p_{a\delta} p_{\gamma\delta}^{n-2} p_{\delta b}\} = S_1 + S_2 + S_3 + S_4.$$

Pero $S_2 = 0$ por hipótesis de inducción, $S_1 = 0$, ya que $\delta \in E_0$ y $b \in E \setminus E_0$, $S_3 = 0$, pues $a \in E_0$ y $\delta \in E \setminus E_0$ y $S_4 = 0$ pues $\delta \in E_0$ y $b \in E \setminus E_0$.[2]

[1]Note que la ecuación $\vec{y}P = \vec{y}$ es la adjunta de la ecuación $P\vec{y} = \vec{y}$, lo que prueba (vii) si y sólo si (viii) de otro modo.

[2]Algunos textos toman a 3.2 como definición de irreducibilidad.

3.4. Comunicación, Cerradura y Periodicidad

Hemos visto bajo qué condiciones el corrimiento de Markov es ergódico. Para ver condiciones sobre las cuales este corrimiento tiene propiedades mezclantes necesitaremos los siguientes conceptos sobre los estados de una cadena de Markov. Retomaremos el enunciado de 3.2 sobre la irreducibilidad de una cadena de Markov.

Definición 3.11. Sea P una matriz estocástica con conjunto finito de estados E. Sean $a, b \in E$ decimos que el estado b es **accesible** desde el estado a, si existe $n = n(a, b) \in \{0, 1, 2, \dots\}$ tal que $p^n_{ab} > 0$ (denotado $a \to b$). Decimos que a y $b \in E$ se "**comunican**" si $a \to b$ y $b \to a$ (denotado $a \leftrightarrow b$).

Proposición 3.2. *La relación $\leftrightarrow$ es una relación de equivalencia.*

Demostración. Claramente $a \leftrightarrow a$ (debido a que $p^0_{aa} = 1$) y $a \leftrightarrow b$ si y sólo si $b \leftrightarrow a$. Supongamos que $a \leftrightarrow b$ y $b \leftrightarrow c$ probaremos que $a \leftrightarrow c$. Como en particular $a \to b$ y $b \to c$, entonces existen $m, n \in \mathbb{N} \cup \{0\}$ tales que: $p^m_{ab} > 0$ y $p^n_{ac} > 0$, entonces:

$$p^{m+n}_{ac} = \sum_{\gamma} p^m_{a\gamma} p^n_{\gamma c} \geq p^m_{ab} p^n_{ac} > 0, \quad \text{por lo tanto} \quad a \to c,$$

y como $c \to b \to a$ entonces análogamente probamos que $c \to a$. □

Definición 3.12. Denotaremos por $[a] = \{b \in E : a \leftrightarrow b\}$ a la $\leftrightarrow$-clase de equivalencia de a en E. Nótese que P es irreducible si y sólo si $[a] = E$ para algún $a \in E$.

A continuación damos una caracterización (que de paso justificará el nombre) de subconjunto cerrado de E).

Teorema 3.3. *Sea $E_0 \subset E$ un subconjunto no vacío entonces: E_0 es cerrado si y sólo si siempre que $a \in E_0$ y $a \to b$, entonces $b \in E_0$ (es decir una vez que el "sistema" entre a E_0 no lleva a estados fuera de E_0).*

Demostración.

(Necesidad). Sea $E_0 \subset E$ cerrado y sea P_0 la "submatriz estocástica generada" por E_0, es decir, $P_0 = (p_{ac})$ con $a, c \in E_0$. Supongamos que $a \to b$, entonces existe $n \geq 0$ tal que $p^n_{ab} > 0$, pero tanto P^n como P^n_0 son estocásticas, entonces por lo que $b \in E_0$, pues de otro modo el a **renglón** de P^n sumaría más que 1.

(Suficiencia). Supongamos que la condición es cierta, pero que E_0 **no** es cerrado, entonces $\sum_{c \in E_0} p_{ac} < 1$ para algún $a \in E_0$ entonces existe $b \in E \setminus E_0$ tal que $p_{ab} > 0$; así pues $a \to b$ y $b \notin E_0$, contrario a la hipótesis. □

Definición 3.13. Para toda $A \subset E$ ($A \neq \emptyset$) denotamos por $\overline{A} = \{c \in E : a \to c$ para alguna $a \in A\}$ y lo llamamos la P**-cerradura de** A.

Lema 3.2. *La P-cerradura de A cumple las siguientes condiciones:*

(i) $\overline{A}$ *es cerrado.*

(ii) $A \subset \overline{A}$. *Si* $A \subset B \subset E$, *entonces* $\overline{A} \subset \overline{B}$.

(iii) A *es cerrado si y sólo si* $A = \overline{A}$ *(en particular* $\overline{\overline{A}} = \overline{A}$*).*

(iv) $\overline{A}$ *es el "menor subconjunto" cerrado de* E *que contiene a* A.

(v) $\overline{(A \cup B)} = \overline{A} \cup \overline{B}$ *y* $\overline{(A \cap B)} \subset \overline{A} \cap \overline{B}$.

Demostración.

(i) Supongamos que existe un $a_0 \in \overline{A}$ tal que $\sum_{b \in \overline{A}} p_{a_0 b} < 1$, como P es estocástica existe un $b_0 \in E \setminus \overline{A}$ tal que $p_{a_0 b_0} > 0$, entonces $a_0 \to b_0$; por lo tanto $b_0 \in \overline{A}$, lo cual no es posible.

(ii) Sea $a \in A$, como $p^0_{aa} = 1$, entonces $a \to a$, por lo tanto $a \in \overline{A}$. Ahora sea $a_0 \in \overline{A}$, entonces existe $a \in A$ tal que $a \to a_0$, como $a \in B$, entonces existe un $b \in B$ tal que $b \to a_0$, por lo tanto $a_0 \in \overline{B}$.

(iii) Por el inciso anterior basta con probar que si A es cerrado, entonces $\overline{A} \subset A$. Supongamos que existe un $a_0 \in \overline{A}$ tal que $a_0 \notin A$, entonces existe $a \in A$ tal que $a \to a_0$, como A es cerrado por la Observación 3.9 se sabe que $p^n_{aa_0} = 0$ para toda $n \geq 0$, lo cual no es posible.

(iv) Supongamos que $A \subset E_0 \subset E$ y E_0 es cerrado, entonces por (ii) y (iii) $\overline{A} \subset E_0$.

(v) Como $\overline{A} \subset \overline{A \cup B}$ y $\overline{B} \subset \overline{A \cup B}$ se tiene que $\overline{A} \cup \overline{B} \subset \overline{A \cup B}$, sea $c \in \overline{A \cup B}$, entonces existen $d \in A \cup B$ tal que $d \to c$, por lo que $c \in \overline{A}$ o $c \in \overline{B}$, lo que prueba la igualdad. Por último, como $A \cap B \subset A$, por (ii) $\overline{A \cap B} \subset \overline{A}$, intersecarse la contención anterior con $\overline{B}$ se tiene el resultado. ☐

Observando las propiedades anteriores es justificable el uso de la palabra cerradura.

Definición 3.14. Sea E un conjunto finito y P una matriz estocástica con E como conjunto de estados. Sea $a \in E$ definimos el **periodo** de a (denotado $per(a)$), y definido como sigue: $per(a) =$m.c.d.$\{n \in \mathbb{N} : p^n_{aa} > 0\}$ (máximo común divisor) si $\{n \in \mathbb{N} : p^n_{aa} > 0\} \neq \emptyset$. Si $per(a) = 1$ entonces diremos que a es **aperiódico**. Diremos que P es **aperiódica** si $per(a) = 1$ para toda $a \in E$.

Teorema 3.4. *El periodo es una propiedad de clase, es decir, si* $per(a) = d$ *y* $b \in [a]$ *entonces* $per(b) = d$.

Demostración. Por hipótesis existen $j, l \in \{0, 1, \dots\}$ tales que $p^j_{ab} > 0$ y $p^l_{ba} > 0$. Supongamos que $n \in \mathbb{N}$ es tal que $p^n_{bb} > 0$, entonces

$$p^{j+n+l}_{aa} = \sum_{\delta,\gamma} p^j_{a\delta} p^n_{\delta\gamma} p^l_{\gamma a} \geq p^j_{ab} p^n_{bb} p^l_{ba} > 0,$$

por lo que $d|j + n + l$, pero como

$$p^{j+l}_{aa} \geq p^j_{ab} p^l_{ba} > 0, \text{ entonces } d|j + l,$$

así que $d|n$, por lo que $d|per(b)$. Como los argumentos son simétricos en a y b, obtenemos de manera similar que $per(b)|d$, por lo tanto $d = per(b)$. ☐

Toda aquella propiedad que compartan los elementos de una misma clase la llamaremos **propiedad de clase**.

Observación 3.10. Para cada $a \in E$ existe un **conjunto finito** $H_a \subset \mathbb{N}$ tal que $per(a) =$m.c.d.$\{h \in H_a\}$. Esto simplifica la definición anterior.

Demostración. Sea $L_a = \{n \in \mathbb{N} : p^n_{aa} > 0\}$, sea $h_1 = \text{mín}\{n : n \in L_a\}$ entonces $per(a)|h_1$. Si $per(a) = h_1$ tomamos $H_a = \{h_1\}$. Si $per(a) < h_1$, llamamos $h_2 = \text{mín}\{n : n \in L_a, h_1 \nmid n\}$ (note que h_2 existe), y sea $d_2 =$m.c.d.(h_1, h_2), entonces $per(a)|d_2$. Si $per(a) = d_2$ tomamos $H_a = \{h_1, h_2\}$. Si $per(a) < d_2$ llamamos $h_3 = \text{mín}\{n : n \in L_a, d_2 \nmid n\}$ y $d_3 =$m.c.d.(h_1, h_2, h_3), así que $per(a)|d_3$. Nótese que de este modo se obtiene una sucesión d_j ($d_1 = h_1$) de naturales tal que $d_1 > d_2 > \dots > d_j \geq per(a)$ la cual no puede ser infinita, entonces existe j tal que $\{n \in L_a, d_j \nmid n\} = \emptyset$ (que ocurre precisamente cuando $d_j = per(a)$) y ponemos $H_a = \{h_1, \dots, h_j\}$. ☐

Teorema 3.5. *Sea $a \in E$ y $d = per(a)$ $(d \geq 1)$. Denotamos $L_a = \{n \in \mathbb{N} : p^n_{aa} > 0\}$, si $L_a \neq \emptyset$ entonces L_a contiene a todos los naturales de la forma kd, $k \in \mathbb{N}$, si k es suficientemente grande. En particular si $d = 1$, L_a contiene un segmento final de naturales.*

Demostración. Por la Observación 3.10 existe un conjunto finito

$$\{h_1, h_2, \dots, h_j\} \subset L_a, \quad \text{tal que} \quad d = \text{m.c.d.}(h_1, \dots, h_j).$$

Como es bien sabido, existe una combinación lineal entera de los h_i igual a d, es decir, existen $a_1, \dots, a_j \in \mathbb{Z}$ tal que $d = \sum_{\nu=1}^j a_\nu h_\nu$. Sea $M = h_1 + \cdots + h_j$ y $n \geq M$; entonces por el algoritmo de la división: $n = Mq + r$ $0 \leq r < M$, $q \in \mathbb{N}$, así pues $dn = dqM + rd = \sum_{\nu=1}^j (qd + a_\nu r) h_\nu$. Ahora bien, $h_\nu \in L_a$ para toda ν y como L_a es **claramente** cerrada bajo sumas, entonces $dn \in L_a$. En cuanto garanticemos que n se puede elegir de tal modo que $(qd + a_\nu r) \in \mathbb{N}$ para todo $\nu = 1, \dots, j$, para lo cual basta que: $[n/M]d = qd \geq \text{máx}_\nu |a_\nu| M$ pues:

$$qd \geq \max_\nu |a_\nu| M > \max_\nu |a_\nu| r \geq \max_\nu \{-a_\nu\} r > -a_\nu r \quad \text{para todo } \nu = 1, \dots, j.$$

Por lo tanto $dn \in L_a$ si n es suficientemente grande. ☐

3.5. Propiedades Mezclantes del Corrimiento de Markov

Estamos en posición de examinar cuándo un corrimiento de Markov es débilmente mezclante y fuertemente mezclante. Hemos visto que si P es irreducible entonces T es ergódico. Para que T sea fuertemente mezclante o débilmente mezclante necesitaremos pedirle una condición mas a P: la de ser aperiódica. También relacionaremos estas propiedades con la matriz Q.

Teorema 3.6. *Sea P una matriz estocástica con conjunto finito de estados E, y sea π una distribución inicial estacionaria para P. Las siguientes condiciones son equivalentes para $T = M(P, \pi)$:*

(i) T es débilmente mezclante.

(ii) $p^n_{ab} \to \pi_b$ para todo $a, b \in E$.

(iii) T es fuertemente mezclante.

(iv) Existe $m \in \mathbb{N}$ tal que $p^n_{ab} > 0$ para toda $n \geq m$, y para todo $a, b \in E$.

(v) P es irreducible y aperiódica.

Demostración.

(i) implica (ii). Sea $A = \{\omega : \omega(0) = a\}$ y $B = \{\omega : \omega(0) = b\}$. Como T es débilmente mezclante existe $J_a = J(a, b) \subset \mathbb{N}$ con $d(J_a) = 0$ tal que $\lim_{J_a \not\ni n \to \infty} \mu(A \cap T^{-n}(B)) = \mu(A)\mu(B)$, o bien $\lim_{J_a \not\ni n \to \infty} \pi_a p^n_{ab} = \pi_a \pi_b$. Por lo tanto $\lim_{J_a \not\ni n \to \infty} p^n_{ab} = \pi_b$. Sea $J = \bigcup_{a \in E} J_a$, entonces $d(J) = 0$ y dada $\varepsilon > 0$ existe $n \notin J$, suficientemente grande, tal que $|p^n_{ab} - \pi_b| < \varepsilon$ para toda $a \in E$, entonces:

$$\begin{aligned} \left|p^{n+1}_{ab} - \pi_b\right| = \left|\sum_{\delta \in E} p_{a\delta} p^n_{\delta b} - \pi_b\right| = \left|\sum_{\delta \in E} p_{a\delta} p^n_{\delta b} - \sum_{\delta \in E} p_{a\delta} \pi_b\right| &= \left|\sum_{\delta \in E} p_{a\delta} (p^n_{\delta b} - \pi_b)\right| \\ &\leq \sum_{\delta \in E} p_{a\delta} \left|p^n_{\delta b} - \pi_b\right| \\ &< \varepsilon, \end{aligned}$$

por lo tanto $p^n_{ab} \to \pi_b$ para toda $a, b \in E$.

(ii) implica (iii). Sean $A = Z(r_1, \ldots, r_l; a_1, \ldots, a_l)$ y $B = Z(n_1, \ldots, n_m; b_1, \ldots, b_m)$. Si $k + n_1 > r_l$ entonces:

$$\mu(A \cap T^{-k}(B)) = \mu(Z(r_1, \ldots, r_l, n_1, \ldots, n_m; a_1, \ldots, a_l, b_1, \ldots, b_m)) =$$

$$(\pi_{a_1} p^{r_2 - r_1}_{a_1 a_2} \cdots p^{r_l - r_{l-1}}) \underbrace{p^{k+m_1-r_l}_{a_l b_1}} p^{n_2 - n_1}_{b_1 b_2} \cdots p^{n_m - n_{m-1}}_{b_{m-1} b_m} \underset{k \to \infty}{\longrightarrow} \mu(A)\mu(B),$$

lo cual prueba que T es fuertemente mezclante.

(iii) implica (iv). Como T es fuertemente mezclante, entonces $p^n_{ab} \to \pi_b$ para toda $a, b \in E$; así que existe $m \in \mathbb{N}$ tal que $p^n_{ab} > (1/2)\pi_b > 0$ para toda $n \geq m$ y $a, b \in E$.

(iv) implica (v). Por hipótesis existe $m \in \mathbb{N}$ tal que $p^n_{aa} > 0$ para toda $n \geq m$ y $a \in E$, por lo que

$$\{m, m+1, \dots\} \subset L_a \quad \text{para toda } a \in E, \quad \text{así que} \quad per(a) = 1 \quad \text{para toda} \quad a \in E,$$

por lo tanto P es aperiódica. Como $p^m_{ab} > 0$ para toda $a, b \in E$ entonces $a \leftrightarrow b$ para toda $a, b \in E$, por lo tanto P es irreducible.

(v) implica (ii). Este resultado está contenido en el siguiente teorema. □

Teorema 3.7 (Convergencia exponencial)**.** *Sea P una matriz irreducible y aperiódica (con conjunto finito de estados E) y π una distribución inicial estacionaria para P. Entonces existen constantes $k > 0$ y $\rho \in (0, 1)$ tales que $|p^n_{ab} - \pi_b| \leq k\rho^n$ para toda n y para todo $a, b \in E$.*

Demostración. Para $n \in \mathbb{N} \cup \{0\}$, sea $m^{(n)}_b = \text{mín}_a\{p^n_{ab}\}$ y $M^{(n)}_b = \text{máx}_a\{p^n_{ab}\}$. Claramente $m^{(n)}_b \leq M^{(n)}_b$ para toda b y n, como P es estócastica:

$$m^{(n+1)}_b = \min_a \left\{ \sum_\delta p_{a\delta} p^{(n)}_{\delta b} \right\} \geq \min_a \left\{ \sum_\delta p_{a\delta} m^{(n)}_b \right\} = m^{(n)}_b \text{ y}$$

$$M^{(n+1)}_b = \max_a \left\{ \sum_\delta p_{a\delta} p^{(n)}_{\delta b} \right\} \leq \max_a \left\{ \sum_\delta p_{a\delta} M^{(n)}_b \right\} = M^{(n)}_b.$$

Luego entonces tenemos la siguiente situación

$$0 \leq m^{(1)}_b \leq \cdots \leq m^{(n)}_b \leq m^{(n+1)}_b \leq \cdots \leq M^{(n+1)}_b \leq M^{(n)}_b \leq \cdots \leq M^{(1)}_b \leq 1.$$

Supongamos (provisionalmente) que $p_{ab} > 0$ para toda $a, b \in E$, y sea $\delta = \text{mín}_{a,b}\, p_{ab} > 0$, entonces $1 = \sum_{b \in E} p_{ab} \geq \text{card}(E)\delta$, por lo tanto $0 < \delta \leq 1/\text{card}(E)$.

Fijemos estados $a, b \in E$. Denotamos por Σ' la suma sobre aquellos $c \in E$ tal que $p_{ac} \geq p_{bc}$, y por Σ'' la suma sobre aquellos $c \in E$ tal que $p_{ac} < p_{bc}$ entonces:

$$0 = 1 - 1 = \sum_{c \in E} p_{ac} - \sum_{c \in E} p_{bc} = \sum_c{}' (p_{ac} - p_{bc}) + \sum_c{}'' (p_{ac} - p_{bc}) \cdots (*),$$

y como $\sum' p_{bc} + \sum'' p_{ac} \geq \delta \text{card}(E)$ tenemos que:

$$\sum{}' (p_{ac} - p_{bc}) = (1 - \sum{}'' p_{ac}) - \sum{}' p_{bc} \leq 1 - \text{card}(E)\delta \cdots (**).$$

Luego entonces:

$$p^{(n+1)}_{ad} - p^{(n+1)}_{bd} = \sum_c (p_{ac} - p_{bc}) p^{(n)}_{cd} \leq \sum{}' (p_{ac} - p_{bc}) M^{(n)}_d + \sum{}'' (p_{ac} - p_{bc}) m^{(n)}_d$$

$$\overset{\text{por}(*)}{=} \sum{}' (p_{ac} - p_{bc})(M_d^{(n)} - m_d^{(n)}) \overset{\text{por}(**)}{\leq} (1 - \text{card}(E)\delta)(M_d^{(n)} - m_d^{(n)}).$$

Como $a, b \in E$ son arbitrarios tenemos que:

$$M_d^{(n+1)} - m_d^{(n+1)} \leq (1 - \text{card}(E)\delta)(M_d^{(n)} - m_d^{(n)}) \quad \text{para toda } n.$$

Usando inducción obtenemos que $M_d^{(n)} - m_d^{(n)} \leq (1 - \text{card}(E)\delta)^n \cdots (* * *)$ (puesto que $M_d^{(n)} - m_d^{(n)} \leq 1$ para $n = 0, 1, \ldots$ Así pues, las sucesiones convergentes $(m_d^{(n)})_{n\in\mathbb{N}}$ y $(M_d^{(n)})_{n\in\mathbb{N}}$ convergen al mismo límite, y como $m_d^{(n)} \leq p_{ad}^{(n)} \leq M_d^{(n)}$ para toda n, entonces $\lim_{n\to\infty} p_{ad}^{(n)}$ existe y es independiente de $d \cdots (* * **)$.

Como P es **irreducible**, por el Teorema 3.2, se tiene que $\lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1} p_{ad}^{(k)} \to \pi_d$. Por lo tanto $p_{ad}^{(n)} \to \pi_d$ para toda $a, d \in E$. Más aún, de (***) obtenemos $\left|p_{ad}^{(n)} - \pi_d\right| \leq (1 - \text{card}(E)\delta)^n$. Si $1 = \text{card}(E)\delta$ entonces $p_{ad}^{(n)} = \pi_d$ para toda n, y no hay nada que hacer, por lo que supondremos que $\delta\text{card}(E) < 1$. Sea $k = 1$ y $\rho = (1 - \text{card}(E)\delta) \in (\,0\,,1\,)$, esto prueba el caso en que $p_{ab} > 0$ para toda $a, b \in E$.

En el caso general, las hipótesis sobre P implican que existe $m > 1$ tal que $p_{ab}^{(m)} > 0$ para toda $a, b \in E$ (es decir (v) implica (iv)). Por lo ya probado (aplicado a la matriz estocástica irreducible y aperiódica P^m) obtenemos $M^{(mt)} - m^{(mt)} \leq \rho^t$. Tomamos $k_0 = 1/\rho$ y reemplazamos ρ con $\rho^{\frac{1}{m}} = \rho_0$. Si $s = mn$ por lo ya probado

$$|p_{ab}^s - \pi_b| \leq \rho^{\frac{s}{m}} < \rho^{-1}\rho^{\frac{s}{m}} = k_0\rho_0^s \quad \text{pues} \quad (1 < \rho^{-1}).$$

Para $s \in (mn, m(n+1))$, tenemos que $M_d^{(s)} - m_d^{(s)} \leq M_d^{(mn)} - m_d^{(mn)} < \rho^n$ por lo que

$$|p_{ab}^s - \pi_b| < \rho^n = \rho^{\frac{m(n+1)}{m} - 1} = k_0\rho^{\frac{m(n+1)}{m}} = k_0\rho_0^{m(n+1)} < k_0\rho_0^s,$$

por lo tanto $|p_{ab}^s - \pi_b| < k_0\rho_0^s$ para toda $a, b \in E$ y $s \geq 1$. □

Observación 3.11. Supongamos que desconocemos la existencia de una distribución inicial estacionaria π para P, entonces por (****) existe una matriz R cuyas columnas tienes elementos iguales tal que $P^n \to R$ $(n \to \infty)$. Claramente $RP = R$ entonces cada vector renglón de R resuelve la ecuación $\vec{x}P = \vec{x}$, por lo que cada vector renglón de R (los cuales son **todos** iguales) es una distribución inicial estacionaria de P, pues:

$$\sum_{b\in E} r_{ab} = \lim_{n\to\infty} \sum_{b\in E} p_{ab}^{(n)} = 1 \quad \text{y} \quad r_{ab} \geq 0 \quad \text{para toda } a, b.$$

De hecho la existencia de una **única** distribución inicial estacionaria para P queda garantizada con sólo pedir que P sea irreducible. Esto es consecuencia del teorema sobre matrices no-negativas (es decir $D = (d_{ab})$ con $d_{ab} \geq 0$) que a continuación enunciamos.

Teorema 3.8 (Perron-Frobenius). *Sea D=(d_{ab}) una matriz de orden $k \times k$ no-negativa, entonces:*

(i) Existe un valor propio positivo λ tal que $|\lambda'| \leq \lambda$ para todo valor propio λ' de D.

(ii) $\min_a \left(\sum_b d_{ab}\right) \leq \lambda \leq \max_a \left(\sum_b d_{ab}\right)$ *para toda* a.

(iii) Correspondiente a λ *existe un vector renglón no-negativo* $\vec{u}$ *y un vector columna no-negativo* $\vec{v}$ *tal que* $\vec{u}D = \lambda\vec{u}$ *y* $D\vec{v} = \lambda\vec{v}$.

