# Teoría de la Información Matemática

Alejandro Maass, Servet Martínez

# Índice general

# Capítulo 1

# Preliminares

## 1.1. Sistemas Básicos Estudiados.

Estudiaremos dos tipos de sistemas dinámicos:

- Sistemas dinámicos abstractos (s.d.a.): $(X, \mathcal{B}, \mu, T)$, donde

- $(X, \mathcal{B}, \mu)$ es un espacio de probabilidad,
- $T : X \to X$ es una transformación $\mathcal{B} - \mathcal{B}$ medible y
- $\mu$ es una medida $T-$ invariante, es decir, $T\mu(B) = \mu(T^{-1}(B)) = \mu(B)$, $(\forall B \in \mathcal{B})$ ..

- Sistemas dinámicos topológicos (s.d.t.): $(X, T)$, donde

- $X$ es un espacio métrico compacto y
- $T : X \to X$ es una transformación continua.

Veremos en el capítulo 5 que dado un s.d.t. $(X, T)$ siempre existe una medida $T-$ invariante $\mu$ definida sobre los borelianos de $X$. Luego el s.d.t. $(X, T)$ se puede mirar como un s.d.a. $(X, \mathcal{B}(X), \mu, T)$.

En la mayor parte de este texto asumimos que $T$, en ambos casos, es biyectiva (bimedible y bicontinua respectivamente), o al menos sobreyectiva. En el siguiente ejercicio mostramos en el caso topológico una manera de asociar sistemas biyectivos a cualquier s.d.t. sobreyectivo. La construcción que se presenta se puede realizar igualmente en el caso medible.

**Ejercicio:** *Sea $(X,T)$ un s.d.t. sobreyectivo. Existe un s.d.t. $(\widehat{X},\widehat{T})$ biyectivo (que se llama extensión natural de $(X,T)$) y una transformación continua y sobreyectiva $\pi:\widehat{X}\to X$ tal que $T\circ\pi(\widehat{x})=\pi\circ\widehat{T}(\widehat{x})$ en cada $\widehat{x}\in\widehat{X}$.*

La construcción es la siguiente, definimos

$$\widehat{X}=\left\{\widehat{x}=(x_0,x_1,...)\in X^{\mathbb{N}}:T(x_i)=x_{i-1},i\geq 1\right\}$$

y $\widehat{T}(\widehat{x})=(Tx_0,x_0,x_1,...)$ . Probar como ejercicio que $\left(\widehat{X},\widehat{T}\right)$ es un s.d.t. biyectivo. ♠

**Ejemplos:**

Los sistemas que presentamos a continuación servirán de ejemplos de base en el texto.

**1. Sistemas simbólicos.**

Sea $A$ un conjunto finito (que llamamos alfabeto). Sea $K=\mathbb{N}\vee\mathbb{Z}$. El conjunto

$$X=A^K=\{(x_i)_{i\in K}:\forall i\in K,x_i\in A\}$$

es un espacio métrico compacto con la métrica

$$d(x,y)=\sum_{i\in K}\frac{\delta(x_i,y_i)}{2^{|i|+1}}$$

donde para $x=(x_i)_{i\in K},y=(y_i)_{i\in K}$ se define $\delta(x_i,y_i)=\begin{cases}1 & x_i\neq y_i\\ 0 & x_i=y_i\end{cases}$ . Intuitivamente, si $x$ e $y$ son iguales en playas grandes en torno al origen, entonces están cerca. Es decir dado $\epsilon>0$ existe $N\in\mathbb{N}$ tal que

$$d(x,y)\leq\epsilon\Longleftrightarrow x_i=y_i,i\in K\cap\{-N,...,N\}.$$

Una base fundamental para esta topología son los cilindros. Dados símbolos $a_1,\ldots,a_m\in A$ y enteros $i_1<\ldots<i_m$ en $K$ se define el cilindro asociado a estas secuencias como

$$C(i_1,\ldots,i_m:a_1,\ldots,a_m)=\{x=(x_i)_{i\in K}\in A^K:x_{i_1}=a_1,\ldots,x_{i_m}=a_m\}.$$

Si $i_1 = k, i_2 = k+1, \ldots, i_m = k+m-1$ entonces usamos la notación

$$C(i_1, \ldots, i_m : a_1, \ldots, a_m) = [a_1 \ldots a_m]_k.$$

Se prueba (hacer como ejercicio) que todo cilindro se puede escribir como una reunión finita de cilindros del tipo $[a_1 \ldots a_m]_k$.

Definamos además la función shift $T : X \to X$ por $T(x)_i = x_{i+1}$, donde $x = (x_i)_{i \in K}$, $i \in K$. Usualmente se usa $\sigma$ en lugar de $T$ para referirse al shift.

**Observaciones:**

- si $K = \mathbb{Z}$ entonces $T$ es biyectiva,

- si $K = \mathbb{N}$, $|A| > 1$ entonces $T$ es sobreyectiva pero no biyectiva,

- si $K = \mathbb{N}$ la extensión natural de $\left(A^{\mathbb{N}}, T\right)$ es $\left(A^{\mathbb{Z}}, T\right)$ (ejercicio).

Luego $\left(A^K, \sigma\right)$ es un s.d.t. para $K = \mathbb{N}$ o $\mathbb{Z}$.

Necesitamos algunas nociones de Lenguajes Formales. El primer elemento es la noción de palabra: una palabra en el alfabeto $A$ es una secuencia finita de símbolos de $A$, digamos $w = w_1 \ldots w_m$. El largo de la palabra es $|w| = m$. El conjunto de palabras en el alfabeto $A$ se denota $A^+$. A este conjunto se le agrega la palabra vacía $\lambda$ y el conjunto resultante se denota $A^*$. Dadas dos palabras $u, w \in A^+$, decimos que $u$ es subpalabra de $w$ y anotamos $u \sqsubseteq w$ si. $|u| \leq |w|$ y existe $i \in \{1, \ldots, |w|\}$ tal que $u = w_i \ldots w_{i+|u|-1}$.

Si miramos $A^K$ a través de las palabras $\omega_0 ... \omega_{n-1} \in A^+$, nos damos cuenta que

$$\forall \omega_0 ... \omega_{n-1} \in A^+, \exists x \in A^K, x_0 ... x_{n-1} = \omega_0 ... \omega_{n-1}.$$

Por esta razón a $\left(A^K, \sigma\right)$ se le llama fullshift (unilateral si $K = \mathbb{N}$ y bilateral si $K = \mathbb{Z}$). Veamos una \'ultima notaci\'on. Dado un punto $x = (x_i)_{i \in K} \in A^K$ y enteros $i \leq j$ en $K$, anotaremos $x[i,j] = x_i \ldots x_j$.

Es posible sacar palabras de $A^K$ y obtener s.d.t. interesantes. La noción que se obtiene es la de subshift. Sea $X \subseteq A^K$ tal que $\sigma(X) = X$, es decir $\sigma : X \to X$ está bien definido. Cuando $K = \mathbb{Z}$ es biyectiva y si $K = \mathbb{N}$ es sobreyectiva. Si además $X$ es cerrado (para la topología producto o equivalentemente respecto de la métrica anterior), entonces decimos que $(X, \sigma)$ es un *subshift* de $\left(A^K, \sigma\right)$.

**Ejemplo:** *Sea* $X = \left\{ x \in \{0,1\}^{\mathbb{Z}} : \forall i \in \mathbb{Z},\, x[i,i+1] \neq 11 \right\}$*, entonces* $(X,\sigma)$ *es un subshift.*

Probemos que es shift invariante ($\sigma(X) = X$). Si $x = (x_i)_{i\in\mathbb{Z}} \in X,$ en cada $i \in \mathbb{Z}$ se tiene $\sigma(x)_i\, \sigma(x)_{i+1} = x_{i+1}x_{i+2} \neq 11$. Además $\forall x \in X, \sigma^{-1}(x) \in X$ pues $\sigma^{-1}(x)_i\, \sigma^{-1}(x)_{i+1} = x_{i-1}x_i \neq 11$. Luego, $\forall x \in X, \exists y \in X, \sigma(y) = x$.

Para probar la cerradura basta observar que $X = \left(\bigcup_{i\in\mathbb{Z}} [11]_i\right)^c$. ♠

En el ejemplo anterior, $X$ resulta de sacar la palabra 11 del conjunto de palabras finitas asociadas a $X$. En general, para un subshift $(X,\sigma)$ definimos su conjunto de palabras por

$$L(X) = \left\{ \omega \in A^+ : \exists x \in X, x_0 ... x_{|\omega|-1} = \omega \right\}.$$

Entonces en el ejemplo anterior

$$L(X) = A^+ \setminus \{\omega \in A^+ : 11 \sqsubseteq \omega\}.$$

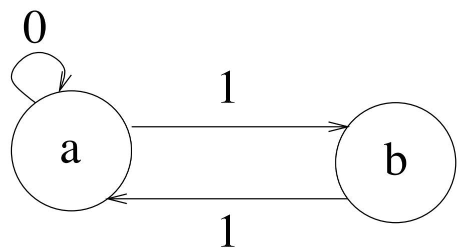
Sea $W \subseteq A^+$. Definimos $\langle W \rangle = \{w \in A^+ : \exists u \in W, u \sqsubseteq w\}$. Siempre que $L(X) = A^+ \setminus \langle W \rangle$ con $W \subseteq A^+, |W| < +\infty$, decimos que $X$ es un subshift de tipo finito. El ejemplo anterior es un subshift de tipo finito.

**Ejemplo**: *El siguiente subshift no es de tipo finito*

$$X = \{x \in \{0,1\}^{\mathbb{Z}} : \forall i,j \in \mathbb{Z}, (i < j \wedge x[i,j] = 01..,10 \Rightarrow j - i + 1\, es\, par)\}$$

*(entre 2 ceros no consecutivos hay un número par de unos).*

Se prueba que $X$ es subshift (ejercicio) y que $X$ no es subshift de tipo finito. Las siguientes palabras no están en $L(X)$: $D = \{01^i0 : i \text{ es impar}\}$. Si $W$ es un subconjunto finito de $\{0,1\}^+$, tal que $L(X) = \{0,1\}^+ \setminus \langle W \rangle$, entonces $\forall \omega \in D, \exists \overline{\omega} \in W$, tal que $\omega = u\overline{\omega}v$ con $u, v \in \{0,1\}^*$. Como $W$ es finito existen infinitos $\omega \in D$ asociados con el mismo $\overline{\omega}$, lo que implica que $\overline{\omega} \in L(X)$, que es una contradicción. El subshift $X$ tiene la propiedad siguiente: $L(X)$ está dado por el siguiente grafo finito: $L(X)$ son las palabras que se escriben al recorrer los caminos finitos sobre los arcos de este grafo. Todos los subshift con esta propiedad se llaman **Sóficos.** Pruebe como ejercicio que el subshift $X \subseteq \{a,b\}^{\mathbb{Z}}$ tal que $L(X)$ son las palabras finitas que se escriben al seguir los caminos de vertices en el grafo del ejemplo es un subshift de tipo finito.




♠

**Ejemplo:** Veamos ahora un ejemplo de subshift no sófico. Considere la substitución

$\tau_F : \{0,1\} \to \{0,1\}^+$ tal que $\tau_F(0) = 01, \tau_F(1) = 0$. La substitución $\tau_F$ se puede extender a $\{0,1\}^*$ y $\{0,1\}^{\mathbb{N}}$ por concatenación, es decir

$$\tau_F(a_0 \dots a_{n-1}) = \tau_F(a_0)\,\tau_F(a_1)\dots\tau_F(a_{n-1})$$

y

$$\tau_F((x_i)_{i\in\mathbb{N}} = \tau_F(x_0x_1\dots) = \tau_F(x_0)\,\tau_F(x_1)\dots$$

Vamos a construir a partir de $\tau_F$ un subshift. Sea $\overline{x} \in \{0,1\}^{\mathbb{N}}$ y considere el punto $x = 0\overline{x}$. Se tiene que $\exists! x_{\tau_F}$ tal que $\tau_F^i(x) \underset{i\to+\infty}{\longrightarrow} x_{\tau_F}$ y $\tau_F(x_{\tau_F}) = x_{\tau_F}$. Además $x_{\tau_F}$ es independiente de $\overline{x}$ (ejercicio). El subshift asociado a $\tau_F$ es $X = \overline{\{\sigma^i(x_{\tau_F}) : i \in \mathbb{N}\}}$.

**Ejercicios:** *Probar que $(X,\sigma)$ es un subshift y que no es sófico.*

Indicación: El hecho que $(X,\sigma)$ no es sófico viene del siguiente resultado general. Si un subshift $(X,\sigma)$ es minimal, es decir, $\overline{O(x)} = \overline{\{\sigma^i(x) : i \in \mathbb{N}\}} = X$ en cada $x \in X$ y $X$ es infinito, entonces $X$ no es sófico. Para ello se observa que un subshift minimal infinito $(X,\sigma)$ no tiene puntos periódicos para el shift, es decir, no existen puntos $x \in X$ tal que $\sigma^p(x) = x$ para algún $p \geq 1$. Si $(X,\sigma)$ fuese sófico entonces en el grafo que define $L(X)$ existe un ciclo (pues $X$ es infinito) y esto produce un punto periódico. Probar entonces que $(X,\sigma)$ no tine puntos periódicos.

## 2. Rotaciones en grupos compactos.

Sea $(G,*)$ un grupo topológico y $a \in G$. Se define $T_a : G \to G$ por $T_a g = g * a$. El sistema $(G, T_a)$ es un $s.d.t$ y si le ponemos la medida de Haar (que resulta ser la única invariante) lo estudiamos como $s.d.a.$

## 1.2. Equivalencia de Sistemas.

**Definición**. *Conjugación Topológica y Métrica .*

- *Decimos que los s.d.t.* $(X,T)$ *e* $(Y,S)$ *son conjugados si existe una biyección continua* $\pi : X \to Y$ *tal que* $S \circ \pi = \pi \circ T$.
- *Decimos que los s.d.a.* $(X,\mathcal{B},\mu,T)$ *y* $(Y,\mathcal{A},\nu,S)$ *son conjugados si existe* $\pi : X \to Y$ *biyectiva, bimedible tal que* $S \circ \pi = \pi \circ T$ *y* $\pi\mu = \nu$.
- Si en las definiciones anteriores $\pi$ es sólo sobreyectiva, decimos que $Y$ es factor de $X$, o que $X$ es extensión de $Y$.

**Definición**. *Noción de disyunción.*

- *Sean* $(X,T),(Y,S)$ *s.d.t. Decimos que* $X \perp Y$ *o que* $X$ *e* $Y$ *son disjuntos si cada vez que* $(X,T)$ *y* $(Y,S)$ *son factores de un mismo s.d.t.* $(Z,R)$ *por* $\pi : (Z,R) \to (X,T)$ *y* $\pi' : (Z,R) \to (Y,S)$, *entonces existe* $\alpha : (Z,R) \to (X \times Y, T \times S)$ *tal que* $\pi = p_X \circ \alpha, \pi' = p_Y \circ \alpha$, *donde* $p_X$ *y* $p_Y$ *son las proyecciones canónicas.*
- *Decimos que los s.d.a.* $(X,\mathcal{B},\mu,T)$ *y* $(Y,\mathcal{A},\nu,S)$ *son disjuntos o que* $(X,\mathcal{B},\mu,T) \perp (Y,\mathcal{A},\nu,S)$ *si cambiamos en el caso anterior lo continuo por medible.*

**Proposición**. *Probar que si* $(X,T)$ *y* $(Y,S)$ *son s.d.t. disjuntos entonces* $(X,T)$ *es minimal o* $(Y,S)$ *es minimal, en el sentido que*

$$(\forall x \in X), \overline{O(x)} = \overline{\{T^i(x) : i \in \mathbb{Z}\}} = X \vee (\forall y \in Y), \overline{O(y)} = \overline{\{S^i(y) : i \in \mathbb{Z}\}} = Y$$

*(se asumen* $S,T$ *homeomorfismos).*

**Demostración:** Si no fuese así, entonces existen $x_0 \in X$ y $y_0 \in Y$ tales que $\overline{O(x_0)} \neq X$ y $\overline{O(y_0)} \neq Y$. Además $\left(\overline{O(x_0)}, T\right)$ y $\left(\overline{O(y_0)}, S\right)$ son s.d.t. Denotemos por $\overline{O(x_0)} = A$, $\overline{O(y_0)} = B$. Sea $C = (A \times Y) \cup (X \times B)$. Se tiene que $T \times S(C) = C$, luego $(C, T \times S)$ es un s.d.t., tal que $p_X(C) = X$, $p_Y(C) = Y$, $T \circ p_X = p_X \circ (T \times S)$ y $S \circ p_Y = p_Y \circ T \times S$ . Como $X$ e $Y$ son disjuntos existe un factor $\alpha : (C, T \times S) \to (X \times Y, T \times S)$ tal que

$p_X = p_X \circ \alpha,\ p_Y = p_Y \circ \alpha$, lo que implica que $A = X$ o bien $B = Y$. Esto es una contradicción. ♠

**Ejercicio**: *Probar que el subshift* $(X, \sigma)$ *asociado a la substitución* $\tau_F(0) = 01, \tau_F(1) = 0$ *es disjunto de* $\left(\{0,1\}^{\mathbb{N}}, \sigma\right)$.

# Capítulo 2

# Teoremas Ergódicos.

## 2.1. Teoremas de Recurrencia.

Dado un s.d.a. $(X, \mathcal{B}, \mu, T)$, nos interesa estudiar propiedades del estilo siguiente

- dado $A \in \mathcal{B}$, ¿existe $n \geq 1$ tal que $T^{-n}(A) \cap A \neq \phi$ ?

- dado $A \in \mathcal{B}$, ¿exiten infinitos $n \geq 1$ tales que $T^{-n}(A) \cap A \neq \phi$ ?

La respuesta a estas preguntas marca el inicio de la teoría ergódica.

