



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE CIENCIAS

SOBRE EL TEOREMA DE KOLMOGOROV-SINAI

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN MATEMÁTICAS

P R E S E N T A :

AZUARA ROMERO FABIÁN

TUTOR

DR. CARLOS ALFONSO CABRERA OCAÑAS



CIUDAD UNIVERSITARIA, CDMX, 2021



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*«La felicidad es inversamente proporcional a la distancia que hay entre lo que soy y lo que quiero ser. Soy matemático, toda mi vida quise ser matemático, por lo tanto: soy inmensamente feliz»*

Alberto Barajas Celis

# Agradecimientos

Dedico este trabajo a mi hermano Andrés y a Isabel mi madre, que el final de esta etapa sea el comienzo de nuevas alegrías.

Al Dr. Carlos Cabrera quien me orientó de manera paciente y dedicada durante todo el proceso de esta tesis.

# Índice general

<b>Agradecimientos</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Preliminares.</b>	<b>6</b>
2.1. Particiones Medibles . . . . .	6
2.2. Álgebras de particiones . . . . .	9
2.3. Propiedades sobre $L_2$ . . . . .	11
2.4. Sistemas canónicos de medida . . . . .	12
<b>3. Entropía de particiones medibles.</b>	<b>16</b>
<b>4. Entropía condicional media.</b>	<b>28</b>
<b>5. Espacios de particiones.</b>	<b>37</b>
<b>6. Entropía de un endomorfismo.</b>	<b>42</b>
6.1. Teorema de Kolmogorov-Sinai . . . . .	48
<b>7. Aplicación del teorema de Kolmogorov-Sinai.</b>	<b>50</b>
7.1. Construcción de las medidas de Markov y de Bernoulli. . . . .	50
7.2. Sobre el teorema de Kolmogorov-Sinai. . . . .	60
<b>8. Conclusiones.</b>	<b>63</b>
<b>Bibliografía</b>	<b>65</b>

# 1 Introducción

El término entropía es usado en múltiples campos de la ciencia que no siempre están relacionados con la física o las matemáticas (como la sociología) en donde su carácter cuantitativo se deja a un lado y se le asocia de manera más intuitiva una noción de desorden, caos o pérdida de la diversidad, entre otros significados.

Uno de los objetivos de este trabajo es tener una interpretación más natural de esta palabra sin perder de vista que es una magnitud cuantitativa, cuyo valor nos da información sobre un determinado fenómeno. Para ello, tomaremos como punto de partida y como antecedente de todo este trabajo la implementación de la teoría de la probabilidad en su definición.

Nuestro objetivo principal en esta tesis es dar una demostración del teorema de Kolmogorov-Sinai, explicando con detalle qué es la entropía dinámica y el porqué este teorema juega un papel fundamental para su cálculo de manera económica, siendo un resultado importante dentro del contexto de la teoría ergódica y un buen punto de partida para el alumno egresado de la licenciatura en matemáticas con intereses en teoría ergódica.

Para dar inicio con nuestra labor, debemos recordar que el matemático alemán Rudolf Clausius fue el primero en utilizar el término entropía, asociándolo con base en el trabajo de Nicolas Léonard Sadi Carnot a la energía que no se puede convertir en calor.

Tiempo después James Maxwell y Ludwig Boltzmann trabajaron en un modelo atómico de calor basado en probabilidad. Esto lo hacían considerando la función de distribución del estado fase de una de las partículas de un gas, llamada microestado, que en conjunto con el total de partículas forman la distribución del total llamada macroestado. Esta idea permite tomar sobre el estado fase una partición finita de pequeñas celdas y asociar un vector  $p$  con entradas  $p_i$  en el que se muestra la proporción de partículas en la celda  $i$ .

Con base en lo anterior, Boltzmann concebía a la entropía como una medida de la incertidumbre sobre el estado en el que permanece un sistema conociendo solo  $p_i$  y lo

media con el logaritmo del número de estados (microestados) que puede tomar esta distribución.

Para  $n$  partículas esto es:

$$\frac{\prod_i (np_i)!}{n!}.$$

Cálculando el límite cuando  $n \rightarrow \infty$  y usando la fórmula de Stirling se tiene:

$$\lim_{n \rightarrow \infty} \frac{\prod_i (np_i)!}{n!} = \sum_i p_i \log(p_i). \quad (1.1)$$

Basandose en lo anterior, Claude Elwood Shannon inició su trabajo considerando alfabetos con símbolos finitos, donde cada elemento tenía una probabilidad  $p_i$  de ser elegido en una extracción al azar; buscando una medida para la incertidumbre al momento de hacer una elección, demostró que a gran escala la parte derecha de la ecuación (1.1) era la única función continua de  $p_i$  que crecía en  $n$  cuando la distribución era uniforme. Interpretó la aditividad de esta función como el resultado de la suma de distribuciones de probabilidad independientes. Además, notó que en procesos estacionarios con una cantidad finita de estados, las probabilidades dependen de sus precedentes y que el efecto de la longitud de la cantidad de símbolos a elegir es capturado por la distribución conjunta de información.

El trabajo de Shannon es la piedra angular en lo que hoy se conoce como teoría de la información y dentro de este contexto, al igual que Boltzmann, entendía a la entropía como una medida de incertidumbre. Para entender un poco más de su trabajo partamos de lo siguiente:

De manera similar a Boltzmann, consideramos un vector de probabilidad  $p$ , el cual es una sucesión finita de números reales entre 0 y 1, es decir,  $p = (p_1, p_2, p_3, \dots, p_n)$  donde cada  $p_i \in [0, 1]$ . Luego, definimos la entropía de Shannon para el vector  $p$  como sigue:

$$H(p) = - \sum_i^n p_i \log_2 p_i.$$

Por otro lado, consideramos una partición  $\xi$  de un espacio de probabilidad con medida  $\mu$ , de manera que los conjuntos  $\{A_1, A_2, A_3, \dots, A_n\}$  son disjuntos dos a dos y su unión es el espacio total. Entonces las probabilidades  $p_i = \mu(A_i)$  forman un vector de probabilidad, por lo que al asociar la entropía de Shannon a este se tiene lo siguiente:

$$H_\mu(\xi) = H(p_\xi).$$

Esta última idea será formalizada en el capítulo 3 y estudiaremos varias de sus propiedades.

En este punto Shannon definió la información asociada a un conjunto  $A$  como se muestra a continuación:

$$I(A) = -\log_2(\mu(A))$$

por lo que de manera formal, suponiendo que  $I(A_i)$  es constante para todos los puntos  $\omega \in A_i$  definimos la función de información  $I_\xi(\omega)$  asociada a la partición  $\xi = \{A_1, A_2, A_3, \dots, A_n\}$  como se muestra en la siguiente ecuación:

$$I_\xi(\omega) = \sum_{i=1}^n -\log_2(\mu(A_i))1_{A_i}(\omega).$$

Es gracias a esa última definición que podemos considerar a la entropía como la información promedio de la partición.

$$H_\mu(\xi) = E_\mu(I_\xi).$$

Con base en lo ya expuesto deberíamos notar que a partir de una partición  $\xi$  y de los elementos  $\omega$  del espacio de probabilidad, la información que se obtiene es la respuesta a la pregunta ¿en cuál  $A_i$  se encuentra el elemento  $\omega$  dado? También debe observarse que, a menos que la partición solo contenga dos elementos, la respuesta a esta pregunta es no binaria, sin embargo, sí podemos reemplazar estas preguntas por unas que lo sean de la forma ¿Me encuentro en el conjunto  $A_k$ ?, cuya respuesta solo puede tomar los valores sí o no. En conclusión, la entropía debería ser interpretada como la cantidad esperada (esperanza) de información necesaria para localizar un punto en la partición.

Adicionalmente, debemos tener en cuenta que un bit de información es equivalente a la respuesta de una pregunta binaria, con esto, cada bit debería interpretarse como una pregunta realizada con el fin de encontrar el lugar de  $\omega$ . Todo esto debería quedar más claro con el siguiente ejemplo:

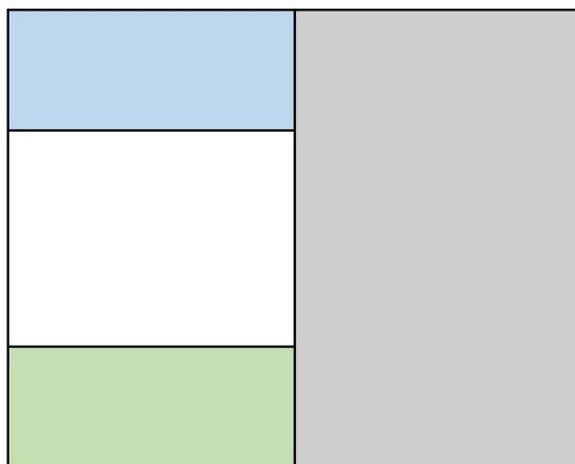


Figura 1.1: Partición de un cuadrado unitario.

Consideremos a la figura anterior, que representa una partición del cuadrado unitario, donde el elemento  $A_1$  corresponde a la sección azul,  $A_2$  a la blanca,  $A_3$  a la verde y

$A_4$  a la gris. De modo que las probabilidades para estos conjuntos están dadas por  $\frac{1}{8}$ ,  $\frac{1}{4}$ ,  $\frac{1}{8}$  y  $\frac{1}{2}$  respectivamente.

La función de información para cada uno de ellos está dada por:

$$I(A_1) = -\log_2 \left( \frac{1}{8} \right) = 3$$

$$I(A_2) = -\log_2 \left( \frac{1}{4} \right) = 2$$

$$I(A_3) = -\log_2 \left( \frac{1}{8} \right) = 3$$

$$I(A_4) = -\log_2 \left( \frac{1}{2} \right) = 1$$

con esto, la entropía para esta partición, llamémosla  $\xi$ , es la siguiente:

$$H_\mu(\xi) = \frac{1}{8} * 3 + \frac{1}{4} * 2 + \frac{1}{8} * 3 + \frac{1}{2} * 1 = \frac{7}{4}$$

Luego, notemos que con la anterior podemos formular las siguientes preguntas secuencialmente:

i) ¿ $\omega$  está en el lado izquierdo del cuadrado?

La respuesta no localiza a  $\omega$  en  $A_4$  usando un solo bit.

ii) ¿ $\omega$  se localiza en el centro del cuadrado?

La respuesta sí localiza a  $\omega$  en  $A_2$  usando 2 bits.

iii) ¿ $\omega$  está en la parte superior?

Las respuestas sí o no localizan a  $\omega$  en  $A_1$  o  $A_3$  respectivamente.

En este ejemplo la función de información en cada punto coincide con el número de preguntas necesarias para llegar al conjunto en el que se encuentra y  $\frac{7}{4}$  representa el valor esperado de preguntas a realizar tomando un punto arbitrario.

Otra manera de interpretar este concepto es considerando el caso en el que un observador conoce la distribución de una variable aleatoria  $X$  antes de realizar su experimento, entonces, la incertidumbre sobre el valor esperado de la información que él está perdiendo para tener un evento cierto es exactamente  $H_\mu(\xi)$ .

Con estas ideas Shannon abrió el camino para que otros matemáticos como Khinchin o Kolmogorov continuaran con el estudio de la entropía a partir de particiones de espacios de probabilidad y que después permitió el surgimiento de lo que hoy se conoce como entropía dinámica.

Gracias a lo anterior, hemos dado las bases para que el lector tenga una noción probabilística del significado de entropía y de su cálculo.

Un comentario muy importante sobre lo que se presentará a continuación está relacionado con el estilo o perspectiva desde la que abordaremos los capítulos, pues este trabajo está basado en su mayoría en [1], por lo que si bien la exposición que hemos dado de la entropía de Shannon es muy útil, en el trabajo de Rokhlin la construcción de la entropía para una partición parte desde nociones conjuntistas. Esto es importante ya que si se revisa [6] se puede encontrar una versión sintetizada de los elementos necesarios para demostrar el teorema de Kolmogorov-Sinai, sin embargo, el enfoque constructivo de Rokhlin, si bien es un poco confuso al inicio debido a la gran cantidad de cuentas sobre particiones medibles, esto aporta un gran valor en el entendimiento de los resultados. Además, dado que esta tesis está pensada para alumnos que recién terminan con sus cursos de licenciatura en matemáticas, considero bastante adecuado familiarizarse con esta construcción. Salvo, cuando llegue el momento de definir a las particiones medibles (que serán la base de nuestro trabajo), optaremos por la definición dada en [6], debido a que es más moderna y nos ahorra el trabajo de construirla para después probar que cumple con lo que aquí se especifica.

Para terminar con esta sección, daremos una explicación muy corta de los capítulos a trabajar.

- Capítulo 2 En este daremos los elementos necesarios de teoría de la medida para abordar este trabajo.
- Capítulo 3 Iniciaremos con la definición de entropía para una partición medible, lo cual es formalizar lo que hemos expuesto para la entropía de Shannon.
- Capítulo 4 Continuaremos con una extensión de la entropía de particiones medibles para espacios cocientes de estas. El valor conceptual de este capítulo está en entender cómo funciona la entropía sobre los representantes de las clases definidas en una partición con respecto a otra partición medible.
- Capítulo 5 Haremos un breve estudio sobre las propiedades que verifica el conjunto que contiene a las particiones medibles con entropía finita.
- Capítulo 6 Esta sección tiene como objetivo dar una nueva definición de entropía sobre particiones invariantes bajo determinadas transformaciones y por fin llegar al concepto de entropía dinámica. La segunda parte de este capítulo dará la demostración del teorema de Kolmogorov-Sinai, la cual, con todo lo que se habrá presentado hasta ese momento, será breve.
- Capítulo 7 En este último apartado realizaremos una reflexión sobre la importancia de este resultado y algunas de sus implicaciones.

Dicho todo lo anterior, iniciamos con la exposición de esta tesis.

## 2 Preliminares.

El objetivo de esta sección es presentar los elementos necesarios para el desarrollo de este trabajo en las siguientes secciones. Para la comprensión de este texto se requieren conocimientos básicos de un primer curso de teoría de la medida o al menos que el lector conozca las definiciones de álgebra,  $\sigma$ -álgebra, función medible, espacio medible, medida de Lebesgue, espacio de Lebesgue, álgebra de Borel e integral con respecto a una medida. Además, es recomendable que conozca las primeras definiciones de un curso de probabilidad.

Iniciaremos este capítulo presentando algunas nociones para funciones entre espacios medibles, nuestro interés en futuras secciones recaerá en la transformaciones que van de un espacio en sí mismo que preservan medida.

**Definición 2.1.** Sean  $X$  y  $Y$  espacios medibles y  $f : X \rightarrow Y$  una función, llamaremos a  $f$  un **homomorfismo** entre espacios medibles si para todo  $A \subset Y$  medible  $f^{-1}[A]$  es medible y tiene la misma medida que  $A$ .

Un **isomorfismo** de espacios medibles es un homomorfismo uno a uno cuya inversa también es un homomorfismo. Si los espacios  $X$  y  $Y$  son el mismo, un homomorfismo será llamado **endomorfismo** y un isomorfismo **automorfismo**.

**Definición 2.2.** Dos espacios medibles serán llamados *isomorfos* si es posible construir un isomorfismo entre ellos.

Como se mencionó en la introducción, el estudio de la entropía desde una perspectiva de probabilidad requiere una partición del espacio. Por fines prácticos abordaremos esta partición desde una noción moderna, es decir, consideraremos como punto de partida en la definición de partición medible la  $\sigma$ -álgebra de Borel, sin embargo, como se puede leer en [1], esta también puede ponerse en términos de las bases de un espacio de medida.

### 2.1. Particiones Medibles

**Definición 2.3** (Partición medible). Llamaremos *partición medible de un conjunto  $M$  medible* a una partición  $\xi$  de  $M$  a lo más numerable de subconjuntos de Borel.

Es importante señalar que pese a que en esta definición estamos indicando una cantidad a lo más numerable de subconjuntos de Borel, esta restricción puede eliminarse

para considerar una cantidad no numerable de estos, sin embargo, para fines de este trabajo omitiremos este caso.

La notación para los casos triviales de nuestras particiones:

- i) Una partición de  $M$  en conjuntos unitarios es denotada por  $\epsilon$ .
- ii) La partición trivial tiene como único elemento a  $M$  y es denotada por  $\nu$ .

**Definición 2.4.** *Llamaremos subpartición medible de  $\xi$ , a una partición  $\zeta$  de  $M$  tal que:*

- 1)  $\zeta$  es una partición medible de  $M$ .
- 2) Para cualquier  $U \in \zeta$  existe  $V \in \xi$  tal que  $U \subset V$ .

Escribiremos  $\xi \leq \zeta$  si  $\zeta$  es una subpartición de  $\xi$ .

Cuando se habla de subparticiones de una partición hacemos referencia a una partición más fina que la original. Esta idea de obtener refinamientos a partir de una partición es crucial para la comprensión de este trabajo, pues, como se irá observando más adelante, contar con un mayor nivel de granularidad aumenta la entropía calculada, esto debido a que al tener más conjuntos resulta intuitivo saber que hay una mayor cantidad de preguntas a realizar para identificar la posición de un elemento  $\omega$ .

Las siguientes definiciones muestran que para cualquier cantidad de particiones de un espacio de probabilidad existen dos particiones que acotan inferior y superiormente a todas las demás; esto nos ayudará en un futuro a determinar las particiones óptimas de trabajo.

**Definición 2.5.** *Para cualquier sistema de particiones medibles  $\{\xi_\alpha\}$  definimos el producto  $\vee_\alpha \xi_\alpha$ , como la partición medible  $\xi$  que satisface dos condiciones:*

- 1)  $\xi_\alpha \leq \xi$  para todo  $\alpha$ .
- 2) Si  $\xi_\alpha \leq \xi'$  para todo  $\alpha$  entonces  $\xi \leq \xi'$ .

El producto  $\vee_{\alpha=1}^n \xi_\alpha$  es también denotado por  $\xi_1 \xi_2 \xi_3 \dots \xi_n$ . Cuando se tiene  $\vee_{\alpha=1}^\infty \xi_\alpha$  lo podemos escribir como  $\xi_1 \xi_2 \xi_3 \dots$

**Definición 2.6.** *Para cualquier sistema de particiones medibles  $\{\xi_\alpha\}$  definimos la intersección  $\wedge_\alpha \xi_\alpha$  como la partición medible  $\xi$  que satisface dos condiciones:*

- 1)  $\xi \leq \xi_\alpha$  para todo  $\alpha$ .

2) Si  $\xi' \leq \xi_\alpha$  para todo  $\alpha$  entonces  $\xi' \leq \xi$ .

**Observación 2.1.** Para cualquier familia  $\{\xi_\alpha\}$  de particiones medibles el producto y la intersección existen.

*Demostración.* Probaremos primero el caso del producto.