(iv) Si D *es irreducible (es decir para todo* a, b *existe* $n = n(a, b) \in \mathbb{N}$ *tal que* $d_{ab}^{(n)} > 0$*) entonces* λ *es un valor propio simple. Los vectores propios izquierdo y derecho* $\vec{u}$ *y* $\vec{v}$ *son estrictamente positivos, y* λ *es el* ***único*** *valor propio de* D *con vectores propios (izquierdos y derechos) que son no-negativos.*

Demostración. La demostración de este resultado se encuentra en Gantmacher [9, págs. 64-65]. □

Observación 3.12. Si $D = P$ es estocástica entonces por Teorema (ii) se tiene que $\lambda = 1$. Si P es además irreducible, normalizando al vector propio $\vec{u}$ obtenemos una distribución inicial estacionaria para P, la cual es **única** pues λ es simple. Otro modo de ver que tal $\vec{u}$ es único es considerar el corrimiento de Markov $T = M(P, \vec{u})$, como P es irreducible se tiene que T es ergódico y por Teorema 3.2: $u_b = \lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1} p_{ab}^k$ para toda b, por lo tanto $\vec{u}$ está determinado de manera única.

3.6. Una aplicación al motor de búsqueda de Google

En esta sección, se muestra una aplicación de la Teoría Ergódica al motor de búsqueda de Google (www.google.com). El motor de búsqueda de Google fue creado por Sergey Brin and Lawrence Page. Este motor es una serie de algoritmos que se utilizan para la búsqueda de información en Internet. Uno de sus principales componentes es el PageRank. Este es una medida de la importancia de una página web basada en la democracia de la web. Sin embargo, la obtención del PageRank no es sencilla. La aplicación está basada en el Teorema 3.7, el cual dará una buena aproximación del PageRank (ver Brin [5, pág. 103]).

Internet ofrece una gran cantidad de información. Podríamos comparar la búsqueda de información en Internet con la búsqueda de un libro en una biblioteca sin catálogo. La búsqueda de información en Internet es realizada por un motor de búsqueda (search engine en inglés). La principal tarea de estos motores de búsqueda es la recopilación de información de las páginas web, procesar y almacenar esta información en una base de datos y producir de esta base de datos una lista de páginas web relevante a un arreglo de una o más palabras. La recopilación de información es realizada por programas llamados **rastreadores** (crawlers) que ‶rastrean″ la web siguiendo los enlaces que hay en las páginas web. La información recabado por los rastreadores es analizada y codificada por el **indexador** (indexer), el cual produce para cada página un conjunto de palabras [incluyendo su posición en la página, tipo de fuente y tamaño (mayúsculas o minúsculas)] y los registros de todos los enlaces de esta página web a otras páginas. Lo cual crea el **indexador hacia delante** (forward index), es decir, a cada página web el indexador hacia delante le asocia un conjunto que podemos pensar como un conjunto de vectores cuya primera entrada es una palabra que aparece en el documento. La segunda la posición donde aparece esta en la página; la tercera su tipo de fuente; la cuarta si está en mayúsculas o minúsculas; y la última el número de enlaces que hay de esta página a otras páginas web.

El clasificador (sorter) arregla esta información por palabras, es decir, realiza una especie de proceso inverso para crear el índice hacia atrás (inverted index); es decir, a cada conjunto de palabras les asigna un conjunto de páginas, o sea, en qué páginas recopiladas aparecen estas palabras. El buscador (searcher) utiliza el índice hacia atrás para generar una lista de páginas web relevantes a un conjunto de palabras.

Una de las principales diferencias del motor de búsqueda de Google con otros motores de búsqueda, es la idea de que no todas las páginas obtenidas por el índice hacia atrás son relevantes para el conjunto de palabras, es decir, el orden en que aparece el resultado de la búsqueda es importante. El resultado de una búsqueda puede contener alrededor de 10000 páginas web, sin embargo sólo 50 serán consultadas por el usuario. Google utiliza dos características de la página web para determinar el orden de aparición de las páginas encontradas; la relevancia de la página respecto al conjunto de palabras buscado y el **PageRank** de la página. La relevancia está basada en la posición relativa, la fortificación y la frecuencia de las palabras buscadas en la página, ésta por sí sola no produce buenos resultados. Por ejemplo, si buscamos la palabra Internet en un motor de búsqueda antiguo da resultados cuya primera página está en chino y no contiene más palabras en inglés que Internet, incluso algunos motores de búsqueda actuales dan resultados apenas relevantes cuando buscamos términos comunes.

Google utiliza cadenas de Markov para asignarle un rango a las páginas web. La colección de las páginas web y los vínculos entre ellas son vistos como una gráfica dirigida G, en la cual los vértices son las páginas web y las aristas son los hipervínculos que existen de una página web a otra página web (desde la página web donde aparece hacia la página web que enlaza). En la actualidad existen alrededor de 1.5 billones de páginas web con 10 veces más hipervínculos[3].

Numeramos los vértices con números enteros $i = 1, 2, \ldots, N$ donde N es el número de páginas en la web, consideramos $\overline{G}$ la gráfica que se obtiene de añadir a G el vértice 0 con aristas desde él hacia todos los demás vértices y viceversa. Sean $a_{ij} = 1$ si existe una arista del vértice i al vértice j en $\overline{G}$ y $C(i)$ =número de aristas que salen del vértice i. Notemos que $C(i) > 0$ para toda $i \in \overline{G}$ ya que existe una arista del vértice 0 a los demás vértices. Fijemos un parámetro $d \in (0, 1)$ (por ejemplo $d = .85$). Sea P la matriz dada como sigue: $p_{ii} = 0$ para toda $i \geq 0$, para $i, j > 0$ con $i \neq j$ sean

[3]Un hipervínculo es un enlace de una página web a otra.

$$P_{0i} = \frac{1}{N} \qquad P_{i0} = \begin{cases} 1 & \text{si} \quad C(i) = 1 \\ \\ 1-d & \text{si} \quad C(i) \neq 1 \end{cases} \quad P_{ij} = \begin{cases} 0 & \text{si} \quad a_{ij} = 0 \\ \\ \dfrac{d}{C(i)-1} & \text{si} \quad b_{ij} = 1. \end{cases}$$

La matriz P es estocástica, irreducible y aperiódica. Por la Observación 3.11 existe una única distribución inicial estacionaria π para P. Google interpreta a π_i como el PageRank de la página web i. Google utiliza éste junto con el factor de relevancia de la página para determinar en qué posición se encontrará en el resultado de la búsqueda.

Más aún, por el Teorema 3.7, para cualquier distribución inicial π' sobre los vértices de $\overline{G}$, la sucesión $\{\pi' P^n\}$ converge exponencialmente a π. Por lo que podemos encontrar una buena aproximación de π calculando $\pi' P^n$, donde π' es la distribución uniforme. Esta aproximación para encontrar π es mucho más eficiente que tratar de encontrar un vector propio de una matriz de 1.5 billones de columnas y renglones.

Capítulo 4

Entropía

En este capítulo introduciremos el concepto de entropía, el cual está relacionado con medir la **incertidumbre** de un experimento aleatorio. Por ejemplo, pensemos en un experimento con posibles resultados A_1, A_2, con p_1 y p_2 sus respectivas probabilidades. Este experimento tiene un grado de incertidumbre acerca de su resultado, es decir, no tenemos la certeza de que al realizar el experimento obtengamos A_1 ó A_2. Supongamos que solamente conocemos los valores p_1 y p_2. Si el experimento es tal que $p_1 = p_2$ (los eventos son equiprobables), entonces el grado de incertidumbre de este experimento intuitivamente debe ser mayor que si el experimento es tal que $p_1 \neq p_2$. Para ilustrar lo anterior supongamos $p_1 > p_2$, entonces la probabilidad de obtener el evento A_1 es mayor que la probabilidad de obtener el evento A_2, intuitivamente el evento A_1 será más frecuente que A_2, por lo que la incertidumbre sobre el resultado del experimento disminuye. Con base en unas propiedades que se consideraron para esta **medida de incertidumbre** es que surge la definición de entropía (ver Khinchin [14, págs. 2-12]).

La entropía es una propiedad invariante bajo **isomorfismo** (vér Definición 2.1, pág. 61). Una de las principales motivaciones por la cual Kolmogorov introdujo el concepto de entropía en la teoría ergódica fue para dar solución al siguiente problema: ¿El corrimiento bilateral de de Bernoulli con parámetros (1/2,1/2) y el corrimiento bilateral de Bernoulli con parámetros (1/3,1/3,1/3) son isomorfos? Se puede mostrar que ambos corrimientos tienen espectro de Lebesgue numerable (ver Walters [23, pág. 65]), por lo tanto sabemos que estos corrimientos son espectralmente isomorfos (ver Teorema 2.1, pág. 64). Sin embargo no fue hasta que Kolmogorov introdujo la entropía que se pudo resolver. Este problema estuvo abierto por 15 años. Kolmogorov mostró que estos corrimientos tienen entropía $\log(2)$ y $\log(3)$ respectivamente y por tanto no son isomorfos (lo cual veremos en este capítulo).

La definición de entropía de una transformación preservadora de medida está dada en tres partes: la entropía de una sub-σ-álgebra finita de $\mathscr{F}$, la entropía de una transformación T relativa a una sub-σ-álgebra finita, y la entropía de T. Antes de comenzar con las definiciones veremos algunas propiedades de las sub-σ-álgebras finitas de $\mathscr{F}$.

4.1. Particiones y sub-σ-álgebras

A partir de ahora $(\Omega, \mathscr{F}, \mu)$ denotará un espacio de probabilidad. Antes de pasar a la noción de entropía necesitaremos las siguientes definiciones:

Definición 4.1. Una partición de $(\Omega, \mathscr{F})$ es una colección disjunta de elementos de $\mathscr{F}$ cuya unión es Ω.

Observaciones 4.1.

(1) Estaré interesado en particiones finitas que denotaremos con letras griegas, por ejemplo, $\xi = \{A_1, \ldots, A_k\}$.

(2) Si ξ es una partición finita de $(\Omega, \mathscr{F})$ entonces la colección de todos los elementos de $\mathscr{F}$ que son uniones de elementos de ξ junto con el vacío, forman una sub-σ-álgebra finita de $\mathscr{F}$, denotándola como $\mathscr{A}(\xi)$.

Demostración. $\Omega \in \mathscr{A}(\xi)$ ya que $\bigcup_{i=1}^{k} A_i = \Omega$ por ser partición. Si $A, B \in \mathscr{A}(\xi)$, entonces $A = \bigcup_{i \in N} A_i$ (disjunta) para algún $N \subset \{1, \ldots, k\}$, y $B = \bigcup_{i \in M} A_i$ (disjunta) para algún $M \subset \{1, \ldots, k\}$; así que $A \setminus B = \bigcup_{i \in S} A_i$ donde $S = N \setminus M$ ya que los A_i don disjuntos dos a dos. Por último, todas las colecciones numerables de elementos de ξ, en algún momento se repiten porque ξ es finito, entonces $\bigcup_{n=1}^{\infty} A_n = \bigcup_{i \in N} A_i$ para algún $N \subset \{1, \ldots, k\}$. Por lo tanto $\mathscr{A}(\xi)$ es una sub-σ-álgebra finita de $\mathscr{F}$. □

(3) Si $\mathscr{C} = \{C_i : i = 1, \ldots, n\}$ es una sub-σ-álgebra finita de $\mathscr{F}$ entonces los conjuntos no vacíos de la forma $E_1 \cap \cdots \cap E_n$, donde $E_i = C_i$ ó $E_i = \Omega \setminus C_i$, forman una partición finita de $(\Omega, \mathscr{F})$ que denotaremos por $\xi(\mathscr{C})$.

Demostración. Si $A = E_1 \cap \cdots \cap E_n$ y $B = E_1' \cap \cdots \cap E_n'$ tal que $A \neq B$, entonces $A \cap B = \emptyset$, ya que $E_i \cap E_i' = \emptyset$ para alguna $i \in \{1, 2 \ldots, n\}$. Por último, basta probar que Ω está contenido en la unión de los elementos de la partición; pero si $\omega \in \Omega$, se tiene que $w \in C_i$ ó $w \in \Omega \setminus C_i$ para toda $i \in \{1, \ldots, n\}$, por lo tanto ω pertenece a la unión de los elementos de la partición. □

(4) Considerando las observaciones anteriores, tenemos que $\mathscr{A}(\xi(\mathscr{C})) = \mathscr{C}$ y $\xi(\mathscr{A}(\eta)) = \eta$. Por lo que tenemos una correspondencia uno a uno entre particiones finitas y las sub-σ-álgebras finitas de $\mathscr{F}$, lo cual nos será de utilidad en este capítulo.

Definición 4.2. Sean ξ y η dos particiones finitas de $(\Omega, \mathscr{F})$. Diremos que $\xi \leq \eta$ si cada elemento de ξ es una unión de elementos de η (es decir, η es un **refinamiento** de ξ). Tenemos que $\xi \leq \eta$ si y sólo si $\mathscr{A}(\xi) \subseteq \mathscr{A}(\eta)$, y $\mathscr{A} \subseteq \mathscr{C}$ si y sólo si $\xi(\mathscr{A}) \leq \xi(\mathscr{C})$.

Definición 4.3.

(a) Sean $\xi = \{A_1, \ldots, A_n\}$, $\eta = \{C_1, \ldots, C_m\}$ dos particiones finitas de $(\Omega, \mathscr{F})$. Su **junta** es la partición
$$\xi \vee \eta = \{A_i \cap C_j : 1 \leq i \leq n,\ 1 \leq j \leq m\}.$$

(b) Sean $\mathscr{A}$ y $\mathscr{C}$ dos sub-σ-álgebras finitas de $\mathscr{F}$. Definimos $\mathscr{A} \vee \mathscr{C} = \sigma(\mathscr{A} \cup \mathscr{B})$.

Observaciones 4.2.

(1) Si $\mathscr{A}$, $\mathscr{C}$ son sub-σ-algebras finitas de $\mathscr{F}$. Entonces $\mathscr{A} \vee \mathscr{C} = \sigma(\mathscr{A} \cup \mathscr{C})$ consiste de todos los conjuntos que son uniones finitas y disjuntas de conjuntos de la forma $A \cap C$, $A \in \mathscr{A}$, $C \in \mathscr{C}$. La prueba de lo anterior es análoga a la prueba de la Observación 4.1.2.

(2) Por la observación anterior tenemos que $\xi(\mathscr{A} \vee \mathscr{C}) = \xi(\mathscr{A}) \vee \xi(\mathscr{C})$ y $\mathscr{A}(\xi \vee \eta) = \mathscr{A}(\xi) \vee \mathscr{A}(\eta)$.

Definición 4.4. Sean $(\Omega, \mathscr{F})$ un espacio medible, $T : \Omega \to \Omega$ una transformación medible, $k \in \mathbb{N}$ fijo y $(n \geq 0)$ fijo. Si $\xi = \{A_1, \ldots, A_k\}$ es una particion de $(\Omega, \mathscr{F})$, $T^{-n}(\xi)$ denota la partición $\{T^{-n}(A_1), \ldots, T^{-n}(A_k)\}$. Si $\mathscr{A}$ es una sub-σ-álgebra de $\mathscr{F}$, $T^{-n}(\mathscr{A})$ denota la sub-σ-álgebra $\{T^{-n}(A) : A \in \mathscr{A}\}$.

Sea $n \geq 0$ fijo. Como la imagen inversa preserva las operaciones entre conjuntos, obtenemos:

(i) $\xi(T^{-n}(\mathscr{A})) = T^{-n}(\xi(\mathscr{A}))$.

(ii) $\mathscr{A}(T^{-n}(\xi)) = T^{-n}(\mathscr{A}(\xi))$.

(iii) $T^{-n}(\mathscr{A} \vee \mathscr{C}) = T^{-n}(\mathscr{A}) \vee T^{-n}(\mathscr{C})$.

(iv) $T^{-n}(\xi \vee \eta) = T^{-n}(\xi) \vee T^{-n}(\eta)$.

(v) Si $\xi \leq \eta$ entonces $T^{-n}(\xi) \leq T^{-n}(\eta)$.

(vi) Si $\mathscr{A} \subseteq \mathscr{C}$ entonces $T^{-n}(\mathscr{A}) \subseteq T^{-n}(\mathscr{C})$.

Definición 4.5. Si $\mathscr{C}$, $\mathscr{D}$ son sub-σ-álgebras de $\mathscr{F}$ (no necesariamente finitas) se escribe $\mathscr{C} \overset{\circ}{\subset} \mathscr{D}$ si para cada $C \in \mathscr{C}$ existe $D \in \mathscr{D}$ tal que $\mu(D \bigtriangleup C) = 0$, y $\mathscr{C} \overset{\circ}{=} \mathscr{D}$ si $\mathscr{C} \overset{\circ}{\subset} \mathscr{D}$, y $\mathscr{D} \overset{\circ}{\subset} \mathscr{C}$. Si ξ, η son dos particiones finitas, entonces $\xi \overset{\circ}{=} \eta$ se refiere a $\mathscr{A}(\xi) \overset{\circ}{=} \mathscr{A}(\eta)$.

Observación 4.3. Sean $\mathscr{C}, \mathscr{D}$ dos sub-σ-álgebras finitas tal que $\mathscr{C} \overset{\circ}{=} \mathscr{D}$, si $\xi(\mathscr{C}) = \{C_1, \ldots, C_p, C_{p+1}, \ldots, C_q\}$, donde $\mu(C_i) > 0$ para $1 \leq i \leq p$ y $\mu(C_i) = 0$ para $p+1 \leq i \leq q$; entonces $\xi(\mathscr{D}) = \{D_1, \ldots, D_p, D_{p+1}, \ldots, D_s\}$ donde $\mu(C_i \bigtriangleup D_i) = 0$ para $1 \leq i \leq p$ y $\mu(D_i) = 0$ para $p+1 \leq i \leq s$.

4.2. Entropía de una partición

La expresión $0 \cdot \log(0)$ será considerada como cero, y $\log(x)$ se referirá al logaritmo natural de x. Continuando con la idea que se introdujo en el principio podemos interpretar a los elementos de una partición $\xi = \{A_1, \ldots, A_n\}$ de un espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$ como los posibles resultados de un experimento con $\mu(A_n)$, la probabilidad del evento A_n. Además como ξ es una partición, estos eventos son mutuamente excluyentes. Como hemos mencionado, queremos que la entropía de este experimento sea un cierto número $H(\xi)$ que describa la **incertidumbre** acerca del resultado de este experimento. En otras palabras, $H(\xi)$ medirá la cantidad de incertidumbre removida (o la información obtenida) después de realizar el experimento que representa ξ. Tomaremos indistintamente al concepto de entropía como la incertidumbre que existe en el resultado de un experimento, o como la información obtenida después de realizar éste. Queremos que $H(\xi)$ sólo dependa de los números $\{\mu(A_1), \ldots, \mu(A_n)\}$, es decir, $H(\mu(A_1), \ldots, \mu(A_n))$. Ahora pasamos a dar la definición de entropía de una partición:

Definición 4.6. Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad y $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$ con $\xi(\mathscr{A}) = \{A_1, \ldots, A_k\}$. La **entropía** de $\mathscr{A}$ (o de $\xi(\mathscr{A})$) es el número

$$H(\mathscr{A}) = H(\xi(\mathscr{A})) = -\sum_{i=1}^{k} \mu(A_i)\log(\mu(A_i)).$$

Observaciones 4.4.

(1) Si $\mathscr{A} = \{\Omega, \emptyset\}$ entonces $H(\mathscr{A}) = 0$. Aquí $\mathscr{A}$ representa el resultado de ‶cierto″ experimento por el que no hay incertidumbre sobre el resultado.

(2) Si $\xi(\mathscr{A}) = \{A_1, \ldots, A_k\}$ donde $\mu(A_i) = (1/k)$ para toda i, entonces

$$H(\mathscr{A}) = -\sum_{i=i}^{k} \frac{1}{k}\log\left(\frac{1}{k}\right) = \log(k).$$

Más adelante mostraremos que $\log(k)$ es el valor máximo para la entropía de una partición con k conjuntos. La mayor incertidumbre sobre el resultado de un experimento debería ocurrir cuando los eventos son equiprobables.

(3) $H(\mathscr{A}) \geq 0$.

(4) Si $\mathscr{A} \overset{\circ}{=} \mathscr{C}$ entonces $H(\mathscr{A}) = H(\mathscr{C})$. Lo anterior dice que si el experimento asociado a $\mathscr{A}$ y a $\mathscr{C}$, tienen los mismos resultados modulo cero, entonces tienen la misma cantidad de incertidumbre.

(5) Si $T : \Omega \to \Omega$ es una transformación medible y preservadora de medida, entonces $H(T^{-1}(\mathscr{A})) = H(\mathscr{A})$, es decir, la información obtenida después de realizar el experimento $\mathscr{A}$ es la misma que la obtenida del experimento asociado a la partición $\{T^{-1}(A_1), \ldots, T^{-1}(A_k)\}$.

(6) Si $\mathscr{A} = \{A_1, A_2\}$, $\mathscr{B} = \{B_1, B_2\}$, $\mathscr{C} = \{C_1, C_2\}$ particiones de $(\Omega, \mathscr{F})$ tales que $\mu(A_1) = .99$, $\mu(A_2) = .01$, $\mu(B_1) = .7$, $\mu(B_2) = .3$ y $\mu(C_1) = .5$, $\mu(C_2) = .5$, entonces $H(\mathscr{A}) < H(\mathscr{B}) < H(\mathscr{C})$. Lo anterior nos dice lo que intuitivamente esperaríamos de la entropía; esto pues como el resultado del experimento asociado a $\mathscr{A}$ es casi siempre A_1, entonces la entropía de $\mathscr{A}$ debe ser menor que la entropía de $\mathscr{B}$, ya que intuitivamente será más difícil predecir el resultado del experimento $\mathscr{B}$. De la misma manera será más difícil predecir el resultado del experimento $\mathscr{C}$ que el de $\mathscr{B}$.

A continuación veremos que esta definición de entropía es una medida adecuada de incertidumbre. Varias propiedades de la entropía son consecuencia del siguiente resultado:

Teorema 4.1. *La función $\phi : [0, \infty) \to \mathbb{R}$ definida por*

$$\phi(x) = \begin{cases} 0 & \text{si } x = 0 \\ x \log(x) & \text{si } x \neq 0 \end{cases}$$

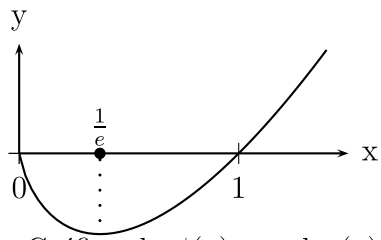

Figura 4.1: Gráfica de $\phi(x) = x\log(x)$. Note que $1/e$ es el mínimo de la función,

es convexa, es decir, $\phi(\alpha x + \beta y) \leq \alpha\phi(x) + \beta\phi(y)$ si $x, y \in [0, \infty)$, α, $\beta \geq 0$, $\alpha + \beta = 1$; la igualdad se da solamente cuando $x = y$ o $\alpha = 0$ o $\beta = 0$. Por inducción obtenemos

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

si $x_i \in [0, \infty)$, $\alpha_i \geq 0$, $\sum_{i=1}^{k} \alpha_i = 1$; la igualdad se da sólo cuando todas las x_i, correspondientes a α_i distintas de cero, son iguales.

Demostración. Si derivamos $\phi(x)$ obtenemos

$$\phi'(x) = 1 + \log(x)$$

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

Sean α, β fijos. Supongamos que $y > x$. Por el Teorema del Valor Medio tenemos que $\phi(y) - \phi(\alpha x + \beta y) = \phi'(z)\alpha(y - x)$ para algún z tal que

$$\alpha x + \beta y < z < y \quad \text{y} \quad \phi(\alpha x + \beta y) - \phi(x) = \phi'(w)\beta(y - x),$$

para algún w tal que $x < w < \alpha x + \beta y$. Como $\phi'' > 0$, tenemos que $\phi'(z) > \phi'(w)$, por lo que

$$\beta(\phi(y) - \phi(\alpha x + \beta y)) = \phi'(z)\alpha\beta(y - x) > \phi'(w)\alpha\beta(y - x) = \alpha(\phi(\alpha x + \beta y) - \phi(x)).$$

Por lo tanto $\phi(\alpha x + \beta y) < \alpha\phi(x) + \beta\phi(y)$ si $x, y > 0$. Claramente es válido para x, $y \geq 0$ y $x \neq y$. $\square$

Corolario 4.1.1. *Si* $\xi = \{A_1, \dots, A_k\}$ *entonces* $H(\xi) \leq \log(k)$ *y* $H(\xi) = \log(k)$ *sólo cuando* $\mu(A_i) = 1/k$ *para toda* i.

Demostración. Considere $\alpha_i = 1/k$ y $x_i = \mu(A_i)$, $1 \leq i \leq k$. □

Observación 4.5. El corolario anterior nos asegura que esta definición de entropía tiene un máximo en las particiones ξ con k elementos, y éste se alcanza cuando todos los eventos son equiprobables. Esta es una propiedad que intuitivamente debe tener una buena medida de incertidumbre.