**Teorema**. *Sea* $(X, \mathcal{B}, \mu, T)$ *un s.d.a. y* $A \in \mathcal{B}$ *con* $\mu(A) > 0$. *Entonces,*

*Recurrencia de Poincaré débil*

$$\exists n \geq 1, \mu\left(T^{-n}(A) \cap A\right) > 0.$$

*Recurrencia de Poincaré fuerte*

$$\forall m \in \mathbb{N}, \exists n \geq m, \mu\left(T^{-n}(A) \cap A\right) > 0.$$

**Demostración**: Sea $F = \{x \in A : \forall n > 0, T^n x \notin A\}$. Analicemos los iterados de $F$ : $F \cap T^{-1}(F) = \phi$, pues si no existe $x \in F$ tal que $Tx \in F$. Además $F \cap T^{-i}(F) = \phi$ pues en el caso contrario existe $x \in F$ tal que

$T^i x \in F$. Por otro lado, se tiene que $T^{-i}(F) \cap T^{-j}(F) = \phi$ para $i < j$, pues $T^{-(j-i)}F \cap F = \phi$. Concluimos que $\{F, T^{-1}(F), ..., T^{-i}(F), ...\}$ es una familia disjunta de conjuntos medibles tal que

$$\mu\left(\bigcup_{i\in\mathbb{N}} T^{-i}F\right) = \sum_{i\in\mathbb{N}} \mu\left(T^{-i}F\right) = \sum_{i\in\mathbb{N}} \mu(F) \leq 1.$$

Luego $\mu(F) = 0$. Esto prueba el Teorema de Poincaré en su versión débil. La versión fuerte queda de ejercicio. ♠

## 2.2. Teoremas Ergódicos.

Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. y $f : X \to \mathbb{R}$ una función medible. Si tomamos $x \in X$, vamos a llamar observables de $x$ por $f$ a cualquier $f(T^i x)$ con $i \in \mathbb{N}$. Equivalentemente, decimos que $\{f(T^i x)\}_{i\in\mathbb{N}}$ es la observación por $f$ de la órbita de $x$. Un problema básico de la teoría ergódica es saber si existe el límite cuando $N$ tiende a infinito de $\dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} f \circ T^n$ en algún sentido y caracterizar dicho límite. Por muchos años se supuso en ciertos casos lo que se llamaba la "hipótesis ergódica", que decía que para muchos $x \in X$ el límite en $N$ de $\dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} f(T^n x)$ existe y es $\int_X f d\mu$. Al inicio de los años 30 aparecieron resultados matemáticos:

1) Von Neumann probó el resultado en $L^2(X, \mathcal{B}, \mu)$.

2) Birkhoff probó el resultado puntual.

### 2.2.1. Teorema Ergódico de Von Neumann.

**Teorema.** *Sea $U : H \to H$ una contracción lineal del espacio de Hilbert $H$, es decir, si $f \in H$, $\|Uf\| \leq \|f\|$, donde $\|\cdot\|$ es la norma en $H$. Sea $M = \{f \in H : Uf = f\}$ y $P : H \to M$ la proyección. Entonces $\forall f \in H$,*

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n f \underset{N\to\infty}{\longrightarrow} Pf.$$

**Demostración**: Definamos $N = \overline{\langle\{g - Ug : g \in H\}\rangle}$. Probemos que $M = N^{\perp}$ :

<u>$N^{\perp} \subseteq M$ :</u> Sea $h \in N^{\perp}$, luego

$$\begin{aligned} ||Uh - h||^2 &= \langle Uh - h, Uh - h\rangle \\ &= \langle Uh, Uh\rangle - \langle Uh, h\rangle - \langle h, Uh\rangle + \langle h, h\rangle \\ &\leq ||h||^2 - \langle h, U^*h\rangle - \langle U^*h, h\rangle + ||h||^2 \end{aligned}$$

Por otro lado en cada $g \in H$ se tiene $\langle h, g - Ug\rangle = 0$, luego $\langle h, g\rangle = \langle U^*h, g\rangle$. Esto prueba que $h = U^*h$. Usando la desigualdad anterior concluimos que $||Uh - h|| = 0$ y $h = Uh$. Esto prueba que $N^{\perp} \subseteq M$.

Observemos que $U^*$ también es una contracción:

$$\begin{aligned} ||U^*f||^2 &= \langle U^*f, U^*f\rangle \\ &= \langle UU^*f, f\rangle \\ &\leq ||UU^*f|| \cdot ||f|| \\ &\leq ||U^*f|| \cdot ||f|| \end{aligned}$$

pues $U$ es una contracción. De aquí deducimos que $||U^*f|| \leq ||f||$.

Sea $h \in M$. Entonces $Uh = h$ y $U^*h = h$. Basta tomar

$$||U^*h - h||^2 \leq ||h||^2 - \langle h, Uh\rangle - \langle Uh, h\rangle + ||h||^2.$$

Para $g \in H$ se tiene que

$$\langle h, g - Ug\rangle = \langle h, g\rangle - \langle h, Ug\rangle = \langle h, g\rangle - \langle U^*h, g\rangle = 0.$$

Además si $f = \lim_{i\to\infty} (g_i - Ug_i)$, donde $g_i \in H$, $i \in \mathbb{N}$ se tiene que

$$\begin{aligned} \langle h, f\rangle &= \langle h, f - (g_i - Ug_i)\rangle + \langle h, g_i - Ug_i\rangle \\ &\leq ||h|| \cdot ||f - (g_i - Ug_i)|| + 0 \leq \epsilon \end{aligned}$$

para cada $\epsilon$ en un $i$ suficientemente grande. Concluimos que $N^{\perp} = M$.

Probemos el teorema para $f = g - Ug$, $g \in H$. Notar que $Pf = 0$ pues $N^{\perp} = M$. En efecto,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n f\right\| &= \left\|\frac{1}{N}\sum_{n=0}^{N-1}(U^n g - U^{n+1}g)\right\| \\ &= \frac{1}{N}\left\|(U^N g - g)\right\| \\ &\leq \frac{2\|g\|}{N}\end{aligned}$$

lo que tiende a cero cuando $N$ va a infinito.

Si $f \in N$, se deduce directamente por continuidad que:

$$\left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n f\right\| \to_{N\to\infty} 0.$$

Finalmente si $f \in H$, $f = Pf + f_0, f_0 \in N$, luego

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n f - Pf\right\| &= \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n (f - Pf)\right\| \\ &= \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n f_0\right\|,\end{aligned}$$

lo que tiende a cero pues $f_0 \in N$. ♠

Para el caso de s.d.a. el teorema se enuncia como sigue.

**Teorema.** *Sean* $(X, \mathcal{B}, \mu, T)$ *un s.d.a.,* $f \in L^2(X, \mathcal{B}, \mu)$ *y* $\mathcal{I} = \{B \in \mathcal{B} : \mu(T^{-1}(B)\Delta B) = 0\}$ *($\sigma-$álgebra de conjuntos invariantes), luego*

$$\left\|\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n - \mathbb{E}(f|\mathcal{I})\right\|_2 \longrightarrow_{n\to\infty} 0$$

**Demostración**: Hay que traducir el teorema anterior,

1) $H = L^2(X, \mathcal{B}, \mu)$

2) $U : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{B}, \mu)$, definido como $Uf = f \circ T$ está bien definido pues como $T\mu = \mu$ entonces

$$\int_X |f|^2 \, d\mu = \int_X |f \circ T|^2 \, d\mu.$$

**Ejercicio**: Si $T\mu = \mu$ y $f$ es $\mu-$ integrable entonces $\int\limits_X f d\mu = \int\limits_X f \circ T d\mu.$

3) $\|Uf\|_2^2 = \int\limits_X |(f \circ T)|^2 d\mu = \int\limits_X |f|^2 d\mu = \|f\|_2^2$, luego $U$ es una contracción.

4) Sea $M = \{f \in L^2(X, \mathcal{B}, \mu) : Uf = f \circ T = f\}$. Para terminar probemos que $P = \mathbb{E}(\cdot \mid \mathcal{I})$. Verifiquemos que

$$\int_B Pf d\mu = \int_B f d\mu, \qquad \forall f \in L^2(X, \mathcal{B}, \mu), \forall B \in \mathcal{I}.$$

Sea $f = Pf + f_0$ donde $f_0 \in N$ (usamos la notación del teorema anterior). Luego,

$$\begin{aligned}
\int_B Pf d\mu &= \int_B f d\mu - \int_B f_0 d\mu \\
&= \int_B f d\mu - \int_X f_0 \cdot 1_B d\mu \\
&= \int_B f d\mu
\end{aligned}$$

pues $1_B \in M$ (verificar como ejercicio), $f_0 \in N$ y $N^\perp = M$.

Hay que verificar que $h = Pf$ es $\mathcal{I}-$ medible. Sea $A \in \mathcal{B}(\mathbb{R})$. Se tiene

$T^{-1}(h^{-1}(A)) = (h \circ T)^{-1}(A) = h^{-1}(A)$, lo que prueba que $Pf$ es $M-$medible.
♠

### 2.2.2. Teorema Ergódico Puntual (Teorema de Birkhoff).

Dado un s.d.a. $(X, \mathcal{B}, \mu, T)$ y $x \in X$, queremos ver la existencia del límite cuando $N$ tiende a infinito de

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x),$$

donde $f : X \to \mathbb{R}$ es una función medible. Hay varias aplicaciones importantes.

1) Supongamos $A \in \mathcal{B},\ x \in X$ y $f = 1_A$. Luego la existencia y conocimiento del límite en $N$ de

$$\frac{1}{N}\sum_{n=0}^{N-1} 1_A\left(T^n x\right) = \frac{\sharp\left\{0 \le n < N : T^n x \in A\right\}}{N}$$

permite calcular la frecuencia de entrada a $A$ de los iterados del punto $x$.

2) Otra aplicación es un Teorema de Grandes Números. Sean $\left(Y_i\right)_{i\in\mathbb{N}}$ variables aleatorias independientes, idénticamente distribuidas en $\{0,1\}$ con distribución $\pi = (\pi_0, \pi_1)$, y consideremos $X = \{0,1\}^{\mathbb{Z}}, \mu = \pi^{\mathbb{Z}}$ (medida producto), $T$ el shift y $f : X \to \{0,1\}$ dada por $f(x) = x_0$ en cada $x = (x_i)_{i\in\mathbb{Z}} \in X$. Luego, la existencia y conocimiento del límite en $N$ de

$$\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right) = \frac{1}{N}\sum_{n=0}^{N-1} x_n$$

permitiría deducir un Teorema de Grandes Números para variables independientes de Bernoulli.

3) **Problema de Gelfand:** ¿Cual es la frecuencia del dígito 7 en la secuencia del primer dígito en base 10 de $2^n, n \ge 0$ ? Es decir, si $\omega = 1,2,4,8,1,3,6,1,2, \ldots..$ es dicha secuencia y $P_7(n) = \sharp\{0 \le i < n : \omega_i = 7\}$ buscamos $\lim_{n\to\infty} \frac{P_7(n)}{n}$. Definamos $\alpha = \log_{10} 2$ y observemos que $P_7(n) = \sum_{j=0}^{n-1} 1_{I_7}(j \cdot \alpha \, mod\, 1)$, donde $I_7 = [\log_{10} 7, \log_{10} 8)$. La expresión $j \cdot \alpha \, mod\, 1$ corresponde al iterado $j-$ ésimo de 0 para la rotación en ángulo $\alpha$ en $[0,1]\, mod\, 1$. Si calculamos el límite obtenemos $\log_{10}\left(\frac{8}{7}\right)$.

Los teoremas ergódicos puntuales pasan por desigualdades llamadas maximales (en todo lo que sigue el sistema $T$ es invertible). La demostración del Teorema siguiente es de Garsia (1965).

**Teorema Maximal.** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. (invertible) y $f : X \to \mathbb{R}$ una función medible tal que $\int_X |f|\, d\mu < +\infty$. Si definimos $f^*(x) =_{N\ge 1}^{\sup}$*

$\frac{1}{N}\sum_{n=0}^{N-1} f\left(T^n x\right)$, *entonces*

$$\int_{\{x\in X: f^*(x)>0\}} f d\mu \geq 0.$$

**Demostración**: Para cada $n \geq 1$ definimos $f_n = \sup_{1\leq j\leq n} \sum_{k=0}^{j-1} f\circ T^k$. Claramente, $f = f_1 \leq f_2 \leq ....$ y se tiene la igualdad

$$\{x\in X : f^*(x) > 0\} = \bigcup_{n\geq 1}\{x\in X : f_n(x) > 0\}.$$

Probemos la siguiente propiedad: $f_n \leq f + f_n^+ \circ T$ para todo $n \geq 1$. Para $n = 1$, $f_1 = f \leq f + f^+ \circ T = f + f_1^+ \circ T$. Sea $1 < j \leq n$, entonces

$$\sum_{k=0}^{j-1} f\circ T^k = f + (\sum_{k=0}^{j-2} f\circ T^k)\circ T \leq f + (f_{j-1}\circ T) \leq f + f_n\circ T \leq f + f_n^+\circ T.$$

Como $f_1 = f \leq f + f_n^+ \circ T$ deducimos que $f_n \leq f + f_n^+ \circ T$. Entonces,

$$\begin{aligned} \int_{\{x\in X: f_n(x)>0\}} f d\mu &\geq \int_{\{x\in X: f_n(x)>0\}} (f_n - f_n^+\circ T) d\mu \\ &= \int_{\{x\in X: f_n(x)>0\}} f_n^+ d\mu - \int_{\{x\in X: f_n(x)>0\}} f_n^+\circ T d\mu \\ &\geq \int_X f_n^+ d\mu - \int_X f_n^+ \circ T d\mu = 0. \end{aligned}$$

Luego $\int_{\{x\in X: f_n(x)>0\}} f d\mu \geq 0$, y usando el teorema de convergencia dominada concluimos que $\int_{\{x\in X: f^*(x)>0\}} f d\mu \geq 0$. ♠

**Corolario**. *Sea* $\alpha \in \mathbb{R}$, *luego* $\int_{\{x\in X: f^*(x)>\alpha\}} f d\mu \geq \alpha \cdot \mu\left\{x\in X : f^*(x) > \alpha\right\}$.

**Demostración**: Tomar $g = f - \alpha$ y probar que $g^* = f^* - \alpha$. ♠

**Corolario**. *Sean* $Y \subseteq X$ *y* $\alpha \in \mathbb{R}$. *Si* $T(Y) \subseteq Y$ *y* $T : Y \to Y$ *es medible, entonces*

$$\int_{\{x\in Y: f^*(x)>\alpha\}} f d\mu \geq \alpha \cdot \mu\{x\in Y : f^*(x) > \alpha\}.$$

**Demostración:** ejercicio ♠

**Teorema Ergódico de Birkhoff (en $L^1$).** *Sean $(X,\mathcal{B},\mu,T)$ un s.d.a. y $f\in L^1(X,\mathcal{B},\mu)$. Se tiene,*

(1) $\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n=\overline{f}$ *existe* $\mu-c.s.$

(2) $\overline{f}\circ T=\overline{f}\qquad \mu-c.s.$

(3) $\overline{f}\in L^1(X,\mathcal{B},\mu)$ *y* $\left\|\overline{f}\right\|_1\leq\|f\|_1$.

(4) $\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n\longrightarrow_{L^1}\overline{f}$.

(5) *Si* $\mathcal{I}=\{A\in\mathcal{B}:\mu(A\Delta T^{-1}(A))=0\}$, *entonces* $\overline{f}=\mathbb{E}(f|\mathcal{I})$.

**Demostración**: Hacemos la demostración con $T$ invertible y con $L^1(X,\mathcal{B},\mu)$ real.

(1) Sean $\alpha,\beta\in\mathbb{R}$, $\alpha<\beta$ y definamos

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

Probaremos que $\mu(E_{\alpha,\beta})=0$. Esto basta pues si tomamos un conjunto denso en $\mathbb{R}^2$

$$\left\{(\alpha_i,\beta_i)\in\mathbb{Q}^2:\, i\in\mathbb{N},\alpha_i<\beta_i\right\}$$

se tendrá que

$$\mu\left\{x\in X:\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)\text{ no existe}\right\}=0.$$

Observemos que $T(E_{\alpha,\beta})\subseteq E_{\alpha,\beta}$ pues los límites liminf y limsup no cambian si tomo $x$ o bien $Tx$. En efecto, basta considerar la igualdad siguiente,

$$\frac{(N+1)}{N}.\frac{1}{N+1}\sum_{n=0}^{N} f(T^n x)-\frac{1}{N}f(x)=\frac{1}{N}\sum_{n=0}^{N-1} f(T^n(Tx)).$$

Además $E_{\alpha,\beta}\subseteq\{x\in X:f^*(x)\geq\beta\}$ (recordar $f^*$del Teorema Maximal). Luego de los dos corolarios anteriores se deduce

$$\begin{aligned}\int_{\{x\in E_{\alpha\beta}:f^*(x)\geq\beta\}} f d\mu &\geq \beta\cdot\mu(\{x\in X: f^*(x)\geq\beta\}\cap E_{\alpha\beta}\})\\ &= \beta\cdot\mu(E_{\alpha,\beta}).\end{aligned}$$

Por otro lado, $E_{\alpha,\beta}\subseteq\{x\in X:(-f)^*\geq-\alpha\}$, luego usando los corolarios anteriores

$$\int\limits_{\left\{x\in E_{\alpha,\beta}:(-f)^*\geq-\alpha\right\}}(-f)d\mu\geq(-\alpha)\,\mu\left(E_{\alpha,\beta}\right).$$

Concluimos $\beta\cdot\mu\left(E_{\alpha,\beta}\right)\leq\displaystyle\int_{E_{\alpha,\beta}} f d\mu\leq\alpha\cdot\mu\left(E_{\alpha,\beta}\right)$. Como $\beta>\alpha$ necesariamente $\mu\left(E_{\alpha,\beta}\right)=0$.

(2) Queda de ejercicio.