Definimos

$$P = \left\{ \bigcap_{\alpha} C_{\alpha} \mid C_{\alpha} \in \xi_{\alpha} \right\}.$$

Entonces:

- i) Sea  $\xi_k \in \{\xi_\alpha\}$ . Luego, consideremos  $Q \in P$ , por la definición de  $P$  tenemos que  $Q \subset C_k \in \xi_k$ . Por lo tanto  $P$  refina a todo elemento de  $\{\xi_\alpha\}$ .
- ii) Sea  $P'$  una partición distinta de  $P$  que refina a cada elemento de  $\{\xi_\alpha\}$ . Queremos probar que  $P \leq P'$ . Entonces:  
sea  $S \in P'$  entonces para cada  $\alpha$  existe  $R_\alpha \in \xi_\alpha$  tal que  $S \subset R_\alpha$ .

Luego,  $S = \bigcap_{\alpha} R_{\alpha} \in P$ . Por lo tanto  $P \leq P'$ .

Ahora continuaremos con el caso de la intersección. Para ello definimos:

$$P = \left\{ \bigcup_{\alpha} C_{\alpha} \mid C_{\alpha} \in \xi_{\alpha} \right\}.$$

Entonces:

- i) Sea  $\xi_k \in \{\xi_\alpha\}$ . Luego, consideremos  $Q \in \xi_k$ , por la definición de  $P$  tenemos que  $Q \subset R \in P$ . Por lo tanto  $\xi_\alpha$  refina a todo elemento de  $P$ .
- ii) Sea  $P'$  una partición distinta de  $P$  que es refinada por cada elemento de  $\{\xi_\alpha\}$ . Queremos probar que  $P' \leq P$ . Entonces:  
sea  $S_\alpha \in \xi_\alpha$  para cada  $\alpha$ , entonces como  $\xi_\alpha$  es un refinamiento de  $P'$  para toda  $\alpha$  entonces  $\bigcup_{\alpha} S_{\alpha} \in P'$  pero  $\bigcup_{\alpha} S_{\alpha} = R \in P$ , de modo que  $P' \leq P$ .

□

A continuación, de manera simplificada introducimos un tipo de convergencia de sucesiones de particiones medibles construidas a partir de las operaciones anteriores.

Los símbolos  $\xi_n \nearrow \xi$  indican que:

$$\xi_1 \leq \xi_2 \leq \dots \text{ y } \bigvee_{n=1}^{\infty} \xi_n = \xi$$

De manera similar,  $\xi_n \searrow \xi$  indica que:

$$\xi_1 \geq \xi_2 \geq \dots \text{ y } \bigwedge_{n=1}^{\infty} \xi_n = \xi$$

En [1] encontrará más detalles de la siguiente propiedad.

**Observación 2.2.** *Para cualquier partición medible  $\xi$  existe una sucesión de particiones finitas  $\{\xi_n\}$  tales que  $\xi_n \nearrow \xi$ .*

Las siguientes definiciones tienen como objetivo darnos una notación para identificar conjuntos formados a partir de elementos de una partición medible.

**Definición 2.7.** *Subconjuntos de  $M$  que son uniones de elementos de una partición  $\xi$  son llamados  $\xi$ -conjuntos.*

**Definición 2.8.** *Las particiones medibles  $\epsilon, \eta$ , son llamadas independientes si para cualquier  $\epsilon$ -conjunto  $A$  y cualquier  $\eta$ -conjunto  $B$*

$$\mu(A \cap B) = \mu(A)\mu(B).$$

**Definición 2.9.** *El conjunto  $B$  es independiente de la partición  $\epsilon$  si la ecuación anterior se verifica para cualquier  $\epsilon$ -conjunto  $A$ .*

Cerraremos esta sección introduciendo el concepto de partición inducida, el cual hace referencia al cómo queda particionado un conjunto a partir de una partición  $\xi$  dada. Como se verá más adelante a través de esta definición es posible construir particiones en un conjunto a partir de otra partición dada.

**Definición 2.10.** *Si  $\xi$  es una partición medible de  $M$  y  $C$  es un subconjunto de  $M$ , entonces denotamos por  $\xi_C$  a la partición de  $C$  inducida por  $\xi$ , definida como:*

$$\xi_C = \{E \cap C \mid E \in \xi\}.$$

A partir de lo que resta del capítulo seguiremos presentando varios conceptos junto con algunas de sus propiedades y relaciones entre ellos, sin embargo, si bien todos son resultados interesantes, para nuestros fines solo algunos nos serán útiles más adelante. Pese a esto último, es necesario tener un desarrollo previo para su adecuada comprensión.

Lo anterior es dicho dado que la parte que mayor referencias tendrá a lo largo del texto es la que habla sobre los espacios canónicos de medida y sobre la cual profundizaremos en el apartado 2.4.

## 2.2. Álgebras de particiones

Una vez que se ha establecido la forma en la que particionaremos los conjuntos en los que trabajaremos nos interesa observar cómo se relacionan con las álgebras del espacio.

**Definición 2.11.** *Decimos que dos conjuntos  $A$  y  $B$  son iguales módulo 0, denotado por  $A = B \text{ mod } 0$  si  $A$  y  $B$  son iguales salvo en un conjunto de medida 0.*

**Observación 2.3.** *La relación  $A \sim B$  si  $A = B \bmod 0$  es de equivalencia. Por lo tanto, las clases de equivalencia inducidas son clases de conjuntos que son iguales salvo en conjuntos de medida 0. El conjunto de estas clases lo denotaremos por  $\kappa$ .*

Las operaciones de unión, intersección numerable y substracción de conjuntos funcionan igual en las clases haciendo de  $\kappa$  un álgebra, de modo que, para cualquier parte de  $\kappa$  que sea cerrada con respecto a estas operaciones será llamada un subálgebra de  $\kappa$ .

Como la intersección de álgebras es también un álgebra, entonces, la intersección de  $\bigcap_{\alpha} \kappa_{\alpha}$  de cualquier sistema de subálgebras  $\kappa_{\alpha}$  de  $\kappa$  es un álgebra de  $\kappa$ .

De manera similar al caso de las particiones, la siguiente definición nos permite tener una noción de convergencia sobre refinamientos de subálgebras.

**Definición 2.12.** *La suma  $\bigvee_{\alpha} \kappa_{\alpha}$  de subálgebras  $\kappa_{\alpha}$  es definida como la intersección de todas las subálgebras que contienen todos los  $\kappa_{\alpha}$ .*

- 1) Si  $\kappa_1 \subset \kappa_2 \subset \dots$ , y  $\bigvee \kappa_n = \kappa'$  escribimos  $\kappa_n \nearrow \kappa'$ .
- 2) Si  $\kappa_1 \supset \kappa_2 \supset \dots$ , y  $\bigwedge \kappa_n = \kappa'$  escribimos  $\kappa_n \searrow \kappa'$ .

En este punto solo hemos hablado de álgebras de manera muy general, sin embargo, la siguiente definición y los comentarios después de esta tienen como objetivo dar una conexión entre el concepto de partición medible y el de álgebra, lo cual se logra construyendo esta última a través de los elementos de la partición.

**Definición 2.13.** *Para cualquier partición medible  $\xi$ , denotamos por  $\kappa(\xi)$  el subálgebra de  $\kappa$  que consiste de las clases de  $\xi$ -conjuntos medibles.*

Como se menciona en [1] las subálgebras de  $\kappa$  están en correspondencia uno a uno con las clases de particiones medibles, pues:

- i) Si  $\kappa(\xi) = \kappa(\xi')$  entonces  $\xi = \xi'$ .
- ii) Para toda subálgebra de  $\kappa$  existe una partición medible  $\xi$  tal que  $\kappa(\xi)$  es subálgebra.

También en [1] podemos encontrar las siguientes propiedades:

- i)  $\kappa(\xi) \subset \kappa(\xi')$  si y solo si  $\xi \leq \xi'$ .
- ii)  $\kappa(\bigvee_{\alpha} \xi_{\alpha}) = \bigvee_{\alpha} \kappa(\xi_{\alpha})$ .
- iii)  $\kappa(\bigwedge_{\alpha} \xi_{\alpha}) = \bigcap_{\alpha} \kappa(\xi_{\alpha})$ .

Como resulta natural pensar, las equivalencias entre las nociones de convergencia para particiones medibles y álgebras son equivalentes:

- i)  $\kappa(\xi_n) \nearrow \kappa(\xi)$  es equivalente a  $\xi_n \nearrow \xi$ .

ii)  $\kappa(\xi_n) \searrow \kappa(\xi)$  es equivalente a  $\xi_n \searrow \xi$ .

Para terminar con esta sección definiremos la distancia entre conjuntos medibles como sigue:

**Definición 2.14.** La distancia  $\rho(A, B)$  entre los conjuntos medibles  $A, B$  en un espacio de probabilidad con medida  $\mu$  está definida como:

$$\rho(A, B) = \mu[(A \cup B) - (A \cap B)].$$

La función  $\rho$  hace de  $\kappa$  un espacio métrico completo y separable. Las operaciones de unión, intersección y substracción son continuas con respecto a esta métrica.

## 2.3. Propiedades sobre $L_2$

Esta sección tiene como propósito mostrar la relación entre el álgebra de una partición y sus propiedades en  $L_2$ .

**Definición 2.15.** Llamaremos  $L_2(M)$  al espacio de funciones cuadro integrables definidas en  $M$ , en el que consideraremos a las operaciones de suma y multiplicación como las habituales. Definimos el producto escalar  $f, g \in L_2(M)$  como:

$$\langle f, g \rangle = \int fg \, d\mu.$$

Por último, denotaremos por  $\|f\|$  a la norma de  $f$ .

Para cualquier partición medible  $\xi$  denotamos por  $L_2(M, \xi)$  al subespacio de  $L_2$  que consiste de las funciones que son constantes sobre los elementos de  $\xi$ .

De esto último podemos concluir que  $L_2(M, \xi)$  contiene a las funciones características de los conjuntos de  $\kappa(\xi)$  y es generado por estas funciones.

Así como hicimos una relación entre las equivalencias de las propiedades de álgebras con respecto a las de las particiones también tenemos algo análogo para las funciones de  $L_2$  y las particiones medibles. Estas propiedades podemos consultarlas en [1].

- i)  $L_2(M, \kappa) = L_2(M, \kappa')$  si y sólo si  $\xi = \xi'$ .
- ii)  $L_2(M, \kappa) \subset L_2(M, \kappa')$  si y solo si  $\xi \leq \xi'$ .
- iii)  $\xi_n \nearrow \xi$  es equivalente a  $L_2(M, \kappa) \nearrow L_2(M, \kappa')$ .
- iv)  $\xi_n \searrow \xi$  a  $L_2(M, \kappa) \searrow L_2(M, \kappa')$ .

Para el caso en el que la partición  $\xi$  sea  $\epsilon$  o  $\nu$  se tiene lo siguiente:

i)  $L_2(M, \epsilon) = L_2(M)$ .

ii)  $L_2(M, \nu) = C(M)$ .

Donde  $C(M)$  es el espacio de constantes.

Como se mencionó con anterioridad el siguiente apartado contiene información de gran relevancia para lograr cubrir nuestro objetivo, pues será referenciada en resultados importantes de los capítulos 4 y 5.

Antes de iniciar con el siguiente apartado daremos la siguiente definición, la cual es necesaria para entender lo que se presentará a continuación.

**Definición 2.16.** *El espacio factor o espacio cociente de  $M$  con respecto a la partición medible  $\xi$  es el espacio medible cuyos puntos son los elementos de  $\xi$  y la medida  $\mu_\xi$  definida como sigue:*

*Sea  $p$  el mapeo que toma cada punto  $x \in M$  y lo manda al elemento de  $\xi$  que lo contiene; dicho esto, un conjunto  $Z \in \xi$  es medible si  $p^{-1}(Z)$  es medible en  $M$ . Entonces, se define:*

$$\mu_\xi(Z) = \mu(p^{-1}(Z)).$$

*Denotamos a este espacio factor por  $M/\xi$ .*

**Observación 2.4.** *El mapeo  $p$  es un homomorfismo de  $M$  en  $M/\xi$  debido a que preserva medidas. Este homomorfismo es conocido como proyección.*

**Observación 2.5.** *Note que  $\mu_\xi$  es una medida de probabilidad y los elementos del espacio factor  $M/\xi$  definen una  $\sigma$ -álgebra.*

## 2.4. Sistemas canónicos de medida

En este apartado enunciaremos algunas propiedades de medidas condicionales, esto debido a que en el capítulo 4 se presentará el concepto de entropía condicional media, sin embargo, antes de llegar a esa sección y abordar dicho tema, es necesario puntualizar varios hechos, el primero, como se observará a continuación, es que dados dos conjuntos medibles, uno contenido en el otro, la medida del conjunto más pequeño será la misma si se considera la partición del espacio total o una inducida sobre el conjunto que lo contiene. Para formalizar esto iniciaremos con las siguientes dos definiciones:

**Definición 2.17.** Sean  $M$  un espacio de Lebesgue con medida  $\mu$  y sea  $C \subset M$  medible: si  $\mu(C) > 0$  y  $C \in \xi$ , definimos:

$$\mu_C = \frac{\mu(X)}{\mu(C)}.$$

Llamamos a  $C$  un subespacio de Lebesgue con medida de Lebesgue  $\mu_C$

**Definición 2.18.** Un sistema canónico de medidas o sistema de medidas condicionales pertenecientes a una partición  $\xi$  es una sistema de medidas  $\{\mu_C\}$ ,  $C \in \xi$  que satisfacen dos condiciones:

- 1)  $\mu_C$  es una medida de Lebesgue en  $C \in \xi$ .
- 2) Para cualquier conjunto medible  $X \subset M$ , el conjunto  $X \cap C$  es medible en  $C$  para casi todos los puntos  $C \in M/\xi$ . La función  $\mu_C(X \cap C)$  es medible en  $M/\xi$  para:

$$\mu(X) = \int_{M/\xi} \mu_C(X \cap C) d\mu_C.$$

Como se puede ver en [1] toda partición medible tiene un sistema canónico de medidas y cualesquiera dos sistemas que pertenecen a la misma partición son idénticos módulo 0.

Sean  $\xi, \zeta$  particiones medibles tales que  $\zeta \leq \xi$  y sean  $A$  un elemento de  $\xi$  y  $C$  un elemento de  $\zeta$  que contiene a  $A$ . Como elemento de la partición  $\xi$  de  $M$ ,  $A$  es un espacio de Lebesgue con medida  $\mu_A$ .

Por otro lado, como elemento de la partición  $\xi_C$  de  $C$  con medida  $\mu_C$ ,  $A$  es un espacio de Lebesgue con medida  $(\mu_C)_A$ . La unicidad de los sistemas canónicos de medidas, implica que  $(\mu_C)_A = \mu_A$  para casi todo  $A \in M/\xi$ , esta propiedad es llamada la transitividad de los sistemas canónicos de medida.

De la definición de sistema canónico de medida se sigue que si  $f$  es una función integrable en  $M$ , entonces para casi todos los puntos de  $C \in M/\xi$ , la sección  $f_C$  definida por la formula:

$$f_C(x) = \begin{cases} f(x) & \text{si } x \in C \\ 0 & \text{si } x \notin C. \end{cases} \quad (2.1)$$

es integrable en  $C$  y:

$$\int_M f(x) d\mu = \int_{M/\xi} d\mu + \int_C f_C(x) d\mu_C.$$

Si  $f \in L_2(M)$ , entonces la integral está en  $L_2(M/\xi)$  y la función:

$$g(x) = \int_{p(x)} f_{p(x)}(y) d\mu_{p(x)}.$$

corresponde al isomorfismo canónico entre  $L_2(M/\xi)$  y  $L_2(M, \xi)$  y puede ser considerado como el resultado de promediar  $f$  con los elementos de  $\xi$ . El detalle de esto y de los elementos que componen el resto del capítulo podemos consultarlos en [1].

El operador  $E_\xi : L_2(M) \rightarrow L_2(M, \xi)$  definido como

$$E_\xi f(x) = g(x).$$

es llamado el operador de promedio sobre  $C$ .

Como  $E_\xi$  es un operador ortogonal en  $L_2(M, \xi)$ , entonces, para cualquier función  $f \in L_2(M)$ :

$$f = E_\xi f + (f - E_\xi f) \text{ con } E_\xi f \in L_2(M, \xi).$$

Es suficiente probar que para cualquier  $f \in L_2(M)$ , la función  $(f - E_\xi f)$  es ortogonal en  $L_2(M, \xi)$ , esto es:

$$\langle (f - E_\xi f), g \rangle = 0 \text{ si } f \in L_2(M) \text{ y } g \in L_2(M, \xi).$$

Sean  $f_1, g_1$  funciones en  $L_2(M/\xi)$ , relacionando a  $E_\xi f$  y  $\bar{g}$  mediante el isomorfismo canónico entre  $L_2(M, \xi)$  y  $L_2(M/\xi)$ . Tenemos que:

$$\begin{aligned} \langle f, g \rangle &= \int_{M/\xi} d\mu_\xi \int_C f(x) \overline{g(x)} d\mu_C \\ &= \int_{M/\xi} g_1(C) d\mu_\xi \int_C f(x) d\mu_C \\ &= \int_{M/\xi} g_1(C) f_1(C) d\mu_\xi \\ &= \int_M \overline{g(x)} E_\xi f(x) d\mu \\ &= \langle E_\xi f, g \rangle. \end{aligned}$$

Sean  $\xi_1, \xi_2, \dots$  una sucesión de particiones medibles de  $M$ . De acuerdo al apartado anterior tenemos que  $E_{\xi_n} f \rightarrow E_\xi f$  en  $L_2(M)$  para cualquier función  $f \in L_2(M)$ .

Denotamos por  $p_n$  y  $p$  a los operadores proyección de  $M$  a  $M/\xi_n$  y  $M/\xi$  respectivamente.

Si  $f$  es una función característica de un conjunto  $X \subset M$  entonces:

$$E_{xi_n} f(x) = \mu_{p_n(x)}(X \cap p_n(X))$$

$$E_{xi} f(x) = \mu_{p(x)}(X \cap p(X)).$$

Consecuentemente si  $\xi_n \nearrow \xi$  o  $\xi_n \searrow \xi$  entonces existe una sucesión de funciones  $\mu_{p_n(x)}(X \cap p_n(X))$  que tiende a  $E_{xi} f(x) = \mu_{p(x)}(X \cap p(X))$  en  $L_2$ .