4.3. Entropía Condicional

La entropía condicional no se requiere para la definición de entropía de una transformación, pero será de utilidad para dar propiedades de la entropía que ayudarán a mostrar que la definición de entropía es una buena media de incertidumbre. Sean $\mathscr{A}$, $\mathscr{C}$ dos sub-σ-álgebras finitas de $\mathscr{F}$ y $\xi(\mathscr{A}) = \{A_1, \dots, A_k\}$, $\xi(\mathscr{C}) = \{C_1, \dots, C_m\}$.

Definición 4.7. La **entropía de $\mathscr{A}$ dado $\mathscr{C}$** es el número:

$$\begin{aligned} H(\xi(\mathscr{A})/\xi(\mathscr{C})) = H(\mathscr{A}/\mathscr{C}) &= -\sum_{j=1}^{m} \mu(C_j) \sum_{i=1}^{k} \frac{\mu(A_i \cap C_j}{\mu(C_j))} \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right) \\ &= -\sum_{i,j} \mu(A_i \cap C_j) \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right). \end{aligned}$$

Omitiendo los términos donde $\mu(C_j) = 0$.

Entonces para obtener $H(\mathscr{A}|\mathscr{C})$ consideramos el espacio de probabilidad $(C_j, \mathscr{F} \cap C_j, \mu')$, donde $\mu'(F \cap C_j) = \mu(F \cap C_j)/\mu(C_j)$ para toda $F \in \mathscr{F}$. Ahora consideramos la entropía de la partición $\xi = \{A_1, \dots, A_k\}$ con respecto a este espacio, entonces:

$$H(\xi) = -\sum_{i=1}^{k} \mu'(A_i) \log(\mu'(A_i)) = -\sum_{i=1}^{k} \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right),$$

tomando el promedio ponderado de estos números con respecto a las medidas de C_j obtenemos $H(\mathscr{A}|\mathscr{C})$, es decir, $H(\mathscr{A}|\mathscr{C})$ mide la incertidumbre del resultado de $\mathscr{A}$ dado el resultado del experimento $\mathscr{C}$. Note que esta definición es la esperanza de los números

$$-\sum_{i=1}^{k} \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right),$$

con respecto al experimento asociado a $\mathscr{C}$.

Sea $\mathscr{N} = \{\emptyset, \Omega\}$, entonces $H(\mathscr{A}|\mathscr{N}) = H(\mathscr{A})$, es decir, si ocurre el experimento $\mathscr{N}$ no se obtiene información acerca del experimento $\mathscr{A}$.

Observaciones 4.6.

(1) $H(\mathscr{A}/\mathscr{C}) \geq 0$.

(2) Si $\mathscr{A} \overset{\circ}{=} \mathscr{D}$ entonces $H(\mathscr{A}/\mathscr{C}) = H(\mathscr{D}/\mathscr{C})$, es decir, la información contenida en $\mathscr{A}$ es la misma que la contenida en $\mathscr{D}$, incluso si sabemos la información contenida de realizar cualquier experimento $\mathscr{C}$.

(3) Si $\mathscr{C} \overset{\circ}{=} \mathscr{D}$ entonces $H(\mathscr{A}/\mathscr{C}) = H(\mathscr{A}/\mathscr{D})$, es decir, la información que aportan los experimentos $\mathscr{C}$ y $\mathscr{D}$ al experimento $\mathscr{A}$, es la misma.

Los siguientes resultados ayudarán con las ideas intuitivas de $H(\xi)$:

Teorema 4.2. *Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad. Sean $\mathscr{A}$, $\mathscr{C}$, $\mathscr{D}$ sub-σ-álgebras finitas de $\mathscr{F}$ entonces:*

(i) $H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) = H(\mathscr{A}/\mathscr{D}) + H(\mathscr{C}/\mathscr{A} \vee \mathscr{D})$.

(ii) $H(\mathscr{A} \vee \mathscr{C}) = H(\mathscr{A}) + H(\mathscr{C}/\mathscr{A})$.

(iii) Si $\mathscr{A} \subseteq \mathscr{C}$ entonces $H(\mathscr{A}/\mathscr{D}) \leq H(\mathscr{C}/\mathscr{D})$.

(iv) Si $\mathscr{A} \subseteq \mathscr{C}$ entonces $H(\mathscr{A}) \leq H(\mathscr{C})$.

(v) Si $\mathscr{C} \subseteq \mathscr{D}$ entonces $H(\mathscr{A}/\mathscr{C}) \geq H(\mathscr{A}/\mathscr{D})$.

(vi) $H(\mathscr{A}) \geq H(\mathscr{A}/\mathscr{D})$.

(vii) $H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) \leq H(\mathscr{A}/\mathscr{D}) + H(\mathscr{C}/\mathscr{D})$.

(viii) $H(\mathscr{A} \vee \mathscr{C}) \leq H(\mathscr{A}) + H(\mathscr{C})$.

(ix) Si T es una transformación medible y preservadora de la medida entonces:

$$H(T^{-1}(\mathscr{A})/T^{-1}(\mathscr{C})) = H(\mathscr{A}/\mathscr{C}), \ y$$

(x) $H(T^{-1}(\mathscr{A})) = H(\mathscr{A})$.

Observación 4.7. Intuitivamente (i) dice que la información obtenida de realizar el experimento $\mathscr{A} \vee \mathscr{C}$, sabiendo la información obtenida de realizar el experimento $\mathscr{D}$; es igual a la información del experimento $\mathscr{A}$, sabiendo la información del experimento $\mathscr{D}$; más la información del experimento $\mathscr{C}$, sabiendo la información del experimento $\mathscr{A} \vee \mathscr{C}$. Los demás incisos los podemos interpretar de manera similar.

Demostración. Sean $\xi(\mathscr{A}) = \{A_i\}$, $\xi(\mathscr{C}) = \{C_j\}$, $\xi(\mathscr{D}) = \{D_k\}$. Suponemos sin pérdida de generalidad que todos los conjuntos tienen medida positiva (ya que si $\xi(\mathscr{A}) = \{A_1, \ldots, A_k\}$ con $\mu(A_i) > 0$, $1 \le i \le r$ y $\mu(A_i) = 0$, $r < i \le k$ podemos reemplazar $\xi(\mathscr{A})$ por

$$\{A_, \ldots, A_{r-1}, A_r \cup A_{r+1} \cup \cdots \cup A_k\}$$

(ver Observación 4.6(2) y 4.6(3)).

(i)$H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) = -\sum_{i,j,k} \mu(A_i \cap C_j \cap D_k) \log\left(\frac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)}\right)$. Pero

$$\frac{\mu(A_i \cap C_j \cap D_k)}{\mu(D_k)} = \frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)},$$

cuando $\mu(A_i \cap D_k) \neq 0$, pero el lado izquierdo se vuelve cero, por lo que no es necesario considerarlo; por lo tanto

$$\begin{aligned}
H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) &= -\sum_{i,j,k} \mu(A_i \cap C_j \cap D_k) \log\left(\frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right) \\
&\quad -\sum_{i,j,k} \mu(A_i \cap C_j \cap D_k) \log\left(\frac{\mu(A_i \cap C_j \cap D_k)}{\mu(A_i \cap D_k)}\right) \\
&= -\sum_{i,k} \mu(A_i \cap D_k) \log\left(\frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right) + H(\mathscr{C}/\mathscr{A} \vee \mathscr{D}) \\
&= H(\mathscr{A}/\mathscr{D}) + H(\mathscr{C}/\mathscr{A} \vee \mathscr{D}).
\end{aligned}$$

(ii) Utilice $\mathscr{D} = \mathscr{N} = \{\emptyset, \Omega\}$ en el inciso anterior.

(iii) Como $\mathscr{A} \subseteq \mathscr{C}$, entonces $\mathscr{A} \vee \mathscr{C} = \mathscr{C}$ utilizando (i) tenemos que

$$H(\mathscr{C}/\mathscr{D}) = H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) = H(\mathscr{A}/\mathscr{D}) + H(\mathscr{C}/\mathscr{A} \vee \mathscr{D}) \ge H(\mathscr{A}/\mathscr{D}).$$

(iv) Utilice $\mathscr{D} = \mathscr{N}$ en el inciso anterior.

(v) Sean i, j fijos y

$$\alpha_k = \frac{\mu(D_k \cap C_j)}{\mu(C_j)}, \qquad x_k = \frac{\mu(A_i \cap D_k)}{\mu(D_k)}.$$

Por el Teorema 4.1

$$\phi\left(\sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right) \le \sum_k \frac{\mu(D_k \cap C_j)}{\mu(C_j)} \phi\left(\frac{\mu(A_i \cap D_k)}{\mu(D_k)}\right),$$

pero como $\mathscr{C} \subseteq \mathscr{D}$ entonces el lado izquierdo es igual a

$$\phi\left(\frac{\mu(A_j \cap C_j)}{\mu(C_j)}\right) = \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right).$$

Multiplicando ambos lados de la desigualdad por $\mu(C_j)$, y sumando sobre i y j obtenemos:

$$\begin{aligned}\sum_{i,j}\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\right) &\leq \sum_{i,j,k}\mu(D_k\cap C_j)\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\log\left(\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\right)\\ &= \sum_{i,k}\mu(D_k)\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\log\left(\frac{\mu(A_i\cap D_k)}{\mu(D_k)}\right),\end{aligned}$$

así que $-H(\mathscr{A}/\mathscr{C})\leq -H(\mathscr{A}/\mathscr{D})$. Por lo tanto $H(\mathscr{A}/\mathscr{D})\leq H(\mathscr{A}/\mathscr{C})$.

(vi) Utilice $\mathscr{C}=\mathscr{N}$ en el inciso anterior.

(vii) Considerando el inciso (i) obtenemos la desigualdad utilizando el inciso (v).

(viii) Utilice $\mathscr{D}=\mathscr{N}$ en el inciso anterior.

(ix),(x) Son claras a partir de las definiciones. □

Teorema 4.3. *Sean $\mathscr{A}$, $\mathscr{C}$ dos sub-σ-álgebras finitas de $\mathscr{F}$. Entonces*

(i) $H(\mathscr{A}/\mathscr{C})=0$ (es decir $H(\mathscr{A}\vee\mathscr{C})=H(\mathscr{C})$) si y sólo si $\mathscr{A}\overset{\circ}{\subset}\mathscr{C}$.

(ii) $H(\mathscr{A}/\mathscr{C})=H(\mathscr{A})$ (es decir $H(\mathscr{A}\vee\mathscr{C})=H(\mathscr{A})+H(\mathscr{C})$) si y sólo si $\mathscr{A}$ y $\mathscr{C}$ son independientes (es decir $\mu(A\cap C)=\mu(A)\mu(C)$ siempre que $A\in\mathscr{A}$, $C\in\mathscr{C}$).

Demostración. Sea $\xi(\mathscr{A})=\{A_1,\ldots,A_n\}$, $\xi(\mathscr{C})=\{C_1,\ldots,C_m\}$. Sin pérdida de generalidad podemos suponer que todos estos conjuntos tienen medida positiva.

(i) Como $\mathscr{A}\overset{\circ}{\subset}\mathscr{C}$ entonces para cada i y para cada j se tiene que $\mu(A_i\cap C_j)=\mu(C_j)$ o $\mu(A_i\cap C_j)=0$. Lo cual implica $H(\mathscr{A}/\mathscr{C})=0$. Supongamos que $H(\mathscr{A}/\mathscr{C})=0$, entonces:

$$0=-\sum_{i=1}^{n}\sum_{j=1}^{m}\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\right),$$

pero

$$-\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j}{\mu(C_j)}\right)\geq 0,$$

debemos tener

$$\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j}{\mu(C_j)}\right)=0,$$

para cada i, j. Por lo que $\mu(A_i\cap C_j)=\mu(C_j)$ o $\mu(A_i\cap C_j)=0$. Por lo tanto $\mathscr{A}\overset{\circ}{\subset}\mathscr{C}$.

(ii) Si $\mathscr{A}$ y $\mathscr{C}$ son independientes podemos observar de la definición de $H(\mathscr{A}/\mathscr{C})$ que $H(\mathscr{A}/\mathscr{C})=H(\mathscr{A})$. Para probar el recíproco supongamos que $H(\mathscr{A}/\mathscr{C})=H(\mathscr{A})$. Entonces

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\mu(A_i\cap C_j)\log\left(\frac{\mu(A_i\cap C_j)}{\mu(C_j)}\right)=-\sum_{i=1}^{n}\mu(A_i)\log(\mu(A_i))(*).$$

Si fijamos i y aplicamos el Teorema 4.1 con $\alpha_j = \mu(C_j)$ y $x_j = \mu(A_j \cap C_j)/\mu(C_j)$ obtenemos que

$$-\sum_{j=1}^{m} \mu(A_i \cap C_j) \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right) \leq -\mu(A_i) \log(\mu(A_i))(^{**}),$$

donde sólo se da la igualdad cuando $\mu(A_i \cap C_j)/\mu(C_j)$ no depende de j. Si a_i denota esta constante, entonces sumando las relaciones anteriores tenemos que $\mu(A_i \cap C_j) = a_i\mu(C_j)$ sobre las j tenemos que $a_i = \mu(A_i)$. Así pues la igualdad se da solamente cuando $\mu(A_i \cap C_j) = \mu(A_i)\mu(C_j)$. Como la relación (*) dice que la igualdad se da en la desigualdad (**), para cada i tenemos que $\mu(A_i \cap C_j) = \mu(A_i)\mu(C_j)$ para toda i, j. Por lo tanto $\mu(A \cap C) = \mu(A)\mu(C)$ siempre que $A \in \mathscr{A}$, $C \in \mathscr{C}$. □

Observación 4.8. El inciso (i) del teorema anterior se puede interpretar de la siguiente manera: conociendo la información del experimento $\mathscr{C}$ podemos conocer la información del experimento $\mathscr{A}$. En otras palabras, realizando el experimento $\mathscr{C}$ ya no hay incertidumbre acerca del resultado del experimento $\mathscr{A}$.

El inciso (ii) del teorema dice que la información aportada por el experimento $\mathscr{C}$ no da información acerca del experimento $\mathscr{A}$, siempre y cuando $\mathscr{A}$ y $\mathscr{C}$ sean independientes. Recordando la Definición 3.1 y su interpretación, la propiedad anterior es una de las propiedades esperadas para una medida de incertidumbre. Es decir, si la correncia de $A \in \mathscr{A}$ no depende de la ocuurencia de $C \in \mathscr{C}$ para todo $A \in \mathscr{A}$, $C \in \mathscr{C}$; entonces la cantidad de incertidumbre contenida en $\mathscr{A}$, no debe verse alterada por el conocimiento de la ocurrencia de $\mathscr{C}$.

Con las propiedades y resultados anteriores ya podemos decir que la definición de entropía es una buena medida de incertidumbre. Ahora pasamos a ver un espacio donde la entropía nos dará una métrica.

Teorema 4.4. *Denotemos como ∇ al espacio de todas las sub-σ-álgebras finitas de $\mathscr{F}$, en el que dos álgebras finitas $\mathscr{A}$, $\mathscr{C}$ son idénticas si $\mathscr{A} \overset{\circ}{=} \mathscr{C}$. Si definimos*

$$d(\mathscr{A}, \mathscr{C}) = H(\mathscr{A}/\mathscr{C}) + H(\mathscr{C}/\mathscr{A}),$$

entonces d es una métrica en ∇.

Demostración. Tenemos que $d(\mathscr{A}, \mathscr{C}) \geq 0$, donde la igualdad se da si y sólo si $\mathscr{A} \overset{\circ}{=} \mathscr{C}$ (Teorema 4.3). También

$$H(\mathscr{A}/\mathscr{D}) \leq H(\mathscr{A} \vee \mathscr{C}/\mathscr{D}) = H(\mathscr{C}/\mathscr{D}) + H(\mathscr{A}/\mathscr{C} \vee \mathscr{D}) \leq H(\mathscr{C}/\mathscr{D}) + H(\mathscr{A}/\mathscr{C}),$$

de manera análoga $H(\mathscr{D}/\mathscr{A}) \leq H(\mathscr{C}/\mathscr{A}) + H(\mathscr{D}/\mathscr{C})$. Por lo tanto:

$$d(\mathscr{A}, \mathscr{D}) \leq d(\mathscr{A}, \mathscr{C}) + d(\mathscr{C}, \mathscr{D}).$$ □

También podemos definir la entropía condicional $H(\mathscr{A}/\mathscr{B})$ cuando $\mathscr{A}$ es una sub-σ-álgebra finita de $\mathscr{F}$, y $\mathscr{B}$ es una sub-σ-álgebra de $\mathscr{F}$ arbitraria. Para realizar lo utilizamos la función esperanza condicional $E(\cdot/\mathscr{B}) : L_1(\Omega, \mathscr{F}, \mu) \to L_1(\Omega, \mathscr{B}, \mu)$. Si $\mathscr{C}$ es una sub-σ-álgebra finita de $\mathscr{F}$ con $\xi(\mathscr{C}) = \{C_1, \ldots, C_m\}$ entonces:

$$E(f/\mathscr{C})(x) = \sum_{i=1}^{m} \chi_{C_i}(x) \frac{1}{\mu(C_i)} \int_{C_i} f d\mu.$$

Si $\mathscr{A}$ es otra sub-σ-álgebra finita y $\xi(\mathscr{A}) = \{A_1, \ldots, A_k\}$ entonces:

$$\begin{aligned} H(\mathscr{A}/\mathscr{C}) &= -\sum_{i,j} \mu(A_i \cap C_j) \log\left(\frac{\mu(A_i \cap C_j)}{\mu(C_j)}\right) \\ &= -\sum_{i=i}^{k} \int \chi_{A_i} \log\left(E\left(\chi_{A_i}/\mathscr{C}\right)\right) d\mu \\ &= -\int \sum_{i=1}^{k} E\left(\chi_{A_i}/\mathscr{C}\right) \log\left(E\left(\chi_{A_i}/\mathscr{C}\right)\right) d\mu. \end{aligned}$$

Lo cual nos lleva a la siguiente definición:

Definición 4.8. Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad; si $\mathscr{A}$ es una sub-σ-álgebra finita de $\mathscr{F}$ con $\xi(\mathscr{A}) = \{A_1, \ldots, A_k\}$, y $\mathscr{B}$ es una sub-σ-álgebra arbitraria de $\mathscr{F}$, la **entropía de $\mathscr{A}$ dado $\mathscr{B}$** es el número

$$H(\mathscr{A}/\mathscr{B}) = -\int \sum_{i=1}^{k} E(\chi_{A_i}/\mathscr{B}) \log(E(\chi_{A_i}/\mathscr{B}))\, d\mu.$$

Observación 4.9. Como $E(\cdot/\mathscr{B})$ es un operador lineal positivo y $\sum_{i=1}^{k} \chi_{A_i} = 1$, obtenemos que $0 \leq E(\chi_{A_i}/\mathscr{B})(x) \leq 1$ casi dondequiera, por lo que:

$$-\sum_{i=1}^{k} E(\chi_{A_i}/\mathscr{B})(x) \log(E(\chi_{A_i}/\mathscr{B})(x)) \leq k \max_{t \in [0,1]} (-t \log(t)) = ke.$$

Por lo tanto $H(\mathscr{A}/\mathscr{B})$ es finita.

La función $\sum_{i=1}^{k} E(\chi_{A_i}/\mathscr{B}) \log(E(\chi_{A_i}/\mathscr{B}))$ se conoce como **función información** en teoría de la información. En esta teoría la entropía, por lo observado anteriormente, sería la esperanza de la función información. Lo anterior justifica por qué podemos interpretar a la entropía como una medida de la información contenida en el resultado de un experimento.

La entropía que utilizamos en este trabajo es conocida como entropía de Shannon y tiene que ver con un problema en esta teoría. Podemos mostrar que las propiedades listadas en el Teorema 4.2 se cumplen para esta definición más general de la entropía condicional. Sin embargo podemos deducir del Teorema 4.2 (en el caso de que $(\Omega, \mathscr{F}, \mu)$ tenga una base numerable) utilizando un teorema límite. Para ello necesitaremos la siguiente definición: Si $\{\mathscr{B}_n\}$ es una familia de sub-σ-álgebras de $\mathscr{F}$, $\bigvee_{n=1}^{\infty} \mathscr{B}_n$ denotará la sub-σ-álgebra más pequeña que contiene a todas las $\mathscr{B}_n$.

Lema 4.1. *Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad y $\{\mathscr{B}_n\}$ una sucesión creciente de sub-σ-álgebras de $\mathscr{F}$. Denotamos $\bigvee_{n=1}^{\infty} \mathscr{B}_n$ por $\mathscr{B}$. Para cada $f \in L_2(\Omega, \mathscr{F}, \mu)$ se tiene que*

$$\|E(f/\mathscr{B}_n) - E(f/\mathscr{B})\|_2 \to 0.$$

Demostración. Recordemos que $E(\cdot/\mathscr{B}_n)$ es la proyección ortogonal de $L_2(\Omega, \mathscr{F}, \mu)$ sobre $L_2(\Omega, \mathscr{B}_n, \mu)$ (ver Shiryaev [21, pág. 237]). Sea $A \in \mathscr{B}$, como $\{\mathscr{B}_n\}$ es una sucesión creciente, se tiene que $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathscr{B}_n$ es un álgebra que genera a $\mathscr{B}$. Por el Lema B.1 existe una sucesión $\{A_n\}$ tal que $A_n \in \mathscr{B}_n$ y $\mu(A_n \triangle A) \to 0$. Como $E(\chi_A/\mathscr{B}_n)$ es la función más cercana en $L_2(\Omega, \mathscr{B}_n, \mu)$ a χ_A tenemos que

$$\|E(\chi_A/\mathscr{B}_n) - \chi_A\|_2^2 \leq \|\chi_{A_n} - \chi_A\|_2^2 = \mu(A_n \triangle A) \to 0.$$

Como las combinaciones lineales finitas de funciones características son densas en $L_2(\Omega, \mathscr{F}, \mu)$ tenemos que $\|E(h/\mathscr{B}_n) - h\|_2 \to 0$ para toda $h \in L_2(\Omega, \mathscr{F}, \mu)$. Por lo tanto, si $f \in L_2(\Omega, \mathscr{F}, \mu)$, sea $h = E(f/\mathscr{B})$ por lo anterior se tiene que $\|E(h/\mathscr{B}_n) - h\|_2 \to 0$, pero:

$$\|E(h/\mathscr{B}_n) - h\|_2 = \|E(E(f/\mathscr{B})/\mathscr{B}_n) - E(f/\mathscr{B})\|_2 = \|E(f/\mathscr{B}_n) - E(f/\mathscr{B})\|_2,$$

donde la última igualdad es consecuencia de que $E(E(f/\mathscr{B})/\mathscr{B}_n) = E(f/\mathscr{B}_n)$. □

Observación 4.10. El mismo resultado también es válido para una sucesión decreciente $\{\mathscr{B}_n\}$ de sub-σ-álgebras con $\bigcap_{n=1}^{\infty} B_n = \mathscr{B}$.

Teorema 4.5. *Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad. Sea $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$, y $\{\mathscr{B}_n\}_{n=1}^{\infty}$ una sucesión creciente de sub-σ-algebras de $\mathscr{F}$ con $\bigvee_{n=1}^{\infty} \mathscr{B}_n = \mathscr{B}$. Entonces $H(\mathscr{A}/\mathscr{B}_n) \to H(\mathscr{A}/\mathscr{B})$ (el análogo para particiones finitas también es cierto).*

Demostración. Sea $\xi(\mathscr{A}) = \{A_1, \ldots, A_m\}$. Por el lema anterior sabemos que para cada i

$$\|E(\chi_{A_i}/\mathscr{B}_n) - E(\chi_{A_i}/\mathscr{B})\|_2 \to 0.$$

Por lo tanto $E(\chi_{A_i}/\mathscr{B}_n)$ converge en medida a $E(\chi_{A_i}/\mathscr{B})$ (ver Bartle [3, pág. 69]), así que,

$$-\sum_{i=1}^{m} E(\chi_{A_i}/\mathscr{B}_n) \log(E(\chi_{A_i}/\mathscr{B}_n)),$$

converge en medida a $-\sum_{i=1}^{m} E(\chi_{A_i}/\mathscr{B}) \log(E(\chi_{A_i}/\mathscr{B}))$. Ahora como todas estas funciones están acotadas por ke debido al Teorema de Convergencia Dominada en Medida (ver Teorema B.9, pág. 135), se tiene que éstas también convergen en $L_1(\mu)$. Por lo tanto $H(\mathscr{A}/\mathscr{B}_n) \to H(\mathscr{A}/\mathscr{B})$. □

Observaciones 4.11.

(1) El mismo resultado también es válido para una sucesión decreciente $\{\mathscr{B}_n\}$ de sub-σ-álgebras con $\bigcap_{n=1}^{\infty} \mathscr{B}_n = \mathscr{B}$.