(3)

$$\begin{aligned}\int_X \left|\lim_{N\to\infty}\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} f\circ T^n(x)\right| d\mu(x) &\leq \int_X \liminf_{N\to\infty}\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1}|f|\circ T^n(x)\,d\mu(x)\\ &\leq \liminf_{N\to\infty}\int_X\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1}|f|\circ T^n(x)\,d\mu(x)\\ &= \int_X|f|\,d\mu(x)<+\infty.\end{aligned}$$

(4) Sean $f\geq0$ y $0\leq g\leq f, g$ acotada. Luego,

$$\begin{aligned}\left\|\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} f\circ T^n-\overline{f}\right\|_1 &\leq \left\|\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} f\circ T^n-\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} g\circ T^n\right\|_1\\ &+\left\|\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} g\circ T^n-\overline{g}\right\|_1+\left\|\overline{f}-\overline{g}\right\|_1\\ &\leq 2\left\|f-g\right\|_1+\left\|\tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} g\circ T^n-\overline{g}\right\|_1.\end{aligned}$$

Tomando $\|f-g\|_1$ pequeño y usando el teorema de convergencia dominada en el otro término concluimos

$$\left\|\frac{1}{N}\sum_{n=0}^{N-1} f\circ T^n-\overline{f}\right\|_1\underset{N\to\infty}{\longrightarrow}0$$

(5) Sean $n\geq1, k\in\mathbb{Z}$ y $A\in\mathcal{I}$. Definimos $A_{n,k}=\left\{x\in A:\frac{k}{2^n}\leq\overline{f}(x)<\frac{k+1}{2^n}\right\}$. Claramente $T:A_{n,k}\to A_{n,k}$ está bien definido y es inyectivo. Además

$\mu(A\Delta \cup_{k\in\mathbb{Z}} A_{n,k}) = 0$. En $A_{n,k}$, $f^*(x) > \frac{k}{2^n} - \epsilon$ para $\epsilon$ fijo y $(-f)^*(x) > -\left(\frac{k+1}{2^n}\right)$. Luego (de los corolarios anteriores)

$$\int_{A_{n,k}} f d\mu \geq \left(\frac{k}{2^n} - \epsilon\right) \mu\left(A_{n,k}\right)$$

y

$$\int_{A_{n,k}} (-f)\, d\mu \geq -\left(\frac{k+1}{2^n}\right) \mu\left(A_{n,k}\right).$$

Concluimos

$$\left|\int_{A_{n,k}} f d\mu - \int_{A_{n,k}} \overline{f} d\mu\right| \leq \frac{1}{2^n} \mu\left(A_{n,k}\right).$$

Finalmente sumando sobre $k$ se obtiene

$$\left|\int_A f d\mu - \int_A \overline{f} d\mu\right| \leq \frac{1}{2^n} \mu\left(A\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Como $\overline{f}$ es $\mathcal{I}-$ medible y $A \in \mathcal{I}$ obtenemos $\overline{f} = \mathbb{E}(f \mid \mathcal{I})$. ♠

**Observación:**

1) si $f \in L^1_{\mathbb{C}}(X, \mathcal{B}, \mu)$ el teorema es cierto pues se descompone $f$ en parte real e imaginaria como $f = f_1 + i f_2$.

2) si $\mathcal{I} =_\mu \{X, \phi\}$ ($\mathcal{I}$ es trivial), entonces $\mathbb{E}\left(f \mid \mathcal{I}\right) = \int_X f d\mu$.

3) La recíproca de (2): sea $A \in \mathcal{I}$ y $1_A \in L^1\left(X, \mathcal{B}, \mu\right)$, luego

$$\frac{1}{N}\sum_{n=0}^{N-1} 1_A\left(T^n x\right) = \frac{1}{N}\sum_{n=0}^{N-1} 1_{T^{-n}A}\left(x\right) = 1_A\left(x\right) \underset{N\to\infty}{\longrightarrow} \mu\left(A\right) \mu - c.s.$$

lo que implica $\mu\left(A\right) = 0 \vee \mu\left(A\right) = 1$.

# Capítulo 3

# Sistemas Ergódicos y Mezcladores.

## 3.1. Sistemas Ergódicos.

**Definición.** *Decimos que un s.d.a.* $(X, \mathcal{B}, \mu, T)$ *es ergódico si*

$$\mathcal{I} = \left\{ A \in \mathcal{B} : \mu(T^{-1}(A) \Delta A) = 0 \right\}$$

*es trivial, es decir si* $A \in \mathcal{I}$ *entonces* $\mu(A) = 0 \vee \mu(A) = 1$.

En este caso $\mathbb{E}(f \mid \mathcal{I}) = \int_X f d\mu$. Usando el Teorema Ergódico puntual se deduce que el sistema es ergódico (a veces decimos que $\mu$ es ergódica) si

$$\forall f \in L^1(X, \mathcal{B}, \mu), \frac{1}{N} \sum_{n=0}^{N-1} f(T^n(x)) \underset{N \to \infty}{\longrightarrow} \int_X f d\mu.$$

**Proposición.** *Un s.d.a.* $(X, \mathcal{B}, \mu, T)$ *es ergódico sí y sólo sí toda función medible* $f : X \to \mathbb{C}$ *tal que* $f = f \circ T$ *es constante* $\mu-$ *casi seguramente.*

**Demostración:**

$(\Leftarrow)$ Sea $A \in \mathcal{I}$ y $f = 1_A$. Claramente $f$ es medible y como $\mu(A \Delta T^{-1}(A)) = 0$ entonces $f \circ T = 1_{T^{-1}(A)} = f$ $\mu-$casi seguramente, de lo que concluimos que $f =$ cte $\mu-$ casi seguramente. Es decir, $1_A =_\mu 1 \vee 0$ y $\mu(A) = 0 \vee 1$.

($\Rightarrow$) para $\mu$ casi todo punto $x \in X$ si $f =_\mu f \circ T$ entonces

$$\frac{1}{N}\sum_{n=0}^{N-1} f \circ T^n (x) = f(x) = \int_X f d\mu = cte.$$

♠

**Teorema.** *Sea $(X, \mathcal{B}, \mu, T)$ un sistema dinámico abstracto.*

*(1) El sistema es ergódico sí y sólo sí todo valor propio de $U_T : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{B}, \mu)$ definido por $U_T f = f \circ T$ es simple.*

*(2) Si el sistema es ergódico entonces todo valor propio de $U_T$ es simple y su módulo es 1. Además $G = \{\lambda \in \mathbb{C} : \lambda$ es valor propio de $U_T\}$ es un subgrupo de $\mathbb{K} = \{z \in \mathbb{C} : |z| = 1\}$.*

*(3) Si $G \subseteq \mathbb{C}$ es un subgrupo de $\mathbb{K}$ entonces existe un s.d.a. $(X, \mathcal{B}, \mu, T)$ ergódico tal que $G$ es su subgrupo de valores propios (no haremos la demostración).*

**Demostración:** Por simplicidad anotaremos $L^2 = L^2(X, \mathcal{B}, \mu)$.

(1) ($\Leftarrow$) Si 1 es valor propio entonces existe $f \neq 0 \in L^2$ tal que $f \circ T =_\mu f$, y 1 como función de $L^2$ es vector propio. Luego $f = cte \cdot 1$ y $(X, \mathcal{B}, \mu, T)$ es ergódico.

($\Rightarrow$) Si $T$ es ergódico, $f \circ T =_\mu f$ implica $f =_\mu cte$. Luego $\lambda = 1$ es valor propio simple.

(2) Sea $\lambda \in \mathbb{C}$ valor propio. Entonces existe $f \in L^2 \setminus \{0\}$, $f \circ T =_\mu \lambda \cdot f$. De aquí

$$\int_X |f|^2 d\mu = \int_X (|f \circ T|)^2 \, d\mu = |\lambda|^2 \int_X |f|^2 d\mu$$

de lo que concluimos que $|\lambda| = 1$.

Sean $f, g \in L^2 \setminus \{0\}$ y $\alpha, \beta$ valores propios asociados a $f$ y $g$ $(|\alpha| = |\beta| = 1)$. Sea $\overline{g} = \frac{f}{g}$. Vamos a probar que $\overline{g} \circ T = \frac{\alpha}{\beta}\overline{g}$. En efecto,

$$\overline{g} \circ T = \tfrac{f \circ T}{g \circ T} = \tfrac{\alpha f}{\beta g} = \tfrac{\alpha}{\beta}\overline{g}$$

y $G$ es no vacío pues 1 es valor propio. Luego $G$ es subgrupo. Sean $f, g \in L^2$ asociados a $\lambda$ . Luego $\frac{1}{g}$ es vector propio asociado a $\lambda^{-1}$y $\frac{f}{g}$ es vector propio asociado a 1. La ergodicidad se deduce del hecho que $\frac{f}{g} =_\mu cte$ implica $f =_\mu g \cdot cte$. ♠

**Ejemplo:** Sea $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Probemos que $Tx = x + \alpha \mod 1$ en $\Pi^1 = [0,1]mod1$ y $\mu$ la medida de lebesgue es ergódico. Para ello calculamos las funciones $T-$ invariantes. Sea $f \in L^2(\Pi^1, \mathcal{B}(\Pi^1), \mu)$. Esta se escribe usando la descomposición de Fourier como $f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi inx}$, luego

$$f \circ T(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi in(x+\alpha)} = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i\alpha} e^{2\pi inx}.$$

Si $f =_\mu f \circ T$ entonces $a_n \left( \underbrace{1 - e^{2\pi in\alpha}}_{\neq 0} \right) = 0$ para $n \neq 0$ pues $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Luego $a_n = 0,\ n \neq 0$ y $f =_\mu cte$. De aquí deducimos que 1 es valor propio en $L^2(\Pi^1, \mathcal{B}(\Pi^1), \mu)$ y $T$ es ergódico. ♠

El ejemplo anterior sirve para intuir el resultado del problema del Gelfand. Recordemos que se quiería calcular

$$\lim_{n\to\infty} \left| \frac{\{0 \leq i \leq n-1 : \omega_i = 7\}}{n} \right|.$$

Tomemos $\alpha = log_{10}7$ y definamos $I_7 = [log_{10}7, log_{10}8)$. La transformación definida en $\Pi^1$ por $Tx = x + \alpha\, mod\, 1$ satisface,

$$\frac{1}{N} \sum_{i=0}^{N-1} 1_{I_7}\left(T^i(0)\right) = \frac{1}{N} \sum_{i=0}^{N-1} 1_{I_7}(i \cdot \alpha\, mod 1) = \frac{\sharp\{0 \leq i \leq N-1 : i \cdot \alpha \text{ mod1} \in I_7\}}{N}.$$

Como $T$ es ergódica se tiene que si 0 pertenece al conjunto de convergencia el límite será $log_{10}\frac{8}{7}$. ♠

**Proposición.** *Un sistema dinámico abstracto $(X, \mathcal{B}, \mu, T)$ es ergódico sí y sólo sí para $f, g \in L^2(X, \mathcal{B}, \mu)$ se tiene,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \left\langle U_T^k f, g \right\rangle = \langle f, 1 \rangle \cdot \langle 1, g \rangle .$$

**Demostración:** Denotemos $L^2 = L^2(X, \mathcal{B}, \mu)$.

$(\Rightarrow)$ Asumamos que $T$ es ergódico. Se tiene,

$$\frac{1}{N}\sum_{n=0}^{N-1}\langle U_T^n f, g\rangle = \left\langle \frac{1}{N}\sum_{n=0}^{N-1} U_T^n f, g\right\rangle \underset{N\to\infty}{\longrightarrow} \langle \mathbb{E}(f \mid \mathcal{I}), g\rangle$$

y $\langle \mathbb{E}(f \mid \mathcal{I}), g\rangle = \langle < f, 1 >, g\rangle = \langle f, 1\rangle \cdot \langle 1, g\rangle$.

$(\Leftarrow)$ Sean $f \in L^2$ y $g \in L^2$. Por el Teorema Ergódico puntual y de Von Neumann se tiene que $\dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} U_T^n f$ converge a $\mathbb{E}(f \mid \mathcal{I})$ puntualmente y en $L^2$. Luego$\left\langle \frac{1}{N}\sum_{n=0}^{N-1} U_T^n f, g\right\rangle$ converge a $\langle \mathbb{E}(f \mid \mathcal{I}), g\rangle$ en cada $g \in L^2$. De aquí deducimos que $\langle f, 1\rangle = \mathbb{E}(f \mid \mathcal{I})$ y que $T$ es ergódico. ♠

**Proposición.** *Un sistema dinámico abstracto* $(X, \mathcal{B}, \mu, T)$ *es ergódico sí y sólo sí para todo* $A, B \in \mathcal{B}$ *se tiene*

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

**Demostración**: Tarea. ♠

**Ejercicio:** Probar que en la proposición anterior basta con $A$ y $B$ en una semiálgebra que genera $B$ .

**Ejemplo:** El sistema $\left(\{0,1\}^{\mathbb{Z}}, T, \mu_\pi\right)$ es ergódico.Donde la medida $\mu_\pi$ es la medida producto en $\{0,1\}^{\mathbb{Z}}$ definida por $\pi = (\pi_0, \pi_1)$ tal que $\pi_0 + \pi_1 = 1$ y $\pi_0, \pi_1 \in (0,1)$, y $T$ es el shift.

Usamos el ejercicio anterior: sean $A = [x_0...x_{l-1}]_a$ y $B = [y_0...y_{m-1}]_b$. Si $k$ es grande $(\{a+k, ..., a+k+l-1\} \cap \{b, ..., b+m-1\} = \phi)$ se tendrá,

$$\begin{aligned}\mu_\pi\left(T^{-k}A \cap B\right) &= \mu_\pi\left([x_0...x_{l-1}]_a \cap [y_0...y_{m-1}]\right)\\ &= \pi_{x_0}...\pi_{x_{l-1}} \cdot \pi_{y_0}...\pi_{y_{m-1}}\\ &= \mu_\pi(A)\cdot\mu_\pi(B).\end{aligned}$$

Entonces $\frac{1}{N}\sum_{k=0}^{N-1}\mu\left(T^{-k}A\cap B\right)\underset{N\to\infty}{\longrightarrow}\mu(A)\cdot\mu(B)$, lo que prueba la ergodicidad. ♠

## 3.2. Sistemas Mezcladores.

Sea $(X,\mathcal{B},\mu)$ un s.d.a. La ergodicidad dice que $\forall A,B\in\mathcal{B}$,

$$\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(T^{-n}A\cap B\right)\underset{N\to\infty}{\longrightarrow}\mu(A)\,\mu(B).$$

Es decir, $\left|\frac{1}{N}\sum_{n=0}^{N-1}\mu\left(T^{-n}A\cap B\right)-\mu(A)\cdot\mu(B)\right|\underset{N\to\infty}{\longrightarrow}0$. Una cota superior directa,

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

Luego si $\left|\mu\left(T^{-n}A\cap B\right)-\mu(A)\cdot\mu(B)\right|$ converge a 0 se tiene la ergodicidad. La condición anterior tiene interés en Teoría Ergódica, lo que motiva la definición siguiente.

**Definición.** *Sea $(X,\mathcal{B},\mu,T)$ un sistema dinámico abstracto.*

(1) *$T$ es débilmente mezclador si en todo $A,B\in\mathcal{B}$ se tiene*

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

(2) *$T$ es mezclador si en todo $A,B\in\mathcal{B}$ se tiene*

$$\mu\left(T^{-n}A\cap B\right)\underset{n\to\infty}{\longrightarrow}\mu(A)\,\mu(B).$$

Claramente $T$ mezclador implica $T$ débilmente mezclador lo que a su vez implica $T$ ergódico.

**Teorema.** *Sea $(X,\mathcal{B},\mu,T)$ un s.d.a. Son equivalentes,*

1) *$T$ es débilmente mezclador,*

2) $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ *es ergódico,*

3) $\forall A, B \in \mathcal{B}, \quad \exists\, E \subseteq \mathbb{N}$ *de densidad* $0$, *tal que*

$$\lim_{\substack{n\to\infty \\ n\notin E}} \mu\left(T^{-n}A \cap B\right) = \mu(A)\,\mu(B)\,, \ \textit{donde } d(E) = \lim_{n\to\infty} \frac{|E \cap \{0 \le j \le n\}|}{n}$$

*es la densidad de* $E$.

4) $\forall f, g \in L^2(X, \mathcal{B}, \mu)\,, \ \lim_{N\to\infty} \dfrac{1}{N} \displaystyle\sum_{n=0}^{N-1} |\langle U_T^n f, g\rangle - \langle f, 1\rangle \langle 1, g\rangle| = 0.$

5) *Las únicas funciones propias de* $U_T$ *son las constantes.*

Para probar el teorema usamos el lema siguiente.

**Lema.** *Si* $f : \mathbb{R} \to \mathbb{R}$ *es no-negativa y acotada entonces,*

$$\frac{1}{N} \sum_{n=0}^{N-1} f(n) \underset{N\to\infty}{\longrightarrow} 0 \Leftrightarrow \exists\, E \subseteq \mathbb{N}$$

*de densidad cero tal que* $f(n) \underset{\substack{n\to\infty \\ n\in E^c}}{\longrightarrow} 0.$

**Demostración:** $(\Leftarrow)$ Supongamos que $\exists\, E \subseteq \mathbb{N}$ de densidad cero tal que $f(n)$ converge a $0$ en $E^c$. Luego,

$$\begin{aligned} \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} f(n) = \tfrac{1}{N} \sum_{\substack{n \le N-1 \\ n\in E}} f(n) + \tfrac{1}{N} \sum_{\substack{n \le N-1 \\ n\in E^c}} f(n) \quad &\le\ K \tfrac{|E\cap\{n\le N-1\}|}{N} \\ &+ \tfrac{1}{N} \textstyle\sum_{\substack{n \le N-1 \\ n\in E^c}} f(n)\,. \end{aligned}$$

De lo que se deduce el resultado.

$(\Rightarrow)$ Definamos $E_m = \left\{n \in \mathbb{N} : f(n) > \frac{1}{m}\right\}$, luego

$$\begin{aligned} \tfrac{|E_m\cap\{n\le N-1\}|}{N} \quad &= \tfrac{1}{N} \textstyle\sum_{k=0}^{N-1} 1_{E_m}(k) \\ &\le \tfrac{m}{N} \textstyle\sum_{k=0}^{N-1} f(k) \underset{N\to\infty}{\longrightarrow} 0. \end{aligned}$$

Luego $E_m$ tiene densidad cero.