Para cualquier partición medible  $\eta$  con una medida discreta condicional  $\mu_\beta$  (esto es, con a lo más una cantidad numerable de elementos  $\beta$ -mod 0) existen finitas o numerables particiones medibles  $\xi$  tales que

$$\eta\xi = \epsilon.$$

Aún más,  $\xi$  puede ser elegida tal que para una indexación  $C_1, C_2, \dots$  de sus elementos las medidas condicionales de sus intersecciones de puntos únicos  $C_i \cap \beta$  de elementos con cada elemento de  $\beta$  de  $\eta$  forman una sucesión decreciente:

$$\mu_\beta(C_1 \cap \beta) \geq \mu_\beta(C_2 \cap \beta) \geq \dots$$

Con lo anterior, hemos terminado de dar los conceptos y algunos preliminares que nos servirán como base en este trabajo.

Con el objetivo de terminar este capítulo, se hace el recordatorio de la importancia de tener clara la noción de partición medible y algunas de sus propiedades como las de convergencia que mencionamos en su respectiva sección. Esto nos ayudará a tener un desarrollo más natural sobre todo lo que se construirá en los siguientes capítulos ya que al menos en los dos siguientes daremos definiciones y propiedades del comportamiento de la entropía sobre estas.

### 3 Entropía de particiones medibles.

En este capítulo se presentará la definición de entropía con respecto a una partición medible y se demostrarán varias propiedades que serán de utilidad para las siguientes secciones.

Antes de iniciar es importante hacer un par de comentarios sobre lo que se presentará a continuación:

Primero, el lector observará que la definición de entropía requiere de una función logaritmo, tomando como base la definición de Rokhlin se usará en base 2, sin embargo, los resultados son igualmente válidos para logaritmos en otras bases.

Segundo, a partir de ahora y en lo que resta del texto trabajaremos en un espacio de probabilidad  $M$  con medida  $\mu$ , por lo que cada vez que se hable de una partición medible se entenderá que es sobre este espacio. Además, se consideran estas particiones medibles como a lo más numerables.

La estructura de este apartado muestra inicialmente la definición de entropía para una partición medible, posteriormente se prueban algunas propiedades generales de esta, como lo son la no negatividad o monotonía con respecto a la propiedad de refinamiento.

Con base en las propiedades anteriores se demostrarán resultados de convergencia, en los que se observa que si una familia de particiones converge de manera monótona a una partición, su entropía también lo hará. Por último, mostraremos una generalización de la desigualdad triángular con el producto de particiones.

Una vez dicho todo lo anterior, se inicia con la definición de entropía para una partición.

**Definición 3.1.** *Sea  $\xi$  una partición medible de un espacio  $M$  y sean  $C_1, C_2, \dots$  los elementos de  $\xi$  con medida positiva; definimos la entropía de  $\xi$  como sigue:*

$$H(\xi) = \begin{cases} -\sum_k \mu(C_k) \log_2(\mu(C_k)) & \text{si } \mu(M - \bigcup_k C_k) = 0. \\ +\infty & \text{si } \mu(M - \bigcup_k C_k) > 0. \end{cases} \quad (3.1)$$

En la definición anterior se considera que:

$$\log_2 \cdot 0 = \infty \quad \text{y} \quad 0 \cdot \log_2(0) = 0$$

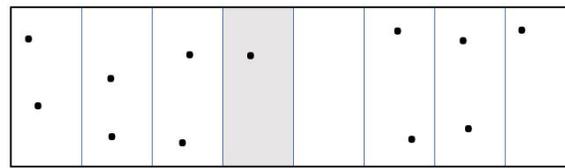
Sea  $x \in M$ , denotaremos por  $m(x; \xi)$  a la medida del elemento en  $\xi$  que contiene a  $x$ .

De esta manera, por la definición de integral con respecto a una medida podemos escribir (3.1) como sigue:

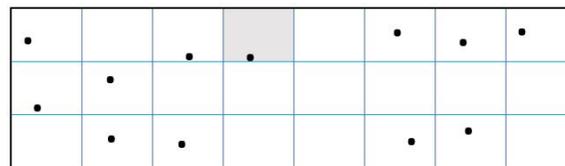
$$H(\xi) = - \int_M \log_2(m(x; \xi)) d\mu.$$

A continuación mostraremos dos propiedades básicas de la entropía, la primera es que es una función no negativa, esto debido a que no podemos considerar una incertidumbre negativa.

La segunda es que si consideramos dos particiones medibles  $\xi$  y  $\eta$  donde  $\eta$  es una supartición de  $\xi$  denotado por  $\xi \leq \eta$  de acuerdo a nuestra definición 2.4, entonces para un determinado punto de  $M$  la medida del conjunto que lo contiene en  $\xi$  es mayor que el que lo contiene en  $\eta$  como se muestra a continuación:



(a) Distribución en la partición  $\xi$ .



(b) Distribución en la partición  $\eta$ .

Figura 3.1: Particiones.

De manera intuitiva podemos esperar que la entropía del punto en el conjunto gris sea mayor en la partición  $\eta$  ya que se cuenta en proporción con menor información que en  $\xi$  por lo que la incertidumbre en este caso es más grande.

Lo anterior se formaliza con la siguiente proposición.

**Proposición 3.1.** *Dada una partición  $\xi$  medible para el espacio de probabilidad  $M$  se verifica lo siguiente:*

- i)  $H(\xi) \geq 0$ ;  $H(\xi) = 0$  sí y solo sí  $\xi = \nu$ .
- ii) Si  $\xi \leq \eta$ , entonces  $H(\xi) \leq H(\eta)$ . Si  $\xi \leq \eta$  y  $H(\xi) = H(\eta) < \infty$ , entonces  $\xi = \eta$ .

*Demostración.* i) Como  $\mu$  es una medida de probabilidad, solo toma valores entre 0 y 1, de modo que para todo conjunto  $C \in \xi$

$$\log_2(\mu(C)) \leq 0$$

lo que implica que para todo  $x \in M$ :

$$\begin{aligned} \log_2(m(x; \xi)) &\leq 0 \\ -\log_2(m(x; \xi)) &\geq 0 \end{aligned}$$

por lo que se concluye que:

$$H(\xi) = - \int_M \log_2(m(x; \xi)) d\mu \geq 0.$$

Por otra parte, note que si  $H(\xi) = 0$ , como

$$-\log_2(m(x; \xi)) \geq 0$$

entonces, para todo  $x \in M$ :

$$-\log_2(m(x; \xi)) = 0$$

pero esto pasa sí y solo sí

$$m(x; \xi) = 1$$

sí y solo sí

$$m(x; \xi) = \mu(M).$$

Luego, como  $\xi$  es una partición de  $M$  y esto sucede para todo  $x \in M$ , entonces lo anterior pasa sí y solo sí

$$\xi = \nu.$$

ii) Si  $\xi \leq \eta$ , consideremos algún  $x \in M$ , luego para  $B \in \eta$  tal que  $x \in B$ , existe  $A \in \xi$  tal que  $B \subset A$ , lo que implica que  $\mu(A) \geq \mu(B)$ .

De lo anterior se sigue que, para todo  $x \in M$ ,  $m(x; \xi) \geq m(x; \eta)$ .

Luego, por la monotonía de la función logaritmo:

$$\log_2(m(x; \xi)) \leq \log_2(m(x; \eta)).$$

Entonces

$$- \int_M \log_2(m(x; \eta)) d\mu \geq - \int_M \log_2(m(x; \xi)) d\mu$$

y, por tanto

$$H(\eta) \geq H(\xi).$$

Además, si  $H(\eta) = H(\xi)$ , se tiene que:

$$-\int_M \log_2(m(x; \eta))d\mu + \int_M \log_2(m(x; \xi))d\mu = 0$$

de modo que:

$$\int_M \log_2 \left( \frac{m(x; \xi)}{m(x; \eta)} \right) d\mu = 0$$

luego, como:

$$\frac{m(x; \xi)}{m(x; \eta)} \geq 1$$

el integrando es mayor o igual que 0, por lo que:

$$\left( \log_2 \frac{m(x; \xi)}{m(x; \eta)} \right) = 0$$

para toda  $x \in M$ , lo cual ocurre si:

$$m(x; \xi) = m(x; \eta)$$

lo que al suceder para toda  $x \in M$  implica que:

$$\xi = \eta$$

□

Ahora se presentan dos propiedades que relacionan la convergencia de una familia de particiones medibles con la convergencia de sus respectivas entropías límite.

### Propiedades de convergencia

**Proposición 3.2.** *i) Sea  $\{\xi_n\}$  una familia de subparticiones medibles de  $M$  tales que  $\xi_n \nearrow \xi$  entonces  $H(\xi_n) \nearrow H(\xi)$ .*

*ii) Sea  $\{\xi_n\}$  una familia de subparticiones medibles de  $M$  tales que  $\xi_n \searrow \xi$  entonces  $H(\xi_n) \searrow H(\xi)$ .*

*Demostración.* i) Por la hipótesis  $\xi_n \nearrow \xi$  se tiene que:

$$\xi_1 \leq \xi_2 \leq \xi_3 \dots \leq \xi$$

lo cual implica:

$$m(x; \xi_1) \geq m(x; \xi_2) \geq m(x; \xi_3) \dots \geq m(x; \xi).$$

De esta manera, tenemos una sucesión monótona decreciente y acotada inferiormente por  $m(x; \xi)$ , de modo que casi en todas partes:

$$m(x; \xi_n) \searrow m(x; \xi).$$

Del hecho anterior y de que la integral es un operador monótono se sigue lo siguiente:

$$-\int_M \log_2 m(x; \xi_n) d\mu \nearrow -\int_M \log_2 m(x; \xi) d\mu$$

y, entonces

$$H(\xi_n) \nearrow H(\xi).$$

ii) Como  $\xi_n \searrow \xi$  se tiene que:

$$\xi_1 \geq \xi_2 \geq \xi_3 \dots \geq \xi,$$

lo cual implica:

$$m(x; \xi_1) \leq m(x; \xi_2) \leq m(x; \xi_3) \dots \leq m(x; \xi).$$

Similarmente al inciso anterior, tenemos una sucesión monótona creciente y acotada superiormente por  $m(x; \xi)$ , de modo que casi en todas partes:

$$m(x; \xi_n) \nearrow m(x; \xi).$$

Con esto, podemos concluir de la siguiente manera:

$$-\int_M \log_2 m(x; \xi_n) d\mu \searrow -\int_M \log_2 m(x; \xi) d\mu,$$

es decir,

$$H(\xi_n) \searrow H(\xi).$$

□

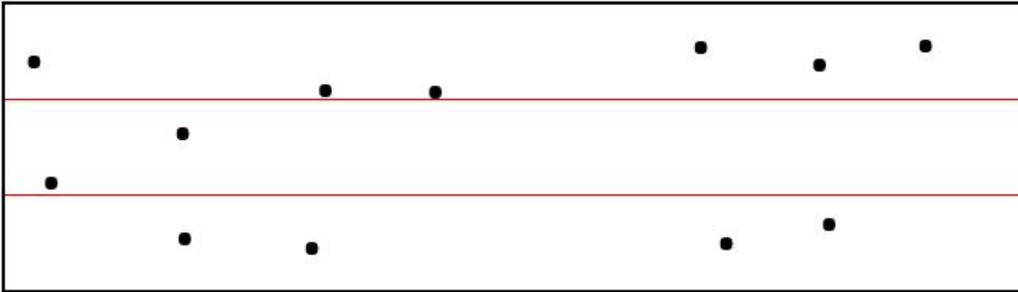


Figura 3.2: Distribución en la partición  $\zeta$ .

Para continuar, consideremos una partición medible  $\zeta$  de un espacio de probabilidad  $M$  en tres partes iguales como se muestra en la figura 3.2, y aplicando la definición 3.1 se tiene que:

$$H(\zeta) = -\frac{1}{3} \log_2 \left( \frac{1}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right) - \frac{1}{3} \log_2 \left( \frac{1}{3} \right).$$

$$H(\zeta) = -\log_2\left(\frac{1}{3}\right).$$

$$H(\zeta) = \log_2(3).$$

Adicionalmente, consideremos otra partición  $\delta$  también en tres partes no necesariamente iguales. Para simplificar los cálculos tomémosla con conjuntos de medidas  $\{\frac{6}{8}, \frac{1}{8}, \frac{1}{8}\}$  como en la figura 3.3.

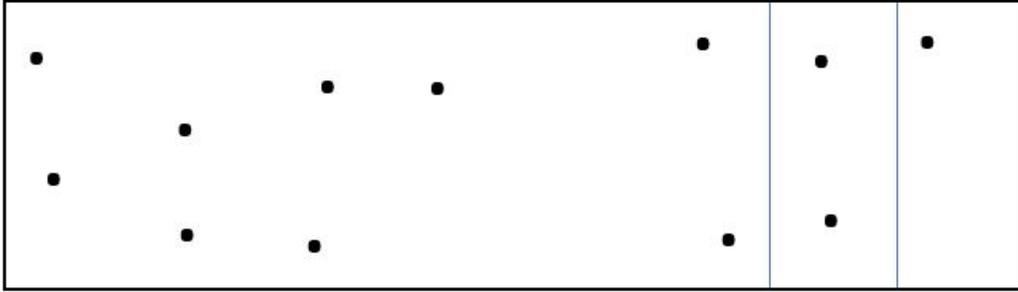


Figura 3.3: Distribución en la partición  $\delta$ .

Apliquemos nuevamente 3.1:

$$H(\delta) = -\frac{6}{8}\log_2\left(\frac{6}{8}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right).$$

$$H(\delta) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{2}{3}\log_2\left(\frac{1}{3}\right).$$

$$H(\delta) = -\frac{3}{4}\log_2(3) + \frac{2}{3}\log_2(3) + \frac{3}{4}\log_2(4) - \frac{2}{3}\log_2(1).$$

$$H(\delta) = -\frac{1}{12}\log_2(3) + \frac{3}{2}.$$

Por último, como  $\log_2(3) > \frac{3}{2}$ , entonces:

$$H(\zeta) > H(\delta).$$

Con lo anterior se observa que la entropía varía con la medida de los conjuntos de la partición que se esté considerando, pues como se observa en este caso, cuando esta fue uniforme se obtuvo una mayor entropía que cuando consideramos un conjunto con probabilidad del 75 %.

Esto último tiene sentido, pues cuando tomamos una distribución uniforme de probabilidad para un determinado evento, sabemos que cada posible resultado tiene igual probabilidad que los demás, de hecho, este es el escenario con más incertidumbre, ya que de modificar un poco la probabilidad de algún resultado estaríamos ganando más información sobre el suceso.

En conclusión, la distribución uniforme (finita) es el escenario con mayor entropía y esto lo demostraremos a continuación:

**Proposición 3.3.** *La entropía de una partición medible de  $n$  conjuntos es menor o igual a  $\log_2(n)$ . La entropía es igual a  $\log_2(n)$  sí y sólo si cada elemento de la partición tiene medida  $\frac{1}{n}$ .*

*Demostración.* Sean  $\xi$  una partición medible de  $M$  en  $n$  elementos y  $p_1, p_2, p_3, \dots, p_n$  las medidas de los elementos de la partición  $\xi$ .

Definimos la función:

$$\begin{aligned}\phi : R^+ &\rightarrow R \\ \phi(x) &= x \log_2(x).\end{aligned}$$

Observamos que esta aplicación es estrictamente convexa, pues, para todo,  $x \in R > 0$

$$\phi''(x) = \frac{1}{x \ln(2)} > 0.$$

Por tanto, si  $\{a_1, a_2, a_3, \dots, a_n\}$  es un conjunto de números reales no negativos tales que  $\sum_{j=1}^n a_j = 1$ , entonces se verifica que:

$$\phi\left(\sum_{j=1}^n a_j x_j\right) \leq \sum_{j=1}^n a_j \phi(x_j).$$

Haciendo  $a_j = \frac{1}{n}$  y  $x_j = p_j$  para todo  $j \in \{1, 2, 3, \dots, n\}$  se tiene

$$\phi\left(\sum_{j=1}^n \frac{1}{n} p_j\right) \leq \sum_{j=1}^n \frac{1}{n} \phi(p_j).$$

Entonces

$$\frac{1}{n} \sum_{j=1}^n p_j \log_2\left(\frac{1}{n} \sum_{j=1}^n p_j\right) \leq \frac{1}{n} \sum_{j=1}^n p_j \log_2(p_j).$$

Como estamos trabajando en espacios de probabilidad  $\sum_{j=1}^n p_j = 1$ , se tiene que

$$\frac{1}{n} \log_2\left(\frac{1}{n}\right) \leq \frac{1}{n} \sum_{j=1}^n p_j \log_2(p_j)$$

lo cual se puede simplificar como

$$\log_2(1) - \log_2(n) \leq \sum_{j=1}^n p_j \log_2(p_j).$$

Luego, usando que  $\log_2(1) = 0$  y despejando

$$-\sum_{j=1}^n p_j \log_2(p_j) \leq \log_2(n)$$

$$H(\xi) \leq \log_2(n).$$

Además, si  $p_j = \frac{1}{n}$  para todo  $j$

$$\begin{aligned} H(\xi) &= \sum_{j=1}^n p_j \log_2(p_j) \\ &= -\sum_{j=1}^n \frac{1}{n} \log_2\left(\frac{1}{n}\right) \\ &= -\frac{1}{n}n(\log_2(1) - \log_2(n)) \\ &= \log_2(n). \end{aligned}$$

De modo que hemos cubierto todo lo que queríamos probar.  $\square$

En esta última sección del capítulo se estudiarán propiedades relativas a la entropía del producto de particiones. Para ello ilustremos a esta nueva partición considerando dos particiones como las siguientes:

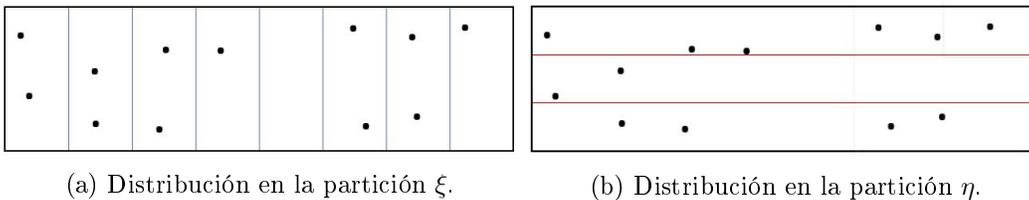


Figura 3.4: Particiones  $\xi$  y  $\eta$ .

Su producto es la partición más fina que es refinamiento para ambas y en este caso sería como se ilustra en la figura 3.5.