(2) Cuando $(\Omega, \mathscr{F}, \mu)$ tiene base numerable (ver Observación 1.29, pág. 51), los resultados del Teorema 4.2 son ciertos para una sub-σ-álgebra $\mathscr{D}$ de $\mathscr{F}$ de la siguiente manera: elegimos una sucesión creciente $\{\mathscr{D}_n\}_{n=1}^{\infty}$ de sub-σ-álgebras finitas tales que $\mathscr{D}_n \uparrow \mathscr{D}$ y utilizando el teorema anterior. (Notemos que $\mathscr{D}_n = \sigma(E_1, \ldots, E_n)$ donde $\{E_n\}$ es la base numerable del espacio).

Ahora tenemos la siguiente extensión del Teorema 4.3.

Teorema 4.6. *Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad y sean $\mathscr{A}$, $\mathscr{B}$ sub-σ-álgebras de $\mathscr{F}$ con $\mathscr{A}$ finita. Entonces:*

(i) $H(\mathscr{A}/\mathscr{B}) = 0$ si y sólo si $\mathscr{A} \overset{\circ}{\subset} \mathscr{B}$.

(ii) $H(A/\mathscr{B}) = H(\mathscr{A})$ si y sólo si $\mathscr{A}$ y $\mathscr{B}$ son independientes.

Demostración. Sea $\xi(\mathscr{A}) = \{A_1, \ldots, A_m\}$.

(i). Si $\mathscr{A} \overset{\circ}{\subset} \mathscr{B}$ entonces $E(\chi_{A_i}/\mathscr{B})(x)$ sólo toma los valores 0,1 c. d., por lo que $H(\mathscr{A}/\mathscr{B}) = 0$. Para el recíproco, si $0 = H(\mathscr{A}/\mathscr{B}) = \int(-\sum_{i=1}^{m} E(\chi_{A_i}/\mathscr{B}) \log(E(\chi_{A_i}/\mathscr{B}))\, d\mu$, entonces como $-E(\chi_{A_i}/\mathscr{B})(x) \log(E(\chi_{A_i}/\mathscr{B})(x)) \geq 0$ tenemos que para cada i, $E(\chi_{A_i}/\mathscr{B})$ toma sólo los valores 0,1 c. d. Por lo tanto $\mathscr{A} \overset{\circ}{\subset} \mathscr{B}$.

(ii). Supongamos que $H(\mathscr{A}/\mathscr{B}) = H(\mathscr{A})$. Sea $B \in \mathscr{B}$ y $\mathscr{C}$ una sub-σ-álgebra finita de $\mathscr{B}$ formada por los conjuntos $\{\emptyset, B, \Omega \setminus B, \Omega\}$, entonces $H(\mathscr{A}) \geq H(\mathscr{A}/\mathscr{C}) \geq H(\mathscr{A}/\mathscr{B}) = H(\mathscr{A})$. Así pues, $H(\mathscr{A}) = H(\mathscr{A}/\mathscr{C})$, por lo que $\mu(A \cap B) = \mu(A)\mu(B)$ para todo $A \in \mathscr{A}$, entonces $\mathscr{A}$ y $\mathscr{B}$ son independientes.

Si $\mathscr{A}$ y $\mathscr{B}$ son independientes entonces para cada $A \in \mathscr{A}$ $E(\chi_A/\mathscr{B}) = \mu(A)$ (ya que $\int_B E(\chi_A/\mathscr{B})\, d\mu = \int_B \chi_A = \mu(A)\mu(B)$ para toda $B \in \mathscr{B}$ y $E(\chi_A/\mathscr{B})$ es la única función $\mathscr{B}$-medible con esta propiedad). Por lo tanto $H(\mathscr{A}/\mathscr{B}) = H(\mathscr{A})$. □

4.4. Entropía de una Transformación Preservadora de la Medida

Recordando que si $\xi(\mathscr{A}) = \{A_1, \ldots, A_m\}$, entonces

$$H(\xi(\mathscr{A})) = H(\mathscr{A}) = -\sum_{i=1}^{m} \mu(A_i) \log(\mu(A_i)),$$

y que los elementos de la partición $\xi(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})) = \bigvee_{i=0}^{n-1} T^{-i}\xi(\mathscr{A})$ son todos los conjuntos de la forma $\bigcap_{i=0}^{n-1} T^{-i}(A_{j_i})$ con $j_i \in \{1, \ldots, m\}$ para $i \in \{0, \ldots, n-1\}$.

Definición 4.9. Supongamos que $T : \Omega \to \Omega$ es una transformación medible que preserva la medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Si $\mathscr{A}$ es una sub-σ-álgebra finita de $\mathscr{F}$ entonces

$$h(T,\, \xi(\mathscr{A})) = h(T,\, \mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right),$$

le llamaremos **la entropía de T con respecto a $\mathscr{A}$**.

Al número $h(T, \mathscr{A})$ lo podemos interpretar como sigue: si al aplicar T lo asociamos con el paso de un día, es decir, al aplicar n veces T, entonces $\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})$ representaría la información obtenida después de realizar el experimento asociado a la partición $\mathscr{A}$, por n días consecutivos. Por lo que $h(T, \mathscr{A})$ es la información promedio por día que se obtiene de realizar el experimento $\mathscr{A}$ por siempre. Finalmente no queda claro que este límite siempre exista. Más adelante veremos que este límite siempre existe.

Observación 4.12. $h(T,\, \mathscr{A}) \geq 0$.

Finalmente estamos en posición de dar la definición de la entropía de una transformación preservadora de medida.

Definición 4.10. Si $T : \Omega \to \Omega$ es una transformación medible que preserva la medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$, entonces se define: $h(T) = \sup h(T,\, \mathscr{A})$; donde el supremo se toma sobre todas las sub-σ-álgebras finitas $\mathscr{A}$ de $\mathscr{F}$ le llamaremos la **entropía de** T. De la misma manera $h(T) = \sup h(T, \xi)$, donde el supremo se toma sobre todas las particiones finitas de $(\Omega, \mathscr{F}, \mu)$.

Si pensamos a una aplicación de T como el paso de un día de tiempo (como lo hemos estado haciendo), entonces $h(T)$ es el "máximo" de la información promedio obtenida de realizar el mismo experimento por siempre.

Observaciones 4.13.

(1) $h(T) \geq 0$. $h(T)$ puede ser $+\infty$.

(2) $h(I_x) = 0$. Si $h(T) = 0$, entonces $h(T,\, \mathscr{A}) = 0$ para toda sub-σ-álgebra $\mathscr{A}$ finita, lo cual implica que $\bigvee_{i=0}^{\infty} T^{-i}(\mathscr{A})$ no "cambia mucho" cuando $n \to \infty$.

Ahora se prueba la existencia de $h(T, \mathscr{A})$. Esto lo haremos de dos formas, una utilizando un resultado de sucesiones de números reales, y la otra utilizando propiedades de la entropía condicional.

Lema 4.2. *Si $\{a_n\}_{n\geq 1}$ es una sucesión de números reales tal que $a_{n+p} \leq a_n + a_p$ para toda $n\,, p$, entonces $\lim_{n\to\infty} a_n/n$ existe y es igual al $\inf_n\{a_n/n\}$. (El límite puede ser $-\infty$, pero si a_n está acotada inferiormente entonces el límite será no negativo).*

Demostración. Fijamos $p > 0$. Si $n \geq p$ entonces podemos escribir $n = kp + i$ con $0 \leq i < p$. Entonces

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

Si $n \to \infty$ entonces $k \to \infty$, por lo que

$$\limsup \frac{a_n}{n} \leq \frac{a_p}{p},$$

así que

$$\limsup \frac{a_n}{n} \leq \inf \left\{ \frac{a_p}{p} \right\}.$$

Pero

$$\inf \left\{ \frac{a_p}{p} \right\} \leq \liminf \frac{a_n}{n},$$

por lo tanto $\lim a_n/n$ existe y es igual al $\inf\{a_n/n\}$. □

Teorema 4.7. *Sea $T : \Omega \to \Omega$ medible que preserva la medida y $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$, entonces $\lim_{n\to\infty}(1/n)H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)$ existe.*

Demostración. Sea $a_n = H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) \geq 0$, entonces

$$\begin{aligned}
a_{n+p} &= H\left(\bigvee_{i=0}^{n+p-1} T^{-i}(\mathscr{A})\right) \\
&\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) + H\left(\bigvee_{i=n}^{n+p-1} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.2(viii)} \\
&= a_n + H\left(\bigvee_{i=0}^{p-1} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.2(x)} \\
&= a_n + a_p.
\end{aligned}$$

Aplicando el lema anterior a la sucesión a_n obtenemos el resultado. □

Teorema 4.8. *Sea $T : \Omega \to \Omega$ una transformación preservadora de medida y $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$, entonces $(1/n)H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)$ **decrece** a $h(T, \mathscr{A})$.*

Demostración. Primero tenemos que mostrar (por inducción) que

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) = H(\mathscr{A}) + \sum_{j=1}^{n-1} H\left(\mathscr{A} \Big/ \bigvee_{i=1}^{j} T^{-i}(\mathscr{A})\right).$$

Para $n = 1$ es claro. Supongamos válido para $n = k$, entonces es válido para $n = k + 1$ ya que

$$\begin{aligned}
H\left(\bigvee_{i=0}^{p} T^{-i}(\mathscr{A})\right) &= H\left(\bigvee_{i=1}^{p} T^{-i}(\mathscr{A} \vee \mathscr{A})\right) \\
&= H\left(\bigvee_{i=1}^{p} T^{-i}(\mathscr{A})\right) + H\left(\mathscr{A} \Big/ \bigvee_{i=1}^{p} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.2(ii)} \\
&= H\left(\bigvee_{i=0}^{p-1} T^{-i}(\mathscr{A})\right) + H\left(\mathscr{A} \Big/ \bigvee_{i=1}^{p} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.2(x)} \\
&= H(\mathscr{A}) + \sum_{j=1}^{p} H\left(\mathscr{A} \Big/ \bigvee_{i=1}^{j} T^{-i}(\mathscr{A})\right) \quad \text{por hipótesis de inducción.}
\end{aligned}$$

Por lo tanto la fórmula anterior es válida para toda n. Utilizando la fórmula y el Teorema 4.2(iii) tenemos que $H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) \geq nH(\mathscr{A}/\bigvee_{i=1}^{n} T^{-i}(\mathscr{A}))$, por lo que

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

Así

$$\frac{1}{n+1}H\left(\bigvee_{i=0}^{n} T^{-i}(\mathscr{A})\right) \leq \frac{1}{n}H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right). \qquad \square$$

Ahora mostraremos que la entropía es una propiedad invariante bajo isomorfismo.

Teorema 4.9. *La entropía es invariante bajo isomorfismo.*

Demostración. Sean T_i $(i = 1, 2)$ transformaciones preservadoras de medida isomorfas actuando sobre espacios de probabilidad $(\Omega_i, \mathscr{F}_i, \mu_i)$ $(i = 1, 2)$ (ver Definición 2.3, pág. 64). Entonces existen M_i $(i = 1, 2)$ con $\mu(M_i) = 1$ $i = 1, 2$ y $\phi : M_1 \to M_2$ invertible, medible y preservadora de medidas. Ya que los conjuntos de medida cero no son relevantes para el cálculo de la entropía de una sub-σ-álgebra, podemos considerar a los espacios $(\Omega_i, \mathscr{F}_i, \mu_i)$ y los espacios $(M_i, \mathscr{F}_i \cap M_i, \mu_i|_{M_i})$ $i = 1, 2$ respectivamente como iguales, y así lo haremos.

Sea $\mathscr{A}_1$ una sub-σ-álgebra finita de $\mathscr{F}_1$ con $\xi(\mathscr{A}_1) = \{A_1, \ldots, A_r\}$. Sean $B_i = \phi(A_i)$ para toda $i \in \{1, \ldots, r\}$, entonces $\eta = \{B_1, \ldots, B_r\}$ es una partición de $(\Omega_2, \mathscr{F}_2, \mu_2)$. Sea $\mathscr{B}_2 = \sigma(\eta)$. Así pues $\bigcap_{i=0}^{n-1} T_2^{-i}(B_{q_i})$, donde $q_i \in \{1, \ldots, r\}$ tiene la misma medida que $\bigcap_{i=0}^{n-1} T_1^{-i}(A_{q_i})$ ya que

$$\mu_1(\bigcap_{i=0}^{n-1} T_1^{-i}(A_{q_i})) = \mu_1(\bigcap_{i=0}^{n-1} T_1^{-i}(\phi^{-1}(B_{q_i}))) = \mu_1(\bigcap_{i=0}^{n-1} \phi^{-1}(T_2^{-i}(B_{q_i}))) =$$

$$\mu_1(\phi^{-1}(\bigcap_{i=0}^{n-1} T_2^{-i}(B_{q_i}))) = \mu_2(\bigcap_{i=0}^{n-1} T_2^{-i}(B_{q_i})).$$

Por lo que $H(\bigvee_{i=0}^{n-1} T_1^{-i}(\mathscr{A}_1)) = H(\bigvee_{i=0}^{n-1} T_2^{-i}(\mathscr{B}_2))$, así que $h(T_1,\mathscr{A}_1) = h(T_2,\mathscr{B}_2)$, entonces $h(T_1) \leq h(T_2)$. Análogamente obtenemos $h(T_1) = h(T_2)$. □

Después de ver propiedades de $h(T,\mathscr{A})$ y $h(T)$, consideraremos el problema de cómo calcular $h(T)$. Estos cálculos y el teorema anterior darán ejemplos de transformaciones que no son isomorfas.

4.5. Propiedades de $h(T,\mathscr{A})$ y $h(T)$

Recordando que

$$h(T,\mathscr{A}) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right).$$

Teorema 4.10. *Supongamos que $\mathscr{A}$, $\mathscr{B}$ son sub-σ-álgebras finitas de $\mathscr{F}$ y T una transformación preservadora de medida del espacio de probabilidad $(\Omega,\mathscr{F},\mu)$. Entonces:*

(i) $h(T,\mathscr{A}) \leq H(\mathscr{A})$.

(ii) $h(T,\mathscr{A}\vee\mathscr{B}) \leq h(T,\mathscr{A}) + h(T,\mathscr{B})$.

(iii) Si $\mathscr{A}\subseteq\mathscr{B}$, *entonces* $h(T,\mathscr{A}) \leq h(T,\mathscr{B})$.

(iv) $h(T,\mathscr{A}) \leq h(T,\mathscr{B}) + H(\mathscr{A}/\mathscr{B})$.

(v) $h(T,T^{-1}(\mathscr{A})) = h(T,\mathscr{A})$.

(vi) Si $k \geq 1$, $h(T,\mathscr{A}) = h(T,\bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A}))$.

(vii) Si T es invertible y $k \geq 1$, entonces

$$h(T,\mathscr{A}) = h\left(T, \bigvee_{i=-k}^{k} T^{i}(\mathscr{A})\right).$$

Demostración.

(i)

$$\begin{aligned}
\frac{1}{n}H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) &\leq \frac{1}{n}\sum_{i=0}^{n-1} H(T^{-i}(\mathscr{A})) \quad \text{Teorema 4.2}\\
&= \frac{1}{n}\sum_{i=0}^{n-1} H(\mathscr{A}) \quad \text{Teorema 4.2(x)}\\
&= H(\mathscr{A}).
\end{aligned}$$

(ii)

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}\vee\mathscr{B})\right) &= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\vee\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)\\ &\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)+H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right) \quad \text{Teorema 4.2(viii).}\end{aligned}$$

(iii) Si $\mathscr{A}\subseteq\mathscr{B}$ entonces

$$\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\subseteq\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B}),\quad n\geq 1,$$

utilizando el Teorema 4.2(iv) obtenemos el resultado.

(iv)

$$\begin{aligned} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) &\leq H\left(\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)\vee\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)\right) \quad \text{Teorema 4.2(iv)}\\ &= H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)+H\left(\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}\right)\Big/\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)\right) \quad \text{Teorema 4.2(ii).}\end{aligned}$$

Pero por el Teorema 4.2(vii)

$$\begin{aligned} H\left(\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)\Big/\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)\right) &\leq \sum_{i=0}^{n-1} H\left(T^{-i}(\mathscr{A})\Big/\left(\bigvee_{j=0}^{n-1} T^{-j}(\mathscr{B})\right)\right)\\ &\leq \sum_{i=0}^{n-1} H(T^{-i}(\mathscr{A})/T^{-i}(\mathscr{B})) \quad \text{Teorema 4.2(v)}\\ &= nH(\mathscr{A}/\mathscr{B}) \quad \text{Teorema 4.2(ix).}\end{aligned}$$

Por lo tanto,

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right)\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{B})\right)+nH(\mathscr{A}/\mathscr{B}).$$

(v)

$$H\left(\bigvee_{i=1}^{n} T^{-i}(\mathscr{A})\right)=H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.2(x), por lo tanto}$$
$$h(T,T^{-1}(\mathscr{A}))=h(T,\mathscr{A}).$$

(vi)

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

(vii)

$$\begin{aligned}
h\left(T, \bigvee_{i=-k}^{k} T^{-i}(\mathscr{A})\right) &= h\left(T, \bigvee_{0}^{2k} T^{-i}(\mathscr{A})\right) \quad \text{por (v)} \\
&= h(T,\mathscr{A}) \quad \text{(vi)}.
\end{aligned}$$

□

Corolario 4.10.1. *Si $\mathscr{A}$, $\mathscr{B}$ son sub-σ-álgebras finitas de $\mathscr{F}$, tenemos que*

$$|h(T,\mathscr{A}) - h(T,\mathscr{B})| \leq d(\mathscr{A},\mathscr{B}), \quad \textit{por lo que,} \quad h(T,\cdot)$$

es una función continua del espacio métrico (∇, d) a $\mathbb{R}$.

Demostración. Por el inciso (iii)

$$|h(T,\mathscr{A}) - h(T,\mathscr{B})| \leq \text{máx}(H(\mathscr{A}/\mathscr{B}), H(\mathscr{B}/\mathscr{A})) \leq d(\mathscr{A},\mathscr{B}).$$ □

Teorema 4.11. *Sea T una transformación medible que preserva la medida del espacio de probabilidad $(\Omega,\mathscr{F},\mu)$, entonces:*

(i) Para $k > 0, h(T^k) = kh(T)$.

(ii) Si T es invertible entonces $h(T^k) = |k|h(T)$ para toda $k \in \mathbb{Z}$.

Demostración.

(i) Mostraremos primero que

$$h(T^k, \bigvee_{i=0}^{k-1} T^{-i}(\mathscr{A})) = kh(T,\mathscr{A}), \quad \text{si } k > 0.$$

Lo anterior es válido puesto que

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{j=0}^{k-1}T^{-kj}\left(\bigvee_{i=0}^{k-1}T^{-i}(\mathscr{A})\right)\right) &= \lim_{n\to\infty}\frac{k}{nk}H\left(\bigvee_{i=0}^{nk-1}T^{-i}(\mathscr{A})\right)\\ &= kh(T,\mathscr{A}).\end{aligned}$$

Por lo tanto,

$$\begin{aligned}kh(T) &= k\cdot\sup_{\mathscr{A}\text{ finita}} h(T,\mathscr{A}) = \sup_{\mathscr{A}} h\left(T^k,\bigvee_{i=0}^{k-1}T^{-i}(\mathscr{A})\right)\\ &\le \sup_{\mathscr{B}} h(T^k,\mathscr{B}) = h(T^k).\end{aligned}$$

También, $h(T^k,\mathscr{A}) \le h(T^k,\bigvee_{i=0}^{k-1}T^{-i}(\mathscr{A})) = kh(T,\mathscr{A})$, por el Teorema 1.7. Por lo tanto, $h(T^k)\le kh(T)$. Utilizando las dos desigualdades obtenemos el resultado.

(*ii*) Basta mostrar que $h(T^{-1}) = h(T)$, y todo lo que necesitamos probar es $h(T^{-1},\mathscr{A}) = h(T,\mathscr{A})$ para toda $\mathscr{A}$ finita. Pero

$$\begin{aligned}H\left(\bigvee_{i=0}^{n-1}T^{-i}(\mathscr{A})\right) &= H\left(T^{-(n-1)}\left(\bigvee_{i=0}^{n-1}T^{i}(\mathscr{A})\right)\right) \quad \text{Teorema 4.2(x)}\\ &= H\left(\bigvee_{j=0}^{n-1}T^{-j}(\mathscr{A})\right). \qquad \square\end{aligned}$$

Con el objetivo de mostrar como $h(T)$ se comporta bien bajo algunas operaciones en las transformaciones preservadora de medida, daré formas para que el cálculo de $h(T)$ sea más simple. Los siguientes resultados ayudarán a comprender bajo qué condiciones $h(T,\mathscr{A}) = 0$. Lo que permitirá concluir que una transformación que no es invertible módulo cero (es decir $T^{-1}(\mathscr{F}) \overset{\circ}{\neq} \mathscr{F}$), debe tener entropía estrictamente positiva.

Teorema 4.12. *Si $\mathscr{A}$ es una sub-σ-álgebra de $\mathscr{F}$ y T es m. y p.m. de $(\Omega,\mathscr{F},\mu)$, entonces*

$$h(T,\mathscr{A}) = \lim_{n\to\infty} H\left(\mathscr{A}\Big/\left(\bigvee_{i=1}^{n}T^{-i}(\mathscr{A})\right)\right) = H\left(\mathscr{A}\Big/\left(\bigvee_{i=1}^{\infty}T^{-i}(\mathscr{A})\right)\right).$$

Demostración. El límite existe ya que el lado derecho es una sucesión decreciente por el Teorema 4.2(v). Sabemos de la prueba del Teorema 4.8 que para $n\ge 1$ se tiene que

$$H\left(\bigvee_{i=0}^{n-1}T^{-i}(\mathscr{A})\right) = H(\mathscr{A}) + \sum_{j=1}^{n-1} H\left(\mathscr{A}\Big/\left(\bigvee_{i=1}^{j}T^{-i}(\mathscr{A})\right)\right).$$

El resultado se obtiene dividiendo entre n y tomando límites, ya que el límite de los promedios es igual al límite original si éste existe. La última igualdad se da debido al Teorema 4.5. $\square$

Estamos en posición de decir para cuales sub-σ-álgebras finitas se tiene $h(T, \mathscr{A}) = 0$.

Corolario 4.12.1. *Sea T una transformación preservadora de medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Sea $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$. Entonces $h(T, \mathscr{A}) = 0$ si y sólo si $\mathscr{A} \overset{\circ}{\subset} \bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A})$.*

Demostración. Por el teorema anterior y el Teorema 4.6. □

Intuitivamente, el resultado anterior dice que la información promedio por día de realizar el experimento $\mathscr{A}$, es cero cuando el resultado del primer día puede ser determinado por la combinación de resultados conocidos en todos los días subsecuentes (el futuro determina el presente).

Corolario 4.12.2. *Sea T una transformación preservadora de medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Entonces $h(T) = 0$ si y sólo si para toda sub-σ-álgebra finita $\mathscr{A}$ de $\mathscr{F}$ se tiene que $\mathscr{A} \overset{\circ}{\subset} \bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A})$.*

Ahora podemos concluir que una transformación preservadora de medida que no es invertible debe tener entropía mayor que cero.

Corolario 4.12.3. *Sea T una transformación preservadora de medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Si $h(T) = 0$, entonces $T^{-1}(\mathscr{F}) \overset{\circ}{=} \mathscr{F}$.*

Demostración. Sea $F \in \mathscr{F}$ y $\mathscr{A}$ la sub-σ-álgebra finita formada por $\{\emptyset, F, \Omega \backslash F, \Omega\}$. Por el corolario anterior tenemos que $\mathscr{A} \overset{\circ}{\subset} \bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A}) \subset T^{-1}(\mathscr{F})$. Como F es un elemento arbitrario de $\mathscr{F}$ se tiene que $\mathscr{F} \overset{\circ}{=} T^{-1}(\mathscr{F})$. □

Se le puede agregar aún más al resultado anterior.

Corolario 4.12.4. *Sea T una transformación preservadora de medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Supongamos que $h(T) = 0$. Si $\mathscr{A}$ es una sub-σ-álgebra de $\mathscr{F}$ tal que $T^{-1}(\mathscr{A}) \overset{\circ}{\subset} \mathscr{A}$, entonces $T^{-1}(\mathscr{A}) \overset{\circ}{=} \mathscr{A}$.*

Demostración. Aplicamos el corolario anterior a $T|_{(\Omega, \mathscr{F}, \mu)}$. □

4.6. Algunos métodos para calcular $h(T)$

Es difícil calcular la entropía de una transformación a partir de la definición ya que se debe calcular $h(T, \mathscr{A})$ para toda $\mathscr{A}$ sub-σ-álgebra finita, y en pocas ocasiones se conoce como son estas sub-σ-álgebras. Lo anterior lleva a buscar formas alternativas para el cálculo de $h(T)$. En esta sección encontraré condiciones para poder asegurar que $h(T,\mathscr{A}) = h(T)$ (y así simplificar el cálculo de $h(T)$). También mostraré bajo qué condiciones una sucesión de sub-σ-álgebras $\{\mathscr{A}_n\}$ finitas cumplen que $h(T) = \lim_{n\to\infty} h(T,\mathscr{A}_n)$. Lo anterior dará métodos para calcular $h(T)$ para algunos ejemplos de transformaciones preservadora de medida, y veré mas propiedades de $h(T)$. De igual manera probaré el teorema de Kolmogorov-Sinai, que es frecuentemente utilizado para el cálculo de $h(T)$, además de varios resultados que facilitarán la obtención de $h(T)$ para nuestros ejemplos.