Formemos una sucesión de enteros $i_0 = 0 < i_1 < i_2 < ...$ tales que para todo $(\forall n \ge 1)$,

$$\frac{1}{N} \sum_{n=0}^{N-1} 1_{E_m}(n) \le \frac{1}{m} \text{ para } N \ge i_{m-1}.$$

Definimos $E = \bigcup_{m\geq 1} E_m \cap [i_{m-1}, i_m]$. Probemos que $E$ tiene densidad cero. Sea $i_{m-1} \leq N \leq i_m$, luego

$$\begin{aligned} \tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} 1_E(n) &= \tfrac{1}{N}\textstyle\sum_{n=0}^{i_{m-1}-1} 1_E(n) + \tfrac{1}{N}\textstyle\sum_{n=i_{m-1}}^{N-1} 1_E(n) \\ &\leq \tfrac{1}{N}\textstyle\sum_{n=0}^{i_{m-1}-1} 1_{E_{m-1}}(n) + \tfrac{1}{N}\textstyle\sum_{n=0}^{N-1} 1_{E_m}(n) \\ &\leq \tfrac{1}{N}\textstyle\sum_{n=0}^{i_{m-1}-1} 1_{E_{m-1}}(n) + \tfrac{1}{m} \\ &\leq \tfrac{1}{m-1} + \tfrac{1}{m}. \end{aligned}$$

Concluimos que $d(E) = 0$. Además $f(n)$ converge a 0 en $E^c$pues si $n \in [i_{m-1}, i_m]$ se tiene $f(n) \leq \frac{1}{m}$. ♠

**Demostración del Teorema:**

(1) $\Rightarrow$ (2) Sean $A, B, C, D, \in \mathcal{B}$. Se tiene,

$$\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu\times\mu\left((T\times T)^{-n}(A\times B)\cap(C\times D)\right)-\mu(A)\mu(B)\mu(C)\mu(D)\right| = (*)$$

$$(*) \quad = \frac{1}{N}\sum_{n=0}^{N-1}\left|\mu\left(T^{-n}A\cap C\right)\mu\left(T^{-n}B\cap D\right)-\mu(A)\mu(B)\mu(C)\mu(D)\right|.$$

Sean $E_1, E_2 \subseteq \mathbb{N}$ conjuntos de densidad cero como en el lema anterior. Sea $f(n) = |\mu(T^{-n}A\cap C) - \mu(A)\mu(C)|$. Usando el lema deducimos

$$\mu\left(T^{-n}A\cap C\right)\mu\left(T^{-n}B\cap D\right) \underset{\substack{n\to\infty \\ n\in E^c}}{\longrightarrow} \mu(A)\mu(B)\mu(C)\mu(D),$$

de lo que concluimos que $(*)$ converge a 0. Como $\{A\times B : A, B\in\mathcal{B}\} = \mathcal{S}$ es semiálgebra que genera $\mathcal{B}$ deducimos que $T\times T$es débilmente mezclador.

(3) $\Rightarrow$ (1) Directo del lema.

(4) $\Leftrightarrow$ (1) usando técnica de $f$ simple, $f\geq 0, ...,$ etc.

(1) $\Rightarrow$ (5) Sea $f\in L^2(X,\mathcal{B},\mu), f\neq 0, \alpha\in\mathbb{C}$ tal que $f\circ T = \alpha f$ y definamos $g(x,y) = f(x)\overline{f(y)}$. Entonces $g(Tx,Ty) = \alpha f(x)\overline{\alpha}f(\overline{y}) = |\alpha|^2 g(x,y)$.

Luego como $|\alpha| = 1$,$g \circ T \times T = g$. Como $T \times T$ es ergódico entonces $g = cte$ y $f = cte\,\mu - cs$. Concluimos que $\alpha = 1$.

$(5) \Rightarrow (1)$ Usar el teorema de Herglotz. ♠

**Teorema de Herglotz.** *Sea $U : H \to H$ una isometría del espacio de Hilbert $H$ y $x \in H$. Entonces, existe una única medida finita $\mu_x$ en $\mathbb{K} = \{z \in \mathbb{C} : |z| = 1\}$ tal que $\forall n \geq 1, \langle U^n x, x\rangle = \displaystyle\int_X z^n d\mu_x$.* ♠

**Ejemplos:**

1) Fullshift con medida uniforme es fuerte mezclador.

2) En $\{1, ..., n\}^{\mathbb{Z}}$ se pone una medida Markoviana $\mu$ dada por la matriz estocástica $A = M_{n\times n}([0,1])$ y el vector estacionario $\pi$. Además $A$ se toma irreducible. Finalmente,

$$\mu\left([\omega_0...\omega_t]_s\right) = \pi_{\omega_0} A_{\omega_0\omega_1}...A_{\omega_{t-1}\omega_t}.$$

Esta medida es shift invariante en la semiálgebra de los cilindros, luego shift invariante en la $\sigma-$ álgebra producto.

# Capítulo 4

# Información y Entropía Métrica

## 4.1. Información.

Veremos en primer lugar la información en un cuadro no dinámico. Sea $(X, \mathcal{B}, \mu)$ un espacio de probabilidad y $\alpha = (A_i : i \in I)$ una partición medible de $X$ (finita o numerable). Es decir, $\forall i \in I, A_i \in \mathcal{B}$, $\forall i, j \in I, i \neq j, \mu(A_i \cap A_j) = 0$, $\mu(\bigcup\limits_{i \in I} A_i) = 1$. A menudo supondremos $\forall i \in I, \mu(A_i) > 0$. Por comodidad usaremos una notación más compacta, $\alpha = (A)$, es decir, los átomos de $\alpha$ se denotan $A \in \alpha$. A menos que digamos lo contrario cada vez que nos refiramos a una partición ella será medible y numerable (o finita). Dadas dos particiones $\alpha, \beta$ decimos que $\alpha$ es más fina que $\beta$ y anotamos $\alpha \succeq \beta$ si dado un átomo $A \in \alpha$ existe otro átomo $B \in \beta$ tal que $A \subseteq B$. También decimos que $\beta$ es más gruesa que $\alpha$.

La función de información es una función

$$I_\alpha : X \to \mathbb{R}_+ \cup \{+\infty\}$$

verificando algunos axiomas básicos:

- $I_\alpha(x) \geq 0$, en cada $x \in X$,
- $I_\alpha(x) = \sum\limits_{A \in \alpha} 1_A(x) \cdot \vartheta(\mu(A))$ para alguna función decreciente

$$\vartheta : [0, 1] \to \mathbb{R}_+ \cup \{\infty\},$$

- si $\alpha$ y $\beta$ son particiones independientes, es decir, para $A \in \alpha$ y $B \in \beta$, $\mu(A \cap B) = \mu(A) \cdot \mu(B)$, entonces en cada $x \in X$ se tieme

  $$I_{\alpha \vee \beta}(x) = I_{\alpha}(x) + I_{\beta}(x),$$

  donde $\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$ es la partición refinada de $\alpha$ y $\beta$.

Analicemos estas condiciones. Para $x \in A \cap B$ se tiene

$$I_{\alpha}(x) = \vartheta(\mu(A)), I_{\beta}(x) = \vartheta(\mu(B)), I_{\alpha \vee \beta}(x) = \vartheta(\mu(A \cap B)).$$

Luego si $\alpha$ y $\beta$ son independientes se tiene que $I_{\alpha \vee \beta}(x) = \vartheta(\mu(A) \cdot \mu(B))$. Por lo tanto $\vartheta(\mu(A) \cdot \mu(B)) = \vartheta(\mu(A)) + \vartheta(\mu(B))$. Si además de las condiciones anteriores pedimos que la función $\vartheta$ sea continua obtenemos que $\vartheta(s) = -log(s)$, salvo multiplicación por constante positiva.

**Definición.** *Consideremos una partición medible numerable $\alpha = (A)$.*

- *La función de información asociada a $\alpha$ se define por*

  $$I_{\alpha}(x) = -\sum_{A \in \alpha} 1_A(x) \ \ \log(\mu(A)).$$

  *En general se escribe $I_{\alpha} = -\sum_{A \in \alpha} 1_A \cdot \log(\mu(A))$.*

- *La entropía asociada a $\alpha$ es la información media*

$$H(\alpha) = \mathbb{E}(I_{\alpha}) = \int_X I_{\alpha} d\mu = -\sum_{A \in \alpha} \mu(A) \cdot \log(\mu(A)).$$

Para denotar la dependencia de $\mu$, se anota $H_{\mu}(\alpha)$ (en vez de $H(\alpha)$). Notaremos $\mathcal{N}$ a la partición trivial

$$\mathcal{N} =_{\mu} \{X, \emptyset\} = \{A \in \mathcal{B} : \mu(A) = 1 \vee \mu(A) = 0\}.$$

Se tiene $I_{\mathcal{N}} = 0,\ H_{\mu}(\mathcal{N}) = 0$.

En lo que sigue definiremos la información y entropía condicional. En cuanto a notación identificaremos una partición $\alpha$ con la $\sigma-$ álgebra engendrada por ella:

$$\sigma(\alpha) = \{\bigcup_{i \in J} A_i : J \subseteq I\}.$$

Lo anterior está bien definido pues $I$ es numerable.

**Definición.** *Sean $\alpha = (A), \beta = (B)$ dos particiones medibles y numerables. La información condicional de $\alpha$ dado $\beta$ es*

$$I_{\alpha|\beta} = -\sum_{B \in \beta} 1_B \cdot \left( \sum_{A \in \alpha} 1_A \cdot \log(\mu\left(A \mid B\right)) \right),$$

*donde $\mu\left(A \mid B\right) = \frac{\mu(A \cap B)}{\mu(B)}$. Hemos asumido que $\mu(B) > 0$ en cada $B \in \beta$. Otra manera de ver la misma expresión es*

$$I_{\alpha|\beta} = -\sum_{B \in \beta} \sum_{A \in \alpha} 1_{A \cap B} \cdot \log(\mu\left(A \mid B\right)).$$

*La entropía condicional es la media de esta cantidad,*

$$H\left(\alpha \mid \beta\right) = \int_X I_{\alpha|\beta} d\mu = -\sum_{B \in \beta} \sum_{A \in \alpha} \mu\left(A \cap B\right) \cdot \log(\mu\left(A \mid B\right)).$$

**Observaciones:**

1. Si $x \in A \cap B$, entonces

$$\begin{aligned} I_{\alpha|\beta}(x) &= -\log(\mu(A|B)) \\ &= -\log(\mu(A \cap B)) - (-\log(\mu(B))) \\ &= I_{\alpha \vee \beta}(x) - I_\beta(x) \end{aligned}$$

luego $I_{\alpha \vee \beta} = I_{\alpha|\beta} + I_\beta$ y deducimos que $H\left(\alpha \vee \beta\right) = H\left(\alpha \mid \beta\right) + H\left(\beta\right)$.

2. Si $\alpha$ y $\beta$ son $\mu$- independientes, se tiene $\mu\left(A \mid B\right) = \mu\left(A\right)$ para cada $A \in \alpha, B \in \beta$, luego

$I_{\alpha|\beta} = I_\alpha, H\left(\alpha \mid \beta\right) = H\left(\alpha\right)$ y $H\left(\alpha \vee \beta\right) = H\left(\alpha\right) + H\left(\beta\right)$.

3. Si $\beta = \mathcal{N}$ es la partición trivial entonces $I_{\alpha|\beta} = I_\alpha, H\left(\alpha \mid \beta\right) = H\left(\alpha\right)$, pues $\mathcal{N}$ es $\mu$- independiente de toda partición $\alpha$.

4. Observemos que se tiene $\mathbb{E}(1_A \mid \beta) = \sum_{B\in\beta} 1_B \cdot \mu(A \mid B)$, luego la información condicional puede ser escrita como $I_{\alpha|\beta} = -\sum_{B\in\beta}\sum_{A\in\alpha} 1_{A\cap B}\cdot\log(\mathbb{E}(1_A \mid \beta))$. Esto permite dar posteriormente una definición más general.

5. Si $\alpha$ es $\sigma(\beta)-$ medible, o equivalentemente si $\alpha\vee\beta = \beta$ se tiene $\mathbb{E}(1_A \mid \beta) = 1_A$. Como $\log(1) = 0$ se concluye que $I_{\alpha|\beta} = 0, H(\alpha \mid \beta) = 0$.

6. Mostremos la recíproca de la propiedad anterior. Supongamos $I_{\alpha|\beta} = 0$ $\mu$ -casi seguramente o de manera equivalente $H(\alpha \mid \beta) = 0$. Probemos que $\alpha$ es $\sigma(\beta)$- medible, $\mu$ -casi seguramente. Si $I_{\alpha|\beta} = 0$, $\mu$ -c.s., entonces en cada $A \in \alpha$ y $B \in \beta$, $1_{A\cap B}\mathbb{E}(1_A \mid \beta) = 1_{A\cap B}$, $\mu$ -c.s. Luego, sumando en $B \in \beta$, resulta

$$1_A\mathbb{E}(1_A \mid \beta) = 1_A \ (*).$$

Luego

$$\begin{aligned}
\int_X \mathbb{E}(1_A \mid \beta)\, d\mu &= \int_X 1_A\mathbb{E}(1_A \mid \beta)\, d\mu + \int_X 1_{A^c}\mathbb{E}(1_A \mid \beta)\, d\mu \\
\int_X 1_A d\mu &= \int_X 1_A d\mu + \int_X 1_{A^c}\mathbb{E}(1_A \mid \beta)\, d\mu \ (**).
\end{aligned}$$

Por (**) concluímos que $\mathbb{E}(1_A \mid \beta) = 1_A\mathbb{E}(1_A \mid \beta)$ y por (*) $\mathbb{E}(1_A \mid \beta) = 1_A$, $\mu$ -c.s. Concluimos que $A$ es $\sigma(\beta)-$ medible $\mu-$ c.s.

**Propiedades Básicas.**

*Sean $\alpha, \beta, \gamma$ particiones numerables medibles. Entonces*

*1. $I_{\alpha\vee\beta|\gamma} = I_{\alpha|\gamma} + I_{\beta|\alpha\vee\gamma}$.*

*2. $H(\alpha\vee\beta \mid \gamma) = H(\alpha \mid \gamma) + H(\beta \mid \alpha\vee\gamma)$.*

*3. Si $\gamma \succeq \beta$ ($\gamma\vee\beta = \gamma$), es decir $\gamma$ es más fina que $\beta$, entonces $I_{\alpha\vee\beta|\gamma} = I_{\alpha|\gamma}$ y $H(\alpha\vee\beta|\gamma) = H(\alpha|\gamma)$.*

*4. Si $\alpha \succeq \beta$ entonces $I_{\alpha|\gamma} \geq I_{\beta|\gamma}$ y $H(\alpha|\gamma) \geq H(\beta|\gamma)$.*

*5. Si $\beta \succeq \alpha$ entonces $H(\gamma \mid \alpha) \geq H(\gamma \mid \beta)$. En general no se tiene $I_{\gamma|\alpha} \geq I_{\gamma|\beta}$.*

**Demostración:** Sean $\alpha = (A), \beta = (B), \gamma = (C)$ las particiones medibles y numerables del enunciado.

1. Sea $x \in A \cap B \cap C$ donde $A \in \alpha, B \in \beta, C \in \gamma$, entonces

$$I_{\alpha \vee \beta | \gamma}(x) = -\log(\mu(A \cap B | C)) = -\log(\frac{\mu(A \cap B \cap C)}{\mu(C)}),$$

$$I_{\alpha | \gamma}(x) = -\log(\mu(A \mid C)) = -\log(\frac{\mu(A \cap C)}{\mu(C)}),$$

y

$$I_{\beta | \alpha \vee \gamma}(x) = -\log(\mu(B | A \cap C)) = -\log(\frac{\mu(B \cap A \cap C)}{\mu(A \cap C)}).$$

Luego $I_{\alpha|\gamma}(x) + I_{\beta|\alpha\vee\gamma}(x) = -\log \frac{\mu(B\cap A\cap C)}{\mu(C)} = I_{\alpha\vee\beta|\gamma}(x)$.

2. Es directa de la propiedad 1.

3. Si $\gamma \succeq \beta$ entonces $\alpha \vee \gamma \succeq \beta$. De lo que deducimos que $I_{\beta|\alpha\vee\gamma} = 0$. Para concluir basta aplicar la propiedad 1.

4. Si $\alpha \succeq \beta$ entonces por la propiedad 1 tenemos que

$$\underbrace{I_{\alpha\vee\beta|\gamma}}_{=I_{\alpha|\gamma}} = I_{\beta|\gamma} + \underbrace{I_{\alpha|\beta\vee\gamma}}_{\geq 0}$$

de donde concluimos que $I_{\alpha|\gamma} \geq I_{\beta|\gamma}$.

5. La propiedad 5 la probaremos en un contexto más general. ♠

**Definición.** *Sean $(X, \mathcal{B}, \mu)$ un espacio de probabilidad, $\alpha$ una partición, numerable y medible, y $\mathcal{A} \subseteq \mathcal{B}$ una sub $\sigma-$ álgebra de $\mathcal{B}$.*

- *La función de información de $\alpha$ dado $\mathcal{A}$ se define como*

$$I_{\alpha|\mathcal{A}} = -\sum_{A\in\alpha} 1_A \cdot \log(\mathbb{E}(1_A|\mathcal{A})).$$

- *La entropía condicional de $\alpha$ dado $\mathcal{A}$ es $H(\alpha \mid \mathcal{A}) = \int\limits_X I_{\alpha|\mathcal{A}} d\mu$.*

Cuando sea necesario especificar la medida $\mu$ anotamos $I_\mu(\alpha|\mathcal{A})$ y $H_\mu(\alpha|\mathcal{A})$ respectivamente. La definición anterior generaliza la definición con particiones pues $I_{\alpha|\beta} = I_{\alpha|\sigma(\beta)}$.

**Proposición.** *Si* $\mathcal{A}_1 \supseteq \mathcal{A}_2$ *entonces* $H(\alpha \mid \mathcal{A}_1) \leq H(\alpha \mid \mathcal{A}_2)$.

La proposición anterior queda incluída en la proposición precedente pues

$$\alpha \succeq \beta \Rightarrow \sigma(\alpha) \supseteq \sigma(\beta).$$

En la demostración usaremos la Desigualdad de Jensen.