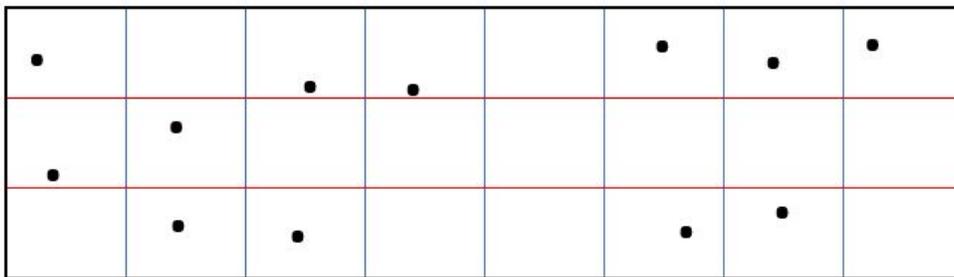


Figura 3.5: Producto de particiones

Lo siguiente es demostrar algunas propiedades de este, como un tipo de desigualdad triangular que relaciona la entropía del producto de dos particiones medibles con la suma de las entropías de las dos particiones. Una vez realizado esto se extenderá al caso cuando el número de particiones con las que se trabaja tiende a infinito.

Estas proposiciones tienen un alto valor operativo, ya que en las siguientes secciones recurrimos en más de una ocasión a su aplicación.

**Proposición 3.4** (Desigualdad triangular). *Para cualesquiera dos particiones medibles  $\xi, \eta$  de un espacio  $M$  se tiene lo siguiente:*

$$H(\xi\eta) \leq H(\xi) + H(\eta)$$

*siempre que  $H(\xi), H(\eta) < \infty$ . La igualdad se cumple cuando las particiones son independientes.*

*Demostración.* Sean  $p_i, q_j, r_{ij}$  las medidas de los conjuntos de las particiones  $\xi, \eta$  y  $\xi\eta$  respectivamente. Luego, la medida conjunta de  $r_{ij}$  verifica lo siguiente:

$$\sum_j r_{ij} = p_i, \sum_i r_{ij} = q_j.$$

Ahora, definimos  $\phi$  nuevamente como

$$\begin{aligned} \phi : R^+ &\rightarrow R \\ \phi(x) &= x \log_2(x). \end{aligned}$$

Por lo anterior sabemos que la siguiente desigualdad se cumple para cualesquiera  $a_1, a_2, \dots, a_n \geq 0$  tales que  $\sum_{j=1}^n a_j = 1$

$$\phi\left(\sum_{j=1}^n a_j x_j\right) \leq \sum_{j=1}^n a_j \phi(x_j).$$

Entonces, hacemos  $j$  fijo y  $a_i = p_i$  y  $x_i = \frac{r_{ij}}{p_i}$ , así

$$\phi\left(\sum_i p_i \frac{r_{ij}}{p_i}\right) \leq \sum_i p_i \phi\left(\frac{r_{ij}}{p_i}\right).$$

Por tanto

$$\phi\left(\sum_i r_{ij}\right) \leq \sum_i p_i \phi\left(\frac{r_{ij}}{p_i}\right).$$

Entonces

$$\sum_i r_{ij} \log_2\left(\sum_i r_{ij}\right) \leq \sum_i p_i \frac{r_{ij}}{p_i} \log_2\left(\frac{r_{ij}}{p_i}\right),$$

de modo que

$$q_j \log_2(q_j) \leq \sum_i r_{ij} \log_2\left(\frac{r_{ij}}{p_i}\right)$$

y, entonces,

$$q_j \log_2(q_j) \leq \sum_i r_{ij} (\log_2(r_{ij}) - \log_2(p_i)).$$

Luego, sumando sobre todas las  $j$

$$\begin{aligned} \sum_j q_j \log_2(q_j) &\leq \sum_j \sum_i r_{ij} \log_2(r_{ij}) - \sum_j \sum_i r_{ij} \log_2(p_i) \\ - \sum_j \sum_i r_{ij} \log_2(r_{ij}) &\leq - \sum_j q_j \log_2(q_j) - \sum_i \log_2(p_i) \sum_j r_{ij} \\ - \sum_j \sum_i r_{ij} \log_2(r_{ij}) &\leq - \sum_j q_j \log_2(q_j) - \sum_i p_i \log_2(p_i) \\ H(\xi\eta) &\leq H(\xi) + H(\eta). \end{aligned}$$

Si además, nuestras particiones son independientes, es decir, se cumple que

$$q_j = \frac{r_{ij}}{p_i}$$

entonces de la desigualdad

$$q_j \log_2(q_j) \leq \sum_i r_{ij} \log_2(r_{ij}) - \sum_i r_{ij} \log_2(p_i)$$

se tiene que

$$\frac{r_{ij}}{p_i} \log_2\left(\frac{r_{ij}}{p_i}\right) \leq \sum_i r_{ij} \log_2(r_{ij}) - \sum_i r_{ij} \log_2(p_i)$$

lo que implica que

$$\frac{r_{ij}}{p_i} \log_2(r_{ij}) - \frac{r_{ij}}{p_i} \log_2(p_i) \leq \sum_i r_{ij} \log_2(r_{ij}) - \sum_i r_{ij} \log_2(p_i).$$

Descomponiendo todo en términos de  $p_i, q_j$  el lado derecho de la desigualdad

$$\frac{r_{ij}}{p_i} \log_2(r_{ij}) - \frac{r_{ij}}{p_i} \log_2(p_i) \leq \sum_i p_i q_j \log_2(p_i) + \sum_i p_i q_j \log_2(q_j) - \sum_i p_i q_j \log_2(p_i)$$

simplificando

$$\begin{aligned} \frac{r_{ij}}{p_i} \log_2(r_{ij}) - \frac{r_{ij}}{p_i} \log_2(p_i) &\leq q_j \log_2(q_j) \\ \frac{r_{ij}}{p_i} \log_2(r_{ij}) &\leq q_j \log_2(q_j) + q_j \log_2(p_i) \\ r_{ij} \log_2(r_{ij}) &\leq p_i q_j \log_2(q_j) + p_i q_j \log_2(p_i). \end{aligned}$$

Sumando sobre todas las  $i, j$

$$\sum_i \sum_j r_{ij} \log_2(r_{ij}) \leq \sum_j q_j \log_2(q_j) + \sum_i p_i \log_2(p_i)$$

Multiplicando por  $-1$  en ambos lados se tiene

$$-\sum_i \sum_j r_{ij} \log_2(r_{ij}) \geq -\sum_j q_j \log_2(q_j) - \sum_i p_i \log_2(p_i)$$

que equivale a

$$H(\xi\eta) \geq H(\xi) + H(\eta).$$

De modo que, si  $\xi, \eta$  son independientes, entonces

$$H(\xi\eta) = H(\xi) + H(\eta).$$

□

**Proposición 3.5.** *Para cualquier sucesión finita o infinita numerable de particiones medibles  $\xi_1, \xi_2, \xi_3 \dots$*

$$H(\vee \xi_k) \leq \sum H(\xi_k).$$

*Demostración.* Si el producto es finito, entonces, por la desigualdad triangular aplicada al producto  $\vee_{k=1}^n \xi_k$

$$H(\vee_{k=1}^n \xi_k) \leq \sum_{k=1}^{n-1} H(\xi_k) + H(\xi_n).$$

Así, aplicando  $n - 2$  veces más se concluye lo siguiente

$$H(\vee_{k=1}^n \xi_k) \leq \sum_{k=1}^n H(\xi_k).$$

Por otra parte, si el producto no es finito, definimos

$$\zeta_1 = \xi_1.$$

$$\zeta_n = \vee_{k=1}^n \xi_k.$$

$$\zeta = \vee_{k=1}^{\infty} \xi_k.$$

Luego, como  $\zeta_k \nearrow \zeta$  implica que  $H(\zeta_k) \nearrow H(\zeta)$ , entonces la entropía de

$$H(\vee_{k=1}^{\infty} \xi_k).$$

está bien definida.

Por último, partiendo del resultado para el caso finito, calculamos límites.

$$\lim_{n \rightarrow \infty} H(\bigvee_{k=1}^n \xi_k) \leq \lim_{n \rightarrow \infty} \sum_{k=1}^n H(\xi_k)$$

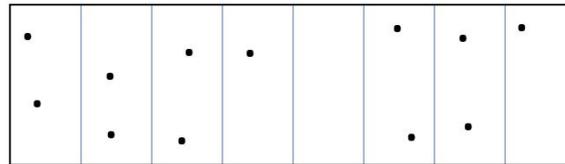
$$H(\bigvee_{k=1}^{\infty} \xi_k) \leq \sum_{k=1}^{\infty} H(\xi_k).$$

□

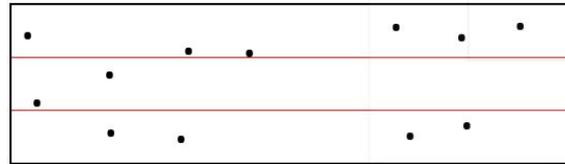
## 4 Entropía condicional media.

En este capítulo definiremos la entropía de una partición medible condicionada a otra partición medible. Antes de dar inicio a las definiciones necesarias se explicará de manera breve y con algunas imágenes este concepto con el objetivo de tener una idea intuitiva de lo que se trabajará más adelante.

De inicio, considere dos particiones medibles  $\xi$  y  $\eta$  de un espacio de probabilidad  $M$ , como se muestra en la siguiente figura:



(a) Distribución en la partición  $\xi$ .



(b) Distribución en la partición  $\eta$ .

Figura 4.1: Particiones  $\xi$  y  $\eta$ .

En la partición  $\eta$  pondremos nuestra atención en la parte superior y realizaremos la intersección de este conjunto con los elementos de  $\xi$ , de esta forma se formaran unas celdas como las siguientes:

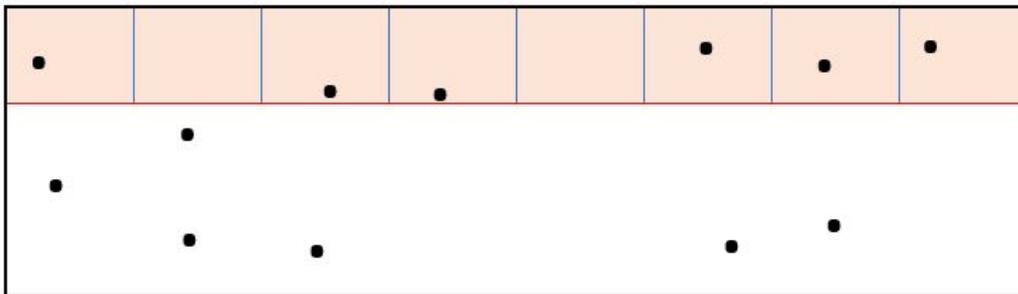


Figura 4.2: Partición inducida.

A las celdas formadas en la figura 4.2 posteriormente las definiremos como **partición inducida** de un conjunto sobre una partición.

La idea central en este capítulo consiste en tomar dos particiones medibles y realizar para cada elemento de alguna de ellas la partición inducida con respecto a la otra.

Las celdas formadas por la repetición de este proceso definirán una nueva partición y utilizaremos la definición de entropía del capítulo anterior para calcular la entropía de esta.

Lo anterior nos permitirá en el capítulo 6 realizar mediciones de entropía con respecto a una partición determinada por un endomorfismo y dará el sustento para definir entropía dinámica.

A partir de ahora la estructura de este capítulo es muy similar a la del anterior, lo cual es natural debido a que ahora se estudia el concepto de entropía en los espacios cocientes  $M/\xi$  donde  $\xi$  es una partición medible. Recordemos que la definición de dichos espacios fue dada en 2.16.

Como ahora mediremos la entropía de los representantes de las clases, hace falta un paso más, pues, dado un conjunto  $B \subset M$  sus puntos no necesariamente tienen que caer en el mismo elemento de la partición, por lo que para identificar los representantes de las clases de  $B$  en  $\xi$  consideramos las particiones dadas por la definición 2.10.

Con esto último, hemos dado los ingredientes necesarios para comenzar con la definición de entropía condicional y exhibir algunas de sus propiedades.

**Definición 4.1.** *Si  $\xi, \eta$  son particiones medibles de un espacio de probabilidad  $M$ , entonces para cualquier  $B \in M/\eta$ , la partición  $\xi_B$  tiene una entropía  $H(\xi_B)$  bien definida.*

*Esta es una función medible no negativa sobre el espacio factor  $M/\eta$  a la que llamaremos **la entropía condicional de  $\xi$  con respecto a  $\eta$** .*

*Su integral en  $M/\eta$ , ya sea finita o infinita, es llamada la **entropía condicional media de  $\xi$  con respecto a  $\eta$**  y se denota por  $H(\xi/\eta)$  definida como sigue:*

$$H(\xi/\eta) = \int_{M/\eta} H(\xi_B) d\mu_\eta.$$

**Definición 4.2** (Definición equivalente). *Sea  $B(x)$  el elemento que contiene a  $x \in M$  en la partición  $\eta$ . Denotaremos por  $m(x; \xi/\eta)$  la medida del elemento de la partición  $\xi_{B(x)}$  que contiene a  $x$ , entonces:*

$$H(\xi/\eta) = - \int_M \log_2 m(x; \xi/\eta) d\mu_\eta. \quad (4.1)$$

Como puede observarse esta fórmula tiene la ventaja de que el integrando no depende de  $\eta$ .

*Demostración.* Como  $\xi$  es una partición medible de  $M$  y  $B \in \eta$  es medible, entonces  $\xi_B$  es una partición medible en  $B$ ; luego, por la definición de entropía para una partición medible tenemos que:

$$H(\xi_B) = - \int_B \log_2 m(x; \xi) d\mu_B.$$

entonces

$$H(\xi/\eta) = \int_{M/\eta} H(\xi_B) d\mu_\eta = - \int_{M/\eta} d\eta_\eta \int_B \log_2 m(x; \xi_B) d\mu_B. \quad (4.2)$$

Recordemos que de la definición de sistema canónico de medidas, se tiene que si una función  $f$  es integrable en  $M$  para casi todo  $C \in M/\xi$ , la sección  $f_C$  definida por:

$$f_C(x) = \begin{cases} f(x) & \text{si } x \in C \\ 0 & \text{en otro caso} \end{cases} \quad (4.3)$$

es integrable en  $C$  y

$$\int_M f(x) d\mu = \int_{M/\xi} d\mu_\xi \int_C f_C(x) d\mu_C.$$

De modo que del lado derecho de (2) se puede concluir que:

$$H(\xi/\eta) = - \int_{M/\eta} d\eta_\eta \int_B \log_2 m(x; \xi_B) d\mu_B = - \int_B \log_2 m(x; \xi) d\mu_B.$$

que es lo que se quería probar. □

Ahora, se presentarán algunas propiedades generales de la entropía media condicional. Como se observará, muchas de ellas son consecuencia de lo visto en el capítulo anterior.

### Propiedades de la entropía media condicional.

**Proposición 4.1.** Sean  $\xi, \eta, \zeta$  particiones medibles de  $M$ , si  $\zeta \leq \eta$  entonces

$$H(\xi/\eta) = H(\xi\zeta/\eta).$$

*Demostración.* Sea  $B \in \eta$ , entonces:

1.  $\xi_B$  es la partición de  $B$  inducida por  $\xi$ .
2.  $(\xi\zeta)_B$  es la partición de  $B$  inducida por  $\xi\zeta$ .

Donde:

$$\xi_B = \{E \cap B | E \in \xi\}.$$

$$(\xi\zeta)_B = \{(E \cap Z) \cap B | E \in \xi, Z \in \zeta\}.$$

Sea  $A \in (\xi\zeta)_B$  entonces existen conjuntos  $X_1 \in \xi, X_2 \in \zeta$  que verifican:

$$A = X_1 \cap X_2 \cap B.$$

Además, como

$$X_1 \cap X_2 \subset X_1 \in \xi.$$

Entonces

$$A \in \xi_B$$

lo que implica que

$$\xi_B \leq (\xi\zeta)_B.$$

Por otro lado, sea  $A' \in \xi_B$ , luego, existe  $X_3 \in \xi$  tal que

$$A' = X_3 \cap B.$$

Luego, como  $\zeta \leq \eta$  entonces  $B \in \zeta$ .

Además,  $A'$  puede escribirse como:

$$A' = (X_3 \cap B) \cap B$$

de esta forma

$$A' \in (\xi\zeta)_B.$$

De esta manera:

$$(\xi\zeta)_B \leq \xi_B.$$

Por lo que se concluye que

$$\xi_B = (\xi\zeta)_B.$$

Lo que implica que

$$\int_{M/\eta} H(\xi_B) d\mu_\eta = \int_{M/\eta} H((\xi\zeta)_B) d\mu_\eta$$

$$H(\xi/\eta) = H(\xi\zeta/\eta).$$

□

**Proposición 4.2.** Sean  $\xi, \eta$  particiones medibles de  $M$ , entonces

$$H(\xi/\eta) \geq 0.$$

Además  $H(\xi/\eta) = 0$  si y solo si  $\xi \leq \eta$ .

*Demostración.* Probaremos solo la primera parte:

como  $\xi_B$  es una partición medible entonces  $H(\xi_B) \geq 0$ , por lo tanto

$$H(\xi/\eta) = \int_{M/\eta} H(\xi_B) d\mu_\eta \geq 0.$$

□

**Proposición 4.3.** Sean  $\xi, \eta$  particiones medibles de  $M$  y  $\xi'$  una subpartición de  $\xi$ , es decir  $\xi \leq \xi'$ , entonces  $H(\xi/\eta) \leq H(\xi'/\eta)$ .

*Demostración.* Como  $\xi_B$  y  $\xi'_B$  son particiones medibles de  $B$ , por lo visto en el capítulo anterior

$$H(\xi_B) \leq H(\xi'_B)$$

luego:

$$\begin{aligned} \int_{M/\eta} H(\xi_B) d\mu_\eta &\leq \int_{M/\eta} H(\xi'_B) d\mu_\eta \\ H(\xi/\eta) &\leq H(\xi'/\eta). \end{aligned}$$

□

**Proposición 4.4.** Sean  $\xi, \eta, \zeta$  particiones medibles. Si  $H(\xi/\zeta) < \infty$  y  $H(\eta/\zeta) < \infty$ , se verifica que:

$$H(\xi\eta/\zeta) \leq H(\xi/\zeta) + H(\eta/\zeta).$$

La igualdad se cumple si  $\xi, \eta$  son independientes con respecto a  $\zeta$ .

*Demostración.* Sea  $C \in M/\zeta$ , entonces para las particiones medibles

$$\xi_C \eta_C, \xi_C, \eta_C$$

aplicamos la desigualdad triangular para particiones medibles, lo que da como resultado la siguiente expresión:

$$H(\xi_C \eta_C) \leq H(\xi_C) + H(\eta_C).$$

Integrando sobre  $M/\zeta$

$$\begin{aligned} \int_{M/\zeta} H(\xi_C \eta_C) d\mu_C &\leq \int_{M/\zeta} H(\xi_C) d\mu_C + \int_{M/\zeta} H(\eta_C) d\mu_C \\ H(\xi\eta/\zeta) &\leq H(\xi/\zeta) + H(\eta/\zeta). \end{aligned}$$

Además, si  $\xi, \eta$  son independientes con respecto a  $\zeta$ , se tiene que  $\xi_C, \eta_C$  son independientes, de modo que

$$H(\xi_C \eta_C) = H(\xi_C) + H(\eta_C).$$

$$H(\xi\eta/\zeta) = H(\xi/\zeta) + H(\eta/\zeta).$$

□

De manera similar a lo presentado en el capítulo anterior, en este se encuentran este par de resultados que relacionan la convergencia de particiones medibles con la convergencia de su entropía condicional media.