El Teorema 4.13 será de utilidad para esta sección. Para demostrarlo necesitaremos el siguiente resultado:

Lema 4.3. *Sea $r \geq 1$ un entero fijo. Para toda $\varepsilon > 0$, existe $\delta = \delta(\varepsilon, r) > 0$ tal que si $\xi = \{A_1,\ldots,A_r\}, \eta = \{C_1,\ldots,C_r\}$ son cualesquiera dos particiones de $(\Omega,\mathscr{F},\mu)$ en r conjuntos con $\sum_{i=1}^r \mu(A_i \vartriangle C_i) < \delta$, entonces $H(\xi/\eta) + H(\eta/\xi) < \varepsilon$.*

Demostración. Sea $\varepsilon > 0$ dado. Escogemos $\delta > 0$ tal que $\delta < \frac{1}{3}$ y

$$-r(r-1)\delta\log(\delta) - (1-\delta)\log(1-\delta) < \varepsilon/2.$$

Donde esta δ existe por continuidad por la derecha de la función $x\log(x)$. Sea ζ la partición formada por los conjuntos $A_i \cap C_j$ $(i \neq j)$ y $\bigcup_{i=1}^r (A_i \cap C_i)$. Entonces $\xi \vee \eta = \eta \vee \zeta$, como $A_i \cap C_j \subset \bigcup_{n=1}^r (A_n \vartriangle C_n)$ $(i \neq j)$ tenemos que

$$\mu(A_i \cap C_j) < \delta \ (i \neq j) \text{ y } \mu\left(\bigcup_{i=1}^r (A_i \cap Ci)\right) > 1-\delta.$$

Así que $H(\zeta) < r(r-1)\delta\log(\delta) - (1-\delta)\log(1-\delta) < \varepsilon/2$. Por lo que:

$$H(\eta) + H(\xi/\eta) = H(\xi \vee \eta) = H(\eta \vee \zeta) \leq H(\eta) + H(\zeta) < H(\eta) + \varepsilon/2,$$

por lo tanto $H(\xi/\eta) < \varepsilon/2$. Análogamente (ya que $\xi \vee \eta = \xi \vee \zeta$) tenemos que $H(\eta/\xi) < \varepsilon/2$. □

Observación 4.14. Si $\widetilde{\nabla}_r$ denota el espacio de todas las particiones finitas ordenadas de $(\Omega,\mathscr{F},\mu)$ en r conjuntos con $r \geq 1$, donde dos particiones $\xi = \{A_1,\ldots,A_r\}$ $\eta = \{C_1,\ldots,C_r\}$, las consideraremos como iguales si $\mu(A_i \vartriangle C_i) = 0$ para toda i. Definimos el espacio métrico $(\widetilde{\nabla}_r, \rho)$ donde $\rho(\xi,\eta) = \sum_{i=0}^r \mu(A_i \vartriangle C_i)$ si $\xi = \{A_1,\ldots,A_r\}$ $\eta = \{C_1,\ldots,C_r\}$. Si (∇, d) denota el espacio métrico introducido en el Teorema 4.4 entonces el lema dice que la función identidad de $(\widetilde{\nabla}_r, \rho)$ a (∇, d) es uniformemente continua.

Teorema 4.13. *Sea $(\Omega,\mathscr{F},\mu)$ un espacio de probabilidad y $\mathscr{F}_0$ un álgebra tal que $\sigma(\mathscr{F}_0) \overset{\circ}{=} \mathscr{F}$. Sea $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$. Entonces para toda $\varepsilon > 0$ existe una sub-σ-álgebra finita de $\mathscr{B}, \mathscr{B} \subseteq \mathscr{F}_0$ tal que $H(\mathscr{B}/\mathscr{A}) + H(\mathscr{A}/\mathscr{B}) < \varepsilon$.*

Demostración. Sea $\xi(\mathscr{A}) = \{A_1, \ldots, A_r\}$, $\varepsilon > 0$. Elegimos $\delta > 0$ correspondiente a r y ε en el lema anterior. Es suficiente mostrar que para toda $\sigma > 0$ podemos encontrar una partición $\{C_1, \ldots, C_r\}$ con $C_i \in \mathscr{F}_0$ y $\mu(A_i \cap C_i) < \sigma$ para toda i. Para mostrar lo anterior elegimos $\lambda > 0$ tal que $\lambda(r-1)[1+r(r-1)] < \sigma$, para cada i elegimos $F_i \in \mathscr{F}_0$ tal que $\mu(A_i \vartriangle F_i) < \lambda$. Si $i \neq j$ entonces $F_i \cap F_j \subset (F_i \vartriangle A_i) \cup (F_j \vartriangle A_j)$ por lo cual $\mu(F_i \cap F_j) < 2\lambda$. Sea $N = \bigcup_{i\neq j}(F_i \cap F_j)$, tenemos que $\mu(N) < r(r-1)\lambda$, sean $C_i = F_i \setminus N$ para $1 \leq i < r$ y $C_r = \Omega \setminus \bigcup_{i=1}^{r-1} C_i$ es una partición de Ω, y cada $C_i \in \mathscr{F}_0$. Si $i < r$ entonces:

$$C_i \vartriangle A_i \subset (F_i \vartriangle A_i) \cup N, \quad \text{por lo que,}$$

$$\mu(C_i \vartriangle A_i) < \lambda[1 + r(r-1)] < \sigma.$$

Mientras, $C_r \vartriangle A_r \subset \bigcup_{i=1}^{r-1}(C_i \vartriangle A_i)$, por lo tanto

$$\mu(C_r \vartriangle A_r) < (r-1)\lambda[1 + r(r-1)] < \sigma. \qquad \square$$

Si $\{A_n\}$ es una sucesión de sub-σ-álgebras de $\mathscr{F}$, entonces $\bigvee_n \mathscr{A}_n$ denota la sub-σ-álgebra generada por $\{A_n\}$, es decir, $\bigvee_n \mathscr{A}_n$ es la intersección de todas las sub-σ-álgebras de $\mathscr{F}$ que contienen a cada $\mathscr{A}_n$. Usaremos $\bigcup_n \mathscr{A}_n$ para denotar la colección de conjuntos que pertenecen a algún A_n. Si $\{A_n\}_{n=1}^{\infty}$ es una sucesión **creciente** entonces $\bigcup_n \mathscr{A}_n$ es un álgebra, pero no necesariamente una σ-álgebra.

Corolario 4.13.1. *Si $\{\mathscr{A}_n\}$ es una sucesión creciente de sub-σ-álgebras finitas de $\mathscr{F}$ y $\mathscr{C}$ es una sub-σ-álgebra finita tal que $\mathscr{C} \overset{\circ}{\subset} \bigvee_n \mathscr{A}_n$, entonces $H(\mathscr{C}/\mathscr{A}_n) \to 0$ cuando $n \to \infty$.*

Demostración. Si $\mathscr{F}_0 = \bigcup_{j=1}^{\infty} \mathscr{A}_j$ entonces $\mathscr{F}_0$ es un álgebra y $\mathscr{C} \overset{\circ}{\subset} \sigma(\mathscr{F}_0)$ por hipótesis. Sea $\varepsilon > 0$, por el Teorema 4.13 existe una sub-σ-álgebra finita $\mathscr{B}_\varepsilon$ de $\mathscr{F}_0$ tal que $H(\mathscr{C}/\mathscr{B}_\varepsilon) < \varepsilon$. Pero $\mathscr{B}_\varepsilon \subseteq \mathscr{A}_{j_0}$ para algún j_0 ya que $\mathscr{B}_\varepsilon$ es finita. Si $j \geq j_0$ por el Teorema 4.2(v) $H(\mathscr{C}/\mathscr{A}_j) \leq H(\mathscr{C}/\mathscr{A}_{j_0}) \leq H(\mathscr{C}/\mathscr{B}_\varepsilon) < \varepsilon$. Por lo tanto $H(\mathscr{C}/\mathscr{A}_n) \to 0$. $\square$

Los principales métodos para calcular $h(T)$ están dados en los siguientes dos resultados.

Teorema 4.14 (Teorema de Kolmogorov-Sinai)**.** *Sea T una transformación invertible, medible que preserva la medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$, y sea $\mathscr{A}$ una sub-σ-álgebra finita de $\mathscr{F}$ tal que $\bigvee_{n=-\infty}^{\infty} T^n(\mathscr{A}) \overset{\circ}{=} \mathscr{F}$. Entonces $h(T) = h(T, \mathscr{A})$.*

Demostración. Sea $\mathscr{B} \subseteq \mathscr{F}$ finita, debemos mostrar que $h(T, \mathscr{B}) \leq h(T, \mathscr{A})$, para $n \geq 1$,

$$\begin{aligned} h(T,\mathscr{B}) &\leq h\left(T, \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) + H\left(\mathscr{B} \middle/ \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) && \text{Teorema 4.10(iv)} \\ &= h(T,\mathscr{A}) + H\left(\mathscr{B} \middle/ \bigvee_{i=-n}^{n} T^i(\mathscr{A})\right) && \text{Teorema 4.10(vii).} \end{aligned}$$

Sea $\mathscr{A}_n = \bigvee_{i=-n}^{n} T^i(\mathscr{A})$, es suficiente mostrar que $H(\mathscr{B}/\mathscr{A}_n)$ tiende a cero cuando $n \to \infty$, lo cual se cumple por el Corolario 4.13.1. $\square$

Un resultado similar también es cierto cuando T no necesariamente es invertible:

Teorema 4.15. *Sea T es una transformación preservadora de medida (pero no necesariamente invertible) del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$. Si $\mathscr{A}$ es una sub-σ-álgebra finita de $\mathscr{F}$ tal que $\bigvee_{i=0}^{\infty} T^{-i}(\mathscr{A}) \overset{\circ}{=} \mathscr{F}$, entonces $h(T) = h(T, \mathscr{A})$.*

Demostración. Esta prueba es análoga a la del teorema anterior reemplazando $\bigvee_{i=-n}^{n} T^{i}(\mathscr{A})$ por $\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})$ y utilizando el Teorema 4.10(vi). $\square$

El siguiente resultado es útil para mostrar que una transformación tiene entropía cero. Lo utilizaremos más adelante para mostrar que la entropía de una rotación es cero.

Corolario 4.15.1. *Si T es una transformación invertible y preservadora de medida del espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$, y $\bigvee_{i=0}^{\infty} T^{-i}(\mathscr{A}) \overset{\circ}{=} \mathscr{F}$ para alguna sub-σ-álgebra finita $\mathscr{A}$, entonces $h(T) = 0$.*

Demostración. Por el Teorema 4.15 tenemos que

$$\begin{aligned} h(T) &= h(T, \mathscr{A}) \\ &= \lim_{n\to\infty} H\left(\mathscr{A} \middle/ \bigvee_{i=1}^{n} T^{-i}(\mathscr{A})\right) \quad \text{Teorema 4.12.} \end{aligned}$$

Pero $\bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A}) \overset{\circ}{=} T^{-1}(\mathscr{F}) \overset{\circ}{=} \mathscr{F}$, sean $\mathscr{A}_n = \bigvee_{i=1}^{n} T^{-i}(\mathscr{A})$, entonces $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \cdots$ y $\bigvee_{n=1}^{\infty} \mathscr{A}_n \overset{\circ}{=} \mathscr{F}$. Por el Corolario 4.13.1 tenemos que $H(\mathscr{A}/\mathscr{A}_n) \to 0$, por lo tanto $h(T) = 0$. $\square$

Observación 4.15. La entropía también puede ser definida para particiones numerables como sigue: si $\xi = \{A_1, A_2, \dots\}$ entonces

$$H(\xi) = -\sum_i \mu(A_i) \log(A_i),$$

(la cual puede ser infinita). Se puede probar que $h(T) = \sup h(T, \xi)$, donde el supremo es tomado sobre las particiones numerables ξ tales que $H(\xi) < \infty$.

Definición 4.11. Una partición numerable ξ de Ω se llamará un **generador** para una transformación preservadora de medida invertible, si

$$\bigvee_{n=-\infty}^{\infty} T^n \sigma(\xi) \overset{\circ}{=} \mathscr{F},$$

se puede probar que si ξ es un generador con $H(\xi) < \infty$ entonces $h(T) = h(T, \xi)$ (ver Parry [16, pág. 16]).

Ahora probaremos algunos resultados que facilitarán el cálculo de $h(T)$ para nuestros ejemplos.

Teorema 4.16. *Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad. Si $\mathscr{F}_0$ es una sub-σ-álgebra de $\mathscr{F}$ tal que $\sigma(\mathscr{F}_0) \overset{\circ}{=} \mathscr{F}$, entonces para toda transformación T medible y preservadora de medida se tiene que*

$$h(T) = \sup h(T, \mathscr{A}),$$

donde el supremo se toma sobre todas las sub-σ-álgebras finitas $\mathscr{A}$ de $\mathscr{F}_0$.

Demostración. Sean $\varepsilon > 0$, $\mathscr{A} \subseteq \mathscr{F}$ finita. Por el Teorema 4.13 existe un álgebra finita $\mathscr{B}$ tal que $H(\mathscr{A}/\mathscr{B}_\varepsilon) < \varepsilon$, por lo que

$$\begin{aligned} h(T, \mathscr{A}) &\leq h(T, \mathscr{B}_\varepsilon) + H(\mathscr{A}/\mathscr{B}_\varepsilon) \quad \text{Teorema 4.10(iv)} \\ &\leq h(T, \mathscr{B}_\varepsilon) + \varepsilon. \end{aligned}$$

Por lo tanto $h(T, \mathscr{A}) \leq \varepsilon + \sup\{h(T, \mathscr{B}) : \mathscr{B} \subset \mathscr{F}_0, \mathscr{B}\text{finita}\}$, así que

$$h(T) \leq \sup\{h(T, \mathscr{B}) : \mathscr{B} \subset \mathscr{F}_0, \mathscr{B} \quad \text{finita}\}.$$

Es fácil ver la otra desigualdad. □

Cuando es difícil encontrar un generador para una transformación preservadora de medida el siguiente resultado es útil para calcular la entropía.

Teorema 4.17. *Sean $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad y $\{\mathscr{A}_n\}_{n=1}^{\infty}$ una sucesión creciente de sub-σ-álgebras finitas de $\mathscr{F}$ tal que $\bigvee_{n=1}^{\infty} \mathscr{A}_n \overset{\circ}{=} \mathscr{F}$. Si T es una transformación medible que preserva la medida, entonces $h(T) = \lim_{n\to\infty} h(T, \mathscr{A}_n)$.*

Demostración. Notemos que $h(T, \mathscr{A}_n)$ es una sucesión creciente por el Teorema 4.10(iii). También $\mathscr{F}_0 = \bigcup_{n=1}^{\infty} \mathscr{A}_n$ es un álgebra y $\sigma(\mathscr{F}_0) \overset{\circ}{=} \mathscr{F}$. Por el Teorema 4.16

$$h(T) = \sup\{h(T, \mathscr{B}) : \mathscr{B} \subset \mathscr{F}_0\,, \mathscr{B} \quad \text{finita}\},$$

si $\mathscr{B} \subset \mathscr{F}_0$ finita entonces $\mathscr{B} \subseteq \mathscr{A}_{n_0}$ para algún n_0. Por lo que

$$h(T, \mathscr{B}) \leq h(T, \mathscr{A}_{n_0}),$$

lo cual implica que $h(T) \leq \lim_{n\to\infty} h(T, \mathscr{A}_n)$, por lo tanto $h(T) = \lim_{n\to\infty} h(T, \mathscr{A}_n)$. □

El Teorema 4.16 nos permitirá probar la siguiente fórmula para el cálculo de la entropía del producto directo de dos transformaciones preservadoras de medida.

Teorema 4.18. *Sean $(\Omega_1, \mathscr{F}_1, \mu_1)$ y $(\Omega_2, \mathscr{F}_2, \mu_2)$ espacios de probabilidad, y T_1, T_2 transformaciones medibles que preservan la medida del espacio de probabilidad respectivamente. Entonces*

$$h(T_1 \times T_2) = h(T_1) + h(T_2).$$

Demostración. Si $\mathscr{A}_1 \subset \mathscr{F}_1$, $\mathscr{A}_2 \subset \mathscr{F}_2$ finitas, entonces $\mathscr{A}_1 \otimes \mathscr{A}_2$ es finita, donde

$$\xi(\mathscr{A}_1 \otimes \mathscr{A}_2) = \{A_1 \times A_2 : A_1 \in \xi(\mathscr{A}_1),\ A_2 \in \xi(\mathscr{A}_2)\}.$$

Si $\mathscr{F}_0$ denota el álgebra formada por las uniones finitas de rectángulos medibles entonces $\sigma(\mathscr{F}_0) = \mathscr{F}_1 \otimes \mathscr{F}_2$, por definición de $\mathscr{F}_1 \otimes \mathscr{F}_2$ y por el Teorema 4.16,

$$h(T_1 \times T_2) = \sup\{h(T_1 \times T_2, \mathscr{B}) : \mathscr{B} \subset \mathscr{F}_0,\ \mathscr{B} \text{ finita}\}.$$

Pero $\mathscr{B}$ es finita y $\mathscr{B} \subseteq \mathscr{F}_0$, entonces $\mathscr{B} \subseteq \mathscr{A}_1 \otimes \mathscr{A}_2$ para algunas $\mathscr{A}_1 \subseteq \mathscr{F}_1$, $\mathscr{A}_2 \subseteq \mathscr{F}_2$ finitas. Así pues, por el Teorema 4.10(v),

$$h(T_1 \times T_2) = \sup\{h(T_1 \times T_2, \mathscr{A}_1 \otimes \mathscr{A}_2) : \mathscr{A}_1 \subset \mathscr{F}_1,\ \mathscr{A}_2 \subset \mathscr{F}_2; \mathscr{A}_1,\ \mathscr{A}_2 \text{ finitas}\}.$$

Tenemos que

$$\begin{aligned}
h\left(\bigvee_{i=0}^{n-1} (T_1 \times T_2)^{-i}(\mathscr{A}_1 \otimes \mathscr{A}_2)\right) &= H\left(\left(\bigvee_{i=0}^{n-1} T_1^{-i}(\mathscr{A}_1)\right) \times \left(\bigvee_{i=0}^{n-1} T_2^{-i}(\mathscr{A}_2)\right)\right) \\
&= -\sum (\mu_1 \otimes \mu_2)(C_k \times D_j) \cdot \log((\mu_1 \otimes \mu_2)(C_k \times D_j)) \\
&= -\sum \mu_1(C_k)\mu_2(D_j) \log(\mu_1(C_k)\mu_2(D_j)) \\
&= -\sum \mu_1(C_k)\mu_2(D_j)[\log(\mu_1(C_k)) + \log(\mu_2(D_j))] \\
&= -\sum \mu_1(C_k) \log(\mu_1(C_k) - \mu_2(D_j)\mu_2(D_j)) \\
&= H\left(\bigvee_{i=0}^{n-1} T_1^{-i}(\mathscr{A}_1)\right) + H\left(\bigvee_{i=0}^{n-1} T_2^{-i}(\mathscr{A}_2)\right),
\end{aligned}$$

donde $\{C_k\}$ son elementos de $\xi(\bigvee_{i=0}^{n-1} T_1^{-i}(\mathscr{A}_1))$, y $\{D_j\}$ son elementos de $\xi(\bigvee_{i=0}^{n-1} T_2^{-i}(\mathscr{A}_2))$. Por lo que $h(T_1 \times T_2, \mathscr{A}_1 \otimes \mathscr{A}_2) = h(T_1, \mathscr{A}_1) + h(T_2, \mathscr{A}_2)$. Por lo tanto

$$h(T_1 \times T_2) = h(T_1) + h(T_2).$$

□

Observación 4.16. Podemos extender el resultado anterior para el producto directo de cualquier número finito de transformaciones preservadoras de medida.

4.7. Ejemplos

Ahora calcularemos la entropía de algunos de nuestros ejemplos.

Ejemplo 4.1. Si $I : ((\Omega, \mathscr{F}, \mu) \to (\Omega, \mathscr{F}, \mu))$ es la identidad, entonces $h(I) = 0$, ya que $h(I, \mathscr{A}) = \lim_{n\to\infty}(1/n)H(\mathscr{A}) = 0$. También se tiene que si $T^p = I$ para algún $p \neq 0$ entonces $h(T) = 0$ puesto que $0 = h(T^p) = |p| \cdot h(T)$.

Teorema 4.19. *Cualquier rotación, $T(z) = az$, de S^1 tiene entropía cero.*

Demostración.

Caso 1: Supongamos que $\{a^n : n \in \mathbb{Z}\}$ no es denso, es decir, a es una raíz de la unidad, entonces $a^p = 1$ para algún $p \neq 0$; por lo que $T^p = a^p z = z$, por lo tanto $h(T) = 0$.

Caso 2: Supongamos que $\{a^n : n \in \mathbb{Z}\}$ es denso en S^1, entonces $\{a^n : n < 0\}$ es denso en S^1. Sea $\xi = \{A_1, A_2\}$, donde $A_1 = \{e^{i\theta} : \theta \in [0, \pi)\}$ y $A_2 = \{e^{i\theta} : \theta \in [\pi, 2\pi)\}$, para $n > 0\,, T^{-n}(\xi)$ está formada por semicircunferencias con extremos en a^{-n} y $-a^{-n}$. Como $\{a^n : n < 0\}$ es denso se tiene que $\bigvee_{n=0}^{\infty} \overset{\circ}{=} \mathscr{F}$, por lo tanto $h(T) = 0$. □

Teorema 4.20. *El corrimiento bilateral (unilateral) de Bernoulli con parámetros* $(p_0, \ldots, p_{k-1})$ *tiene entropía*

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

Demostración. Sea $A_i = \{\{x_k\} : x_0 = i\}$, $0 \leq i \leq k-1$, entonces $\xi = \{A_0, \ldots, A_{k-1}\}$ es una partición de Ω. Por la definición de la σ-álgebra producto, $\mathscr{F}$, tenemos que

$$\bigvee_{i=-\infty}^{\infty} T^i(\mathscr{A}) = \mathscr{F}.$$

Por el Teorema de Kolmogorov-Sinai, tenemos

$$h(T) = \lim_{n\to\infty} \frac{1}{n} H(A \vee T^{-1}(\mathscr{A}) \vee \cdots \vee T^{-(n-1)}(\mathscr{A})).$$

Un elemento típico de $\xi(A \vee T^{-1}(\mathscr{A}) \vee \cdots \vee T^{-(n-1)}(\mathscr{A}))$ es de la forma:

$$A_{i_0} \cap T^{-1}(A_{i_1}) \cap \cdots \cap T^{-(n-1)}(A_{i_{n-1}}) = \{\{x_n\} : x_0 = i_0, x_1 = i_1, \ldots, x_{n-1} = i_{n-1}\},$$

que tiene medida $p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}$. Por lo que

$$\begin{aligned} H(A \vee T^{-1}(\mathscr{A}) \vee \cdots \vee T^{-(n-1)}(\mathscr{A})) &= -\sum (p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}) \cdot \log(p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}}) \\ &= -\sum_{i_0,\ldots,i_{n-1}=0}^{k-1} (p_{i_0} \cdot p_{i_1} \cdots p_{i_{n-1}})[\log(p_{i_0}) + \cdots + \log(p_{i_{n-1}})] \\ &= -n \sum_{i=0}^{k-1} p_i \cdot \log(p_i). \end{aligned}$$

Por lo tanto $h(T) = h(T, \mathscr{A}) = -\sum_{i=0}^{k-1} p_i \cdot \log(p_i)$. □

Observación 4.17. El corrimiento bilateral de Bernoulli con parámetros (1/2,1/2) tiene entropía log(2), y el corrimiento bilateral de Bernoulli con parámetros (1/3,1/3,1/3) tiene entropía log(3), por lo que estas dos transformaciones no pueden ser isomorfas.

Ejemplo 4.2 (Una transformación con entropía infinita)**.** Sea $I = (\,0\,,1\,]$. Considerando el espacio $(I, \mathscr{B}_{(0,1]}, \mu)$, donde μ la medida de Lebesgue. Sea $\Omega = \prod_{-\infty}^{\infty} I$, con la σ-álgebra producto, la medida producto, $T : \Omega \to \Omega$ dada por $(T(\omega))_i = \omega_{i+1}$, entonces $h(T) = \infty$.