**Desigualdad de Jensen .** *Sea* $\varphi : \mathbb{R} \to \mathbb{R}$ *una función cóncava. Entonces, para* $f \in L^1(X, \mathcal{B}, \mu), \mathcal{A} \subseteq \mathcal{B}$ *sub* $\sigma$ *-álgebra,* $\varphi(f) \in L^1(X, \mathcal{B}, \mu)$, *se tiene*

$$\varphi(\mathbb{E}(f \mid \mathcal{A})) \geq \mathbb{E}(\varphi(f) \mid \mathcal{A}). \qquad (*)$$

**Demostración:** Propuesta. ♠

**Demostración proposición anterior:** Observemos que $\varphi : [0,1] \to \mathbb{R}_+, \varphi(t) = -t \cdot \log(t),$ es cóncava (se define $\varphi(0) = -0\log(0) = 0$ por continuidad). Entonces,

$$\begin{aligned} H(\alpha \mid \mathcal{A}) &= -\sum_{A\in\alpha} \int_A \log(\mathbb{E}(1_A \mid \mathcal{A}))d\mu \\ &= -\sum_{A\in\alpha} \int_X \mathbb{E}(1_A \log(\mathbb{E}(1_A \mid \mathcal{A})) \mid \mathcal{A})\, d\mu \\ &= -\sum_{A\in\alpha} \int_X \log(\mathbb{E}(1_A \mid \mathcal{A}))\mathbb{E}(1_A \mid \mathcal{A})\, d\mu \\ &= \sum_{A\in\alpha} \int_A \varphi(\mathbb{E}(1_A \mid \mathcal{A}))\, d\mu. (**) \end{aligned}$$

Fijemos $A \in \alpha$ y fijemos $f = \mathbb{E}(1_A \mid \mathcal{A}_1)$ en $(*), \mathcal{A} = \mathcal{A}_2$ en $(*)$. Por desigualdad de Jensen queda

$$\varphi(\mathbb{E}(\mathbb{E}(1_A \mid \mathcal{A}_1) \mid \mathcal{A}_2)) \geq \mathbb{E}(\varphi(\mathbb{E}(1_A \mid \mathcal{A}_1)) \mid \mathcal{A}_2)$$

es decir

$$\varphi(\mathbb{E}(1_A \mid \mathcal{A}_2)) \geq \mathbb{E}(\varphi(\mathbb{E}(1_A \mid \mathcal{A}_1)) \mid \mathcal{A}_2).$$

Integrando queda $\int\limits_X \varphi(\mathbb{E}(1_A \mid \mathcal{A}_2))\, d_\mu \geq \int\limits_X \varphi(\mathbb{E}(1_A \mid \mathcal{A}_1))\, d\mu$. Sumando sobre $A \in \alpha$ y utilizando (**) obtenemos $H(\alpha \mid \mathcal{A}_2) \geq H(\alpha \mid \mathcal{A}_1)$. ♠

## 4.2. Resultados Básicos de Martingalas.

Antes de continuar con teoría de información, daremos unos resultados básicos de martingalas..

**Lema Maximal.** *Sea* $(X, \mathcal{B}, \mu)$ *un espacio de probabilidad. Sean* $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \ldots \subseteq \mathcal{B}_N \subseteq \mathcal{B}$ *una secuencia creciente de sub* $\sigma-$ *álgebras de* $\mathcal{B}$*. Entonces* $\forall f \in L^1(X, \mathcal{B}, \mu)$ *se tiene:*

$$\forall \lambda > 0, \mu(E_N) \leq \frac{1}{\lambda} \int\limits_X |f| \, d\mu,$$

*donde* $E_N = \{x \in X : \text{máx}_{1 \leq n \leq N} \mathbb{E}(f \mid \mathcal{B}_n)(x) > \lambda\}$.

**Demostración:** Podemos asumir $f \geq 0$ (si no cambiamos $f$ por $|f|$). Se tiene que $E_N = \bigcup\limits_{n=1}^{N} E_n^{(N)}$ donde

$$E_n^{(N)} = \{x \in X : (\forall k \in \{1, ..., n-1\}, \mathbb{E}(f|\mathcal{B}_k)(x) \leq \lambda) \wedge \mathbb{E}(f \mid \mathcal{B}_n)(x) > \lambda\}.$$

Luego $\mu(E_N) = \sum\limits_{n=1}^{N} \mu\left(E_n^{(N)}\right)$ y $\int\limits_{E_N} f d_\mu = \sum\limits_{n=1}^{N} \int\limits_{E_n^{(N)}} f d_\mu$. Como $E_n^{(N)} \in \mathcal{B}_n$, se tiene que

$$\begin{aligned} \int\limits_{E_N} f d_\mu &= \sum_{n=1}^{N} \int\limits_{E_n^{(N)}} f d_\mu \\ &= \sum_{n=1}^{N} \int\limits_{E_n^{(N)}} \mathbb{E}(f \mid \mathcal{B}_n) \, d_\mu \\ &\geq \sum_{n=1}^{N} \lambda \cdot \mu\left(E_n^{(N)}\right) \\ &= \lambda \cdot \mu(E_N) \end{aligned}$$

Luego $\mu(E_N) \leq \frac{1}{\lambda} \int\limits_{E_N} f d\mu \leq \frac{1}{\lambda} \int\limits_X f d_\mu$. ♠

**Corolario.** *Sea* $(\mathcal{B}_n)_{n \in \mathbb{N}}$ *una suseción creciente de sub sigma álgebras de* $\beta$.

*Entonces* $\forall f \in L^1(\mu), \forall \lambda > 0,$

$$\mu\left\{x \in X : \sup_{n\in\mathbb{N}} \mathbb{E}(f \mid \mathcal{B}_n) > \lambda\right\} \le \frac{1}{\lambda}\int_X |f|\, d\mu$$

**Demostración:** Ejercicio. ♠

**Teorema Creciente de Martingalas.** *Sea* $(X, \mathcal{B}, \mu)$ *un s.d.a. Sea* $(\beta_n)_{n\in\mathbb{N}}$ *una sucesión creciente de sub sigma álgebras de* $\mathcal{B}$ *y* $\mathcal{B}_\infty$ *el límite creciente. Entonces* $\forall f \in L^1(X, \mathcal{B}, \mu)$ *se tiene,*

$$\lim_{n\to\infty} \mathbb{E}(f \mid \mathcal{B}_n) = \mathbb{E}(f \mid \mathcal{B}_\infty) \text{ en } L^1(X, \mathcal{B}, \mu) \text{ y } \mu - c.s.$$

**Demostración:** a) Probemos primero la convergencia en $L^1(X, \mathcal{B}, \mu)$. Se tiene que el resultado es cierto para $f \in \bigcup_{n\in\mathbb{N}} L^1(X, \mathcal{B}_n, \mu)$. En efecto, si $f \in L^1(X, \mathcal{B}_n, \mu)$, $\forall n \le n' \le \infty, \mathbb{E}(f \mid \mathcal{B}_{n'}) = f,$ pues $f = \mathbb{E}(f \mid \mathcal{B}_n)$.

Por otro lado, $\cup_{n\in\mathbb{N}} L^1(X, \mathcal{B}_n, \mu)$ es denso en $L^1(X, \mathcal{B}_\infty, \mu)$

$$(\ \mathcal{B}_\infty = \{B \in \mathcal{B} : \forall \epsilon > 0, \exists n \in \mathbb{N}, \exists B_n \in \mathcal{B}_n \text{ tal que } \mu(B \Delta B_n) < \epsilon\}).$$

Ahora tomemos $f \in L^1(X, \mathcal{B}_\infty, \mu)$, y para $\epsilon > 0$ tomemos algún $n_0 \in \mathbb{N}$, $g \in L^1(X, \mathcal{B}_{n_0}, \mu)$ tal que $\|f - g\|_1 < \epsilon$. Para $n \ge n_0$ se tiene,

$$\begin{aligned} ||\mathbb{E}(f|\mathcal{B}_n) - f||_1 &= ||\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(g|\mathcal{B}_n) + \mathbb{E}(g|\mathcal{B}_n) - g + g - f||_1 \\ &\le ||\mathbb{E}(f - g|\mathcal{B}_n)||_1 + ||g - f||_1 \le 2\epsilon \end{aligned}$$

pues $\mathbb{E}(g \mid \mathcal{B}_n) = g$ si $n \ge n_0$. Luego se tiene el resultado para $f \in L^1(X, \mathcal{B}, \mu)$ pues $\mathbb{E}(f \mid \mathcal{B}_n) = \mathbb{E}(\mathbb{E}(f \mid \mathcal{B}_\infty) \mid \mathcal{B}_n)$, y podemos suponer que $f$ es $\mathcal{B}_\infty -$ medible.

b) Probemos la convergencia $\mu - c.s.$ Como dijimos antes, basta suponer $f \in L^1(\mathcal{B}_\infty, \mu)$.

Sea $\epsilon > 0$,

$$
\begin{aligned}
\mu\left\{\limsup_{n\to\infty} \mid \mathbb{E}(f|\mathcal{B}_n) - f \mid > \sqrt{\epsilon}\right\} &= \mu\{\limsup_{n\to\infty} |\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(g|\mathcal{B}_n) \\
&+\mathbb{E}(g|\mathcal{B}_n) - g + g - f| > \sqrt{\epsilon}\} \\
&\leq \mu\left\{\limsup_{n\to\infty} |\mathbb{E}(f-g \mid \mathcal{B}_n)| > \tfrac{\sqrt{\epsilon}}{2}\right\} \\
&+\mu\left\{|g-f| > \tfrac{\sqrt{\epsilon}}{2}\right\} \\
&\left(\begin{array}{l}\text{pues } \{|f+h|>\epsilon\}\subseteq\left\{|f|>\frac{\epsilon}{2}\right\}\cup\left\{|h|>\frac{\epsilon}{2}\right\} \\ \text{y } \mathbb{E}(g|\mathcal{B}_n)=g \text{ para } n\geq n_0\end{array}\right) \\
&\leq \mu\left\{\sup_{n\in\mathbb{N}} \mid \mathbb{E}(f-g \mid \mathcal{B}_n) \mid > \tfrac{1}{2}\sqrt{\epsilon}\right\} \\
&+\mu\left\{|f-g| > \tfrac{1}{2}\sqrt{\epsilon}\right\} \\
&\leq \tfrac{2}{\sqrt{\epsilon}} \int_X |f-g|\, d\mu + \tfrac{2}{\sqrt{\epsilon}} \int_X |f-g|\, d\mu \leq 4\sqrt{\epsilon},
\end{aligned}
$$

donde usamos el Lema maximal y la desigualdad $\mu\{|f| > \lambda\} \leq \dfrac{1}{\lambda}\displaystyle\int_X |f|\, d\mu.$

El resultado se concluye directamente de la desigualdad anterior. ♠

**Teorema Decreciente de Martingalas.** *Sea* $(X, \mathcal{B}, \mu)$ *un s.d.a. Sea* $(\mathcal{B}_n)_{n\in\mathbb{N}}$ *una suseción decreciente de sub sigma álgebras de* $\mathcal{B}$ *y* $\mathcal{B}_\infty$ *el límite decreciente. Entonces* $\forall f \in L^1(X, \mathcal{B}, \mu)$ *se tiene,*

$$
\lim_{n\to\infty} \mathbb{E}(f \mid \mathcal{B}_n) = \mathbb{E}(f \mid \mathcal{B}_\infty) \;\; \textit{en } L^1(X, \mathcal{B}, \mu) \;\; \textit{y } \mu - \textit{c.s.}
$$

**Demostración:** Sea $V_n = Ker\mathbb{E}(\cdot \mid \mathcal{B}_n) = \{f \in L^1(\mu) : \mathbb{E}(f \mid \mathcal{B}_n) = 0\} = \{f - \mathbb{E}(f \mid \mathcal{B}_n) : f \in L^1(X, \mathcal{B}, \mu)\}$. Observemos que $V_n \subseteq V_{n+1}$, pues $\mathbb{E}(f \mid \mathcal{B}_n) = 0$ implica $\mathbb{E}(f \mid \mathcal{B}_{n+1}) = \mathbb{E}(\mathbb{E}(f \mid \mathcal{B}_n) \mid \mathcal{B}_{n+1}) = 0$.

El resultado del teorema se verifica para $f \in L(X, \mathcal{B}_\infty, \mu) + (\cup_{n\in\mathbb{N}} V_n)$. En efecto, si $f = f_1 + f_2, f_1 \in L^1(X, \mathcal{B}_\infty, \mu), f_2 \in V_{n_0}$, entonces para $n \geq n_0$, $\mathbb{E}(f \mid \mathcal{B}_n) = \mathbb{E}(f_1 \mid \mathcal{B}_n) + \mathbb{E}(f_2 \mid \mathcal{B}_n) = \mathbb{E}(f \mid \mathcal{B}_\infty)$ y $\mathbb{E}(f \mid \mathcal{B}_\infty) = \mathbb{E}(f_1 \mid \mathcal{B}_\infty) + \mathbb{E}(f_2 \mid \mathcal{B}_\infty)$. Como $\mathbb{E}(\mathbb{E}(f_2 \mid \mathcal{B}_n) \mid \mathcal{B}_\infty) = 0$ concluimos que $\mathbb{E}(f \mid \mathcal{B}_n) = \mathbb{E}(f \mid \mathcal{B}_\infty), \forall n \geq n_0$ (en $L^1(X, \mathcal{B}, \mu)$ y $\mu - c.s.$ ).

Probemos la convergencia en $L^1(X, \mathcal{B}, \mu)$. Mostremos primero que $\bigcup\limits_{n\in\mathbb{N}} V_n$ es denso en $V_\infty = Ker(\mathbb{E}(\cdot \mid \mathcal{B}_\infty)) = \{f \in L^1(X, \mathcal{B}, \mu) : \mathbb{E}(f \mid \mathcal{B}_\infty) = 0\}$. Es claro que $\bigcup\limits_{n\in\mathbb{N}} V_n \subseteq \mathcal{B}_\infty$ (pues $\mathcal{B}_n \supseteq \mathcal{B}_\infty$), $V_\infty$ es un *s.e.v.* cerrado de $L^1(X, \mathcal{B}, \mu)$ y $\bigcup_{n\in\mathbb{N}} V_n$ es s.e.v. de $L^1(X, \mathcal{B}, \mu)$. Por Hahn-Banach, basta mostrar que

$\forall \varphi \in \left(L^1(X, \mathcal{B}, \mu)\right)^*, \varphi(f) = 0, \ \forall f \in \bigcup_n V_n$, implica $\varphi(f) = 0, \ \forall f \in V_\infty$. Recordemos que $\varphi \in (L^1(X, \mathcal{B}, \mu))^*$ es inducido por $h \in L^\infty(X, \mathcal{B}, \mu)$. Luego nos queda demostrar que $\int_X f \cdot h d\mu = 0, \forall f \in \bigcup_{n \in \mathbb{N}} V_n$ $\quad (*)$ implica $\int_X f \cdot h d\mu = 0, \forall f \in V_\infty \, (**)$.

La condición $(*)$ se puede escribir $\int_X (f - \mathbb{E}(f \mid \mathcal{B}_n)) \cdot h d\mu = 0, \forall f \in L^1(X, \mathcal{B}, \mu)$, $\forall n \in \mathbb{N}$. En particular, es cierto para $f = h$ que $\int_X (h - \mathbb{E}(h \mid \mathcal{B}_n)) \cdot h d\mu = 0, \forall n \in \mathbb{N}$. Pero siempre se tiene que $\int_X (h - \mathbb{E}(h \mid \mathcal{B}_n)) \mathbb{E}(h \mid \mathcal{B}_n) d\mu = 0$ pues $\mathbb{E}(h\mathbb{E}(h \mid \mathcal{B}_n) \mid \mathcal{B}_n) = \mathbb{E}(h \mid \mathcal{B}_n) \mathbb{E}(h \mid \mathcal{B}_n)$. Restando obtenemos

$$\int_X h^2 - 2\mathbb{E}(h \mid \mathcal{B}_n) h + \mathbb{E}(h \mid \mathcal{B}_n)^2 d\mu = \int_X (h - \mathbb{E}(h \mid \mathcal{B}_n))^2 d\mu = 0.$$

Es decir $h = \mathbb{E}(h \mid \mathcal{B}_n), \forall n \in \mathbb{N}$, luego $h \in L^1(X, \mathcal{B}_n, \mu), \forall n \in \mathbb{N}$, de donde $h \in L^1(X, \mathcal{B}, \mu)$. Luego $\int_X (f - \mathbb{E}(f \mid \mathcal{B}_\infty)) \cdot h d\mu = 0, \forall f \in V_\infty$, pues $h = \mathbb{E}(h \mid \mathcal{B}_\infty)$ y

$$\int_X f \cdot h d\mu = \int_X \mathbb{E}(f \cdot h \mid \mathcal{B}_\infty) = \int_X h\mathbb{E}(f \mid \mathcal{B}_\infty) d\mu.$$

Concluimos que la implicancia $(**)$ se verifica y $\bigcup_{n \in \mathbb{N}} V_n$ es denso en $V_\infty$.

Observemos que $L^1(X, \mathcal{B}, \mu) = L^1(X, \mathcal{B}_\infty, \mu) + V_\infty$ pues $f = \mathbb{E}(f \mid \mathcal{B}_\infty) + (f - \mathbb{E}(f \mid \mathcal{B}_\infty))$. Como $\bigcup_{n \in \mathbb{N}} V_n$ es denso en $V_\infty$ se tiene que para $f \in L^1(X, \mathcal{B}, \mu)$, existen $n_0 \in \mathbb{N}, g \in L^1(X, \mathcal{B}_\infty, \mu) + V_{n_0}$ tal que $\|f - g\|_1 < \epsilon$. Luego para $n \geq n_0$ se tiene

$$\begin{aligned} \|\mathbb{E}(f \mid \mathcal{B}_n) - \mathbb{E}(f \mid \mathcal{B}_\infty)\| &\leq \|\mathbb{E}(f - g \mid \mathcal{B}_n)\|_1 + \|\mathbb{E}(g|\mathcal{B}_n) - \mathbb{E}(g|\mathcal{B}_\infty)\|_1 + \\ &\|\mathbb{E}(g - f \mid \mathcal{B}_\infty)\|_1 \leq 2\epsilon. \end{aligned}$$

Tomando limsup se concluye

$$\lim_{n\to\infty} \|\mathbb{E}(f \mid \mathcal{B}_n) - \mathbb{E}(f \mid \mathcal{B}_\infty)\| = 0.$$

Veamos la convergencia $\mu-$ c.s. Como $\mathcal{B}_1 \supseteq \mathcal{B}_2 \supseteq ... \supseteq \mathcal{B}_N$, se tiene $\mu \left\{ x \in X : \max_{1\leq n \leq N} \; \mathbb{E}(f \mid \mathcal{B}_n) \right.$ $\dfrac{1}{\lambda}\displaystyle\int_X |f| \, d\mu$, luego

$$\mu \left\{ x \in X : \sup_{1\leq n} \; \mathbb{E}(f \mid \mathcal{B}_n) > \lambda \right\} \leq \frac{1}{\lambda} \int_X |f| \, d\mu.$$

Con lo que la demostración de la convergencia $\mu - c.s.$ es enteramente similar al caso creciente usando la función $g$ que acabamos de construir. Dado que $\mathbb{E}(g \mid \mathcal{B}_n) = \mathbb{E}(g \mid \mathcal{B}_\infty)$, tenemos

$$\begin{aligned}
\mu \left\{ \overline{\lim_{n}} |\mathbb{E}(f \mid \mathcal{B}_n) - \mathbb{E}(f \mid \mathcal{B}_\infty)| > \sqrt{\epsilon} \right\} \quad & \leq \mu \left\{ \overline{\lim_{n}} \mid \mathbb{E}(f - g \mid \mathcal{B}_n) > \tfrac{1}{2}\sqrt{\epsilon} \right\} + \\
& \mu \left\{ \mathbb{E}(f - g \mid \mathcal{B}_\infty) > \tfrac{1}{2}\sqrt{\epsilon} \right\} \\
& \leq \mu \left\{ \sup_{n} \mid \mathbb{E}(f - g \mid \mathcal{B}_n) > \tfrac{1}{2}\sqrt{\epsilon} \right\} + \\
& \mu \left\{ \mathbb{E}(f - g \mid \mathcal{B}_\infty) > \tfrac{1}{2}\sqrt{\epsilon} \right\} \\
& \leq \tfrac{2}{\sqrt{\epsilon}} \int_X |f - g| \, d\mu + \tfrac{2}{\sqrt{\epsilon}} \int_X |f - g| \, d\mu \\
& \leq 4\sqrt{\epsilon}.
\end{aligned}$$

♠

**Lema de Chung.** *Sea $(X, \mathcal{B}, \mu)$ un espacio de probabilidad. Sea $\alpha$ una partición (numerable medible) de entropía $H_\mu(\alpha) < \infty$. Sea $(\mathcal{A}_n)_{n\in\mathbb{N}}$ una sucesión creciente de sub $\sigma-$ álgebras de $\mathcal{B}$. Entonces*

$$\int_X \sup_{n\in\mathbb{N}} I_{\alpha|\mathcal{A}_n} d\mu \leq H_\mu(\alpha) + 1.$$

*En particular* $\sup_{n\in\mathbb{N}} I_{\alpha|A_n} \in L^1(X, \mathcal{B}, \mu)$.