**Proposición 4.5.** *Las siguientes afirmaciones son ciertas:*

- i) Sea  $\{\xi_n\}$  una familia de subparticiones de  $M$  tales que  $\xi_n \nearrow \xi$  entonces  $H(\xi_n/\eta) \nearrow H(\xi/\eta)$ .
- ii) Sea  $\{\xi_n\}$  una familia de subparticiones de  $M$  tales que  $\xi_n \searrow \xi$  entonces  $H(\xi_n/\eta) \searrow H(\xi/\eta)$ .

*Demostración.* i) Recordemos el siguiente resultado de la sección anterior para particiones medibles.

Si  $\{\xi_n\}$  es una familia de subparticiones de  $M$  tales que  $\xi_n \nearrow \xi$  entonces  $H(\xi_n) \nearrow H(\xi)$ .

Luego para  $B \in \eta$  consideremos a la familia de subparticiones de  $B$  inducidas por cada  $\xi_n$ , entonces de la convergencia anterior se tiene que:

$$H((\xi_n)_B) \nearrow H(\xi_B).$$

Luego, integrando sobre  $M/\eta$

$$\int_{M/\eta} H((\xi_n)_B) d\mu_\eta \nearrow \int_{M/\eta} H(\xi_B) d\mu_\eta.$$

Por tanto,

$$H(\xi_n/\eta) \nearrow H(\xi/\eta).$$

- ii) La demostración es idéntica al caso anterior, considerando ahora el resultado que establece que si  $\{\xi_n\}$  una familia de subparticiones de  $M$  tales que  $\xi_n \searrow \xi$  entonces  $H(\xi_n) \searrow H(\xi)$ .

Luego para la familia  $\{\xi_n\}$  de subparticiones de  $M$  tales que  $\xi_n \searrow \xi$  entonces  $H(\xi_n) \searrow H(\xi)$ .

Integrando sobre  $M/\eta$ ,

$$H(\xi_n/\eta) \searrow H(\xi/\eta).$$

□

La siguiente proposición muestra un resultado de gran utilidad para realizar cálculos.

**Proposición 4.6.** *Para cualesquiera particiones  $\xi, \eta, \zeta$  se cumple la siguiente igualdad*

$$H(\xi\eta/\zeta) = H(\xi/\zeta) + H(\eta/\xi\zeta). \quad (4.4)$$

*Demostración.* Si  $\xi = \nu$  entonces 4.4 se escribe como

$$H(\xi\eta) = H(\xi) + H(\eta/\xi).$$

Empezaremos probando este caso:

### Observación

Si la unión de conjuntos de medida 0 de  $\xi$  o  $\eta$  es un conjunto de medida mayor que 0 entonces los términos en ambos lados de la ecuación son iguales pues son  $\infty$ .

Por lo anterior consideramos que  $\xi, \eta$  son finitos o a lo más numerables.

Sean  $A_1, A_2, A_3, \dots$  y  $B_1, B_2, B_3, \dots$  elementos de  $\xi$  y  $\eta$  respectivamente.

Definimos

$$\begin{aligned} p_i &= \mu(A_i) \\ q_j &= \mu(B_j) \\ r_{ij} &= \mu(A_i \cap B_j). \end{aligned}$$

Luego, como

$$\begin{aligned} \sum_j r_{ij} &= p_i \\ \sum_i r_{ij} &= q_j. \end{aligned}$$

Entonces

$$\begin{aligned} - \sum_i p_i \log_2(p_i) - \sum_i p_i \sum_j \frac{r_{ij}}{p_i} \log_2\left(\frac{r_{ij}}{p_i}\right) &= - \sum_i p_i \log_2(p_i) - \sum_i \frac{p_i}{p_i} \left( \sum_j r_{ij} \log_2(r_{ij}) - \sum_j r_{ij} \log_2(p_i) \right) \\ &= - \sum_i p_i \log_2(p_i) - \sum_i \sum_j r_{ij} \log_2(r_{ij}) + \sum_i \sum_j r_{ij} \log_2(p_i) \\ &= - \sum_i p_i \log_2(p_i) - \sum_i \sum_j r_{ij} \log_2(r_{ij}) + \sum_i p_i \log_2(p_i) \\ &= - \sum_i \sum_j r_{ij} \log_2(r_{ij}). \end{aligned}$$

De forma resumida tenemos que:

$$- \sum_{i,j} r_{ij} \log_2(r_{ij}) = - \sum_i p_i \log_2(p_i) - \sum_i p_i \sum_j \frac{r_{ij}}{p_i} \log_2\left(\frac{r_{ij}}{p_i}\right).$$

Lo cual es la forma desarrollada de

$$H(\xi\eta) = H(\xi) + H(\eta/\xi).$$

Ahora, consideremos el caso general.

A partir de lo que probamos en el caso anterior tenemos que para  $C \in M/\zeta$

$$H(\xi_C \eta_C) = H(\xi_C) + H(\eta_C/\xi_C).$$

Integrando sobre  $M/\zeta$

$$\int_{M/\zeta} H(\xi_C \zeta_C) d\mu_\zeta = \int_{M/\zeta} H(\xi_C) d\mu_\zeta + \int_{M/\zeta} H(\eta_C/\xi_C) d\mu_\zeta.$$

Lo cual se puede simplificar como:

$$H(\xi\eta/\zeta) = H(\xi/\zeta) + \int_{M/\zeta} H(\eta_C/\xi_C) d\mu_\zeta.$$

Con ello, es claro que nos hace falta demostrar que

$$H(\eta/\xi\zeta) = \int_{M/\zeta} H(\eta_C/\xi_C) d\mu_\zeta.$$

Para ello utilizaremos 4.1 para escribir lo siguiente:

$$H(\eta_C/\xi_C) = - \int_C \log_2(m(x; \eta_C/\xi_C)) d\mu_C$$

$$H(\eta/\xi\zeta) = - \int_M \log_2(m(x; \eta/\xi\zeta)) d\mu.$$

Luego, como la función  $\log_2(m(x; \eta_C/\xi_C))$  es la restricción en  $C$  de la función  $\log_2(m(x; \eta/\xi\zeta))$ , por la transitividad de los sistemas canónicos de medidas

$$H(\eta/\xi\zeta) = \int_{M/\zeta} H(\eta_C/\xi_C) d\mu_\zeta.$$

Así, finalmente se tiene que

$$H(\xi\eta/\zeta) = H(\xi/\zeta) + H(\eta/\xi\zeta).$$

□

Por último, enunciaremos tres resultados que serán de utilidad para futuras secciones, para detalles de su demostración sugerimos consultar [1].

En este primer resultado se da una relación de orden sobre la entropía de los cocientes, de modo que, para el caso particular en el que solo se dan dos particiones, la entropía del cociente es menor que la de alguna de las particiones.

**Proposición 4.7.** *Para cualesquiera particiones medibles  $\xi, \eta, \zeta$ , si  $H(\xi/\eta\zeta) < \infty$ , entonces*

$$H(\xi/\eta\zeta) \leq H(\xi/\zeta).$$

*La igualdad se cumple si y sólo si  $\xi, \eta$  son independientes relativos a  $\zeta$ .*

*En particular para cualquiera particiones medibles  $\xi, \eta$  si  $H(\xi/\eta) < \infty$ , entonces*

$$H(\xi/\eta) \leq H(\xi).$$

*La igualdad se tiene cuando  $\xi, \eta$  son independientes.*

Las siguientes dos proposiciones nos dan una noción de estabilidad de la convergencia de la entropía condicional media sobre una sucesión de particiones que convergen de manera creciente o de decreciente a una partición.

**Proposición 4.8.** *Si  $\eta_n \nearrow \eta$  y  $\xi$  es una partición medible tal que  $H(\xi/\eta_1) < \infty$ , entonces*

$$H(\xi/\eta_n) \searrow H(\xi/\eta).$$

Para la última proposición también sabemos que se cumple su dual.

**Proposición 4.9.** *Si  $\eta_n \searrow \eta$  y  $\xi$  es una partición medible tal que  $H(\xi/\eta_1) < \infty$ , entonces*

$$H(\xi/\eta_n) \nearrow H(\xi/\eta).$$

## 5 Espacios de particiones.

A partir de ahora denotaremos por  $Z$  al conjunto de todas las particiones medibles con entropía finita de un espacio de probabilidad  $M$  y definimos

$$\rho(\xi, \eta) = H(\xi/\eta) + H(\eta/\xi) \quad (5.1)$$

para cualesquiera dos particiones en  $Z$ .

Este capítulo tiene dos objetivos, el primero es demostrar que  $Z$  es un espacio métrico completo separable con el fin de utilizar las propiedades de convergencia que esto implica; el segundo, comprender la última proposición de esta sección debido a que será un resultado clave para dos enunciados que se ocuparán en la demostración del teorema de Kolmogorov-Sinai.

La utilidad de lo que se expondrá a continuación va más allá de nociones algebraicas que nos permitan operar relaciones de particiones y entropía, pues ahora estamos describiendo cómo es el espacio de estas particiones a través de su topología métrica.

La ecuación 5.1 es conocida como métrica de Rokhlin y será la que utilizaremos en este texto. Es importante observar que está dada como la suma de las entropías condicionales de dos particiones en  $Z$  por lo que podemos interpretarla como la entropía acumulada entre las dos particiones, pues para cada elemento de alguna de ellas estamos determinando la entropía con respecto a la otra partición.

En este punto, como ya se mencionó que es una métrica, es de imaginarse que cumpla todas las propiedades que se conocen del análisis, por lo que el primer resultado que se demostrará a continuación es la desigualdad triangular.

**Proposición 5.1.** *Para  $\xi, \eta, \zeta$  particiones de un espacio de probabilidad  $M$  se verifica:*

$$\rho(\xi, \zeta) \leq \rho(\xi, \eta) + \rho(\eta, \zeta).$$

*Demostración.* Por las propiedades del producto de particiones sabemos que:

$$\xi \leq \xi\eta.$$

Luego, por (4.3),

$$H(\xi/\zeta) \leq H(\xi\eta/\zeta).$$

Además, aplicando (4.6) al lado derecho

$$H(\xi/\zeta) \leq H(\xi\eta/\zeta) = H(\eta/\zeta) + H(\xi/\eta\zeta).$$

Luego, por (4.7), se tiene que:

$$H(\xi/\zeta) \leq H(\eta/\zeta) + H(\xi/\eta). \quad (5.2)$$

De manera análoga, para  $H(\zeta/\xi)$  tenemos:

$$H(\zeta/\xi) \leq H(\eta/\xi) + H(\zeta/\eta). \quad (5.3)$$

Finalmente, sumando ambos lados de las desigualdades (5.2) y (5.3)

$$H(\xi/\zeta) + H(\zeta/\xi) \leq H(\eta/\zeta) + H(\xi/\eta) + H(\eta/\xi) + H(\zeta/\eta).$$

Entonces,

$$\rho(\xi, \zeta) \leq \rho(\xi, \eta) + \rho(\eta, \zeta).$$

□

**Observación 5.1.** De la definición de  $\rho$  se tiene que:

$$\rho(\xi, \eta) = H(\xi/\eta) + H(\eta/\xi) = H(\eta/\xi) + H(\xi/\eta) = \rho(\eta, \xi).$$

Así mismo, como  $H(\xi/\eta), H(\eta/\xi) \geq 0$ , entonces

$$\rho(\xi, \eta) \geq 0.$$

Por último, nótese que  $\xi = \eta$  si y sólo si  $\rho(\xi, \eta) = 0$ .

A continuación demostraremos este último hecho:

*Demostración.* Si  $\xi = \eta$  entonces  $H(\eta/\xi), H(\xi/\eta) = 0$  por (4.2). Por lo tanto

$$\rho(\xi, \eta) = H(\xi/\eta) + H(\eta/\xi) = 0 + 0 = 0.$$

Por otra parte, si  $\rho(\xi, \eta) = 0$  como  $H(\xi/\eta), H(\eta/\xi) \geq 0$  entonces

$$H(\xi/\eta), H(\eta/\xi) = 0$$

y por (4.2) se tiene  $\xi = \eta$ . □

Lo anterior tiene como consecuencia que  $(Z, \rho)$  es un espacio métrico.

La siguiente proposición demostrará que este espacio es completo, es decir, que para cada sucesión convergente de elementos  $Z$  su punto límite es un elemento de aquí.

**Proposición 5.2.**  $(Z, \rho)$  es completo.

*Demostración.* Para mostrar este hecho es necesario probar que cualquier sucesión de Cauchy  $\{\xi_1, \xi_2, \xi_3, \dots\}$  converge en  $Z$ . Para esto, es suficiente demostrar que

$$\rho(\xi_n, \xi_{n+p}) < \frac{1}{2^n} \quad (p > 0)$$

Definimos:

$$\xi = \bigwedge_{l=1}^{\infty} \bigvee_{k=l}^{\infty} \xi_k.$$

A continuación probaremos que  $\xi \in Z$  y que  $\rho(\xi, \xi_n) \rightarrow 0$ .

Para  $l > n$ , por (4.6) tenemos:

$$H(\bigvee_{k=l}^{\infty} \xi_k / \bigvee_{k=n}^{l-1} \xi_k) = H(\xi_l / \bigvee_{k=n}^{l-1} \xi_k) + H(\bigvee_{k=l+1}^{\infty} \xi_k / \bigvee_{k=n}^l \xi_k).$$

Sumando sobre todas las  $l$ 's

$$H(\bigvee_{k=l+1}^{\infty} \xi_k / \xi_n) = \sum_{l=n+1}^{\infty} H(\xi_l / \bigvee_{k=n}^{l-1} \xi_k) \leq \sum_{l=n+1}^{\infty} H(\xi_l / \xi_{l-1})$$

entonces:

$$\xi \leq \bigvee_{k=n+1}^{\infty} \xi_k, \quad H(\xi_l / \xi_{l-1}) \leq \rho(\xi_{l-1}, \xi_l) < \frac{1}{2^{n-1}}$$

luego

$$H(\xi / \xi_n) < \sum_{l=n+1}^{\infty} \frac{1}{2^{l-1}} = \frac{1}{2^{n-1}}.$$

Por otro lado, aplicando (4.9):

$$H(\xi / \xi_n) = \lim_{l \rightarrow \infty} H(\xi_n / \bigvee_{k=l}^{\infty} \xi_k)$$

y entonces, para  $l$  suficientemente grande:

$$H(\xi_n / \xi) \leq H(\xi_n / \bigvee_{k=l}^{\infty} \xi_k) + \frac{1}{2^n} \leq H(\xi_n / \xi_l) + \frac{1}{2^n}.$$

Si  $l > n$ , entonces

$$H(\xi_n / \xi_l) \leq \rho(\xi_n, \xi_l) < \frac{1}{2^n}$$

y así se tiene que

$$H(\xi) \leq H(\xi_l) + H(\xi / \xi_l) < H(\xi_l) + 1 < \infty$$

y así,

$$\rho(\xi, \xi_n) \leq H(\xi / \xi_n) + H(\xi_n / \xi) < \frac{1}{2^{n-2}}.$$

Por lo tanto

$$\xi_n \rightarrow \xi$$

□

**Proposición 5.3.** *El conjunto de particiones finitas medibles es denso en  $Z$ .*

*Demostración.* Sea  $\xi \in Z$  y sea  $\{\xi_1, \xi_2, \xi_3, \dots\}$  una sucesión de particiones finitas medibles tales que  $\xi_n \nearrow \xi$  entonces

$$\rho(\xi_n, \xi) = H(\xi/\xi_n) \rightarrow H(\xi/\xi) = 0.$$

□

Las siguientes proposiciones tienen como fin dar una noción de "densidad" sobre las particiones que contienen a las particiones más finas con entropía finita, es decir, cada vez que encontremos una sucesión de refinamientos o una partición extremadamente fina de un conjunto, hay una cantidad no numerable de particiones que las contienen con entropía finita.

Este hecho es importante ya que como se mencionó en el capítulo de **Entropía de particiones medibles**, esta es una función creciente, de modo que se busca poner un límite a los refinamientos con el fin de no gastar demasiada energía haciendo particiones cada vez más finas si la variación de la entropía entre ellas será mínima.

**Proposición 5.4.** *Sea  $\{\xi_1, \xi_2, \xi_3, \dots\}$  una sucesión de particiones medibles. El conjunto de particiones  $\xi \in Z$  tales que  $\xi \leq \xi_n$  para al menos una  $n \geq 1$ , es denso en  $Z$ .*

*Demostración.* Es suficiente demostrar que para cualquier partición medible finita  $\eta \in Z$  y para cualquier número positivo  $\delta$  existe un  $n \geq 1$  y  $\xi \in Z$  tales que

$$\xi \leq \xi_n \text{ y } \rho(\xi, \eta) < \delta.$$

Sean  $\{C_1, C_2, C_3, \dots\}$  elementos de  $\eta$ . Como  $\xi_n \nearrow \epsilon$  entonces  $\kappa(\xi_n) \nearrow \kappa(\epsilon)$ .

Luego, para cualquier número positivo  $\delta'$  existen un  $n$  y una familia de  $\xi_n$ -conjuntos  $\{C'_1, C'_2, C'_3, \dots, C'_{m-1}\}$  tales que para  $i \in \{1, 2, 3, \dots, m-1\}$ :

$$\rho(C_i, C'_i) < \delta'.$$

Con esto, hemos creado una forma de aproximar los conjuntos de la partición  $\eta$  dada con elementos de  $\xi_n$ -conjuntos.

Denotaremos por  $\xi$  la partición de  $M$  en conjuntos  $\{D_1, D_2, D_3, \dots, D_m\}$  definidos como sigue:

$$\begin{aligned} D_1 &= C'_1. \\ D_i &= C'_i - \bigcup_{j=1}^{i-1} C'_j \quad (i = 2, 3, \dots, m-1). \\ D_m &= M - \bigcup_{j=1}^{m-1} C'_j. \end{aligned}$$

De lo anterior  $\xi \leq \xi_n$  y

$$\begin{aligned} \rho(\xi, \eta) &= \sum_{i=1}^m \mu(C_i) \log_2(\mu(C_i)) + \sum_{i=1}^m \mu(D_i) \log_2(\mu(D_i)) \\ &\quad - 2 \sum_{i,j=1}^m \mu(C_i \cap D_j) \log_2(\mu(C_i \cap D_j)). \end{aligned}$$

Como se mencionó en la sección 2.2  $\rho$  es continua en  $\kappa$ . Tomando  $\delta'$  lo suficientemente pequeño se tiene que para cada  $i \in \{1, 2, 3, \dots, m\}$

$$C'_i \rightarrow C_i,$$

de modo que  $\rho(\xi, \eta) \rightarrow 0$ , por lo tanto  $\rho(\xi, \eta) < \delta$ .