Para comprobar lo anterior consideramos:

$$A_{n,i} = \left\{ \{x_j\} : \frac{i-1}{n} < x_0 \leq \frac{i}{n}, n > 0, 1 \leq i \leq n \right\},$$

entonces $\mu(A_{n,i}) = 1/n$ y $\xi_n = \{A_{n,1}, \ldots, A_{n,n}\}$ es un partición de X. Utilizando argumentos similares que en el Teorema 4.20 se tiene que $h(T, \xi_n) = \log(n)$ (utilizando la independencia de $\xi_n, T^{-1}(\xi_n), \ldots, T^{-k}(\xi_n)$), así que $h(T) \geq n$ para toda n, por lo tanto $h(T) = \infty$.

Teorema 4.21. *El corrimiento bilateral (unilateral) de Markov con matriz de transición $P(i,j) = p_{ij}$ y distribución inicial estacionaria $\pi = [\pi_0, \ldots, \pi_{k-1}]$ tiene entropía:*

$$-\sum_{i,j} \pi_i p_{ij} \log(p_{ij}).$$

Demostración. Al igual que en el Teorema 4.20, tenemos que:

$$h(T) = \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A}) \right),$$

el elemento $A_{i_0} \cap T^{-1}(A_{i_1}) \cap \cdots \cap T^{-(n-1)}(A_{i_{n-1}})$ tiene medida $\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}$, de donde se tiene que

$$\begin{aligned}
H(\bigvee_{i=0}^{n-1} T^{-i}(\mathscr{A})) &= -\sum_{i_0, i_1, \ldots, i_{n-1}}^{k-1} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log(\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}}) \\
&= -\sum \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} [\log(\pi_{i_0}) + \log(p_{i_0 i_1}) \cdots + \log(p_{i_{n-2} i_{n-1}})] \\
&= -\sum_{i_0=0}^{k-1} \sum_{i_1, \ldots, i_{n-1}}^{k-1} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log(\pi_{i_0}) \\
&\quad -\sum_{j=0}^{n-2} \sum_{i_j, i_{j+1}}^{k-1} \sum_{\{i_1, \ldots, i_{n-1}\} \setminus \{i_j, i_{j+1}\}}^{k-1} \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \log(p_{i_j i_{j+1}}) \\
&= -\sum_{i_0=0}^{k-1} \pi_{i_0} \log(\pi_{i_0}) - \sum_{j=0}^{n-2} \sum_{i_j, i_{j+1}} \pi_{i_j} p_{i_j i_{j+1}} \log(p_{i_j i_{j+1}}) \\
&= -\sum_{i_0=0}^{k-1} \pi_{i_0} \log(\pi_{i_0}) - (n-1) \sum_{i,j=0}^{k-1} \pi_i p_{ij} \log(p_{ij}),
\end{aligned}$$

donde estas igualdades se obtienen de las relaciones siguientes: $\sum_{i=0}^{k-1} \pi_i p_{ij} = p_j$ y $\sum_{j=0}^{k-1} p_{ij} = 1$. Por lo tanto $h(T) = -\sum_{i,j} \pi_i p_{ij} \log(p_{ij})$. ☐

Hemos visto que si dos transformaciones son isomorfas entonces tienen la misma entropía (ver Teorema 4.9, pág. 102). También noté que la relación $\simeq$ es una relación de equivalencia. Cabe preguntarnos si la entropía es una propiedad invariante completa para la relación $\simeq$. Es decir,

¿será cierto que si dos transformaciones tienen las misma entropía entonces son isomorfas? Es fácil ver que no. Para ilustrar lo anterior consideremos lo siguiente:

Sea T la rotación irracional en la circunferencia unitaria dada por $T(z) = az$. Comparémosla con otra del mismo tipo, sea $S(z) = bz$, otra rotación irracional en la circunferencia unitaria. Como hemos visto en el Teorema 4.19 (pág. 112), ambas transformaciones tienen entropía cero. Sin embargo si elegimos a y b tal que $\{a^n\}_{-\infty}^{\infty} \neq \{b^n\}_{-\infty}^{\infty}$, entonces S y T no son espectralmente isomorfas por la Observación 2.1 (recordando que $\{z^n\}_{n\in\mathbb{Z}}$ es una base ortonormal de vectores propios con valores propios para T y S, $\{a^n\}_{-\infty}^{\infty}$ y $\{b^n\}_{-\infty}^{\infty}$ respectivamente). Por lo tanto $T \not\simeq S$.

Así que la entropía no es una buena propiedad invariante bajo isomorfismo, aunque sería bueno encontrar una familia donde lo fuera. Esto fue un problema que estuvo abierto durante varios años, hasta que Ornstein probó su famoso teorema, (el cual no se probará en este trabajo). Por lo visto anteriormente esta clase de transformaciones deben estar fuera de la clase de transformaciones con entropía cero.

4.8. Automorfismos de Bernoulli y Kolmogorov

En esta sección discutiremos el teorema de Ornstein, el cuál nos dice en qué clase de transformaciones la entropía es una propiedad invariante completa para la relación de isomorfismo. Veremos una generalización de esta clase de transformaciones, así como consecuencias del teorema de Ornstein y otra clase de transformaciones introducidas por Kolmogorov que generalizan la clase anterior. Existen resultados que prueban que estas clases **no** son las mismas, así como prueban que la entropía tampoco es una propiedad invariante completa bajo esta nueva clase de transformaciones, sin embargo estos resultados quedan fuera del alcance de este trabajo (ver Walters [23, págs. 105-113]).

Definición 4.12. Sea $(X, \mathscr{G}, \nu)$ un espacio de probabilidad, $(\Omega, \mathscr{F}, \mu) = \prod_{-\infty}^{\infty}(X, \mathscr{G}, \nu)$ y $T : \Omega \to \Omega$ dada por $T(\{x_n\}) = \{y_n\}$, donde $y_n = x_{n-1}$ $n \in \mathbb{Z}$, entonces T es una transformación preservadora de medida invertible y la llamaremos el **corrimiento de Bernoulli** con **espacio de estados** $(X, \mathscr{G}, \nu)$.

Ejemplo 4.3. El corrimiento bilateral de Bernoulli con parámetros $(p_0, \ldots, p_{k-1})$, donde $X = \{0, 1, \ldots, k-1\}$.

Ejemplo 4.4. El Ejemplo 4.2 (pág. 113), donde $X = (\,0,\,1\,]$.

Ejemplo 4.5. Si T es un corrimiento de Bernoulli con espacio de estados $(X, \mathscr{G}, \nu)$, entonces T^n es un corrimiento de Bernoulli con espacio de estados $(\prod_{i=1}^n X, \bigotimes_{i=0}^n \mathscr{G}, \bigotimes_{i=0}^n \nu)$. Considere $\phi : \prod_{i=-\infty}^{\infty} \prod_{i=1}^n X \to \Omega$, dada por

$$\phi((\ldots, \vec{x}_{-1}, \vec{x}_0, \vec{x}_1, \ldots)) = (\ldots, \vec{x}_{-1_1}, \vec{x}_{-1_2}, \ldots, \vec{x}_{-1_n}, \vec{x}_{0_1}, \ldots, \vec{x}_{0_n}, \vec{x}_{1_1}, \ldots, \vec{x}_{1_n}, \ldots),$$

donde el lugar que ocupa $\vec{x}_{0_1}$ es la posición cero de un $x \in \prod_{i=-\infty}^{\infty} X$.

Ejemplo 4.6. Si T_1 y T_2 son corrimientos de Bernoulli, entonces $T_1 \times T_2$ también lo es, el espacio de estados de $T_1 \times T_2$ es el producto directo de los espacios de estados de T_1 y T_2.

Observación 4.18. El método que utilizamos en el Teorema 4.20 y el Ejemplo 4.2 (páginas 113 y 113, respectivamente) puede ser utilizado para mostrar que si T es un corrimiento de Bernoulli, entonces $h(T) < \infty$ si y sólo si existe una partición numerable ξ de $(X, \mathscr{G}, \nu)$ tal que $H(\xi) < \infty$ y$\mathscr{A}(\xi) \overset{\circ}{=} \mathscr{G}$. En este caso $h(T) = H(\mathscr{G})$.

Teorema 4.22 (Ornstein)**.** *Sean T_1, T_2 corrimientos de Bernoulli cuyos espacios de estados son espacios de Lebesgue.*[1] *Si $h(T_1) = h(T_2)$ entonces T_1 y T_2 son isomorfas.*

La prueba de este profundo teorema se encuentra en D.S. Ornstein [15]. Este resultado reduce el problema de isomorfismo de corrimientos de Bernoulli a sus espacios de estados, ya que la entropía depende sólo del espacio de estados. Por ejemplo, se puede tener un corrimiento bilateral de Bernoulli con un espacio de estados formado por dos puntos, y éste ser isomorfo a un corrimiento de Bernoulli con espacio de estados numerable.

Observación 4.19. Como la función $(p_1, \ldots, p_n) \to \sum p_i \log(p_i)$, definida en

$$\{(p_1, \ldots, p_n) : p_i \geq 0, \sum_{i=1}^{n} p_i = 1\},$$

es continua y este simplejo es conexo su imagen es $[\,0,\, n\,]$, entonces para cada $x > 0$ existe un corrimiento de Bernoulli con entropía x.

Definición 4.13. Una transformación invertible preservadora de medida de un espacio de Lebesgue $(\Omega, \mathscr{F}, \mu)$ es llamado un **automorfismo de Bernoulli** si es isomorfa a un corrimiento de Bernouilli (la palabra automorfismo se utiliza porque es una transformación preservadora de medida invertible, es decir, es biyectiva y además preserva la estructura de un espacio de Lebesgue).

Las observaciones hechas anteriormente aplicadas a los automorfismos de Bernouilli, junto con el Teorema de Ornstein, dicen que dos automorfimos de Bernoulli que tengan la misma entropía son isomorfos. A partir de lo anterior se obtiene lo siguiente:

Corolario 4.22.1.

(i) Todo automorfismo de Bernoulli tiene una raíz n-ésima. (S *es una raíz n-ésima de* T *si* $S^n = T$*, donde* $S^n = S \circ \cdots \circ S$ *n-veces*).

[1]Un espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$ es de Lebesgue si es isomorfo a un espacio de probabilidad que es la unión disjunta de un conjunto $\{x_1, x_2, \ldots\}$ numerable (o finito), con medidas positivas, y del espacio $([0,s], \mathscr{L}([0,s]), \lambda)$, donde $\mathscr{L}([0,s])$ es la σ-álgebra de los Lebesgue medibles del $[0,s]$, λ es la medida de Lebesgue y $s = 1 - \sum_{n=1}^{\infty} p_n$ donde p_n es la medida de x_n.

(ii) Todo automorfismo de Bernoulli es isomorfo a un producto directo de dos automorfismos de Bernoulli.

(iii) Todo automorfismo de Bernoulli T es isomorfo a su inversa.

Demostración.

(i) Sea T un automorfismo de Bernoulli y $n > 0$. Sea S un automorfismo de Bernoulli con $h(S) = (1/n)h(T)$. Entonces S^n es un automorfismo de Bernoulli con entropía $h(T)$, por lo tanto S^n y T son isomorfas.

(ii) Sea T un automorfismo de Bernoulli y $n > 0$. Sea S un automorfismo de Bernoulli con entropía $(1/2)h(T)$, entonces $h(S \times S) = h(T)$. Como $S \times S$ es un automorfismo de Bernoulli, se tiene que $S \times S$ y T son isomorfas.

(iii) T y T^{-1} son automorfismos de Bernoulli con la misma entropía (ver Teorema 4.11). $\square$

Podemos pensar al corrimiento bilateral de Bernoulli como un proceso estocástico independiente, e idénticamente distribuido con conjunto parametral igual a los enteros. Entonces los automorfismos de Bernoulli son como una abstracción de estos procesos. Un ejemplo de automorfismo de Bernoulli es la transformación del panadero. Kolmogorov introdujo la siguiente clase de transformaciones preservadoras de medida como una abstracción de los procesos estocásticos idénticamente distribuidos.

Definición 4.14. Una transformación invertible que preserva la medida T de un espacio de probabilidad $(\Omega, \mathscr{F}, \mu)$ es un **automorfismo de Kolmogorov** (K-automorfismo) si existe una sub-σ-álgebra $\mathscr{K}$ de $\mathscr{F}$ tal que:

(a) $\mathscr{K} \subset T(\mathscr{K})$.

(b) $\bigvee_{n=0}^{\infty} T^n(\mathscr{K}) \stackrel{\circ}{=} \mathscr{F}$.

(c) $\bigcap_{n=0}^{\infty} T^{-n}(\mathscr{K}) \stackrel{\circ}{=} \mathscr{N} = \{\Omega, \emptyset\}$.

Supondre que $\mathscr{N} \stackrel{\circ}{\neq} \mathscr{F}$, ya que si no la identidad sería el único K-automorfismo. Notemos que $\mathscr{K} \stackrel{\circ}{\neq} T(\mathscr{K})$. Usualmente el espacio $(\Omega, \mathscr{F}, \mu)$ será de Lebesgue.

Teorema 4.23. *Todo automorfismo de Bernoulli es un automorfismo de Kolmogorov.*

Demostración. Sea $(X, \mathscr{G}, \nu)$ el espacio de estados de T, sea $G \in \mathscr{G}$, $\widetilde{G} = \{\{x_n\} \in \Omega : x_0 \in G\}$ y $\mathscr{D} = \{\widetilde{G} : G \in G\}$, la cual es llamada la σ-álgebra al tiempo cero. Sea $\mathscr{K} = \bigvee_{i=-\infty}^{0} T^i(\mathscr{D})$. Verifico que $\mathscr{K}$ satisface las condiciones de un automorfismo de Kolmogorov.

(a) $\mathscr{K} = \bigvee_{i=-\infty}^{0} T^i(\mathscr{D}) \subset \bigvee_{i=-\infty}^{1} T^i(\mathscr{D}) = T(\mathscr{K})$.

(b) $\bigvee_{n=0}^{\infty} T^n(\mathscr{K}) = \bigvee_{n=0}^{\infty} \bigvee_{i=-\infty}^{n} T^i(\mathscr{D}) = \mathscr{F}$ por definición de $\mathscr{F}$.

(c) Debemos mostrar que $\bigcap_0^\infty T^{-n}(\mathscr{K}) \stackrel{\circ}{=} \mathscr{N} = \{\Omega, \emptyset\}$. Dada

$$A \in \bigcap_0^\infty T^{-n}(\mathscr{K}) = \bigcap_0^\infty \bigvee_{-\infty}^{-n} T^i(\mathscr{D}),$$

sea $B \in \bigvee_{k=j}^{\infty} T^k(\mathscr{D})$, para algún $j \in \mathbb{Z}$ dado, así pues $A \in \bigvee_{i<j} T^i(\mathscr{D})$; A y B son independientes, por lo tanto $\mu(A \cap B) = \mu(A)\mu(B)$. La colección de todos los conjuntos B tal que $\mu(A \cap B) = \mu(A)\mu(B)$ son una clase monótona (ver Definición B.19, pág. 133), por lo anterior, contiene a $\bigcup_{j=-\infty}^{\infty} \bigvee_{k=j}^{\infty} T^k(\mathscr{D})$. Por lo tanto para toda $B \in \mathscr{F}, \mu(A \cap B) = \mu(A)\mu(B)$ (ver Teorema B.2, pág. 131). Haciendo $B = A$ se tiene que $\mu(A) = \mu(A)^2$, lo cual implica que $\mu(A) = 0$ o $\mu(A) = 1$. Así que $\bigcap_{n=0}^{\infty} T^{-n}(\mathscr{K}) \stackrel{\circ}{=} \mathscr{N}$. □

El siguiente resultado nos dice que los K-automorfismo son espectralmente indistinguibles (este resultado no lo probare).

Teorema 4.24 (Rohlin)**.** *Si* $(\Omega, \mathscr{F}, \mu)$ *es un espacio de probabilidad con una base numerable, entonces todo automorfismo de Kolmogorov* $T : \Omega \to \Omega$ *tiene espectro de Lebesgue numerable.*

Demostración. Ver Walters [23, pág. 109] o Cornfeld [7, págs. 338-340]. □

Observación 4.20. El Teorema anterior nos asegura que cualesquiera dos corrimientos bilaterales de Bernoulli tienen espectro de Lebesgue numerable; por lo tanto cualesquiera dos son espectralmente isomorfos. Sin embargo por la Observación 4.17 no cualesquiera dos son isomorfos.

A continuación daremos un ejemplo conocido de K-automorfismo:

Ejemplo 4.7. Los corrimientos bilaterales de Markov con matriz de transición irreducible y aperiódica son K-automorfismo.

Demostración. Sea $A_i = \{\{x_k\} : x_0 = i\}$, $0 \le i \le k-1$, entonces $\xi = \{A_0, \ldots, A_{k-1}\}$ es una partición de Ω. Sea $\mathscr{K} = \bigvee_{n=0}^{\infty} T^{-n}(\mathscr{A}(\xi))$, así que $\mathscr{K} \subset T(\mathscr{K})$ y $\bigvee_{n=0}^{\infty} T^{-n}(\mathscr{A}(\xi)) = \mathscr{F}$. Por último, recordando el Teorema 3.7, se tiene que $\lim_{n\to\infty} p_{ab}^n = \pi_b$ para toda $a, b \in E$. Sean $A = Z(r_1, \ldots, r_l; a_1, \ldots, a_l)$ y $B = Z(n_1, \ldots, n_m; b_1, \ldots, b_m)$ dos cilindros flacos. Si $n_m - n_1 < r_1$, entonces:

$$\mu(A \cap B) = \pi_{b_1} p_{b_1 b_2} \cdots p_{b_{m-1} b_m} p_{b_m a_1}^{r_1 - n_m - n_1} p_{a_1 a_2} \cdots p_{a_{l-1} a_l},$$

así que

$$\mu(A)\mu(B) \min_{a,b} \left\{ \frac{p_{ab}^{r_1 - n_m - n_1}}{\pi_b} \right\} \le \mu(A \cap B) \le \mu(A)\mu(B) \max_{a,b} \left\{ \frac{p_{ab}^{r_1 - n_m - n_1}}{\pi_b} \right\}.$$

La igualdad anterior permanece válida si $A \in \bigvee_{i=r_1}^{r_1+r_l} T^{-i}(\mathscr{A}(\xi))$ y $B \in \bigvee_{i=n_1}^{n_1+n_l} T^{-i}(\mathscr{A}(\mathscr{F}))$. Para $B \in \bigvee_{i=n_1}^{n_1+n_l} T^{-i}(\mathscr{A}(\xi))$ y $N > n_m - n_1$ fijos, considerando la colección $\mathbb{E}$ de los conjuntos medibles A tales que:

$$\mu(A)\mu(B) \inf_{a,b,r_1 \geq N} \left\{ \frac{p_{ab}^{r_1-n_m-n_1}}{\pi_b} \right\} \leq \mu(A \cap B) \leq \mu(A)\mu(B) \sup_{a,b,r_1 \geq N} \left\{ \frac{p_{ab}^{r_1-n_m-n_1}}{\pi_b} \right\},$$

se tiene que $\mathbb{E}$ es una clase monótona que contiene al álgebra $\bigcup_{r_1=N}^{\infty} \bigcup_{r_l=0}^{\infty} \bigvee_{i=r_1}^{r_1+r_l} T^{-i}(\mathscr{A}(\xi))$, por lo que contiene a la σ-álgebra $\bigvee_{i=N}^{\infty} T^{-i}(\mathscr{A}(\xi)) = T^{-N}(\mathscr{K})$. Sea $\varepsilon > 0$ y eligiendo una n_0 tal que para toda $n \geq n_0$ $|p_{a,b}^n/\pi_b - 1| < \varepsilon$ para todo $a, b \in E$. Entonces si $B \in \bigvee_{i=n_1}^{n_1+n_l} T^{-i}(\mathscr{A}(\mathscr{F}))$, se tiene que $|\mu(A \cap B) - \mu(A)\mu(B)| < \varepsilon$, siempre que $A \in T^{-N}(\mathscr{K})$ y N sea suficientemente grande. En particular, lo anterior es válido para $A \in \bigcap_{N=0}^{\infty} T^{-N}(\mathscr{K})$. Ahora, para $A \in \bigcap_{N=0}^{\infty} T^{-N}(\mathscr{K})$ fijo considero $\mathbb{H}$ la colección de conjuntos medibles B tales que $|\mu(A \cap B) - \mu(A)\mu(B)| < \varepsilon$. Así que $\mathbb{H}$ es una clase mnónota que contiene a $\bigvee_{i=n_1}^{n_1+n_m} T^{-i}(\mathscr{A}(\xi))$ para toda n_1 y $n_m \geq 0$. Por lo que, $\mathbb{H} = \mathscr{F}$, por lo tanto $|\mu(A \cap B) - \mu(A)\mu(B)| < \varepsilon$ para toda $B \in \mathscr{F}$ y $A \in \bigcap_{n=0}^{\infty} T^{-n}(\mathscr{K})$. Así que, $\mu(A \cap B) = \mu(A)\mu(B)$ para toda $B \in \mathscr{F}$ y $A \in \bigcap_{n=0}^{\infty} T^{-n}(\mathscr{K})$, haciendo $B = A$ tenemos que $\mu(A) = 0$ o $\mu(A) = 1$ siempre que $A \in \bigcap_{n=0}^{\infty} T^{-n}(\mathscr{K})$. $\square$

4.9. Un poco de Teoría de la Información

La Teoria de la Informacion fue desarrolada por Claude E. Shannon en [20]. En este artículo Shannon menciona que el problema fundamental de esta teoría es la reprodución de un mensaje total o parcial de un lugar a otro. La Teoría de la Información estudia los sistemas de comunicación, los cuales estan formados por varios componentes. A continuación damos un definición de estos componentes:

Definición 4.15. Un sistema de comunicación está conformado por los siguientes componentes:

(a) **Una fuente de información** es el dispositivo que se encarga de emitir un mensaje o una sucesión de mensajes.

(b) **Un transmisor** codifica los mensajes de tal manera que se puedan transmitir por el canal.

(c) **Un canal** es el medio por el cual el mensaje, que ha sido codificado por el transmisor, viaja hacia el receptor.

(d) **Un receptor** es el que realiza la operación inversa del transmisor, es decir, decodifica el mensaje recibido por el canal.

(e) **El destinatario** es el dispositivo al cual va dirigido el mensaje.

Un ejemplo de sistema de comunicación es una llamada telefónica: una persona A tiene una llamada telefónica con una persona B, los componentes de este sistema son: la fuente de información es la persona A; el transmisor es el teléfono de la persona A; el canal es la línea telefónica; el receptor es el teléfono de la persona B; y el destinatario la persona B (esto en caso de que la persona A sea la que hable y la B la que escuche).

Se puede pensar a una fuente de información discreta como una sucesión de símbolos tomados de manera aleatoria sobre un conjunto E. Por esta razón podemos pensar a la fuente de información como una colección de variables aleatorias, es decir, un proceso estocástico.

Sea E un conjunto finito de símbolos que llamaremos alfabeto, el cual se usará para representar los mensajes de la fuente. El modelo matemático que se adopta para una fuente discreta de información es el siguiente: el espacio de probabilidad $(E^{\mathbb{Z}}, \mathscr{F}, \mu)$, donde $\mathscr{F} = \sigma$-álgebra generada por los cilindros, y μ es una media de probabilidad en $(E^{\mathbb{Z}}, \mathscr{F})$.

Dentro de la clase de las fuentes discretas existe otro tipo de fuente. Dire que una fuente es estacionaria si el corrimiento T en $(E^{\mathbb{Z}}, \mathscr{F}, \mu)$ es una transformación preservadora de medida. De manera análoga una fuente estacionaria es ergódica si el corrimiento T es ergódico.

A partir de estas definiciones, y resultados anteriores, puedo deducir los lemas siguientes:

Lema 4.4. *Una fuente $(E^{\mathbb{Z}}, \mathscr{F}, \mu)$ es estacionaria si y sólo si para todo cilindro Z se tiene que $\mu(T(Z)) = \mu(Z)$.*

Lema 4.5. *Una fuente $(E^{\mathbb{Z}}, \mathscr{F}, \mu)$ es ergódica si y sólo si para todo cilindro Z se tiene que $\text{lím}_{n\to\infty}(1/n)\sum_{i=0}^{n-1}\chi_Z(T^i(x)) = \mu(Z)$ casi dondequiera.*

Observación 4.21. En el caso especial de que la fuente de información sea un corrimiento bilateral de Markov se tiene que $h(T) = -\sum_{i,j}\pi_i p_{ij}\log(p_{ij})$. Claramente esta fuente es estacionaria. Además sabemos que esta fuente es ergódica si y sólo si la matriz de probabilidades de transición es irreducible. Shannon solamente considera este caso y en ocasiones también la del corrimiento de Bernouilli.

EL siguiente teorema es crucial en esta teoría. Hay otra versión más general (ver Parry [17, pág. 39]), sin embargo esta versión involucra la que hizo Shannon.