**Demostración:** Sea $f = \sup_n I_{\alpha|\mathcal{A}_n}$, $f : X \longrightarrow \mathbb{R}_+ \cup \{+\infty\}$. Definamos $F(t) = \mu\{x \in X : f(x) > t\}$, $t \geq 0$. Se tiene

$$\begin{aligned}\int_X f d\mu &= -\int_{0^-}^{\infty} t dF(t) \\ &= -tF(t)\,|_{0^-}^{\infty} + \int_{0^-}^{\infty} F(t)\,dt \\ &\leq \int_{0^-}^{\infty} F(t)\,dt\end{aligned}$$

Ahora bien $f =_n^{\sup} -\sum_{A\in\alpha} 1_A \ \log \ \overbrace{\mu(A \mid \mathcal{A}_n)}^{\mathbb{E}_\mu(1_A|\mathcal{A}_n)} = \sum_{A\in\alpha} 1_A {}^{\sup}_n(- \ \log \mu(A \mid \mathcal{A}_n))$.
Luego

$$\begin{aligned}F(t) &= \sum_{A\in\alpha} \mu(A \cap \{x \in X :_n^{\sup} (- \ \log \ \mu(A \mid \mathcal{A}_n)) > t\}) \\ &= \sum_{A\in\alpha} \mu\left(A \cap \left\{x \in X :_n^{\inf} \mu(A \mid \mathcal{A}_n) > e^{-t}\right\}\right).\end{aligned}$$

Ahora bien $\mu\left(A \cap \left\{x \in X :_n^{\inf} \mu(A \mid \mathcal{A}_n) < e^{-t}\right\}\right) = \sum_{n\in\mathbb{N}} \mu(A \cap B_n)$, donde $B_n = \{x \in X : \mu(A|\mathcal{A}_k) \geq e^{-t}, \forall k < n, \mu(A \mid \mathcal{A}_n) < e^{-t}\}$. Luego

$$\begin{aligned}\mu(A \cap B_n) &= \int_{B_n} 1_A d\mu = \int_{B_n} \mathbb{E}(1_A|\mathcal{A}_n) d\mu \\ &\leq \int_{B_n} e^{-t} d\mu \leq e^{-t} \cdot \mu(B_n)\end{aligned}$$

y

$$\sum_{n\in\mathbb{N}} \mu(A \cap B_n) \leq e^{-t} \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) \leq e^{-t}$$

de donde $\mu\left(A \cap \left\{_n^{\inf} \mu(A \mid \mathcal{A}_n) < e^{-t}\right\}\right) \leq \ \text{mín} \ \{\mu(A), e^{-t}\}$. Concluimos que $F(t) \leq \sum_{A\in\alpha} \ \text{mín} \ \left\{\mu(A), e^{-t}\right\}$ y

$$\begin{aligned}\int_{0^-}^{\infty} F(t)\,dt &\leq \sum_{A\in\alpha} \int_0^{\infty} \ \text{mín} \ \{\mu(A), e^{-t}\}\,dt \\ &= \sum_{A\in\alpha} \left( \int_0^{-\log \mu(A)} \mu(A)\,dt + \int_{- \ \log \ \mu(A)}^{\infty} e^{-t} dt \right) \\ &= \sum_{A\in\alpha} [-\mu(A) \log \ \mu(A) + \mu(A)] \\ &= H_\mu(\alpha) + 1.\end{aligned}$$

Concluimos $\int\limits_X \sup_n I_{\alpha|\mathcal{A}_n} d\mu = \int\limits_X f d_\mu \leq H_\mu(\alpha) + 1.$ ♠

**Corolario.** *Sea* $\alpha$ *una partición (numerable, medible) con* $H_\mu(\alpha) < \infty$*. Sea* $(\mathcal{A}_n)_{n\in\mathbb{N}}$ *una sucesión creciente de* $\sigma-$ *álgebras, con* $\mathcal{A}_\infty = \bigvee\limits_n \mathcal{A}_n$*. Entonces* $I_{\alpha|\mathcal{A}_n} \underset{n\to\infty}{\longrightarrow} I_{\alpha|\mathcal{A}_\infty}$ $\mu$*–c.s. y en* $L^1(X, \mathcal{B}, \mu)$*. Además* $H_\mu(\alpha \mid \mathcal{A}_n) \underset{n\to\infty}{\longrightarrow} H_\mu(\alpha \mid \mathcal{A}_\infty)$*.*

**Demostración:** Para $A \in \alpha$ fijo se tiene $\mu(A \mid \mathcal{A}_n) \underset{n\to\infty}{\longrightarrow} \mu(A \mid \mathcal{A}_\infty)$ $\mu-c.s.$ y en $L^1(X, \mathcal{B}, \mu)$ por teorema creciente de martingalas, donde $\mu(A \mid \mathcal{A}_n) = \mathbb{E}(1_A \mid \mathcal{A}_n)$ y $\mu(A \mid \mathcal{A}_\infty) = \mathbb{E}(1_A \mid \mathcal{A}_\infty)$. Luego

$$-\log\ \mu(A \mid \mathcal{A}_n) \underset{n\to\infty}{\longrightarrow} -\log\ \mu(A \mid \mathcal{A}_\infty)\ \mu - c.s.$$

De aquí se deduce

$$-\sum_{A\in\alpha} 1_A\ \log\ \mu(A \mid \mathcal{A}_n) \underset{n\to\infty}{\longrightarrow} -\sum_{A\in\alpha} 1_A\ \log\ \mu(A \mid \mathcal{A}_\infty)\ \mu - c.s.$$

así $I_{\alpha|\mathcal{A}_n} \underset{n\to\infty}{\longrightarrow} I_{\alpha|\mathcal{A}_\infty}$ $\mu.c.s.$ Como $I_{\alpha|\mathcal{A}_n} \leq f = \sup_k I_{\alpha|\mathcal{A}_k}$, por el Lema de Chung $f \in L^1(X, \mathcal{B}, \mu)$. El teorema de convergencia dominada nos permite concluir, $I_{\alpha|\mathcal{A}_n} \underset{n\to\infty}{\longrightarrow} I_{\alpha|\mathcal{A}_\infty}$ en $L^1(X, \mathcal{B}, \mu)$. Esto implica,

$$H_\mu(\alpha \mid \mathcal{A}_n) = \int\limits_X I_{\alpha|\mathcal{A}_n} d\mu \underset{n\to\infty}{\longrightarrow} \int\limits_X I_{\alpha|\mathcal{A}_\infty} d\mu = H_\mu(\alpha \mid \mathcal{A}_\infty).$$

Además, $(H_\mu(\alpha \mid \mathcal{A}_n))_{n\in\mathbb{N}}$ es decreciente pues $(\mathcal{A}_n)_{n\in\mathbb{N}}$ es creciente. ♠

## 4.3. Información y Entropía en Teoría Ergódica.

**Definición.** *Sea* $(X, \mathcal{B}, \mu, T)$ *un s.d.a. Sea* $\alpha$ *una partición numerable y medible tal que* $H_\mu(\alpha) < \infty$*. La entropía de* $\alpha$ *con respecto a* $T$ *se define como*

$$h_\mu(\alpha, T) = H_\mu\left(\alpha \mid \alpha^-\right),$$

*donde* $\alpha^- = \bigvee_{n\geq 1} T^{-n}\alpha$ *(o la* $\sigma-$ *álgebra engendrada por* $\alpha^-$*).*

Veamos algunas propiedades que explotaremos.

1. $\mathbb{E}(f \circ T \mid T^{-1}\mathcal{A}) = \mathbb{E}(f \mid \mathcal{A}) \circ T$ para $f \in L^1(X, \mathcal{B}, \mu)$, y $\mathcal{A}$ una sub $\sigma-$ álgebra. En efecto, para $A \in \mathcal{A}$,

$$\begin{aligned}
\int_{T^{-1}A} f \circ T d\mu &= \int_X f \circ T \cdot 1_A \circ T d\mu = \int_X (f \cdot 1_A) \circ T d\mu \\
&= \int_X f \cdot 1_A d\mu = \int_A f d\mu = \int_A \mathbb{E}(f \mid \mathcal{A})\, d\mu \\
&= \int_X \mathbb{E}(f \mid \mathcal{A}) \cdot 1_A d\mu = \int_X (\mathbb{E}(f \mid \mathcal{A}) \cdot 1_A) \circ T d\mu \\
&= \int_{T^{-1}A} \mathbb{E}(f \mid \mathcal{A}) \circ T d\mu
\end{aligned}$$

Por otra parte $(\mathbb{E}(f \mid \mathcal{A}) \circ T)^{-1}(C) = T^{-1}\left(\mathbb{E}(f \mid \mathcal{A})^{-1}(C)\right) \in T^{-1}\mathcal{A}$.

2. $I_{\alpha|\mathcal{A}} \circ T = I_{T^{-1}\alpha|T^{-1}\mathcal{A}}; H_\mu(\alpha \mid \mathcal{A}) = H_\mu(T^{-1}\alpha \mid T^{-1}\mathcal{A})$. En efecto, esto proviene de 1:

$$\begin{aligned}
-\sum_{A\in\alpha} 1_A \circ T \ \log\ (\mathbb{E}(1_A \mid \mathcal{A}) \circ T) &= -\sum_{A\in\alpha} 1_{T^{-1}A} \ \log\ \mathbb{E}(1_{T^{-1}A} \mid T^{-1}\mathcal{A}) \\
&= I_{T^{-1}\alpha|T^{-1}\mathcal{A}}
\end{aligned}$$

3. Teníamos $h_\mu(\alpha, T) = H_\mu(\alpha \mid \alpha^-)$. Recordemos que por corolario luego del Lema de Chung se tiene que

$$h_\mu(\alpha \mid T) = \lim_{m\to\infty} H_\mu\left(\alpha \mid \bigvee_{n=1}^{m} T^{-n}\alpha\right).$$

El siguiente teorema es clave en Teoría Ergódica.

**Teorema Ergódico de la Información (Shannon-Mc Millan-Breiman-Chung- (1960))**

*Sean* $(X, \mathcal{B}, \mu, T)$ *un s.d.a. y* $\alpha$ *una partición numerable y medible tal que* $H_\mu(\alpha) < \infty$. *Notemos* $f = I(\alpha \mid \alpha^-)$. *Se tiene:*

*1.* $\lim_{N\to\infty} \frac{1}{N} I\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right) = \mathbb{E}(f \mid \mathcal{I})\, \mu - c.s.$ *y en* $L^1(X, \mathcal{B}, \mu)$,

(*donde* $\mathcal{I} = \{B \in \mathcal{B} : \mu(B \Delta T^{-1}B) = 0\}$ *es la* $\sigma-$ *álgebra de conjuntos invariantes*).

*2.* $\lim_{N\to\infty} \frac{1}{N} H_\mu\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right) = h_\mu(\alpha, T)$.

*3. Si* $T$ *es ergódica se tiene* $\lim_{N\to\infty} \frac{1}{N} I\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right) = h_\mu(\alpha, T) = H_\mu(\alpha \mid \alpha^-)$.

**Demostración.** Basta probar 1), el resto resulta de integrar. Se tiene

$$\begin{aligned}
I(\bigvee_{n=0}^{N-1} T^{-n}\alpha) &= I(\alpha \bigvee \bigvee_{n=1}^{N-1} T^{-n}\alpha) \\
&= I(\alpha| \bigvee_{n=1}^{N-1} T^{-n}\alpha) + I(\bigvee_{n=1}^{N-1} T^{-n}\alpha) \\
&= I(\alpha| \bigvee_{n=1}^{N-1} T^{-n}\alpha) + I(T^{-1}(\bigvee_{n=0}^{N-2} T^{-n}\alpha)) \\
&= I(\alpha| \bigvee_{n=1}^{N-1} T^{-n}\alpha) + I(\bigvee_{n=0}^{N-2} T^{-n}\alpha) \circ T
\end{aligned}$$

Si llamamos $f_0 = I(\alpha)$ y $f_k = I\left(\alpha \mid \bigvee_{n=1}^{k} T^{-n}\alpha\right)$ para $k \geq 1$, entonces

$$I\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right) = f_{N-1} + f_{N-2} \circ T + f_{N-3} \circ T^2 + ... + f_0 \cdot T^{N-1}$$

Luego

$$\begin{aligned}
\left|\tfrac{1}{N} I\left(\bigvee_{n=0}^{N-1} T^{-n}\alpha\right) - \mathbb{E}(f \mid \mathcal{I})\right| &= \left|\tfrac{1}{N} \sum_{k=0}^{N-1} f_{N-1-k} \circ T^k - \mathbb{E}(f \mid \mathcal{I})\right| \\
&\leq \left|\tfrac{1}{N} \sum_{k=0}^{N-1} \left(f_{N-1-k} \circ T^k - f \circ T^k\right)\right| \\
&+ \left|\tfrac{1}{N} \sum_{k=0}^{N-1} f \circ T^k - \mathbb{E}(f \mid \mathcal{I})\right|
\end{aligned}$$

Por hipótesis $H_\mu(\alpha) < \infty$ y por propiedad anterior $H_\mu(\alpha \mid \mathcal{A}) < \infty$ para toda sub $\sigma-$ álgebra $\mathcal{A}$. En particular, $H_\mu(\alpha \mid \alpha^-) < \infty$, es decir, $f \in L^1(X, \mathcal{B}, \mu)$. Aplicando el teorema de Birkhoff obtenemos que

$$\left|\frac{1}{N} \sum_{k=0}^{N-1} f \circ T^k - \mathbb{E}(f \mid \mathcal{I})\right|$$

converge $\mu-$ c.s. y en $L^1(X,\mathcal{B},\mu)$.

Estudiemos el otro término. Definamos $g_k = |f_k - f|$. Así obtenemos

$$\left|\frac{1}{N}\sum_{k=0}^{N-1}(f_{N-1-k}-f)\circ T^k\right| \leq \frac{1}{N}\sum_{k=0}^{N-1} g_{N-1-k}\circ T^k \qquad (*)$$

Probaremos que el segundo término converge a $0, \mu - c.s.$ y en $L^1(X,\mathcal{B},\mu)$. Observemos que $f_k = I\left(\alpha \mid \bigvee_{n=1}^{k} T^{-n}\alpha\right) \underset{k\to\infty}{\longrightarrow} f = I(\alpha \mid \alpha^-) \quad \mu - c.s.$ y en $L^1(X,\mathcal{B},\mu)$ (por el colorario del Lema de Chung). Luego $g_k \underset{k\to\infty}{\longrightarrow} 0 \quad \mu - c.s.$ y en $L^1(X,\mathcal{B},\mu)$. Probemos la convergencia en $L^1(X,\mathcal{B},\mu)$ del segundo término en $(*)$, lo que (por positividad) equivale a mostrar que su integral se va a cero. Se tiene,

$$\frac{1}{N}\int_X -\sum_{k=0}^{N-1} g_{N-1-k}\circ T^k d\mu = \frac{1}{N}\sum_{k=0}^{N-1}\int_X g_{N-1-k}d\mu = \frac{1}{N}\sum_{k=0}^{N-1}\int_X g_k d\mu \underset{N\to\infty}{\longrightarrow} 0$$

pues $\int_X g_k d\mu \underset{k\to\infty}{\longrightarrow} 0.$

Probemos la convergencia $\mu - c.s.$ del segundo término en $(*)$. Sea

$$G_N = \sup_{k\geq N} g_k \leq \left(\sup_{k\geq 1} f_k\right) + f \in L^1(X,\mathcal{B},\mu).$$

Esto último, pues $f \in L^1(X,\mathcal{B},\mu)$ y por el lema de Chung, $\sup_{k\geq 1} f_k \in L^1(X,\mathcal{B},\mu)$. Dado que $g_k \underset{k\to\infty}{\longrightarrow} 0 \quad \mu - c.s.$, se tiene que $G_N \underset{N\to\infty}{\searrow} 0 \quad \mu - c.s.$

Por el teorema de convergencia dominada y dado que $G_N \in L^1(X,\mathcal{B},\mu), G_N$ es decreciente en $N$ se tiene que $G_N \underset{N\to\infty}{\longrightarrow} 0$ en $L^1(X,\mathcal{B},\mu)$.