□

Por último, para terminar el capítulo enunciaremos una generalización del resultado anterior, el cual, como se comentó desde el inicio de este apartado es un resultado importante que nos ayudará a probar dos corolarios que simplifican la demostración del teorema de Kolmogorov-Sinai y que serán presentados en la siguiente sección.

**Proposición 5.5.** *Si  $\xi_n \nearrow \xi$  entonces el conjuntos de particiones en  $Z$  que son más gruesas (o de las que es un refinamiento) que algún  $\xi_n$  es denso en el conjunto de todas las particiones medibles en  $Z$  más gruesas (o de las que es un refinamiento) que  $\xi$ .*

La demostración de esta proposición puede encontrarla en [1].

## 6 Entropía de un endomorfismo.

En este capítulo se presentará el concepto de entropía para un endomorfismo, se probarán algunas de sus propiedades y se realizará la demostración del Teorema de Kolmogorov-Sinai.

Para cumplir los objetivos dados, iniciaremos esta sección definiendo algunos conceptos sobre la invarianza de particiones y endomorfismos que permitirán construir una función  $h$  que será la base para todo lo que resta de este trabajo.

**Definición 6.1.** Una partición medible  $\zeta$  de  $M$  es llamada invariante bajo el endomorfismo  $T$  si  $T^{-1}\zeta \leq \zeta$  y completamente invariante si  $T^{-1}\zeta = \zeta$ .

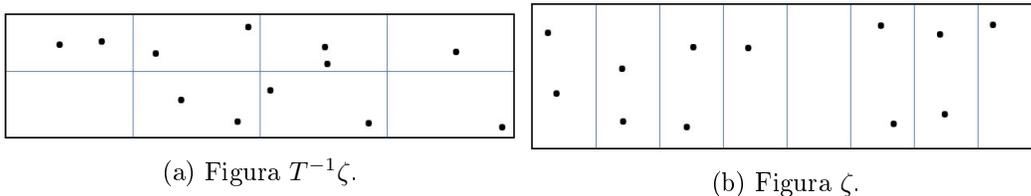


Figura 6.1: Particiones  $T^{-1}\zeta$  y  $\zeta$ .

Con base en lo que se muestra en la figura anterior, es más fácil entender que lo que se busca en este apartado es considerar los cambios de estado que puede tener un sistema. Para ello la función  $T$  juega un papel determinante, pues podemos interpretarla como las condiciones físicas, reglas del juego, condiciones de mercado, fenómenos sociales, marketing, etc que modifican la distribución de un fenómeno observado.

En este punto, podemos decir que nuestro objetivo es definir y entender el cálculo de la entropía de un sistema de acuerdo a los distintos estados que puede tomar, para ello es necesario que se hagan algunos comentarios de 2 condiciones requeridas para la definición 6.1:

- 1) Que  $T$  sea un endomorfismo garantiza que se trabaja con redistribuciones o reacomodos del sistema preservando la probabilidad de cada conjunto. Es por ello que en la figura 6.1 los conjuntos formados en ambas particiones tienen la misma cantidad de puntos.

- 2) La condición de refinamiento de  $\zeta$  con respecto a su imagen inversa es también consecuencia de la invarianza de  $T$ . Sin embargo, también está relacionada con los comentarios del capítulo anterior, en el que se decía que en algún momento es de nuestro interés determinar particiones lo suficientemente "gruesas" para no trabajar con refinamientos excesivos.

**Definición 6.2.** Dado un endomorfismo  $T$ , dentro del conjunto de todas las particiones medibles invariantes más finas que una partición  $\xi$  dada está la más **gruesa**  $\xi^- = \xi_T^-$  definida por la fórmula:

$$\xi^- = \bigvee_{k=0}^{\infty} T^{-k} \xi.$$

La ecuación  $\xi^- = \xi$  es una condición necesaria y suficiente para la invarianza de  $\xi$ .

Similarmente si  $T$  es un automorfismo, la partición:

$$\xi_T = \bigvee_{-\infty}^{\infty} T^k \xi.$$

es la más **gruesa** y completamente invariante partición medible más fina que  $\xi$  y la ecuación

$$\xi_T = \xi.$$

es una condición necesaria y suficiente para la completa invarianza de  $\xi$ .

**Definición 6.3.** Una partición medible  $\xi$  es llamada un generador de un endomorfismo  $T$  si  $\xi_T^- = \epsilon$ .

Una partición medible es llamada doble generador de un automorfismo  $T$  si  $\xi_T = \epsilon$ .

Concluyendo lo que hasta el momento se ha presentado, nuestro interés es calcular la entropía de un sistema al tener distintos cambios de estado. En palabras más simples, la incertidumbre presente en un sistema al pasar de un estado a otro.

Con las definiciones de arriba, ya se cuenta con los elementos necesarios para presentar a la función  $\mathbf{h}$  que se comentó al inicio.

**Definición 6.4.** Sea  $T$  un endomorfismo y  $\xi$  una partición medible de un espacio de probabilidad  $M$ . Definimos **la función  $h$**  como:

$$h(T, \xi) = H(\xi/T^{-1}\xi^-).$$

Apartir de ahora usaremos la notación  $\xi^n$  para referirnos a:

$$\xi_T^n = \bigvee_{k=0}^{k=n-1} T^{-k} \xi, n \geq 1.$$

$$\xi_T^0 = \nu.$$

En este momento, mostraremos tres proposiciones de entropía condicional relacionadas con la función  $h$  sobre los productos de imágenes inversas de una partición  $\xi$  con el fin de dar soporte aritmético a las propiedades de esta función que acabamos de presentar.

**Proposición 6.1.** Si  $\eta \leq \xi$ , entonces

$$H(\xi^n / T^{-n} \eta^-) = \sum_{k=0}^{n-1} H(\xi / T^{-1}(\eta^- \xi^k)).$$

En particular

$$H(\xi^n / T^{-n} \xi^-) = nh(T, \xi).$$

*Demostración.* Como  $\xi^k = \xi T^{-1} \xi^{k-1}$  tenemos:

$$H(\xi^k / T^{-k} \eta^-) = H(T^{-1} \xi^{k-1} / T^{-k} \eta^-) + H(\xi / T^{-k} \eta^- T^{-1} \xi^{k-1}),$$

debido a (4.6).

Además, como  $T$  es un **endomorfismo** y  $\eta \leq \xi$ ,

$$H(T^{-1} \xi^{k-1} / T^{-k} \eta^-) = H(\xi^{k-1} / T^{-(k-1)} \eta^-)$$

y

$$T^{-k} \eta^- T^{-1} \xi^{k-1} = T^{-1}(\eta^- \xi^{k-1}).$$

Entonces

$$H(\xi^k / T^{-k} \eta^-) = H(\xi^{k-1} / T^{-(k-1)} \eta^-) + H(\xi / T^{-1}(\eta^- \xi^{k-1})).$$

Al aplicar inducción sobre  $k$  se tiene lo deseado. □

**Proposición 6.2.** Si  $\eta \leq \xi$  y  $H(\xi / T^- \eta^-) < \infty$  entonces

$$\frac{1}{n} H(\xi^n / T^{-n} \eta^-) \searrow h(T, \xi).$$

*Demostración.* Como  $\xi^n \nearrow \xi^-$  tenemos que  $\eta^- \xi^n \nearrow \xi^-$ , entonces:

$$H(\xi / T^{-1}(\eta^- \xi^n)) \searrow H(\xi / T^{-1} \xi^-) = h(T, \xi).$$

Luego, aplicando la Proposición (6.1) se tiene

$$\frac{1}{n} H(\xi^n / T^{-n} \eta^-) \searrow h(T, \xi).$$

□

**Proposición 6.3.** Si  $\xi \in Z$  entonces,

$$\frac{1}{n} H(\xi^n) \searrow h(T, \xi).$$

*Demostración.* Hacer  $\eta = \nu$  en la prueba anterior. □

A continuación, las siguientes proposiciones serán propiedades exclusivas de esta función  $h$ . Debe tener presente que las más importantes son las últimas dos, pues estas son resultados auxiliares que usaremos directamente en la prueba del teorema de Kolmogorov-Sinai, el resto de las proposiciones son resultados previos que ayudarán a dar un soporte matemático a estos dos resultados, por ello, algunas de ellas cuya demostración requiere la repetición de alguna otra prueba o técnica solo se dará el comentario de lo que se debe hacer para probarla.

**Propiedades de la función  $h$ .**

**Proposición 6.4.**  $h(T, \xi) \leq H(\xi)$ , en particular, si  $\xi \in Z$  entonces  $h(T, \xi) < \infty$ .

*Demostración.* Aplicando (4.7) a la definición de  $h$  se tiene lo siguiente:

$$h(T, \xi) = H(\xi/T^{-1}\xi^-) \leq H(\xi).$$

□

**Proposición 6.5.**

$$h(T, \xi\eta) \leq h(T, \xi) + h(T, \eta).$$

La igualdad se cumple cuando  $\xi^-$  y  $\eta^-$  son independientes.

*Demostración.* Aplicando (4.6) a  $H(\xi\eta/T^{-1}\xi^-T^{-1}\eta^-)$  se tiene que

$$H(\xi\eta/T^{-1}\xi^-T^{-1}\eta^-) = H(\xi/T^{-1}\xi^-T^{-1}\eta^-) + H(\eta/\xi^-T^{-1}\eta^-).$$

Por (4.7),

$$H(\xi/T^{-1}\xi^-T^{-1}\eta^-) \leq H(\xi/T^{-1}\xi^-)$$

y

$$H(\eta/\xi^-T^{-1}\eta^-) \leq H(\eta/T^{-1}\eta^-).$$

Así

$$H(\xi\eta/T^{-1}\xi^-T^{-1}\eta^-) \leq H(\xi/T^{-1}\xi^-) + H(\eta/T^{-1}\eta^-)$$

y

$$h(T, \xi\eta) \leq h(T, \xi) + h(T, \eta).$$

Además, notemos que si  $\xi^-$  y  $\eta^-$  son independientes, por (4.7)

$$H(\xi/T^{-1}\xi^-T^{-1}\eta^-) = H(\xi/T^{-1}\xi^-)$$

y

$$H(\eta/\xi^-T^{-1}\eta^-) = H(\eta/T^{-1}\eta^-),$$

lo cual tiene como consecuencia

$$h(T, \xi\eta) = h(T, \xi) + h(T, \eta).$$

□

**Proposición 6.6.** Si  $n \geq 1$ , entonces  $h(T, \xi) \leq h(T, \xi^n)$ .

*Demostración.* Como  $(\xi^n)^- = \xi^-$  tenemos que

$$H(\xi^n/T^-(\xi^n)^-) = H(\xi^n/T^{-1}\xi^-) \geq H(\xi/T^{-1}\xi^-).$$

□

**Proposición 6.7.**  $h(T^n, \xi^n) = nh(T, \xi)$ .

*Demostración.* Dado que  $(\xi_T^n)_{T^n}^- = \xi_T^-$  se tiene que

$$h(T^n, \xi_T^n) = H(\xi_T^n/T^{-n}(\xi_T^n)_{T^n}^-) = H(\xi_T^n/T_{-1}\xi_T^-),$$

y aplicando la proposición (6.1).

$$h(T^n, \xi^n) = nh(T, \xi).$$

□

**Proposición 6.8.** Si  $T$  es un automorfismo y  $\xi \in Z$ , entonces  $h(T^{-1}, \xi) = h(T, \xi)$ .

*Demostración.* Se sigue de la relación  $\xi_{T^{-1}}^n = T^{n-1}\xi^n$  y la Proposición (6.2) □

**Proposición 6.9.** La función  $h(T, \xi)$  es continua en  $Z$ . Más aún, para cualquier  $\xi \in Z, \eta \in Z$

$$|h(T, \eta) - h(T, \xi)| \leq \rho(\xi, \eta).$$

*Demostración.* Debido a que

$$H(\xi^n) + H(\eta^n/\xi^n) = H(\xi^n\eta^n) = H(\eta^n) + H(\xi^n/\eta^n).$$

se sigue que

$$H(\eta^n) - H(\xi^n) = H(\eta^n/\xi^n) - H(\xi^n/\eta^n).$$

Entonces

$$|H(\eta^n) - H(\xi^n)| \leq H(\eta^n/\xi^n) + H(\xi^n/\eta^n). \quad (6.1)$$

Por otro lado, usando las proposiciones (4.4), (4.7) y (6.8)

$$H(\xi^n/\eta^n) \leq \sum_{k=0}^{n-1} H(T^{-k}\xi/\eta^n) \leq \sum_{k=0}^{n-1} H(T^{-k}\xi/T^{-k}\eta) = nH(\xi/\eta). \quad (6.2)$$

De manera análoga tenemos

$$H(\eta^n/\xi^n) \leq nH(\eta/\xi). \quad (6.3)$$

A partir de (6.1)–(6.3) se sigue que

$$|H(\eta^n) - H(\xi^n)| \leq n\rho(\xi, \eta).$$

Luego dividimos entre  $n$  y usamos la proposición (6.3) para tomar el límite cuando  $n \rightarrow \infty$ , de modo que

$$|h(T, \eta) - h(T, \xi)| \leq \rho(\xi, \eta).$$

□

**Proposición 6.10.** Si  $\xi \leq \eta$  y  $H(\eta/T^{-1}\xi^-) < \infty$  entonces

$$h(T, \xi) \leq h(T, \eta).$$

En particular,  $h(T, \xi)$  es monótono en  $Z$ .

*Demostración.* Como  $\xi \leq \eta$  se tiene

$$\frac{1}{n}H(\xi^n/T^{-n}\eta^-) \leq \frac{1}{n}H(\eta^n/T^{-n}\eta^-).$$

Aplicando la proposición (6.1) al lado derecho y la proposición (6.2) al lado izquierdo de la desigualdad se tiene lo deseado.  $\square$

**Proposición 6.11.** Si  $\xi \leq \eta^-$  y  $\eta \in Z$ , entonces

$$h(T, \xi) \leq h(T, \eta).$$

*Demostración.* Como  $\eta^m \nearrow \eta^-$ , el conjunto de particiones más gruesas que las particiones  $\eta^m$  es denso en el conjunto de particiones medibles más gruesas que  $\eta^-$  por lo visto en (5.5). Por esta razón, es suficiente considerar el caso  $\xi \leq \eta^m$  para algún  $m$ .

Sin embargo, en este caso  $\xi^n \leq (\eta^m)^n = \eta^{m+n-1}$  y

$$h(T, \xi) = \lim_{n \rightarrow \infty} \frac{1}{n}H(\xi^n) \leq \lim_{n \rightarrow \infty} H(\eta^{m+n-1}) = h(T, \eta).$$

$\square$

**Proposición 6.12.** Si  $T$  es un automorfismo y  $\xi \leq \eta_T, \eta \in Z$ , entonces  $h(T, \xi) \leq h(T, \eta)$ .

*Demostración.* Como

$$\eta^m T^{m-1} \eta^m \nearrow \eta_T.$$

es suficiente considerar el caso  $\xi \leq \eta^m T^{m-1} \eta^m$  para algún  $m$ .

En este caso

$$\xi^n \leq (\eta^m T^{m-1} \eta^m)^n = T^{m-1} \eta^{2m+n-2},$$

y

$$h(T, \xi) = \lim_{n \rightarrow \infty} h(\xi^n) \leq \lim_{n \rightarrow \infty} \frac{1}{n}H(\eta^{2m+n-2}) = h(T, \eta).$$

$\square$

Previo a concluir esta sección, definimos la entropía de un endomorfismo  $T$  como sigue:

**Definición 6.5.** Para cualquier endomorfismo  $T$  de un espacio de probabilidad  $M$  definimos, la entropía del endomorfismo  $T$  es:

$$h(T) = \sup h(T, \xi).$$

Donde la cota superior es tomada sobre todas las particiones medibles, o bien sobre todas las particiones invariantes.

En principio este número es positivo o infinito, esto por la proposición (6.4) y de la primera definición de entropía para una partición cuando esta no es finita. Notemos también que el lado derecho en esta última ecuación es el mismo solo si consideramos particiones que están en  $Z$  e incluso si nos restringimos solo a particiones medibles finitas.

Por lo anterior, consideremos cualquier partición medible  $\xi$  de  $M$  y cualquier número positivo  $r$  tal que  $r < h(T, \xi)$  de modo que, demostraremos que existe una partición medible finita  $\eta$  tal que  $h(T, \eta) > r$ .

*Demostración.* Sean  $\xi_1, \xi_2, \dots$  una sucesión de particiones medibles finitas tales que  $\xi_n \nearrow \xi$ . Así  $\xi_n \leq \xi$  para cualquier  $n$ , por lo tanto al aplicar (4.7) a la definición de  $h$  tenemos lo siguiente

$$h(T, \xi_n) = H(\xi_n/T^{-1}\xi_n^-) \geq H(\xi_n/T^{-1}\xi^-).$$

Luego, como el lado derecho de la desigualdad converge a  $H(\xi/T^{-1}\xi^-) = h(T, \xi)$ , entonces  $h(T, \xi_n) > r$  y para un  $n$  suficientemente grande podemos hacer  $\eta = \xi_n$ .  $\square$

## 6.1. Teorema de Kolmogorov-Sinai

El cálculo de la entropía de un endomorfismo  $T$  sobre un espacio de probabilidad  $M$  presenta la dificultad de calcular  $h(T, \xi)$  para cada partición  $\xi$  medible; es por ello que en esta sección demostraremos el teorema de Kolmogorov-Sinai, el cual proporciona una manera más eficiente de realizar esta operación.