Teorema 4.25. *Sean $(\Omega, \mathscr{F}, \mu)$un espacio de probabilidad, $T : \Omega \to \Omega$ una transformación preservadora de medida, y $\xi = \{A_1, \ldots, A_k\}$ una partición finita del espacio con $\mathscr{A} = \mathscr{A}(\xi)$. Si $f = \sum_{i=1}^{k} E(\chi_{A_i}/\bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A}))\log(E(\chi_{A_i}/\bigvee_{i=1}^{\infty} T^{-i}(\mathscr{A})))$ (la función información), entonces :*

$$\frac{1}{n} I\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi)\right) \to E(f/\mathscr{I}) \quad \text{en medida.}$$

Demostración. Sean $f_n = (1/n)I\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi)\right)$ para toda $n \geq 1$. Por propiedad de la esperanza condicional y el Teorema 4.12, se tiene que $f_n \to E(f/\mathscr{I})$ en L_1. Por lo tanto $f_n \to f$ en medida (ver Bartle [3, pág. 69]). $\square$

No sólo se puede probar que converge en medida, más aún, la sucesión anterior también converge casi seguramente (es decir, casi dondequiera). A continuación pruebo este hecho para el corrimiento de Markov.

Lema 4.6. *Sean* $(\Omega, \mathscr{F}, \mu, T)$ *un corrimiento bilateral de Markov y* $\xi = \{A_0, \dots, A_k\}$*, donde* $A_i = \{x_k\}\{x_0 = i\}$*, entonces:*

$$\frac{1}{n}I\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi)\right) \to E(f/\mathscr{I}) \quad \textit{casi dondequiera.}$$

Demostración. Definimos $g_0(\omega) = -\log(\mu(Z(0; x_0(\omega))))$ y

$$g_k(\omega) = -\log\left(\frac{\mu(Z(-k,\dots,0; x_{-k}(\omega),\dots,x_0(\omega)))}{\mu(Z(-k,\dots,-1; x_{-k}(\omega),\dots,x_{-1}(\omega)))}\right) \quad \text{para } k \geq 1.$$

Como

$$\frac{1}{n}I\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi)\right)(\omega) = -\frac{1}{n}\log(\pi_{x_0(\omega)}) - \frac{1}{n}\sum_{i=1}^{n-1}\log(p_{x_{i-1}(\omega)x_i(\omega)}),$$

por lo que:

$$-\frac{1}{n}\log(\pi_{x_0(\omega)}) - \frac{1}{n}\sum_{i=1}^{n-1}\log(p_{x_{i-1}(\omega)x_i(\omega)}) = \frac{1}{n}\sum_{k=0}^{n-1} g_k(T^k(\omega)),$$

pero por el Teorema 3.1 se tiene que g_k no depende de k para $k \geq 1$. Sea g tal función, además $g \in L_1(\mu)$, por el Teorema Ergódico Individual se tiene el resultado. $\square$

De los dos últimos resultados se desprenden los teoremas más importantes de esta teoría. Por último probare uno de los teoremas de codificación de Shannon. Necesitare las siguientes definiciones:

Definición 4.16. Una fuente $(E^{\mathbb{Z}}, \mathscr{F}, \mu, T)$ ergódica se llama sin memoria, si la probabilidad de obtener un símbolo no depende del símbolo anterior. Note que bajo estas hipótesis, las fuentes ergódicas sin memoria son corrimientos de Bernoulli.

Definición 4.17. Si dos fuentes $(E_1^{\mathbb{Z}}, \mathscr{F}_1, \mu_1, T_1)$, $(E_2^{\mathbb{Z}}, \mathscr{F}_2, \mu_2, T_2)$ estacionarias son tales que $T_1 \simeq T_2$, se dice que existe un código invertible.

Teorema 4.26. *Sean* $(\Omega_i, \mathscr{F}_i, \mu_i, T_i)$ *con* $i = 1, 2$ *dos fuentes ergódicas sin memoria. Si* $h(T_1) > \log(s)$ *entonces no existe un código invertible, en caso contrario si existe un código invertible.*

Demostración. Por hipótesis ambas fuentes son automorfismos de Bernoulli, como la entropía es una propiedad completa invariante en estos automorfismos se tiene el resultado. $\square$

Conclusiones

En la tesis utilice a la Teoría Ergódica para obtener diversos resultados de la Teoría de Probabilidad entre otros. Uno de los resultados más importantes que normalmente no se demuestran en la literatura es la Ley Fuerte de los Grandes Números. Resulta que podemos ver al Teorema Ergódico Individual como una generalización de esta ley.

En el trabajo se desarrolló de manera breve uno de los problemas de la Teoría Ergódica, que se denomina interno, el problema de isomorfismo. La importancia de este problema radica en la comprensión de las transformaciones preservadoras de medida. Se dieron dos nociones de isomorfismo para estas transformaciones. Se puede concluir que este problema está lejos de alcanzar una solución. Sin embargo, como se mostró en la tesis se puede dar una solución bastante aceptable.

Es interesante el hecho de que se pueda aplicar esta teoría al motor de búsqueda de Google, ya que relaciona a la teoría con una página de internet que se utiliza frecuantemente hoy en día, por varios usuarios de Internet.

Se puede concluir que la entropía es una buena medida de la cantidad de incertidumbre que existe en el resultado de un experimento, lo que parece increíble ya que la incertidumbre no es tangible. Además de ser una propiedad invariante bajo isomorfismo, la importancia de este hecho radica en la posibilidad de decir cuando dos transformaciones preservadoras de medida no son isomorfas a partir de un número.

Apéndice A

Análisis

Lo que viene fue obtenido de Bartle [1] y [2]. Este apéndice tiene como finalidad recordar definiciones y algunos resultados bien conocidos que se utilizaron en este trabajo. Recordando que podemos definir al "límite superior" de una sucesión acotada (x_n) como el mayor número real que es el límite de alguna subsucesión de (x_n), y definimos al "límite inferior" como el menor número real que es límite de alguna subsucesión de (x_n).

Definición A.1.

(a) Si (x_n) es una sucesión acotada en $\mathbb{R}$ y $m \in \mathbb{N}$, definimos $v_m = \sup\{x_n : n \geq m\}$. Entones la sucesión (v_m) es acotada y decreciente, por lo tanto convergente. Definimos el **límite superior** de (x_n) como el límite de la sucesión (v_m). Es decir

$$\limsup_{n\to\infty} x_n = \lim_{m\to\infty} v_m.$$

(b) Análogamente, definimos $u_m = \inf\{x_n : n \geq m\}$ para $m \in \mathbb{N}$. Entonces la sucesión (u_m) es acotada y creciente, por lo tanto convergente. Definimos el **límite inferior** de (x_n) como:

$$\liminf_{n\to\infty} x_n = \lim_{m\to\infty} u_m.$$

Teorema A.1. *Si (x_n) es una sucesión acotada en $\mathbb{R}$, entonces los siguientes enunciados acerca del número $x^* \in \mathbb{R}$ son equivalentes:*

(i) $x^* = \limsup_{n\to\infty} x_n$.

(ii) *Si $v_m = \sup\{x_n : n \geq m\}$, entonces $x^* = \inf\{v_m : m \in \mathbb{N}\}$.*

(iii) *Si L es el conjunto de $v \in \mathbb{R}$ tal que existe una subsucesión de (x_n) que converge a v, entonces $x^* = \max\{L\}$.*

(iv) *Si V es el conjunto de $v \in \mathbb{R}$ tal que hay un número finito de $n \in \mathbb{N}$ tal que $v < x_n$, entonces $x^* = \min\{V\}$.*

(v) *El número x^* tiene la propiedad de que si $\varepsilon > 0$, entonces hay a lo más un número finito de $n \in \mathbb{N}$ tal que $x^* + \varepsilon < x_n$, pero hay un número infinito tal que $x^* - \varepsilon < x_n$.*

La demostración de este teorema se encuentra en Bartle [2, pág. 133].

Observación A.1. Lo análogo del teorema anterior es válido para el límite inferior si notamos que $\liminf_{n\to\infty} x_n = -\limsup_{n\to\infty} x_n$; por ejemplo como $x^* = \limsup_{n\to\infty} x_n$ si y sólo si $x^* = \inf\{v_m : m \in \mathbb{N}\}$, considerado lo observado y denotando a $x_* = \liminf_{n\to\infty} x_n$, tenemos que $x_* = \liminf_{n\to\infty} x_n$ si y sólo si $x_* = \sup\{-v_m : m \in \mathbb{N}\}$.

Ahora pasamos a ver propiedades sobre estos límites, que son semejantes a las propiedades que cumple el límite de una sucesión.

Propiedades A.1 Sean (x_n) y (y_n) sucesiones acotadas en $\mathbb{R}$, entonces:

(i) $\liminf x_n \leq \limsup x_n$.

(ii) Si $c \geq 0$ y $\gamma \leq 0$, entonces:

- *a)* $\liminf(c \cdot x_n) = c \cdot \liminf x_n$.
- *b)* $\limsup(c \cdot x_n) = c \cdot \limsup x_n$.
- *c)* $\liminf(\gamma \cdot x_n) = \gamma \cdot \limsup x_n$.
- *d)* $\limsup(\gamma \cdot x_n) = \gamma \cdot \liminf x_n$.

(iii) $\liminf x_n + \liminf y_n \leq \liminf(x_n + y_n)$.

(iv) $\limsup(x_n + y_n) \leq \limsup x_n + \limsup y_n$.

(v) Si (x_n) es convergente, entonces:

- *a)* $\liminf(x_n + y_n) = \lim x_n + \liminf y_n$.
- *b)* $\limsup(x_n + y_n) = \lim x_n + \limsup y_n$.

(vi) Si $x_n \geq 0$ y $y_n \geq 0$, entonces:

- *a)* $\liminf(x_n \cdot y_n) \geq (\liminf x_n)(\liminf y_n)$.
- *b)* $\limsup(x_n \cdot y_n) \leq (\limsup x_n)(\limsup y_n)$.

(vii) Si $x_n \geq 0$ y $y_n \geq 0$ y (x_n) es convergente, entonces:

- *a)* $\liminf(x_n \cdot y_n) = (\lim x_n)(\liminf y_n)$.
- *b)* $\limsup(x_n \cdot y_n) = (\lim x_n)(\limsup y_n)$.

(viii) Si $x_n \leq y_n$ para toda $n \in \mathbb{N}$, entonces:

$$\liminf x_n \leq \liminf y_n \quad \text{y} \quad \limsup x_n \leq \limsup y_n.$$

Demostración. Omito las pruebas de (i),(ii),(iii) y (iv), éstas se pueden encontrar en Bartle [2, pág. 134].

(v) Dada (y_{n_k}) una sucesión de (y_n) que converge a $\limsup y_n$. Entonces

$$\lim(x_{n_k} + y_{n_k}) = \lim x_{n_k} + \lim y_{n_k} = \lim x_n + \limsup y_n,$$

ya que $(x_{n_k} + y_{n_k})$ es una subsucesión convergente de $(x_n + y_n)$, concluimos que su límite es menor o igual que $\limsup(x_n + y_n)$, por lo que utilizando (iv) tenemos que

$$\lim x_n + \limsup y_n \leq \limsup(x_n + y_n) \leq \lim x_n + \limsup y_n.$$

La otra prueba es análoga.

(vi) Como $\sup\{x_n \cdot y_n : n \geq m\} \leq \sup\{x_n : n \geq m\} \cdot \sup\{y_n : n \geq m\}$ para toda $m \in \mathbb{N}$, el segundo enunciado se sigue de propiedades de convergencia de sucesiones.

(vii) Sea (y_{n_k}) una subsucesión de (y_n) que converge a $\limsup y_n$. Entonces

$$\lim(x_{n_k} \cdot y_{n_k}) = (\lim x_{n_k}) \cdot (\lim y_{n_k}) = (\lim x_n) \cdot (\limsup y_n),$$

como $x_{n_k} \cdot y_{n_k}$ es una subsucesión convergente de $x_n \cdot y_n$, concluimos que este límite es menor o igual que el $\limsup(x_n \cdot y_n)$, utilizando (vi)se tiene que

$$(\lim x_n) \cdot (\limsup y_n) \leq \limsup(x_n \cdot y_n) \leq (\lim x_n) \cdot (\limsup y_n).$$

(viii) La hipótesis implica que $\sup\{x_n : n \geq m\} \leq \sup\{y_n : n \geq m\}$ para cada $m \in \mathbb{N}$, de donde se sigue que $\limsup x_n \leq \limsup y_n$. □

Observación A.2. En general las desigualdades en los incisos (iii), (iv), (vi), (vii) no pueden ser reemplazadas por igualdades. Si $x_n = (-1)^n$ y $y_n = (-1)^{n+1}$ para $n \in \mathbb{N}$, entonces:

$$\liminf x_n + \liminf y_n = -2 < 0 = \liminf(x_n + y_n) \leq \limsup(x_n + y_n) = 0 < 2 = \limsup x_n + \limsup y_n.$$

También, si $u_{2n-1} = 0$, $u_{2n} = 1$ y $v_{2n-1} = 1$, $v_{2n} = 0$ para $n \in \mathbb{N}$, entonces:

$$\limsup(u_n \cdot v_n) = 0 < 1 = (\limsup u_n)(\limsup v_n).$$

También podemos definir el límite superior e inferior para una sucesión de subconjuntos de un conjunto X, como sigue:

Definición A.2. Sea $\{B_n\}_{n=0}^{\infty} \subset \mathcal{P}(X)$ definimos:

(a) $\limsup_{n\to\infty} B_n = \bigcap_{n=0}^{\infty} \bigcup_{m=n}^{\infty} B_m$.

(b) $\liminf_{n\to\infty} B_n = \bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} B_m$.

Proposición A.1. *Sea* $\{B_n\}_{n=0}^{\infty} \subset \mathcal{P}(X)$, *entonces* $x \in \limsup_{n\to\infty} B_n$ *si y sólo si existen* $n_1 < n_2 < \cdots$ *tal que* $x \in B_{n_i}$ *para toda* $i \geq 1$.

Definición A.3. Un punto x_p de un conjunto convexo C es un punto extremo si no es posible expresar a x_p como una combinación lineal convexa de otros puntos de C distintos a x_p.

Definición A.4. Un espacio vectorial real o complejo con producto interior $(H, \langle , \rangle)$, se dice que es un **espacio de Hilbert**, si éste es completo con la norma $\|f\| = \sqrt{\langle f, f\rangle}$.

Definición A.5. Sea $(H, \langle , \rangle)$ un espacio de Hilbert, se dice que $\{f_n\}_{n=1}^{\infty} \subset H$ es un conjunto ortonormal, si $\|f_n\| = 1$ para toda n y $\langle f_k, f_m\rangle = 0$ para todo $m \neq k$.

En cualquier espacio de Hilbert se cumple que:

Teorema A.2 (Desigualdad de Bessel)**.** *Sea* $(H, \langle , \rangle)$ *un espacio de Hilbert,*$\{f_n\}_{n=1}^{\infty} \subset H$ *un conjunto ortonormal. Entonces, para todo* $f \in H$ *se tiene que:*

$$\sum_{i=1}^{\infty} |\langle f, f_n\rangle|^2 \leq \|f\|^2.$$

Definición A.6. Sea $(H, \langle , \rangle)$ un espacio de Hilbert. $A = \{f_n\}_{n=1}^{\infty} \subset H$ es una base ortonormal, si A es un conjunto ortonormal y el generado de A es denso en H.

Teorema A.3 (Identidad de Parseval)**.** *Sea* $(H, \langle , \rangle)$ *un espacio de Hilbert,*$\{f_n\}_{n=1}^{\infty} \subset H$ *es una base ortonormal. Entonces, para todo* $f \in H$ *se tiene que:*

$$\sum_{i=1}^{\infty} |\langle f, f_n\rangle|^2 = \|f\|^2.$$

Observación A.3. Si consideramos $L_2^{\mathbb{C}}([-\pi,\pi], \mathscr{B}_{[-\pi,\pi]}, \mu)$, se sabe que éste es un espacio de Hilbert, donde el producto interior está dado por $\langle f, g\rangle = \int f\overline{g} d\,\mu$. Además se sabe que el conjunto $\{e^{inx}\}_{n\in\mathbb{Z}}$ es una base ortonormal. Por lo anterior, tenemos que para toda $f \in L_2(\mu)$, se tiene que:

$$f = \sum_{i=-\infty}^{\infty} \langle f, e^{inx}\rangle e^{inx}.$$

A los números $\hat{f}(n) = \langle f, e^{inx}\rangle$ se les llama coeficientes de Fourier de f.

Apéndice B

Teoría de la Medida

En este apéndice recordamos las definiciones y teoremas de Teoría de la Medida que se utilizaron durante este trabajo. Me referiré generalmente a Bartle [3] y Halmos [11].

Definición B.1. Sea $X \neq \emptyset$, una clase no vacía $\mathcal{A} \subset \mathcal{P}(X)$ se llama un álgebra de subconjuntos de X, si:

(a) $X \in \mathcal{A}$.

(b) Si $A, B \in \mathcal{A}$, entonces $A \setminus B \in \mathcal{A}$.

(c) Si $A, B \in \mathcal{A}$, entonces $A \cup B \in \mathcal{A}$.

Definición B.2. Una clase no vacía $\mathscr{F} \subset \mathscr{P}(X)$ se llama una σ-**álgebra** de subconjuntos de X, si:

(a) $X \in \mathscr{F}$.

(b) Si $E, F \in \mathscr{F}$, entonces $E \setminus F \in \mathscr{F}$.

(c) (F_n) una sucesión de elementos de $\mathscr{F}$, entonces $\bigcup\limits_{n=1}^{\infty} F_n \in \mathscr{F}$.

Para la construcción de álgebras y σ-álgebras se utiliza la siguiente proposición:

Proposición B.1. *Sea $\mathfrak{F}$ una clase no vácia de σ-álgebras de subconjuntos de X (álgebras), entonces $\bigcap\{\mathscr{F} : \mathscr{F} \in \mathfrak{F}\}$ es una σ-álgebra (álgebra) de subconjuntos de X.*

La demostración es inmediata a partir de las definiciones.

Teorema B.1. *Sean $X \neq \emptyset$, $\mathbb{G} \subset \mathcal{P}(X)$, entonces existe una* ***única*** *σ-álgebra (álgebra) de subconjuntos de X denotada $\sigma(\mathbb{G})$ $(\mathcal{A}(\mathbb{G}))$ tal que:*

(i) $\mathbb{G} \subset \sigma(\mathbb{G})$ *(o* $\mathbb{G} \subset \mathcal{A}(\mathbb{G})$*).*

(ii) Si $\mathscr{F}$ *(*$\mathcal{A}$*) es una* σ*-álgebra (álgebra) de subconjuntos de* X*, entonces* $\sigma(\mathbb{G}) \subset \mathscr{F}$ *(*$\mathcal{A}(\mathbb{G}) \subset \mathcal{A}$*).*

Para la demostración considérese $\sigma(\mathbb{G}) = \bigcap\{\mathscr{F} : \mathscr{F}$ es una σ-álgebra y $\mathbb{G} \subset \mathscr{F}\}$ (analogamente para el álgebra). Note que la intersección tiene sentido porque se toma sobre una familia distinta del vacío, ya que $\mathcal{P}(X)$ pertenece a esta familia.

Definición B.3. Sea $\mathbb{G} \subset \mathcal{P}(X)$, donde X es un conjunto no vacío, a $\sigma(\mathbb{G})$ ($\mathcal{A}(\mathbb{G})$) la llamaremos **la σ-álgebra generada por** $\mathbb{G}$ (**la álgebra generada por** $\mathbb{G}$). Si $\sigma(\mathbb{G}) = \mathscr{F}$ ($\mathcal{A}(\mathbb{G}) = \mathcal{B}$) al conjunto $\mathbb{G}$ se llama un **generador** de la σ-álgebra $\mathscr{F}$ (álgebra $\mathcal{B}$).

Ejemplo B.1. Para cada $n \in \mathbb{N}$ y $k \in \{0, \cdots, 2^n - 1\}$, defina:

$$D_n(k) = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) \text{ y } \mathcal{D}_n = \{D_n(k) : k \in \{0, \dots, 2^n - 1\}\},$$

$\mathcal{D}_n$ se llama diádicos de rango n. Entonces $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}(\mathcal{D}_n)$ es un álgebra numerable tal que $\sigma(\mathcal{A}) = \mathscr{B}_{[0,1)}$.

Definición B.4. Un **espacio medible** es una pareja $(X, \mathscr{F})$ en la que X es un conjunto no vacío y $\mathscr{F}$ es una σ-álgebra de subconjuntos de X.

Definición B.5. Sea $(X, \mathscr{F})$ un espacio medible. Una **medida** en $(X, \mathscr{F})$ es una función $\mu : \mathscr{F} \to \mathbb{R} \cup \{\infty\}$ con las siguientes propiedades:

(a) $\mu(\emptyset) = 0$.

(b) $\mu(F) \geq 0$ para toda $F \in \mathscr{F}$.

(c) μ es σ-**aditiva**, es decir, si (F_n) es una sucesión de elementos disjuntos de $\mathscr{F}$, entonces:

$$\mu\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mu(F_n).$$

Ejemplo B.2. Sea $X \neq \emptyset$, $\mathscr{F} = \mathcal{P}(X)$, definimos $\mu : \mathcal{P}(X) \to \mathbb{R} \cup \{\infty\}$ como sigue $\mu(F) = |F|$ (cardinalidad de F), entonces μ es una medida llamada **medida de conteo** en X.

Definición B.6. Sea μ una medida en $(X, \mathscr{F})$. Se dice que μ es **finita** si no toma el valor extendido $+\infty$. Si es posible hallar una sucesión (F_n) de elementos de $\mathscr{F}$ tal que $X = \bigcup_{n=1}^{\infty} F_n$ y $\mu(F_n) < +\infty$ para toda $n \in \mathbb{N}$, entonces diremos que μ es σ-**finita**. Claramente toda medida finita es automáticamente σ-finita.

Definición B.7. Un **espacio de medida** es una terna $(X, \mathscr{F}, \mu)$, en la que $(X, \mathscr{F})$ es una espacio medible y $\mu : \mathscr{F} \to \mathbb{R} \cup \{\infty\}$ es una medida.

Definición B.8. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida llamamos a $(X, \mathscr{F}, \mu)$ un **espacio de probabilidad**, si $\mu(X) = 1$.

Proposición B.2. *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida σ-finito, entonces*

$$\mu(F) = \sup_{G \subset F, \mu(G) < \infty} \mu(G) \text{ para toda } F \in \mathscr{F}.$$

En varias ocasiones conocemos cómo son los elementos de un álgebra $\mathcal{A}$, sin embargo, no tenemos la misma suerte para la $\sigma(\mathcal{A})$, pero podemos aproximarnos a los elementos de $\sigma(\mathcal{A})$ con elementos de $\mathcal{A}$. El siguiente lema formaliza lo anterior.

Lema B.1. *Sea $F \in \mathscr{F}$ con $\mu(F) < \infty$. Si $\mathcal{A}$ es un álgebra de conjuntos tal que $\sigma(\mathcal{A}) = \mathscr{F}$ entonces dada $\varepsilon > 0$ existe $A = A(\varepsilon) \in \mathcal{A}$ tal que $\mu(F \triangle A) < \varepsilon$.*

Hay otra manera de describir a los elementos de $\sigma(\mathcal{A})$, donde $\mathcal{A}$ es un álgebra.

Definición B.9. Sean $X \neq \emptyset$, $\mathbb{A} \subset \mathcal{P}(X)$. $\mathbb{A}$ se llama **clase monótona**, si para toda $\{A_n\} \subset \mathbb{A}$ sucesión creciente se tiene que $\bigcup_{n=1}^{\infty} A_n \in \mathbb{A}$ y para toda $\{B_n\} \subset \mathbb{A}$ sucesión decreciente se tiene que $\bigcap_{n=1}^{\infty} B_n \in \mathbb{A}$. Note que la intersección arbitraria de clases monótonas también es una clase monónota, por lo que se puede considerar a la clase monótona generada por una colección de subconjuntos de X.

Teorema B.2. *Sean $X \neq \emptyset$ y $\mathcal{A}$ un álgebra de subconjuntos de X. Entonces $\sigma(\mathcal{A})$ es igual a la clase monóntona generada por $\mathcal{A}$.*

Definición B.10. A la σ-álgebra de subconjuntos de $\mathbb{R}$ generada por los conjuntos abiertos de $\mathbb{R}$, la llamaremos σ-álgebra de Borel y la denotaremos como $\mathscr{B}_{\mathbb{R}}$.

Esta σ-álgebra es utilizada tanto en probabilidad como en Teoría de la Medida. Ésta también es generada por la colección de los intervalos cerrados, semicerrados de subconjuntos de $\mathbb{R}$, incluso también por la colección de intervalos de la forma $(-\infty, a)$, $(-\infty, a]$, (a, ∞), $[a, \infty)$.