Ahora, para $M$ fijo suficientemente grande,

$$\begin{aligned}
\tfrac{1}{N}\sum_{k=0}^{N-1} g_{N-1-k}\circ T^k &= \tfrac{1}{N}\sum_{k=0}^{N-M-1} g_{N-1-k}\circ T^k \\
&+\tfrac{1}{N}\sum_{k=N-M}^{N-1} g_{N-1-k}\circ T^k \\
&\leq \tfrac{1}{N}\sum_{k=0}^{N-M-1} G_M\circ T^k + \tfrac{1}{N}\sum_{k=N-M}^{N-1} G_0\circ T^k.
\end{aligned}$$

Por el Teorema de Birkhoff, se tiene que $\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-M-1}G_M\circ T^k=\mathbb{E}\left(G_M\mid\mathcal{I}\right)$ y por otro lado $\limsup_{N\to\infty}\frac{1}{N}\sum_{k=N-M}^{N-1}G_0\circ T^k=0$. Luego,

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}g_{N-1-k}\circ T^k\leq\mathbb{E}\left(G_M\mid\mathcal{I}\right)$$

para todo $M>0$. Por lo tanto,

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}g_{N-1-k}\circ T^k\leq\limsup_{N\to\infty}\mathbb{E}\left(G_M\mid\mathcal{I}\right)=0.$$

En efecto, como $G_M\underset{M\to\infty}{\searrow}0$ en $L^1(X,\mathcal{B},\mu)$ entonces $\mathbb{E}\left(G_M\mid\mathcal{I}\right)$ decrece $\mu-$c.s. a algún $g\geq 0$. Como $G_M\underset{M\to\infty}{\searrow}0$ en $L^1\left(X,\mathcal{B},\mu\right)$, se tiene $\mathbb{E}\left(G_M\mid\mathcal{I}\right)\underset{M\to\infty}{\rightarrow}0$ en $L^1\left(X,\mathcal{B},\mu\right)$. Se deduce que $\int_X gd\mu=0$, por lo que $g=0\,\mu-$ c.s. ♠

**Corolario.**

$$\Gamma_N\left(\alpha\right)=\left\{A\in\bigvee_{n=0}^{N-1}T^{-n}\alpha:e^{-N(h_\mu(\alpha,T)+\epsilon)}\leq\mu\left(A\right)\leq e^{-N(h_\mu(\alpha,T)-\epsilon)}\right\}$$

verifica $\mu\left(\bigcup_{A\in\Gamma_N(\alpha)}A\right)>1-\epsilon$.*Sea $\left(X,\mathcal{B},\mu,T\right)$ un s.d.a. ergódico y $\alpha$ una partición numerable y medible tal que $H_\mu(\alpha)<\infty$. Luego $\forall\epsilon>0$, $\exists N\left(\epsilon\right)$, tal que $\forall N\geq N\left(\epsilon\right)$, se tiene que la subfamilia de átomos de $\bigvee_{n=0}^{N-1}T^{-n}\alpha$*

$$\Gamma_N\left(\alpha\right)=\left\{A\in\bigvee_{n=0}^{N-1}T^{-n}\alpha:e^{-N(h_\mu(\alpha,T)+\epsilon)}\leq\mu\left(A\right)\leq e^{-N(h_\mu(\alpha,T)-\epsilon)}\right\}$$

*verifica* $\mu\left(\bigcup_{A\in\Gamma_N(\alpha)}A\right)>1-\epsilon$.

***Demostración:*** *Por ser* $T$ *ergódico, se tiene que* $\frac{1}{N}I\left(\bigvee_{n=0}^{N-1}T^{-n}\alpha\right)\underset{N\to\infty}{\longrightarrow}h_\mu(\alpha,T)$ $\mu-$*c.s. Esta convergencia implica la convergencia en probabilidad. Tomemos* $N(\epsilon)$ *tal que* $\forall N\geq N(\epsilon)$

$$\left[\mu\left\{x\in X:\left|\frac{1}{N}I\left(\bigvee_{n=0}^{N-1}T^{-n}\alpha\right)-h_\mu(\alpha,T)\right|>\epsilon\right\}<\epsilon\right]$$

*Sea* $x\in X$ *que verifica* $\left|\frac{1}{N}I\left(\bigvee_{n=0}^{N-1}T^{-n}\alpha\right)(x)-h_\mu(T,\alpha)\right|>\epsilon$ *donde* $T^n x\in A_n(\in\alpha)$. *Luego*

$$\left\{x\in X:\left|\frac{1}{N}I\left(\sum_{n=0}^{N-1}T^{-n}\alpha\right)-h_\mu(\alpha,T)\right|>\epsilon\right\}$$

$$=\cup\left\{A\in\bigvee_{n=0}^{N-1}T^{-n}\alpha:\left|-\frac{1}{N}log\ \mu(A)-h_\mu(\alpha,T)\right|>\epsilon\right\}$$

*de lo que se obtiene el resultado.* ♠

*De ahora en adelante supondremos que* $(X,\mathcal{B},\mu)$ *es un espacio de probabilidad completo y separable, es decir,* $\exists(B_n)_{n\in\mathbb{N}}\subseteq\mathcal{B}$ *tal que* $\sigma\left((B_n)_{n\in\mathbb{N}}\right)=\mathcal{B}(\bmod\ \mu)$. *La demostración de las siguientes propiedades se deja al lector como ejercicio.*

***Propiedades.***

1. Si $(X,\mathcal{B},\mu)$ es un espacio de probabilidad completo y separable entonces $\exists(\alpha_m)_{m\in\mathbb{N}}$ sucesión creciente de particiones finitas (i.e. cada vez más finas) tal que $\sigma(\alpha_m)\underset{m\to\infty}{\longrightarrow}\mathcal{B}(\bmod\ \mu)$.

2. Sea $(X,\mathcal{B},\mu)$ un espacio de probabilidad completo y separable. Sea $\mathcal{A}\subseteq\mathcal{B}$ sub $\sigma-$ álgebra. Entonces $\exists(\alpha_m)_{m\in\mathbb{N}}$ sucesión creciente de particiones finitas (con $\sigma(\alpha_m)\subseteq\mathcal{A}$) tal que $\sigma(\alpha_m)\underset{m\to\infty}{\nearrow}\mathcal{A}$.

*En estos espacios podemos extender los resultados de información y entropía condicional. En el segundo factor donde condicionamos por particiones podemos condicionar por sub* $\sigma-$ *álgebras.*

***Proposición.*** Sea $(X,\mathcal{B},\mu)$ un espacio de probabilidad completo y separable. Sean $\alpha,\beta$ particiones numerables y medibles tales que $H_\mu(\alpha)<\infty$, $H_\mu(\mathcal{B})<\infty$, y $\mathcal{A}$ una sub $\sigma-$ álgebra de $\mathcal{B}$. Entonces

$$H_\mu(\alpha \vee \beta \mid \mathcal{A}) = H_\mu(\alpha \mid \mathcal{A}) + H_\mu(\beta \mid \mathcal{A} \vee \sigma(\alpha))$$

$$I(\alpha \vee \beta \mid \mathcal{A}) = I(\alpha \mid \mathcal{A}) + I(\beta \mid \mathcal{A} \vee \sigma(\alpha))$$

$$H_\mu(\alpha \vee \beta \mid \mathcal{A}) = H_\mu(\alpha \mid \mathcal{A}) + H_\mu(\beta \mid \mathcal{A} \vee \sigma(\alpha))$$

**Demostración:** Sea $(\delta_n)_{n\in\mathbb{N}}$ una sucesión de particiones finitas crecientes a $\mathcal{A}$. Se tiene $I(\alpha \vee \beta \mid \delta_n) = I(\alpha \mid \delta_n) + I(\beta \mid \delta_n \vee \alpha)$. Por el corolario del Lema de Chung, se tiene al hacer $n \to \infty$ que las expresiones anteriores convergen $\mu - c.s.$ a $I(\alpha \vee \beta \mid \mathcal{A}) = I(\alpha \mid \mathcal{A}) + I(\beta \mid \mathcal{A} \vee \sigma(\alpha))$. ♠

Resumamos las propiedades de información y entropía. Sea $(X, \mathcal{B}, \mu)$ un espacio de probabilidad completo y separable. Sean $\alpha, \beta$ particiones de entropía finita y $\mathcal{A}, \mathcal{A}'$ sub $\sigma-$ álgebras de $\mathcal{B}$. Se tiene,

•$H_\mu(\alpha \mid \mathcal{A}) \leq H_\mu(\beta \mid \mathcal{A})$ si $\alpha \preceq \beta$,

•$H_\mu(\alpha \mid \mathcal{A}) \leq H_\mu(\alpha \mid \mathcal{A}')$ si $\mathcal{A} \supseteq \mathcal{A}'$,

• $H_\mu(\alpha \vee \beta \mid \mathcal{A}) = H_\mu(\alpha|\mathcal{A}) + H_\mu(\beta \mid \sigma(\alpha) \vee \mathcal{A})$,

• $H_\mu(\alpha \vee \beta \mid \mathcal{A}) \leq H_\mu(\alpha \mid \mathcal{A}) + H_\mu(\beta \mid \mathcal{A})$,

• $H_\mu(\alpha \mid \mathcal{A}) = H_\mu(T^{-1}\alpha \mid T^{-1}\mathcal{A})$. ♠

**Definición.** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. Una partición $\alpha$ numerable y medible de entropía finita se dice generadora fuerte si $\alpha^- \bigvee \sigma(\alpha) = \mathcal{B}$ (mod $\mu$), donde $\alpha^- = \bigvee_{i\geq 1} T^{-i}\alpha$. En el caso invertible $\alpha$ se dice generadora si $\alpha_T = \mathcal{B}$ (mod $\mu$), donde $\alpha_T = \bigvee\limits_{i\in\mathbb{Z}} T^i\alpha$.*

Habíamos definido, para $\alpha$ partición de entropía finita $h_\mu(\alpha, T) = H_\mu(\alpha \mid \alpha^-)$ y por el teorema ergódico de la información

$$h_\mu(\alpha, T) =^{\text{lím}}_{N\to\infty} 1/N \cdot H\left(\bigvee_{n=0}^{N-1} T^{-i}\alpha\right).$$

**Definición (Kolmogorov- Sinai).** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. La entropía del sistema es*

$h_\mu(T) = \sup\{h_\mu(\alpha, T) : \alpha \text{ partición medible y numerable con } H_\mu(\alpha) < \infty\}$.

**Comentarios:**

- Si $h_\mu(T) = 0$ entonces toda partición $\alpha$ numerable y medible con $H_\mu(\alpha) < \infty$ satisface $H_\mu(\alpha \mid \alpha^-) = 0$. Como $H_\mu(\alpha \mid \mathcal{A}) = 0 \Leftrightarrow \alpha$ es $\mathcal{A}-$ medible, donde $\mathcal{A}$ es una sub $\sigma-$ álgebra de $\mathcal{B}$, entonces $\alpha \subseteq \alpha^-$, es decir, el sistema es "determinista".
- También deducimos que $h_\mu(T) = 0 \Rightarrow T^{-1}\mathcal{B} = \mathcal{B} \pmod{\mu}$. En efecto, si $\exists B \in \mathcal{B} \setminus T^{-1}\mathcal{B}$ consideramos $\alpha = (B, B^c)$. Se tiene que $\alpha \subseteq \alpha^-$. Como $\alpha^- \subseteq T^{-1}\mathcal{B}$ se concluye que $\alpha \subseteq T^{-1}\mathcal{B}$. Luego $B \in T^{-1}\mathcal{B}$, lo que es una contradicción.

**Proposición.** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. Se tiene para todo $n \geq 0$, $h_\mu(T^n) = nh_\mu(T)$. Si el s.d.a. es invertible se tiene que $h_\mu(T^n) = |n|\, h_\mu(T)$ en todo entero $n \in \mathbb{Z}$.*

**Demostración:** Para concluir la segunda igualdad basta probar que $h_\mu(T^{-1}) = h_\mu(T)$. Ahora bien,

$$\begin{aligned} H_\mu(\bigvee_{n=0}^{N-1} (T^{-1})^{-n}\alpha) &= H_\mu(T^{-(N-1)}(\bigvee_{n=0}^{N-1} T^n\alpha)) \\ &= H_\mu(\bigvee_{n=0}^{N-1} T^{-n}\alpha), \end{aligned}$$

de donde se concluye que $h_\mu(\alpha, T^{-1}) = h_\mu(\alpha, T)$ para toda partición $\alpha$.

Si $n = 0$, se deduce que $h_\mu(\alpha, T) = 0$. Sea $n > 0$ y $\alpha$ una partición tal que $H_\mu(\alpha) < \infty$. Entonces,

$$\begin{aligned} h_\mu(\alpha, T^n) &\leq h_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha, T^n\right) \\ &= \lim_{N\to\infty} \tfrac{1}{N} H_\mu\left(\bigvee_{l=0}^{N-1} T^{-nl}\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)\right) \\ &= \lim_{N\to\infty} \tfrac{1}{N} H_\mu\left(\bigvee_{i=0}^{nN-1} T^{-i}\alpha\right) \\ &= \lim_{N\to\infty} \tfrac{n}{nN} H_\mu\left(\bigvee_{i=0}^{nN-1} T^{-i}\alpha\right) \\ &= nh_\mu(\alpha, T). \end{aligned}$$

Deducimos que $h_\mu(\alpha, T^n) \leq n h_\mu(\alpha, T)$. Para la otra desigualdad observemos que acabamos de mostrar que

$$n h_\mu(\alpha, T) = h_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha, T^n\right) \leq h_\mu(T^n),$$

pues $H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leq \sum_{i=0}^{n-1} H_\mu(T^{-i}\alpha) =\leq n H_\mu(\alpha) < \infty.$ ♠

**Lema.** *Sea* $(X, \mathcal{B}, \mu, T)$ *un s.d.a. Si* $\alpha, \beta$ *son particiones de entropía finita entonces*

$$h_\mu(\alpha, T) \leq h_\mu(\beta, T) + H_\mu(\alpha \mid \beta)$$

*y*

$$|h_\mu(\alpha, T) - h_\mu(\beta, T)| \leq H_\mu(\alpha|\beta) + H_\mu(\beta|\alpha).$$

**Demostración:**

$$\begin{aligned} H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) &\leq H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \vee \bigvee_{i=0}^{n-1} T^{-i}\beta\right) \\ &= H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) + H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha \mid \bigvee_{i=0}^{n-1} T^{-i}\beta\right) \\ &\leq H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) + \sum_{i=0}^{n-1} H_\mu(T^{-i}\alpha|T^{-i}\beta) \\ &\leq H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) + nH_\mu(\alpha|\beta). \end{aligned}$$

Dividiendo por $n$ y tomando límite se obtiene el primer resultado. El segundo se deduce directamente del primero. ♠

**Teorema.** *Sea* $(X, \mathcal{B}, \mu, T)$ *un s.d.a. Sea* $(\alpha_n)_{n\in\mathbb{N}}$ *una sucesión creciente de particiones de entropía finita tal que* $\alpha_n \nearrow \mathcal{B}\,(mod \mu)$. *Entonces*

$$h_\mu(T) =_{\lim_{n\to\infty}} h_\mu(T, \alpha_n).$$

**Demostración:** Sea $\delta$ partición con $H_\mu(\delta) < \infty$. De acuerdo al último lema

$$h_\mu(T, \delta) \leq h_\mu(T, \alpha_n) + H_\mu(\delta \mid \alpha_n).$$

Además, como $\alpha_n$ crece a $\mathcal{B}$ se tiene que $H_\mu(\delta \mid \alpha_n) \underset{n\to\infty}{\searrow} H_\mu(\delta \mid \mathcal{B})$. Dado que $h_\mu(T,\delta) \leq \lim_{n\to\infty} h_\mu(T,\alpha_n)$ de donde $h_\mu(T) \leq \lim_{n\to\infty} h_\mu(T,\alpha_n) \leq h_\mu(T)$. ♠

**Proposición.** *Sean $(X_i, \mathcal{B}_i, \mu_i, T_i)$, $i = 1, 2$, s.d.a. separables. Entonces*

$$h_{\mu_1\otimes\mu_2}(T_1 \otimes T_2) = h_{\mu_1}(T_1) + h_{\mu_2}(T_2).$$

**Demostración:** $T_1 \otimes T_2 : X_1 \times X_2 \to X_1 \times X_2$ definida por $T_1 \otimes T_2(x_1, x_2) = (T_1x_1, T_2x_2)$ preserva $\mu_1 \otimes \mu_2$. Tomemos $(\alpha_n^i)_{n\in\mathbb{N}}$ particiones finitas en $X_i$ con $\alpha_n^i \nearrow \mathcal{B}_i \,(\mathrm{mod}\ \mu_i)$, $i = 1, 2$ (por separabilidad ellas existen). Se tiene, de la definición de $\sigma-$ álgebra producto que $\alpha_n^1 \times \alpha_n^2 \nearrow \mathcal{B}_1 \otimes \mathcal{B}_2 \,(\mathrm{mod}\ \mu_1 \otimes \mu_2)$. Por el teorema anterior,

$$\begin{aligned} h_{\mu_1\otimes\mu_2}(T_1 \otimes T_2) &= \lim_{n\to\infty} h_{\mu_1\otimes\mu_2}(T_1 \otimes T_2, \alpha_n^1 \times \alpha_n^2) \\ &= \lim_{n\to\infty} (h_{\mu_1}(T_1, \alpha_n^1) + h_{\mu_2}(T_2, \alpha_n^2)) \\ &= h_{\mu_1}(T_1) + h_{\mu_2}(T_2) \end{aligned}$$

pues

$$H_{\mu_1\otimes\mu_2}\left(\bigvee_{k=0}^{N-1} (T_1 \otimes T_2)^{-k}\left(\alpha_n^1 \times \{X_2\} \bigvee \{X_1\} \times \alpha_n^2\right)\right)$$

$$= H_{\mu_1\otimes\mu_2}\left(\left(\bigvee_{k=0}^{N-1} T_1^{-k}\left(\alpha_n^1\right) \times \{X_2\}\right) \bigvee \left(\{X_1\} \times \bigvee_{k=0}^{N-1} T_2^{-k}\left(\alpha_n^2\right)\right)\right).$$

♠

**Teorema (generador).** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a.*

*a) Si $\alpha, \gamma$ son particiones de entropía finita, entonces*

*a1) Si $\sigma(\alpha) \vee \alpha^- \supseteq \sigma(\gamma)$ entonces $h_\mu(T, \gamma) \leq h_\mu(T, \alpha)$,*

*a2) Suponga $T$ es invertible. Si $T\alpha_T \supseteq \sigma(\gamma)$ entonces $h_\mu(T, \gamma) \leq h_\mu(T, \alpha)$.*

*b) Sea $\alpha$ una partición de entropía finita.*

*b1) Si $\alpha$ es un generador fuerte (es decir $\sigma(\alpha) \vee \alpha^- = \mathcal{B}\,(\ \mathrm{mod}\ \mu)$) entonces*

$$h_\mu(T) = h_\mu(T, \alpha).$$

*b2) Si $T$ es invertible y $\alpha$ es un generador $(\alpha_T = \mathcal{B}(\text{ mod } \mu))$, entonces*

$$h_\mu(T) = h_\mu(T,\alpha).$$

**Demostración:** b)se deduce de a). Probemos a1). Suponemos que $\gamma \preceq \alpha^- \vee \alpha$. Luego

$$h_\mu(T,\gamma) \le h_\mu\left(T, \bigvee_{k=0}^{n-1} T^{-k}\alpha\right) + H_\mu\left(\gamma \mid \bigvee_{k=0}^{n-1} T^{-k}\alpha\right)$$

Sea $\epsilon > 0$. Como $\bigvee_{k=0}^{n-1} T^{-k}\alpha \nearrow \alpha \vee \alpha^-$ y $\gamma$ es $\alpha \vee \alpha^-$medible, se tiene que $\exists n(\epsilon) \in \mathbb{N}$ tal que $\forall n \ge n(\epsilon)$, $H_\mu\left(\gamma \mid \bigvee_{k=0}^{n-1} T^{-k}\alpha\right) < \epsilon$.