La demostración de este teorema utiliza las proposiciones (6.11) y (6.12)

**Teorema 6.1** (Teorema de Kolmogorov-Sinai). *Si  $\xi \in Z$  es un generador para un endomorfismo  $T$  o es un doble generador para un automorfismo  $T$ , entonces*

$$h(T) = h(T, \xi).$$

*Demostración.* Sea  $P$  una partición medible de  $M$ , entonces por la definición de  $h(T)$  tenemos la siguiente desigualdad:

$$h(T) \geq h(T, P).$$

En particular, para  $\xi$  tenemos que

$$h(T) \geq h(T, \xi).$$

Ahora bien, consideremos primero el caso en que  $T$  es un endomorfismo. Luego,  $\xi$  es un generador de  $T$  lo que implica:

$$\xi^- = \bigvee_{k=0}^{\infty} T^{-k}\xi = \epsilon.$$

Así,

$$P \leq \xi^-.$$

Aplicando la proposición 6.11 tenemos que:

$$h(T, P) \leq h(T, \xi).$$

Luego, como esto se cumple para cualquier partición  $P$  entonces por la propiedad de mínima cota superior del supremo tenemos lo siguiente:

$$\sup h(T, P) \leq h(T, \xi).$$

Entonces

$$h(T) \leq h(T, \xi).$$

Ahora se abordará el caso cuando  $T$  es un automorfismo, por lo que ahora  $\xi$  es un doble generador de  $T$ , lo que implica que la partición:

$$\xi_T = \bigvee_{-\infty}^{\infty} T^k \xi = \epsilon.$$

Así,

$$P \leq \xi_T.$$

Utilizando la proposición 6.12:

$$h(T, P) \leq h(T, \xi).$$

Nuevamente, por la propiedad de mínima cota superior del supremo tenemos lo siguiente:

$$\sup h(T, P) \leq h(T, \xi).$$

Entonces

$$h(T) \leq h(T, \xi).$$

Por último, se concluye que

$$h(T) = h(T, \xi).$$

□

Con lo anterior, hemos concluido la demostración del teorema sobre el que hemos dirigido este trabajo. Por esta razón, damos por terminado este capítulo y en el siguiente nos encargaremos de discutir los objetos sobre los que se aplica este resultado.

## 7 Aplicación del teorema de Kolmogorov-Sinai.

En este punto ya hemos cubierto nuestro objetivo de probar el teorema de Kolmogorov-Sinai, sin embargo, a lo largo de este texto se habrá notado que no hemos profundizado en la construcción del algún ejemplo en particular para aplicar este resultado. Esto lo realizaremos en este capítulo comentando dos aspectos que podrían haber quedado al aire bajo los supuestos que hemos trabajado.

El primero de estos es resolver el problema de la existencia de medidas invariantes. La solución a este problema está dada por el teorema de Kyrilov-Bogolubov que se enuncia a continuación:

**Teorema 7.1** (Teorema de Kyrilov-Bogolubov). *Sea  $(X, d)$  un espacio métrico compacto y  $S$  una  $\sigma$ -álgebra. Si  $T : X \rightarrow X$  es una transformación continua, entonces existe una medida  $\mu \in M(X)$  tal que  $\mu$  es invariante bajo  $T$*

Como puede observarse en [7] el espacio de medidas de probabilidad de un espacio métrico compacto es un espacio métrico compacto y convexo. Este hecho, en conjunto con el lema de aproximación y el teorema de la convergencia monótona de teoría de la medida permiten demostrar el teorema (7.1).

El segundo aspecto que cubriremos será la definición de los conceptos de medida de Bernoulli y de Markov, las cuales son esencialmente medidas que cumplen con nuestros supuestos y para las cuales realizaremos un ejemplo numérico. Por esta razón dedicaremos una gran parte de este capítulo a la construcción de dichas medidas.

### 7.1. Construcción de las medidas de Markov y de Bernoulli.

En este momento nuestro primer objetivo es construir una  $\sigma$ -álgebra adecuada en donde definiremos las medidas deseadas. Para ello, iniciaremos esta sección considerando un conjunto  $S_k$  finito, por ejemplo:

$$S_k = \{1, 2, 3, 4, \dots, k\} \text{ con } k \geq 2. \quad (7.1)$$

Nos interesa darle un nombre a las cadenas indexadas de elementos de  $S_k$ , por esto, denotaremos con  $\Sigma$  al conjunto de todas las infinitas sucesiones de símbolos de  $S$  como sigue:

$$\Sigma = \{x = (x_j)_{j=0}^{\infty} | x_j \in S_k\}.$$

De esto último, se puede observar que cada elemento  $x$  es de la forma:

$$x = x_0x_1x_2x_3\dots x_k\dots,$$

donde cada índice es mayor o igual que 0, de modo que, para este tipo de sucesiones, decimos que se están considerando en un solo sentido y las denotamos como  $S_k^N$ .

Para considerar sucesiones en ambos sentidos, denotadas como  $S_k^Z$ , requerimos que:

$$x = \dots x_{-k}\dots x_{-3}x_{-2}x_{-1}x_0x_1x_2x_3\dots x_k\dots$$

En ambos casos el índice 0 representa nuestro punto de partida. Todo lo anterior lo condensamos en la siguiente definición:

**Definición 7.1.** *Sea  $S$  un conjunto finito de  $k$  elementos como en (7.1), entonces denotamos por:*

$$\Sigma_k^+ = S_k^N$$

*al conjunto de sucesiones en un sentido de  $S$ , y como*

$$\Sigma_k = S_k^Z.$$

*al conjunto de sucesiones bilaterales de  $S$ .*

Para tener una noción un poco más natural de estas sucesiones consideraremos dos ejemplos:

- i) En un contexto de teoría de información, nuestro conjunto  $S_k$  estaría conformado por los elementos de un alfabeto, a los que llamaremos símbolos, de esta manera, las sucesiones definidas anteriormente son "palabras", sin embargo, dado que no todas las combinaciones de símbolos tienen sentido, es importante encontrar una manera de identificar cuáles cadenas de símbolos sí tienen un significado.
- ii) En la teoría de procesos estocásticos uno de los primeros conceptos que se definen es el de caminata aleatoria, entonces, para este caso podemos considerar a cada sucesión definida anteriormente como una de estas.

Esta conexión con las caminatas aleatorias será muy importante a partir de este momento, pues de manera intuitiva nos permitirá respaldar lo que expondremos a continuación. Además, gracias a esto podemos retomar el concepto de *matriz adyacente*, el cual para i) nos permite responder cuáles cadenas tienen sentido. Asimismo, en ii) tenemos una representación algebraica de dicha caminata.

Solo como recordatorio, introduciremos la matriz adyacente para nuestras sucesiones en  $S$ .

**Definición 7.2.** Sea  $S_k$  como en (7.1), y sea  $x \in S_k^N$ , donde  $x$  es una sucesión de elementos  $x_i \in S$  tal que  $i \in I \subset N$ , entonces, definimos la matriz  $A \in M(\{0, 1\})_{k \times k}$  como sigue:

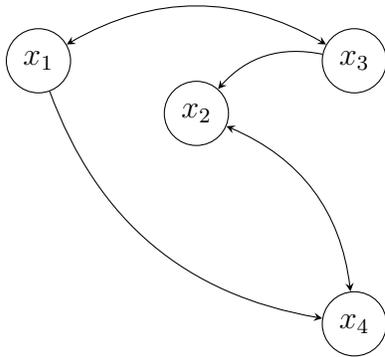
$$a_{i,j} = \begin{cases} 1 & \text{si el elemento } j \text{ puede suceder después del } i \\ 0 & \text{en otro caso.} \end{cases} \quad (7.2)$$

De manera análoga, se tiene lo siguiente para el caso de las sucesiones bilaterales:

**Definición 7.3.** Sea  $S_k$  como en (7.1), y sea  $x \in S_k^Z$ , donde  $x$  es una sucesión de elementos  $x_i \in S$  tal que  $i \in I \subset Z$ , entonces, definimos la matriz  $A \in M(\{0, 1\})_{k \times k}$  como sigue:

$$a_{i,j} = \begin{cases} 1 & \text{si el elemento } j \text{ puede suceder después del } i \\ 0 & \text{en otro caso.} \end{cases} \quad (7.3)$$

Ejemplificando lo anterior, mostraremos la matriz adyacente para la siguiente sucesión lateral:



(a) Sucesión lateral  $x$ .

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad (7.4)$$

(b) Matriz adyacente a  $x$ .

Figura 7.1: Sucesión lateral y su matriz adyacente.

Además, por esto último, podemos observar que las siguientes sucesiones también son recorridos laterales:

- i)  $x_3 \rightarrow x_2 \rightarrow x_4$ .
- ii)  $x_1 \rightarrow x_4 \rightarrow x_2$ .

iii)  $x_2 \rightarrow x_4 \rightarrow x_2$ .

Así, dada una matriz adyacente podemos encontrar otras sucesiones laterales (o bilaterales), lo cual induce la siguiente restricción en  $\Sigma$ .

**Definición 7.4.** *A los siguientes objetos se les conoce como shifts de Markov laterales y bilaterales respectivamente.*

$$1) \Sigma_A^+ = \{(x_n)_{n \in \mathbb{N}} \in \Sigma_k^+ \mid a_{x_n x_{n+1}} = 1 \text{ para toda } n \in \mathbb{N}\}.$$

$$2) \Sigma_A = \{(x_n)_{n \in \mathbb{Z}} \in \Sigma_k^+ \mid a_{x_n x_{n+1}} = 1 \text{ para toda } n \in \mathbb{Z}\}.$$

A continuación, definiremos una función  $\sigma$  para cada caso a las que llamaremos recorridos, de modo que, a partir de un punto  $x_i$  podremos llegar al siguiente  $x_j$ . Esta función es la que nos permitirá recorrer cada elemento de la sucesión.

**Definición 7.5.** *Sea  $A \in M(\{0, 1\})_{k \times k}$  una matriz adyacente y sean  $\Sigma_A^+, \Sigma_A$  las shifts de Markov lateral y bilateral respectivamente, definimos lo siguiente:*

1) *La transformación recorrido lateral es la función:*

$$\sigma : \Sigma_A^+ \rightarrow \Sigma_A^+$$

*tal que*

$$\sigma(x_0 x_1 x_2 x_3 \dots) = (x_1 x_2 x_3 \dots).$$

2) *La transformación recorrido bilateral es la función:*

$$\sigma : \Sigma_A \rightarrow \Sigma_A$$

*tal que*

$$\sigma(\dots x_{-2} x_{-1} x_0 x_1 x_2 \dots) = (\dots x_{-2} x_{-1} x_0 x_1 x_2 \dots).$$

De manera simplificada ambas transformaciones recorren un elemento a la izquierda como sigue:

$$(\sigma(x))_j = x_{j+1}$$

Ahora, dado que nuestro objetivo es construir las medidas de Bernoulli y de Markov, utilizaremos la misma metodología que se usa en [8], que consiste en construir una medida  $\mu$  sobre un álgebra de conjuntos que se extenderá por medio del Teorema de extensión Hahn-Kolmogorov a la  $\sigma$ -álgebra de Borel generada por esta.

Para lograr lo anterior, iniciamos definiendo una función  $n(x, y)$  que nos de el primer índice para el cual dos sucesiones en  $\Sigma_k^+$  son distintas, de forma que:

$$n(x, y) = n,$$

donde,  $x_j \neq y_j$  y  $x_j = y_j$  cuando  $0 \leq j \leq n - 1$ .

Por convención,  $n(x, y) = \infty$  si  $x = y$ .

**Definición 7.6.** Definimos a la función  $d : \Sigma_k^+ \rightarrow \mathbb{R}$  como sigue:

$$d(x, y) = \frac{1}{2^{n(x, y)}}$$

A continuación demostraremos que  $d$  es una métrica en  $\Sigma_k^+$ .

*Demostración.* Por definición  $d \geq 0$ , además de que  $d = 0$  sí y sólo sí  $x = y$ , también por definición sabemos que es  $d$  simétrica, por lo que falta demostrar que verifica la desigualdad triangular:

Sean  $x, y, z \in \Sigma_k^+$  :

a) Si  $x = y = z$  entonces:

$$0 = 0 + 0 = 0,$$

de modo que se cumple lo deseado.

b) Supongamos que  $x, y, z$  son todos distintos. Sean  $n(x, z) = n + 1$  y  $n(y, z) = m + 1$ .

Por lo anterior se cumple que:

$$n(x, z) \geq \min\{n(x, y), n(y, z)\}.$$

Así:

$$\begin{aligned} d(x, z) &= \frac{1}{2^{n(x, z)}} \leq \frac{1}{2^{\min\{n(x, y), n(y, z)\}}} \\ &\leq \frac{1}{2^{n(x, y)}} + \frac{1}{2^{n(y, z)}} = d(x, y) + d(y, z). \end{aligned}$$

c) Supongamos que  $x = y \neq z$ . Entonces se verifica lo siguiente:  $d(x, y) = d(y, x) = 0$  y  $d(x, z) = d(y, z) \geq 0$ . Por lo tanto:

$$d(x, z) = d(y, z) = 0 + d(y, z) = d(x, y) + d(y, z)$$

y

$$0 = d(x, y) = d(y, x) \leq d(x, z) + d(y, z) = 2d(x, z)$$

□

Por último, como los casos  $x = z \neq y$  y  $y = z \neq x$  con análogos a lo que se probó en c) hemos concluido con la demostración.

También se cumple lo siguiente:

**Proposición 7.1.**  $\sigma$  es continua en  $\Sigma_k^+$ .

*Demostración.* Sea  $\epsilon > 0$  y sea  $n \in \mathbb{N}$  tal que  $\frac{1}{2^n} \leq \epsilon$ .

Consideremos  $\delta = \frac{1}{2^{n+1}}$ . Así:

$$n(x, y) > n + 1,$$

pues,  $x, y$  son iguales al menos para los primeros  $n + 1$  elementos.

De lo anterior, se tiene que  $\sigma(x), \sigma(y)$  son iguales para al menos los primeros  $n$  elementos. Por lo tanto:

$$n(\sigma(x), \sigma(y)) > n,$$

lo cual implica que:

$$d(\sigma(x), \sigma(y)) < \frac{1}{2^n} < \epsilon.$$

□

Es importante mencionar que los resultados anteriores se pueden extender a  $\Sigma_k$ . Por esta razón en lo sucesivo se considerarán las definiciones para ambas shifts de Markov. Luego, definiremos unos objetos llamados cilindros que jugarán un papel similar al de los intervalos en el proceso de generación de la  $\sigma$ -álgebra de Borel en  $\mathbb{R}$ .

**Definición 7.7.** *Considere un conjunto  $S_k$  como en (7.1). Definimos:*

1) *Un cilindro  $C_{i_1 i_2 i_3 \dots i_m} \subset \Sigma_k^+$  en  $\Sigma_k^+$  como el conjunto:*

$$C_{i_1 i_2 i_3 \dots i_m} = \{(j_1 j_2 \dots) \in \Sigma_k^+ | j_l = i_l \text{ para toda } l = 1, 2, 3, \dots, m\}.$$

2) *Un cilindro  $C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}} \subset \Sigma_k$  en  $\Sigma_k$  como el conjunto:*

$$C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}} = \{(j_s j_{s+1} \dots) \in \Sigma_k | j_l = i_l \text{ para toda } l = r, r+1, r+2, \dots, r+m\}.$$

En cualquiera de los dos casos anteriores, decimos que el rango del cilindro es  $m$ .

**Proposición 7.2.** *Los cilindros son conjuntos abiertos y cerrados.*

*Demostración.* Solo consideraremos el caso para un cilindro en  $\Sigma_k^+$ .

Sea  $C_{i_1 i_2 i_3 \dots i_m} \subset \Sigma_k^+$  donde:

$$C_{i_1 i_2 i_3 \dots i_m} = \{(j_1 j_2 \dots) \in \Sigma_k^+ | j_l = i_l \text{ para toda } l = 1, 2, 3, \dots, m\}.$$

Sea  $x \in C_{i_1 i_2 i_3 \dots i_m}$  y consideremos  $\epsilon = \frac{1}{2^{m+1}}$ .

Supongamos que  $d(x, y) < \epsilon$ , por lo tanto  $n(x, y) > m + 1$ , es decir  $x, y$  coinciden en al menos los primeros  $m + 1$  elementos, así  $y \in C_{i_1 i_2 i_3 \dots i_m}$ , de modo que  $C_{i_1 i_2 i_3 \dots i_m}$  es abierto.

Ahora, observemos que:

$$\Sigma_k^+ - C_{i_1 i_2 i_3 \dots i_m} = \bigcup C_{i'_1 i'_2 i'_3 \dots i'_m},$$

donde la unión se realiza sobre todas las  $m$ -tuplas que son distintas a los valores del cilindro. Por lo tanto el complemento de  $C_{i_1 i_2 i_3 \dots i_m}$  es una unión finita de conjuntos abiertos, lo que implica que sea abierto y de esta manera  $C_{i_1 i_2 i_3 \dots i_m}$  un conjunto cerrado. □

Denotaremos por  $\gamma$  al álgebra generada por uniones finitas de cilindros y presentamos el siguiente resultado:

**Proposición 7.3.** *Sea  $X$  un espacio métrico compacto y sea  $\beta$  su  $\sigma$ -álgebra de Borel. Sea  $\gamma$  un álgebra de subconjuntos de Borel,  $\gamma \subset \beta$ . Supongamos que para cada  $x_1, x_2 \in X$  distintos, existen conjuntos abiertos disjuntos  $\gamma_1, \gamma_2 \in \gamma$  tales que  $x_1 \in \gamma_1$  y  $x_2 \in \gamma_2$ , entonces  $\gamma$  genera a  $\beta$ .*

Para aplicar esta proposición daremos por hecho la compacidad de  $(\Sigma_k^+, d)$  (la cual puede revisar en [8]) y realizaremos la siguiente observación:

**Observación 7.1.** *Sean  $x, y \in \Sigma_k^+$  distintos, entonces sabemos que existe un  $n \geq 0$  tal que  $x_n \neq y_n$  de modo que  $x, y$  pertenecen a distintos cilindros de rango  $n$ .*

Por lo tanto, el álgebra generada por uniones finitas de cilindros genera el  $\sigma$ -álgebra de Borel sobre  $(\Sigma_k^+, d)$  (la misma idea se puede aplicar en  $(\Sigma_k, d)$ ).

De esta manera hemos construido el  $\sigma$ -álgebra sobre el que definiremos las medidas de Bernoulli y de Markov.

Para iniciar con la construcción de estas medidas definiremos dos conceptos más, el de matriz estocástica y el de par estocástico.

**Definición 7.8.** *Sea  $P \in M_{k \times k}(R)$  una matriz cuadrada de  $k \times k$  tal que  $p_{ij} \geq 0$  para todo  $i, j \in \{1, 2, 3, \dots, k\}$ . Decimos que  $P$  es una matriz estocástica si la suma de sus renglones es 1, es decir si:*

para todo  $i \in \{1, 2, 3, \dots, k\}$  se cumple:

$$\sum_{j=1}^k p_{i,j} = 1.$$

**Definición 7.9.** *Sea  $P \in M_{k \times k}(R)$  una matriz estocástica y sea  $v$  un vector de probabilidad. Decimos que  $(P, v)$  es un par estocástico si:*

$$Pv = v.$$

En este momento, la asociación entre estos últimos dos conceptos con la teoría de procesos estocásticos es directa, pues lo que aquí hemos llamado matriz estocástica es también conocida como matriz de transición, que condensa la información sobre las probabilidades de los distintos cambios de estado dentro de un proceso estocástico.