En varios ejemplos no se utilizó a $\mathscr{B}_{\mathbb{R}}$ si no un subconjunto de ella, por lo que será conveniente introducir la siguiente definición:

Definición B.11. Sea $A \subset \mathbb{R}$, la σ-álgebra de Borel de A, denotada por $\mathscr{B}_A$, se define como sigue:

$$\mathscr{B}_A = \{A \cap B : B \in \mathscr{B}_{\mathbb{R}}\}.$$

Por ejemplo $\mathscr{B}_{(0,1)} = \mathscr{B}_{\mathbb{R}} \cap (0, 1)$.

Definición B.12. Sea $(X, \mathscr{F})$ un espacio medible, a una función $f : X \to \mathbb{R}$ la llamaremos $\mathscr{F}$-medible si $f^{-1}(B) \in \mathscr{F}$ para toda $B \in \mathscr{B}_{\mathbb{R}}$, o si $f^{-1}(I) \in \mathscr{F}$ para toda $I \in \mathcal{I}$, donde $\mathcal{I}$ puede ser cualquier colección de intervalos de $\mathbb{R}$ mencionada en el párrafo anterior. A una función $f : X \to \mathbb{C}$ la llamare $\mathscr{F}$-medible si su parte real y su parte imaginaria son $\mathscr{F}$-medibles.

Definición B.13. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida fijo. Denotaremos por $\mathbb{M}(X, \mathscr{F})$ y $\mathbb{M}^+(X, \mathscr{F})$ al conjunto de funciones reales $f : X \to \mathbb{R}$ que son $\mathscr{F}$-medibles, y al subconjunto de éste consistente de las funciones no-negativas (respectivamente).

Definición B.14. Sea $(X, \mathscr{F})$ un espacio medible. A una función $f \in \mathbb{M}(X, \mathscr{F})$ la llamare s-simple si puede escribirse como $\sum_{i=1}^{n} \alpha_i \chi_{F_i}$, donde todas las α_i son distintas, los conjuntos F_i son ajenos dos a dos, y su unión es igual a X. χ_{F_i} denota la función característica de F_i; es decir, $\chi_{F_i}(x) = 1$ si $x \in F_i$ y $\chi_{F_i}(x) = 0$ si $x \notin F_i$. Las detonaremos con la letra s.

Definición B.15. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y $s \in \mathbb{M}^+(X, \mathscr{F})$ una función s-simple tal que $s = \sum_{i=1}^{n} \alpha_i \mu(F_i)$. Definimos **la integral de s con respecto a μ** denotado como $\int s d\,\mu$ como el número no-negativo (posiblemente infinito) $\sum_{i=1}^{n} \alpha_i \mu(F_i)$, éste número no depende de la representación de s.

Lema B.2. *Sea $(X, \mathscr{F})$ un espacio medible. Si $f \in \mathbb{M}^+(X, \mathscr{F})$, entonces existe una sucesión creciente $\{s_n\} \subset \mathbb{M}^+(X, \mathscr{F})$ de funciones s-simples que converge a f.*

Demostración. Considérese las funciones s_n dadas por:

$$s_n(x) = \begin{cases} \dfrac{i-1}{2^n} & \text{si} \quad \dfrac{i-1}{2^n} \leq f(x) < \dfrac{i}{2^n} \text{ para } i = \{1, 2, \ldots, n2^n\} \\ n & \text{si} \quad f(x) \geq n. \end{cases}$$

□

Definición B.16. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida. Si $f \in \mathbb{M}^+(X, \mathscr{F})$, definimos **la integral de f con respecto a μ** como:

$$\int f d\,\mu = \lim_{n\to\infty} \int s_n d\,\mu.$$

Donde $\{s_n\}$ es la sucesión del lema anterior. Note que esta definición es independiente de la sucesión $\{s_n\}$. Si $F \in \mathscr{F}$, definimos **la integral de f con respecto a μ en $\mathscr{F}$** como:

$$\int_F f d\,\mu = \int f \cdot \chi_F d\,\mu.$$

Teorema B.3 (De la convergencia monótona (1906))**.** *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y (f_n) una sucesión no decreciente de funciones en $\mathbb{M}^+(X, \mathscr{F})$ tal que converge a f, entonces:*

$$\int_F f\, d\mu = \lim_{n\uparrow\infty} \int_F f_n\, d\mu \quad \textit{para toda} \quad F \in \mathscr{F}.$$

Corolario B.3.1 (Lema de Fatou)**.** *Sea $\{f_n\}$ una sucesión de funciones en $\mathbb{M}^+(X, \mathscr{F})$ entonces*

$$\int_F \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_F f_n d\mu \quad \textit{para todo} \quad F \in \mathscr{F}.$$

Definición B.17. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida, $f \in \mathbb{M}(X, \mathscr{F})$, definimos $f^{+}(x) = \text{máx}\{f(x), 0\}$ y $f^{-}(x) = \text{máx}\{-f(x), 0\}$. Note que $f^{+}, f^{-} \in \mathbb{M}^{+}(X, \mathscr{F})$ y $f = f^{+} - f^{-}$. Diremos que f es **integrable con respecto a** μ si $\int f^{+} d\mu < \infty$ y $\int f^{-} d\mu < \infty$. Definimos:

$$\int f d\mu = \int f^{+} d\mu - \int f^{-} d\mu.$$

Diremos que $g : X \to \mathbb{C}$ es integrable si su parte real y su parte imaginaria lo son, y defino:

$$\int g d\mu = \int \Re(g) d\mu + i \int \Im(g) d\mu.$$

Definición B.18. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida. Denotaremos como $\mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$ o $\mathcal{L}^{1}_{\mathbb{R}}(\mu)$ a las funciones en $\mathbb{M}(X, \mathscr{F})$ que sean integrables con respecto a μ. Análogamente denotaremos $\mathcal{L}^{1}_{\mathbb{C}}(X, \mathscr{F}, \mu)$ o $\mathcal{L}^{1}_{\mathbb{C}}(\mu)$ al conjunto de las funciones $g : X \to \mathbb{C}$ que sean integrables.

Teorema B.4 (De la convergencia acotada)**.** *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida finita, $\{f_n\}$ una sucesión de funciones en $\mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$ tal que $f_n(x) \to f(x)$ (casi dondequiera relativo a μ) para alguna $f \in M(X, \mathscr{F})$. Si existe $k > 0$ tal que $|f_n| < k$ casi dondequiera para toda $n > 0$ entonces:*

(i) $f \in \mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$.

(ii) $\int\limits_F f = \lim\limits_{n\to\infty} \int\limits_E f_n$ *para todo* $F \in \mathscr{F}$.

Teorema B.5 (De la convergencia dominada de Lebesgue 1910)**.** *Sea $(X, \mathscr{F}, \mu)$ un espacio de medida, $\{f_n\}$ una sucesión de funciones en $\mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$ tal que $f_n(x) \to f(x)$ (casi dondequiera relativo a μ) para alguna $f \in M(X, \mathscr{F})$. Si existe $g \in \mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$ tal que $|f_n| \leq g$ (casi dondequiera relativo a μ) para toda $n > 0$ entonces:*

(i) $f \in \mathcal{L}^{1}_{\mathbb{R}}(X, \mathscr{F}, \mu)$.

(ii) $\int\limits_F f = \lim\limits_{n\to\infty} \int\limits_E f_n$ *para todo* $F \in \mathscr{F}$.

Definición B.19. Sea $X \neq \emptyset$ y $\mathcal{A} \subset \mathcal{P}(X)$ una álgebra. Una **casi-medida** es una función conjuntista $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ tal que:

(a) $\mu(\emptyset) = 0$.

(b) $\mu(A) \geq 0$, para toda $A \in \mathcal{A}$.

(c) Si (A_n) es una sucesión disjunta de elementos de $\mathcal{A}$ tal que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, entonces: $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{i=1}^{\infty} \mu(A_n)$.

Definición B.20. Una **medida exterior** es una función conjuntista $\rho : \mathcal{P}(X) :\to \overline{\mathbb{R}}$ tal que

(a) $\rho(\emptyset) = 0$.

(b) $\rho(F) > 0$ para toda $F \in \mathcal{P}(X)$.

(c) $\rho(E) \leq \rho(F)$ si $E \subset F \subset X$ (monotonía).

(d) $\rho\left(\bigcup\limits_{n=1}^{\infty} F_n\right) \leq \sum\limits_{n=1}^{\infty} \mu(F_n)$ para toda sucesión $\{E_n\}$ de subconjuntos de X (σ-sub-aditividad).

Como $\rho(\emptyset) = 0$, entonces ρ también es sub-aditiva. Si $\rho(X) < \infty$ entonces ρ se llama finita. Si $X = \bigcup\limits_{n=1}^{\infty} F_n$ con $\rho(F_n) < \infty$ para toda n, entonces la medida exterior se llama σ-finita.

Definición B.21. Sea $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ una casi medida. Definimos $\mu^* : \mathcal{P}(X) \to \mathbb{R} \cup \{\infty\}$ como sigue: $\mu^*(E) = \text{ínf}\{\sum_{n=1}^{\infty} \mu(A_n) : E \subset \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A}\}$ para toda $E \subset X$.

Teorema B.6. $\mu^* : \mathcal{P}(X) \to \mathbb{R} \cup \{\infty\}$ *es una medida exterior tal que* $\mu|_{\mathcal{A}} = \mu$.

Teorema B.7 (De extensión de K. Carathéodory-E.Hopf (1918))**.** *Sea* $\rho : \mathcal{P}(X) \to \overline{\mathbb{R}}$ *una medida exterior, entonces*

(i) $\mathcal{A}^{\rho}$ *es una* σ*-álgebra.*

(ii) $\overline{\rho} = \rho|_{\mathcal{A}^{\rho}} \to \overline{\mathbb{R}}$ *es una medida completa.*

En caso de que $\rho = \mu^*$ *sea la medida exterior generada por una casi medida* $\mu : \mathcal{A} \to \overline{\mathbb{R}}$*, entonces*

(iii) $\sigma(\mathcal{A}) \subset \mathcal{A}^*$.

Teorema B.8 (Teorema de extensión de Hahn)**.** *Sean* $\mu : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$ *una casi media* σ*-finita, entonces existe una única extensión de* μ *a una medida en* $\mathcal{A}^*$.

Definición B.22. Consideramos los conjuntos de la forma $\{(a, b], (-\infty, b], (a, \infty]\}$, sea $\mathcal{A}$ =unión finita y disjunta de los intervalos de la forma anterior. Es fácil ver que $\mathcal{A}$ es un álgebra y $\lambda : \mathcal{A} \to \mathbb{R} \cup \{\infty\}$, donde para cada $A \in \mathcal{A}$ λ le asigna su longitud (∞ si existe un intervalo no acotado en A) es una casi-medida σ-finita, sea $\lambda^* : \mathcal{P}(\mathbb{R}) \to \mathbb{R} \cup \{\infty\}$ la medida exterior generada por λ. Por el teorema anterior existe una $\overline{\lambda} : \mathcal{A}^{\lambda^*} \to \mathbb{R} \cup \{\infty\}$ una medida completa. A $\overline{\lambda}$ la denotaremos como λ, y es conocida como la **medida de Lebesgue**; y a $\mathcal{A}^{\lambda^*}$ la denotaremos como $\mathscr{A}_{\mathbb{R}}^*$ y son conocidos como **los Lebesgue-medibles**. Esta medida fue frecuentemente utilizada en el trabajo.

Definición B.23. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida, a dos funciones $\mathscr{F}$-medibles f,g tales que $\mu(\{x \in X : f(x) \neq g(x)\}) = 0$, diremos que $f = g$ casi dondequiera. En general si una proposición $P(x)$ referente a $x \in X$ se cumple fuera de un conjunto $N \in \mathscr{F}$ de medida cero diremos que $P(x)$ se cumple **casi dondequiera relativa a** μ.

Definición B.24. Sea $(X, \mathscr{F}, \mu)$ un espacio de medida y $p \in [1, \infty)$. Definimos

$$L_p(X, \mathscr{F}, \mu) = \{f \in \mathbb{M}(X, \mathscr{F}) : |f|^p \in L_1(\mu)\}$$

también lo denotaremos como $L_p(\mu)$.

Teorema B.9 (Convergencia dominada en medida)**.** *Sean $(X, \mathscr{F}, \mu)$ un espacio de medida, $p \in [\,1\,, +\infty\,)$ fijo y $(f_n) \subset L_p(\mu)$ una sucesión de funciones tal que $f_n \underset{\mu}{\longrightarrow} f \in M(X, \mathscr{F})$. Si existe $g \in L_1^+(\mu)$ tal que $|f_n|^p \leq g$ casi dondequiera para toda $n \in \mathbb{N}$, entonces $f \in L_p(\mu)$ y $f_n \underset{L_p}{\longrightarrow} f$.*

Teorema B.10 (D. F. Egorov, 1913)**.** *Sea $(X\mathscr{F}, \mu)$ un espacio de medida **finita**, $\{f_n\}$ una sucesión y $f \in \mathbb{M}(X, \mathscr{S})$ tal que $f_n \to f$ casi dondequiera, entonces $f_n \to f$ casi uniformemente.*

Definición B.25. Sea $(X, \mathscr{F})$ un espacio de medida y $\mu_1, \mu_2 : \mathscr{F} \to \mathbb{R} \cup \{\infty\}$ dos medidas en $(X, \mathscr{F})$; diremos que μ_1 y μ_2 son **mutuamente singulares**, denotado $\mu_1 \perp \mu_2$ si existe un $F \in \mathscr{F}$ tal que $\mu_1(F) = 0$ y $\mu_2(X \setminus F) = 0$. Finalmente diremos que μ_2 es **absolutamente continua con respecto a** μ_1, denotado $\mu_2 \ll \mu_1$, si $\mu_2(F) = 0$ siempre que $\mu_1(F) = 0$ donde $F \in \mathscr{F}$.

Definición B.26. Sea $(X, \mathscr{F})$ un espacio medible. Una **medida con signo** en $(X, \mathscr{F})$ es una función conjuntista $\nu : \mathscr{F} \to \mathbb{R}$ tal que:

(a) $\nu(\emptyset) = 0$

(b) ν toma a lo más uno de los valores extendidos $+\infty$ o $-\infty$.

(c) Si (F_i) es una sucesión disjunta de elementos de S, entonces: $\nu\left(\bigcup\limits_{i=1}^{\infty} F_i\right) = \sum\limits_{i=1}^{\infty} \nu(F_i)$, en donde la igualdad anterior debe entenderse como sigue:

Si $|\nu\left(\bigcup_{i=1}^{\infty} F_i\right)| < \infty$ entonces la serie converge absolutamente y si $\nu\left(\bigcup_{i=1}^{\infty} F_i\right) = \infty$ o $-\infty$, entonces la serie converge (en el sentido extendido) a ∞ o $-\infty$ (respectivamente). Diremos que ν es finita si $|\nu(F)| < \infty$ para toda $F \in \mathscr{F}$ y σ-finita si existe (F_n) una sucesión de elementos de $\mathscr{F}$ tal que $X = \bigcup_{n=1}^{\infty} F_n$ y $|\nu(F_n)| < +\infty$ para toda n.

Definición B.27 (Medidas producto)**.** Sean $(X, \mathscr{F})$ y $(Y, \mathscr{G})$ dos espacios de medida. Definimos el **espacio medible producto** denotado $(X \times Y, \mathscr{F} \otimes \mathscr{G})$, donde $\mathscr{F} \otimes \mathscr{G}$ es la σ-álgebra de subconjuntos de $X \times Y$ generada por $\mathscr{F} \times \mathscr{G}$, es decir, $\sigma(\mathscr{F} \times \mathscr{G}) = \mathscr{F} \otimes \mathscr{G}$.

Definición B.28. Sea $H \subset X \times Y$ y $x \in X$, definimos la **x-sección** de H (denotada como H_x), como el subconjunto de Y definido como sigue: $H_x = \{y \in Y : (x, y) \in H\}$. De manera similar podemos definir la **y-sección** de H (denotada como H^y), como el subconjunto de X definido como sigue: $H^y = \{x \in X : (x, y) \in H\}$.

Se puede probar que si $H \in \mathscr{F} \otimes \mathscr{G}$ entonces $H_x \in \mathscr{G}$ y $H^y \in \mathscr{F}$. El siguiente resultado nos garantiza la existencia de una medida en el espacio producto con la propiedad de que la medida de rectángulos medibles es el producto de sus medidas.

Teorema B.11 (De la medida producto). *Sean $(X, \mathscr{F}, \mu)$ y $(Y, \mathscr{G}, \nu)$ dos espacios de medida σ-finita, entonces la función conjuntista $\pi : \mathscr{F} \otimes \mathscr{G} \to \mathbb{R}$ definida por:*

$$\pi(H) = \int \nu(H_x) d\mu = \int \mu(H^y) d\nu$$

*es una medida σ- finita, y es la **única** medida sobre $\mathscr{F} \otimes \mathscr{G}$ con la propiedad de que: $\pi(A \times B) = \mu(A)\nu(B)$ para todo $A \times B \in \mathscr{F} \otimes \mathscr{G}$. Llamaremos a π la medida producto de μ y ν y la denotaremos por $\pi = \mu \otimes \nu$.*

La demostración del teorema anterior se encuentra en Halmos [11].

Teorema B.12. *Sean $(X, \mathscr{F}, \mu)$, $(Y, \mathscr{G}, \nu)$ dos espacios de medida σ-finita, $f \in \mathbb{M}^+(X \times Y, \mathscr{F} \otimes \mathscr{G})$ y $F \times G \in \mathscr{F} \times \mathscr{G}$. Entonces:*

$$\int_{F \times G} f d\,(\mu \otimes \nu) = \int_F \left(\int_G f d\,\nu \right) d\,\mu = \int_G \left(\int_F f d\,\mu \right) d\,\nu.$$

En donde cada término puede ser $+\infty$.

Se puede extender el Teorema de la Medida Producto, para cuando tenemos n espacios de medida σ-finita. Sin embargo cuando tenemos un conjunto numerable de espacios de medida, no es tan sencillo construir el espacio de medida producto. De hecho, sólo podremos considerar el caso en el que se que los espacios sean de probabilidad. El teorema que garantiza dicha construcción es el Teorema de Consistencia de Kolmogorov.

Teorema B.13 (De Consisistencia de Kolmogorov). *Sean $\mu_1, \mu_2, \ldots,$ una sucesión de medidas de probabilidad sobre $(\mathbb{R}, \mathscr{B}_{\mathbb{R}}), (\mathbb{R} \times \mathbb{R}, \mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathbb{R}}), \ldots,$ respectivamente, tales que:*

$$\mu_{n+1}(B \times \mathbb{R}) = \mu_n(B) \text{ para toda } B \in \bigotimes_{i=1}^{n} \mathscr{B}_{\mathbb{R}}, n = 1, 2, \ldots \tag{B.1}$$

Entonces existe una única medida de probabilidad μ sobre $(\prod_{n=1}^{\infty} R, \bigotimes_{n=1}^{\infty} \mathscr{B}_{\mathbb{R}})$ tal que:

$$\mu(B \times \prod_{i=n+1}^{\infty} \mathbb{R}) = \mu_n(B) \text{ para toda } B \in \bigotimes_{i=1}^{n} \mathscr{B}_{\mathbb{R}}, n = 1, 2, \ldots$$

Ejemplo B.3. Sean $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad, $\{f_n\}_{n=1}^{\infty}$ una sucesión de variables aleatorias integrables independientes e idénticamente distribuidas (es decir, $\{f_n\} \subset L_1^{\mathbb{R}}(\mu)$, $\mu(f_n^{-1}(B))$, no depende de n para toda $B \in \mathscr{B}_{\mathbb{R}}$ y

$$\mu(f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \cap \cdots \cap f_n^{-1}(B_n)) = \mu(f_1^{-1}(B_1))\mu(f_2^{-1}(B_2)) \cdots \mu(f_n^{-1}(B_n))$$

para todo n, para toda $B_1, B_2, \cdots B_n \in \mathscr{B}_{\mathbb{R}}$). Defino:

$$\mu_n(B_1 \times B_2 \cdots \times B_n) = \mu(f_1^{-1}(B_1))\mu(f_2^{-1}(B_2)) \cdots \mu(f_n^{-1}(B_n)), \text{ donde } B_1, B_2, \cdots B_n \in \mathscr{B}_{\mathbb{R}},$$

y la extendemos a todos los elementos de $\bigotimes_{i=1}^{n} \mathscr{B}_{\mathbb{R}}$. Así pues, $\mu_1, \mu_2, \ldots$ son medidas de probabilidad sobre $(\mathbb{R}, \mathscr{B}_{\mathbb{R}}), (\mathbb{R} \times \mathbb{R}, \mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathbb{R}}), \ldots,$ que cumplen la propiedad B.1. Por el teorema anterior existe una única medida de probabilidad ν sobre $(\prod_{n=1}^{\infty} R, \bigotimes_{n=1}^{\infty} \mathscr{B}_{\mathbb{R}})$, que coincide con μ_n en los cilindros. Note que en este caso los cilindros son de la forma $B \times \prod_{i=n+1}^{\infty} \mathbb{R}$, donde $B = B_1 \times \cdots \times B_n$ y $B_1, \ldots, B_n \in \mathbb{B}_{\mathbb{R}}$.

Observación B.1. Se puede extender el teorema anterior, para $\mu_1, \mu_2, \ldots,$ medidas de probabilidad tales que cumplen la propiedad B.1 sobre espacios medibles $(\Omega_i, \mathscr{F}_i)$ $i = 1, 2, \ldots,$ donde $(\Omega_i, \mathscr{F}_i)$ es un espacio métrico separable completo y $\mathscr{F}_i$ es la σ-álgebra de Borel. También la propiedad B.1 es equivalente a las condiciones mencionadas en la Observación 1.6.

Definición B.29 (Esperanza Condicional)**.** Sea $(\Omega, \mathscr{F}, \mu)$ un espacio de probabilidad, $f \in L_1(\Omega, \mathscr{F}, \mu)$ y $\mathscr{G}$ una sub-σ-álgebra de $\mathscr{F}$. La esperanza condicional de f dado $\mathscr{G}$ es una función denotada por $E(f/\mathscr{G})$ que cumple:

(a) $E(f/\mathscr{G})$ es $\mathscr{G}$-medible.

(b) $E(f/\mathscr{G}) \in L_1(\Omega, \mathscr{G}, \mu)$.

(c) $\int\limits_G E(f/\mathscr{G}) = \int\limits_G f$ para toda $G \in \mathscr{G}$.

Bibliografía

[1] Bartle Robert G. *Introduction to Real Analysis*. John Wiley & Sons, 2000.

[2] Bartle Robert G. *The Elements of Real Analysis*. John Wiley & Sons, 1976.

[3] Bartle Robert G. *The Elements of Integration and Lebesgue measure*. John Wiley & Sons, 1995.

[4] Billingsley P. *Ergodic theory and information*. Wiley, 1956.

[5] Brin Michael, Stuck Garrett . *Introduction to dynamical systems*. Cambridge University Press, 2002.

[6] Caballero M. E; Rivero V. M; Uribe Bravo G; Velarde C. *Cadenas de Markov: Un enfoque elemental*. Sociedad Matemática Mexicana, 2008.

[7] Cornfeld I.P; Fomin S.V; Sinai Ya.G. *Ergodic theory*. Springer-Verlag, 1982. (Traducido del Russo).

[8] Friedberg Stephen H; Insel Arnold J; Spence Lawrence E. *Linear Algebra*. Prentice Hall, 1989.

[9] Gantmacher F.R; *Applications of the theory of matrices*. Interscience, 1959. (Traducido del Russo).

[10] Halmos. *Introduction to Hilbert Space and Theory of Spectral Multiplicity*. Chelsea, 1957.

[11] Halmos. *Measure Theory*. Springer-Verlag, 1974.

[12] Halmos. *Lectures on Ergodic Theory*. Chelsea, 1956.

[13] Hazewinkel, Michiel (Editor). *Encyclopaedia of Mathematics*. Springer-Verlag, 2002.

[14] Khinchin A. I. *Mathematical foundations of information theory*. Dover Publications, 1957.

[15] Ornstein D. S. *Ergodic theory, Randomness and Dynamical Systems*. Yale University Press, 1974.

[16] Parry William. *Entropy and Generators in Ergodic Theory*. W. A. Benjamin, 1969.

[17] Parry William. *Topics in ergodic theory*. Cambridge University Press, 2004.

[18] Pollicott Mark, Yuri Michiko. *Dynamical systems and ergodic theory*. Cambridge University Press, 1998.

[19] Royden H. L. *Real Analysis*. Macmillan, 1988.

[20] Shannon C. A mathematical theory of communication. *Bell System Tech. J.* **27**, 379-423, 623-656, 1948.

[21] Shiryaev A. N. *Probability*. Springer-Verlag, 1996.

[22] Silva César Ernesto. *Invitation to ergodic theory*. American Mathematical Society (AMS), 2008.

[23] Walters Peter. *An Introduction to Ergodic Theory*. Springer-Verlag, 2000.