Por otro lado, la segunda definición induce inmediatamente la noción de distribución estacionaria. El siguiente resultado enuncia cuándo es posible obtener dicha distribución, pues caracteriza al conjunto de matrices que puede determinar un par estocástico.

**Teorema 7.2** (Teorema de Perron-Frobenius). *Sea  $P \in M_{k \times k}(R)$  una matriz estocástica irreducible, es decir, una matriz para la cual en cada entrada existe  $n > 0$  tal que  $p_{i,j}^n > 0$ . Entonces para  $P$  existe un único vector de probabilidad  $v$ , tal que  $(P, v)$  es un par estocástico.*

La demostración detallada del Teorema (7.2) se puede encontrar en [7].

Antes de dar la definición de las medidas de Markov y de Bernoulli, haremos un comentario más. Estas medidas en principio se definen sobre el álgebra generada por las uniones finitas de cilindros. La extensión de estas se realiza con el teorema de extensión de Hahn-Kolmogorov y el cual enunciaremos sin prueba enseguida.

**Teorema 7.3** (Teorema de extensión de Hahn-Kolmogorov). *Sea  $\gamma$  un álgebra de subconjuntos de  $X$ . Sea  $\mu_0 : \gamma \rightarrow [0, 1]$  que satisface:*

i)  $\mu_0(\emptyset) = 0$ ;

ii) *Dada una colección  $(G_n)_{n \geq 1} \in \gamma$ , si son disjuntos y  $\bigcup_{n=1}^{\infty} G_n \in \gamma$  implica que:*

$$\mu_0\left(\bigcup_{n=1}^{\infty} G_n\right) = \sum_{n=1}^{\infty} \mu_0(G_n).$$

*Entonces existe una única medida de probabilidad  $\mu : \beta(\gamma) \rightarrow [0, 1]$  que es una extensión de  $\mu_0$ .*

Puede encontrar más detalles en [8].

Por todo lo anterior, definimos lo siguiente:

**Definición 7.10.** *Sea  $\beta$  la  $\sigma$ -álgebra de Borel generada por  $\gamma$ , donde  $\gamma$  es el álgebra generada por las uniones finitas de cilindros en las shifts laterales y bilaterales de Markov respectivamente.*

1) *Sea  $(\Sigma_k^+, \beta)$  un espacio medible y sea  $(P, p)$  un par estocástico, definimos la medida de Markov  $\mu_{(P,p)}$  lateral asociada a este par estocástico como sigue:*

$$\mu_{(P,p)}(C_{i_1 i_2 i_3 \dots i_m}) = p_{i_1} p_{i_1 i_2} p_{i_2 i_3} \dots p_{i_{m-1} i_m},$$

*para cada  $m \in \mathbb{N}$  y para todo  $i, i_1, i_2, \dots, i_m \in \{1, 2, \dots, k\}$ .*

2) *Sea  $(\Sigma_k, \beta)$  un espacio medible y sea  $(P, p)$  un par estocástico, definimos la medida de Markov  $\mu_{(P,p)}$  bilateral asociada a este par estocástico como sigue:*

$$\mu_{(P,p)}(C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}}) = p_{i_r} p_{i_r i_{r+1}} p_{i_{r+1} i_{r+2}} p_{i_{r+2} i_{r+3}} \dots p_{i_{r+(m-1)} i_{r+m}}$$

*para cada  $r \in \mathbb{Z}$  y para todo  $i_r, i_{r+1}, i_{r+2}, \dots, i_{r+m} \in \{1, 2, \dots, k\}$ .*

Ahora, definiremos la medida de Bernoulli como sigue:

**Definición 7.11.** *Sea  $\beta$  la  $\sigma$ -álgebra de Borel generada por  $\gamma$ , donde  $\gamma$  es el álgebra generada por las uniones finitas de cilindros en las shifts laterales y bilaterales de Markov respectivamente.*

- 1) *Sea  $(\Sigma_k^+, \beta)$  un espacio medible y sea  $p$  un vector de probabilidad. Definimos la medida de Bernoulli lateral como sigue:*

$$\mu_p(C_{i_1 i_2 i_3 \dots i_m}) = p_{i_1} p_{i_2} p_{i_3} \dots p_{i_m},$$

para cada  $m \in \mathbb{N}$  y  $i_1, i_2, \dots, i_m \in \{1, 2, 3, \dots, k\}$ .

- 2) *Sea  $(\Sigma_k, \beta)$  un espacio medible y sea  $p$  un vector de probabilidad. Definimos la medida de Bernoulli bilateral como sigue:*

$$\mu_p(C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}}) = p_{i_r} p_{i_{r+1}} p_{i_{r+2}} p_{i_{r+3}} \dots p_{i_{r+m}}$$

para cada  $r \in \mathbb{Z}$  y  $i_r, i_{r+1}, i_{r+2}, \dots, i_{r+m} \in \{1, 2, 3, \dots, k\}$ .

Luego, observemos que la medida de Bernoulli es un caso particular de la medida Markov, pues hay que tomar a  $P$  la matriz estocástica como la matriz con el mismo vector  $p$  de la distribución Bernoulli para cada renglón, donde el par estocástico estaría dado por esta matriz y el mismo vector  $p$ . Por lo anterior, basta demostrar que la segunda es invariante bajo la función recorrido  $\sigma$ .

Los siguientes dos resultados nos muestran esta propiedad de invarianza:

**Teorema 7.4.** *Sea  $(\Sigma_k, \beta, \mu_{(P,p)})$  un espacio de medida donde  $\beta$  es como en las definiciones anteriores y sea  $\mu_{(P,p)}$  la medida bilateral de Markov asociada al par estocástico  $(P, p)$ . Entonces, la función de recorrido  $\sigma$  preserva a  $\mu_{(P,p)}$ .*

*Demostración.* Sean  $r \in \mathbb{Z}$  y  $m \in \mathbb{N}$ . Entonces:

$$\sigma^{-1}(C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}}) = (C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}}).$$

Luego,

$$\mu_{(P,p)}(\sigma^{-1}(C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}})) = \mu_{(P,p)}(C_{i_r i_{r+1} i_{r+2} \dots i_{r+m}}).$$

Por lo tanto  $\sigma$  preserva la medida bilateral de Markov. □

**Teorema 7.5.** *Sea  $(\Sigma_k^+, \beta, \mu_{(P,p)})$  un espacio de medida donde  $\beta$  es como en las definiciones anteriores y sea  $\mu_{(P,p)}$  la medida lateral de Markov asociada al par estocástico  $(P, p)$ . Entonces, la función de recorrido  $\sigma$  preserva a  $\mu_{(P,p)}$ .*

*Demostración.* Consideremos:

$$\sigma^{-1}((C_{i_1 i_2 i_3 \dots i_m})) = \bigcup_{j=1}^k (C_{j i_1 i_2 i_3 \dots i_m}) = \{x \in \Sigma_k^+ | x = (j i_1 i_2 \dots i_m \dots), j \in S_k\},$$

de modo que esta última unión de conjuntos es disjunta. Así:

$$\begin{aligned} \mu_{(P,p)}(\sigma^{-1}((C_{i_1 i_2 i_3 \dots i_m}))) &= \mu_{(P,p)}\left(\bigcup_{j=1}^k (C_{j i_1 i_2 i_3 \dots i_m})\right) \\ &= \sum_{j=1}^k \mu_{(P,p)}((C_{i_1 i_2 i_3 \dots i_m})) \\ &= \sum_{j=1}^k p_j p_{j i_1} p_{i_1 i_2} \dots p_{i_{m-1} i_m} \\ &= p_{i_1 i_2} \dots p_{i_{m-1} i_m} \sum_{j=1}^k p_j p_{j i_1} \\ &= p_{i_1 i_2} \dots p_{i_{m-1} i_m} \\ &= \mu_{(P,p)}(C_{i_1 i_2 i_3 \dots i_m}). \end{aligned}$$

Por lo tanto  $\sigma$  preserva la medida bilateral de Markov. □

Para concluir esta sección daremos ejemplos de cada una de estas medidas.

Para el caso de la medida de Markov, de acuerdo a la teoría de procesos estocásticos, la matriz de transición de dos estados es:

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} \quad (7.5)$$

y tiene como distribución estacionaria a :

$$\pi = (\pi_1, \pi_2) = \left(\frac{b}{a+b}, \frac{a}{a+b}\right),$$

de modo que para  $a = \frac{1}{3}$  y  $b = \frac{1}{4}$  se tiene lo siguiente:

$$\begin{aligned} P &= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}; \\ \pi &= \left(\frac{3}{7}, \frac{4}{7}\right) \end{aligned} \quad (7.6)$$

Luego, consideramos el cilindro  $C_{212}$  y calculamos su medida de Markov como

$$\begin{aligned}\mu_{(P,\pi)}(C_{212}) &= \pi_2 p_{21} p_{12} \\ &= \frac{4}{7} \frac{1}{4} \frac{1}{3} = \frac{1}{21}.\end{aligned}$$

Por último, para el caso en el que la medida de Bernoulli tome el vector de probabilidad  $p$  como sigue:

$$p = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{4}\right).$$

Ahora, consideramos el cilindro  $C_{322411}$ , de modo que su medida de Bernoulli está dada por:

$$\mu_p(C_{322411}) = \frac{1}{2} \frac{1}{8} \frac{1}{8} \frac{1}{4} \frac{1}{8} \frac{1}{8} = \frac{1}{8^5}.$$

## 7.2. Sobre el teorema de Kolmogorov-Sinai.

En esta sección realizaremos ejemplos numéricos para las entropías de las medidas de Bernoulli y Markov bajo la función recorrido.

Para ello iniciamos con las observaciones:

**Observación 7.2.** Sean  $\sigma : \Sigma_k \rightarrow \Sigma_k$  la función recorrido y  $\mu_p$  la medida de Bernoulli bilateral  $\Sigma_k$  generada por  $p$ . Consideramos la siguiente partición de cilindros:

$$\xi = \{C_1, C_2, \dots, C_k\}.$$

Así:

$$\bigvee_{j=-n}^n \sigma^{-j} \xi = \{C_{i_{-n} \dots i_n} \mid i_{-n} \dots i_n \in \{1, 2, 3, \dots, k\}\}.$$

Luego, como el álgebra de los cilindros genera la  $\sigma$ -álgebra de Borel sobre  $\Sigma_k$ , por el teorema de Kolmogorov-Sinai, se tiene que:

$$h_{\mu_p}(\sigma) = h_{\mu_p}(\sigma, \xi).$$

**Observación 7.3.** Sea  $\sigma : \Sigma_k \rightarrow \Sigma_k$  la función recorrido y  $\mu_{(P,p)}$  la medida de Markov bilateral asociada al par estocástico  $(P, p)$ . Consideramos la misma partición  $\xi$  que utilizamos para el caso de la medida Bernoulli.

$$\xi = \{C_1, C_2, \dots, C_k\}.$$

Entonces:

$$H_{\mu_{(P,p)}}(\bigvee_{l=0}^{n-1} \sigma^{-l} \xi) = - \sum_{i_0, \dots, i_{n-1}} \mu_{(P,p)}(C_{i_0 \dots i_{n-1}}) \log_2(\mu_{(P,p)}(C_{i_0 \dots i_{n-1}})).$$

Simplificando, tenemos que:

$$H_{\mu_{(P,p)}}(\bigvee_{l=0}^{n-1} \sigma^{-l} \xi) = - \sum_{i=1}^k p_i \log_2(p_i) - (n-1) \sum_{i=1}^k \sum_{j=1}^k p_i p_{ij} \log_2(p_{ij}).$$

De modo que, por la proposición 6.3:

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_{\mu_{(P,p)}}(\bigvee_{l=0}^{n-1} \sigma^{-l} \xi) = h(\sigma, \xi).$$

Y, por el Teorema de Kolmogorov-Sinai:

$$h_{\mu_{(P,p)}}(\sigma) = h_{\mu_{(P,p)}}(\sigma, \xi) = - \sum_{i=1}^k \sum_{j=1}^k p_i p_{ij} \log_2(p_{ij}).$$

A continuación realizaremos unos ejemplos numéricos considerando ambas medidas y la función recorrido, de modo que, solo sustituiremos valores en las ecuaciones obtenidas en las observaciones anteriores.

Para el caso de la medida Markov, consideramos nuevamente la siguiente matriz estocástica y su distribución estacionaria:

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}; \quad (7.7)$$

$$\pi = \left( \frac{3}{7}, \frac{4}{7} \right).$$

Entonces:

$$\begin{aligned} h_{\mu_{(P,p)}}(\sigma) &= - \sum_{i=1}^k \sum_{j=1}^k p_i p_{ij} \log_2(p_{ij}) \\ &= -(\pi_1 p_{11} \log_2(p_{11}) + \pi_1 p_{12} \log_2(p_{12}) + \pi_2 p_{21} \log_2(p_{21}) + \pi_2 p_{22} \log_2(p_{22})) \\ &= -\left( \frac{3}{7} \frac{2}{3} \log_2\left(\frac{2}{3}\right) + \frac{3}{7} \frac{1}{3} \log_2\left(\frac{1}{3}\right) + \frac{4}{7} \frac{1}{4} \log_2\left(\frac{1}{4}\right) + \frac{4}{7} \frac{3}{4} \log_2\left(\frac{3}{4}\right) \right) \\ &= 0.8571429 \end{aligned}$$

Por último, para el caso de la medida de Bernoulli, nuevamente consideramos el vector de probabilidad  $p$  como sigue:

$$p = \left( \frac{1}{8}, \frac{1}{8}, \frac{1}{2}, \frac{1}{4} \right).$$

Como:

$$h_{\mu_p} = - \sum_{i=1}^k p_i \log_2(p_i),$$

entonces:

$$\begin{aligned} h_{\mu_p} &= -\left(\frac{1}{8} \log_2\left(\frac{1}{8}\right) + \frac{1}{8} \log_2\left(\frac{1}{8}\right) + \frac{1}{2} \log_2\left(\frac{1}{2}\right) + \frac{1}{4} \log_2\left(\frac{1}{4}\right)\right) \\ &= 1.75. \end{aligned}$$

Con esto, damos por concluido este capítulo.

## 8 Conclusiones.

A lo largo de este trabajo hemos realizado una construcción conjuntista del concepto de entropía, partiendo fundamentalmente de una noción de incertidumbre.

En este punto, esperamos que el lector entienda a la entropía como una medida de la información que falta para la certidumbre en un fenómeno. Esta interpretación debe quedar cubierta con lo construido en el capítulo 3. De una manera un tanto más detallada, en el capítulo 4 definimos la entropía sobre un espacio cociente, en el que nos interesaba medir el nivel de incertidumbre de una partición condicionada a la información disponible de otra partición.

La idea de condicionar con respecto a una partición tiene dos objetivos:

- i) El primero de ellos, siguiendo la idea para el caso de una partición, es conocer la información disponible a partir de la que se conoce de una partición dada, pues, recordemos que un principio, planteamos a la entropía como el número de preguntas esperadas para identificar un punto dentro de una partición medible, sin embargo, considere ahora, que nos interesa conocer la posición de un punto con respecto a una segunda partición, de modo que, la primera puede ayudarnos a identificar las zonas donde pudieran o no estar localizados nuestros puntos de interés en el espacio total y la segunda, darnos precisión sobre la ubicación de estos dentro de cada elemento de la primera, así, las celdas sobre las que ahora realizaremos la inspección no son las originales de cada una de las particiones y debemos entenderlas como intersecciones de las primeras.
- ii) Gracias al punto i) podemos extender estas ideas sobre las preimágenes de un endomorfismo (o automorfismo)  $T$ , pues como se realizó a partir del capítulo 6, consideramos una partición  $\xi$  y el producto de las aplicaciones  $T^{-1}$ . Esto con el objetivo de considerar todas las posibles combinaciones de estados (posiciones) que puede tener una partícula.

Es aquí cuando llegamos a la parte crucial de esta tesis, donde de manera intuitiva podemos entender a la entropía dinámica como la incertidumbre entre los distintos cambios de estado de un sistema. En términos más didácticos, hasta el momento solo veíamos para cada punto la información necesaria para encontrarlo en un conjunto, sin embargo al considerar a  $T$  como las reglas, condiciones físicas o agentes que cambian el ordenamiento de los puntos, en realidad estamos midiendo la incertidumbre

generada por los distintos movimientos de cada punto, es decir, estamos viendo la incertidumbre generada por cada nuevo ordenamiento de nuestros puntos sin que estos alteren su medida (probabilidad).

Una vez que se entiende lo anterior, el resto de la tesis consistió en establecer cómo calcularla. Es aquí cuando nos adentramos en dos resultados importantes. El primero de ellos es el que justifica la existencia de medidas invariantes. El teorema de Kyrilov-Bogolubov es el que da respuesta a este planteamiento, sin embargo, dado que no es un resultado trivial, la extensidad para desarrollar una prueba de este queda fuera de los alcances de este texto, pues la idea principal es darle protagonismo a la entropía, su comprensión y su cálculo.

El segundo resultado que se aborda, el teorema de extensión de Hahn-Kolmogorov, tiene que ver con la construcción de medidas sobre un álgebra de conjuntos que posteriormente se extienden a un  $\sigma$ -álgebra de Borel. Con ello, hemos realizado la construcción de las medidas Bernoulli y de Markov y una función que las mantiene invariantes.

Esto último es de gran relevancia para este trabajo, ya que nuestro planteamiento se hizo pensando en autorfismos de manera general, estas dos medidas y la función recorrido nos permiten aterrizar nuestras ideas en conceptos que se estudian en un primer curso de procesos estocásticos. En particular, abordamos el cálculo de la entropía dinámica para procesos estocásticos con una distribución estacionaria.

Por todo lo anterior, considero que los objetivos planteados al inicio y durante el desarrollo de esta tesis han sido alcanzados y con ello el lector que no está familiarizado con teoría ergódica puede tener una noción del significado de entropía y las formas de su cálculo.

# Bibliografía

- [1] V. A. Rokhlin, *Lectures on the entropy theory of measure-preserving transformations*. Russian Math Surveys 22, 1967.
- [2] A. Katok, “Fifty years of entropy in dynamics: 1958–2007,” *Journal of modern Dynamics*, 2007.
- [3] E. Glasner, *Ergodic Theory via Joinings*. American Mathematical Society, 1967.
- [4] T. Downarowics, *Entropy in Dynamical Systems*. Cambridge University Press, 2011.
- [5] V. M. and K. Oliveira, *Foundations of Ergodic Theory*. Cambridge University Press, 2016.
- [6] F. Le Maitre, “Notes on the kolmogorov-sinai theorem,” 2017.
- [7] R. Y. M. Pérez, “Compresión, ergodicidad y el teorema de shannon-mcmillan-breiman,” 2016.
- [8] C. Walkden, *Ergodic Theory*. 2018.