Fijemos $n \ge n(\epsilon)$. Observemos que

$$\begin{aligned} h_\mu\left(T, \bigvee_{k=0}^{n-1} T^{-k}\alpha\right) &= \lim_{N\to\infty} \tfrac{1}{N} H_\mu\left(\bigvee_{l=0}^{N-1} T^{-l}\left(\bigvee_{k=0}^{N-1} T^{-k}\alpha\right)\right) \\ &= \lim_{N\to\infty} \tfrac{1}{N} H_\mu\left(\bigvee_{n=0}^{(N+n-2)} T^{-n}\alpha\right) \\ &= h_\mu(T,\alpha), \end{aligned}$$

de lo que se deduce $h_\mu(T,\gamma) \le h_\mu(T,\alpha) + \epsilon$ para cada $\epsilon > 0$ y se concluye. La parte a2) es análoga. ♠

**Ejemplos:** Sea $A$ un alfabeto finito, $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ el shift, con $(\sigma((x_m)_{m\in\mathbb{N}}))_n = x_{n+1} \forall n \in \mathbb{Z}$. Sea $\mathcal{B} = (\mathcal{P}(A))^{\mathbb{Z}}$ la $\sigma-$álgebra producto. Sea $\mu$ una medida $\sigma-$invariante, es decir, $\mu \circ \sigma^{-1} = \mu$. Sigamos denotando $\mathcal{B}$ a su completación con respecto a $\mu$.

La partición de la coordenada 0: $\delta_0 = \{C_a : a \in A\}$ donde $C_a = \left\{x \in A^{\mathbb{Z}} : x_0 = a\right\}$ es un generador.

En efecto, sea $\sigma^{-k}\delta_0 = \left\{\sigma^{-k}C_a : a \in A\right\}$ la partición de la coordenada $k$ (notar que $\sigma^{-k}C_a = \{x \in A^{\mathbb{Z}} : x_k = a\}$), entonces

$$\bigvee_{k=0}^{n-1} \sigma^{-k}\delta_0 = \left\{\bigcap_{k=0}^{n-1} \sigma^{-k}C_{a_k} : (a_0, ...., a_{n-1}) \in A^n\right\}$$

define una familia de particiones crecientes a $\mathcal{B} \mod \mu$.

**1. Shift de Bernoulli:** Consideremos $\pi = (\pi_a : a \in A)$ una medida de probabilidad en $A$ (se puede suponer que $\pi_a > 0$, $\forall_a \in A$). La medida de Bernoulli producto (esto es la medida coordenadas independientes entre sí con marginal $\pi$) que llamamos $\mu_\pi$ es $\sigma$ invariante. Las particiones $\sigma^{-n}\delta_0$ son independientes $\forall n \in \mathbb{N}$. Luego

$$\begin{aligned}
h_{\mu_\pi}(\sigma) &= h_{\mu_\pi}(\sigma, \delta_0) \\
&=_{\lim\limits_{N\to\infty}} \tfrac{1}{N} H_{\mu_\pi}\left(\bigvee_{n=0}^{N-1} \sigma^{-n}\delta_0\right) \\
&=_{\lim\limits_{N\to\infty}} \tfrac{1}{N} \sum_{n=0}^{N-1} H_{\mu_\pi}(\sigma^{-n}\delta_0) \\
&= H_{\mu_\pi}(\delta_0) \\
&= -\textstyle\sum_{a\in A} \mu_\pi(C_a) \log \mu_\pi(C_a) \\
&= -\displaystyle\sum_{a\in A} \pi_a \log \pi_a
\end{aligned}$$

**2. Shift de Markov.** En este caso, la medida $\sigma-$ invariante se construye de la manera siguiente: consideramos una matriz estocástica irreducible

$$P = (P_{a,a'} : a, a' \in A)$$

y $\pi = (\pi_a : a \in A)$ su vector de probabilidad estacionario (es decir tal que $\pi^T P = \pi^T$). La medida de Markov $\mu_P$definida en $A^{\mathbb{Z}}$está dada por

$$\mu_P \left\{x \in A^{\mathbb{Z}} : x_k = a_0, ..., x_{k+l} = a_l\right\} = \pi_{a_0} P_{a_0,a_1} \cdot .... \cdot P_{a_{l-1},a_l}$$

Probar como ejercicio que $\mu_P$ está bien definida y que es una medida $\sigma-$ invariante. Además pruebe que

$$h_{\mu_P}(\sigma) = h_\mu(\sigma, \delta_0) = -\sum_{a\in A}\sum_{b\in A} \pi_a P_{a,b} \log(P_{a,b}).$$

## 4.4. La entropía como invariante.

Sean $(X_i, \mathcal{B}_i, \mu_i, T_i)$ s.d.a. para $i = 1, 2$. Recordemos las nociones de factor y conjugación.

a) Se dice que $(X_2, \mathcal{B}_2, \mu_2, T_2)$ es un factor de $(X_1, \mathcal{B}_1, \mu_1, T_1)$ si $\exists \Phi : X_1 \to X_2$, aplicación $\mathcal{B}_1 - \mathcal{B}_2$ medible tal que: $\Phi(\mu_1) = \mu_2$ (es decir $\mu_1 \circ \Phi^{-1} = \mu_2$), y $\Phi \circ T_1 = T_2 \circ \Phi$, $\mu_1-$c.s.

b) Se dice que $(X_1, \mathcal{B}_1, \mu_1, T_1)$ es conjugado con $(X_2, \mathcal{B}_2, \mu_2, T_2)$ si $\exists \Phi : X_1 \to X_2$ biyectiva y bimedible $\mu_1-$c.s. tal que $\Phi(\mu_1) = \mu_2$ y $\Phi \circ T_1 = T_2 \circ \Phi$ $\mu_1-$c.s. Equivalentemente, $\Phi^{-1} : X_2 \to X_1, \Phi^{-1}(\mu_2) = \mu_1, \Phi^{-1} \circ T_2 = T_1 \circ \Phi^{-1}$ $\mu_2-$c.s.

**Proposición**.

*a) Si* $(X_2, \mathcal{B}_2, \mu_2, T_2)$ *es factor de* $(X_1, \mathcal{B}_1, \mu_1, T_1)$ *entonces* $h_{\mu_2}(T_2) \leq h_{\mu_1}(T_1)$.

*b) Si* $(X_1, \mathcal{B}_1, \mu_1, T_1)$ *y* $(X_2, \mathcal{B}_2, \mu_2, T_2)$ *son conjugados entonces* $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$.

**Demostración**: Sea $\alpha$ una partición en $X_2$ con $H_{\mu_2}(\alpha) < \infty$. Si $\Phi : X_1 \to X_2$ es un factor entre los s.d.a. $(X_1, \mathcal{B}_1, \mu_1, T_1)$ y $(X_2, \mathcal{B}_2, \mu_2, T_2)$ entonces $\Phi^{-1}\alpha = (\Phi^{-1}A : A \in \alpha)$ es la partición de $X_1$ que contiene las preimagenes de los átomos de $\alpha$. Esta partición es de entropía finita pues:

$$\begin{aligned} H_{\mu_1}(\Phi^{-1}\alpha) &= -\sum_{A\in\alpha} \mu_1(\Phi^{-1}A) \log \mu_1(\Phi^{-1}A) \\ &= -\sum_{A\in\alpha} \mu_2(A) \ \log \mu_2(A) \\ &= H_{\mu_2}(\alpha) < \infty \end{aligned}$$

Por el mismo argumento

$$\begin{aligned} h_{\mu_1}(T_1, \Phi^{-1}\alpha) &= \lim_{N\to\infty} \tfrac{1}{N} H_{\mu_1}\left(\bigvee_{n=0}^{N-1} T_1^{-n}(\Phi^{-1}\alpha)\right) \\ &= \lim_{N\to\infty} \tfrac{1}{N} H_{\mu_1}\left(\bigvee_{n=0}^{N-1} \Phi^{-1}\left(T_2^{-n}\alpha\right)\right) \\ &= \lim_{N\to\infty} \tfrac{1}{N} H_{\mu_1}\left(\Phi^{-1}\left(\bigvee_{n=0}^{N-1} T_2^{-n}\right)\right) \\ &= \lim_{N\to\infty} H_{\mu_2}\left(\bigvee_{n=0}^{N-1} T_2^{-n}\alpha\right) = h_{\mu_2}(T_2, \alpha) \end{aligned}$$

luego $h_{\mu_2}(T_2, \alpha) = h_{\mu_1}(T_1, \Phi^{-1}\alpha) \leq h_{\mu_1}(T_1)$. Concluimos $h_{\mu_2}(T_2) \leq h_{\mu_1}(T_1)$.
♠

**Observación:** Los shifts de Bernoulli de medida $\pi^{\mathbb{Z}}$ tienen entropía

$$-\sum_{a\in A} \pi_a \ \log \ \pi_a.$$

Como la entropía es un invariante de conjugación (es decir dos s.d.a. conjugados tienen igual entropía), una condición necesaria para que dos shifts de Bernoulli sean conjugados es que tengan igual entropía. La condición suficiente es también cierta.

**Teorema (Ornstein, 70's).** *Dos shift de Bernoulli son conjugados si tienen la misma entropía.*

La demostración tiene alrededor de 100 páginas y no la incluiremos.

**Definición.** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. La $\sigma-$álgebra de Pinsker $\mathcal{P}(T)$ se define como la más pequeña $\sigma-$álgebra que contiene a*

$$\mathcal{A} = \cup \{\alpha : H_\mu(\alpha) < \infty, h_\mu(\alpha, T) = 0\}$$

*es decir son las particiones de entropía finita de $X$ cuya entropía relativa a $T$ es cero.*

**Proposición.** *$\mathcal{A}$ es un álgebra .*

**Demostración:**

- $X \in \mathcal{A}$ pues $X \in \alpha = \{X\}, h_\mu(\alpha, T) = 0$.

- Sea $A \in \mathcal{A}$. Entonces $A \in \alpha$ con $H_\mu(\alpha) < \infty, h_\mu(\alpha, T) = 0$. La partición $\{A, A^c\}$ es más fina que $\alpha$ luego $h_\mu(\{A, A^c\}, T) = 0$. Esto prueba que $A^c \in \mathcal{A}$.

- Sean $A_1, ..., A_n \in \mathcal{A}$. Entonces $A_i \in \alpha_i$ donde $H_\mu(\alpha_i) < \infty, h_\mu(\alpha_i, T) = 0, i = 1, ..., n$. Luego, $H_\mu\left(\bigvee_{i=1}^{n} \alpha_i\right) \leq \sum_{i=1}^{n} H_\mu(\alpha_i) < \infty$ y $h_\mu\left(\bigvee_{i=1}^{n} \alpha_i, T\right) \leq \sum_{i=1}^{n} h_\mu(\alpha_i, T) = 0$. Esto prueba que $\bigvee_{i=1}^{n} \alpha_i \subseteq \mathcal{A}$, es decir $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$. ♠

**Proposición**. $T^{-1}\mathcal{P}(T) = \mathcal{P}(T)$, *es decir $\mathcal{P}(T)$ es estrictamente $T-$ invariante.*

**Demostración:** Si $A \in \mathcal{A}$ entonces $T^{-1}A \in \mathcal{P}(T)$ pues $h_\mu(\{T^{-1}A, T^{-1}A^c\}, T) = h_\mu(\{A, A^c\}, T) = 0$. Luego $T^{-1}\mathcal{A} \subseteq \mathcal{P}(T)$, de donde $T^{-1}\mathcal{P}(T) \subseteq \mathcal{P}(T)$.

Recíprocamente, si $H_\mu(\alpha) < \infty, h_\mu(\alpha, T) = 0$, entonces $\alpha \subseteq \alpha^- = \bigvee_{i \geq 1} T^{-i}\alpha$. Pongamos $\alpha^- = T^{-1}\left(\bigvee_{i \geq 0} T^{-i}\alpha\right)$. Como $\alpha \subseteq \mathcal{P}(T)$ entonces $T^{-i}\alpha \subseteq \mathcal{P}(T)$ pues $h_\mu\left(T^{-i}\alpha, T\right) = h_\mu(\alpha, T) = 0$ para $i \in \mathbb{N}$. Luego $\bigvee_{i \geq 0} T^{-i}\alpha \subseteq \mathcal{P}(T)$.

Se concluye que existe $\delta \subseteq \bigvee_{i\geq 0} T^{-i}\alpha \subseteq \mathcal{P}(T)$ tal que $\alpha = T^{-1}\delta$ y $\alpha \subseteq T^{-1}\mathcal{P}(T)$. Luego $\mathcal{A} \subseteq T^{-1}\mathcal{P}(T)$, de donde $\mathcal{P}(T) \subseteq T^{-1}\mathcal{P}(T)$. ♠

**Teorema.** *Sea $\alpha$ una partición con $H_\mu(\alpha) < \infty$. Se tiene la equivalencia $h_\mu(\alpha, T) = 0 \Leftrightarrow \alpha \subseteq \mathcal{P}(T)$. Luego $A \in \mathcal{P}(T) \Leftrightarrow h_\mu(\{A, A^c\}, T) = 0$.*

**Demostración:**

$(\Rightarrow)$ definición de $\mathcal{P}(T)$.

$(\Leftarrow)$ $\mathcal{A}$ es el álgebra engendrada por $\mathcal{P}(T)$. Por ser el espacio separable $\exists (\delta_n)_{n\in\mathbb{N}}$ sucesión creciente de particiones finitas con $\delta_n \subseteq \mathcal{A} \wedge \delta_n \nearrow \mathcal{P}(T)$.

Sea $\alpha$ con $H_\mu(\alpha) < \infty, \alpha \subseteq \mathcal{P}(T)$. De un lema anterior, $h_\mu(\alpha, T) \leq h_\mu(\delta_n, T) + H_\mu(\alpha|\delta_n)$ pues $\delta_n \nearrow \mathcal{P}(T)$ y $\alpha \subseteq \mathcal{P}(T)$, con lo que

$$H_\mu(\alpha \mid \mathcal{P}(T)) = 0.$$

♠

La $\sigma-$ álgebra de Pinsker $\mathcal{P}(T) = \{A \in \mathcal{B} : h_\mu(\{A, A^c\}, T) = 0\}$ se llama la parte determinista de $(X, \mathcal{B}, T, \mu)$.

**Teorema.** *Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. invertible. Se tiene la igualdad*

$$\mathcal{P}(T) = \bigvee_{\alpha : H_\mu(\alpha)<\infty} \bigcap_{n\geq 0} T^{-n}\alpha^-$$

Veamos primero unos lemas.

**Lema.**

$$\frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^i\alpha|\delta^-\right) \underset{n\to\infty}{\rightrightarrows} H_\mu\left(\alpha \mid \alpha^-\right) (= h_\mu(\alpha, T)).$$

*Sean $\alpha$ y $\delta$ particiones con $H_\mu(\alpha) < \infty, H_\mu(\delta) < \infty$, tal que $\alpha \preceq \delta$ o $\delta \preceq \alpha$. Entonces*

$$\frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^i\alpha|\delta^-\right) \underset{n\to\infty}{\rightrightarrows} H_\mu\left(\alpha \mid \alpha^-\right) (= h_\mu(\alpha, T)).$$

$\left(\text{en particular, } \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^i\alpha\right) \underset{n\to\infty}{\longrightarrow} h_\mu(\alpha, T) \text{ si } H_\mu(\alpha) < \infty\right).$

***Demostración:***

*a) Supongamos que $\delta \preceq \alpha$, entonces*

$$T^{-n}\left(\delta^- \bigvee \left(\bigvee_{i=0}^{n-1} T^i\alpha\right)\right) = T^{-n}\delta^- \bigvee \left(\bigvee_{i=1}^{n} T^{-i}\alpha\right) \nearrow_{n\to\infty} \alpha^-$$

*pues $T^{-n}\delta^- \subseteq \delta^- \subseteq \alpha^-$. Sea $\alpha^n = \bigvee_{i=0}^{n} T^i\alpha$, entonces*

$$H_\mu\left(\alpha | T^{-n}\left(\delta^- \bigvee (\bigvee_{i=0}^{n-1} T^i\alpha)\right)\right) = H_\mu\left(T^n\alpha \mid \delta^- \bigvee \alpha^{n-1}\right) \searrow H_\mu\left(\alpha \mid \alpha^-\right).$$

*Se concluye que*

$$\begin{aligned} H_\mu(\alpha^n \mid \delta^-) &= H_\mu(T^n\alpha \bigvee \alpha^{n-1} \mid \delta^-) \\ &= H_\mu(T^n\alpha \mid \alpha^{n-1} \bigvee \delta^-) + H_\mu(\alpha^{n-1} \mid \delta^-) \\ &= \sum_{k=0}^{n} H_\mu\left(T^k\alpha \mid \alpha^{k-1} \bigvee \delta^-\right) \end{aligned}$$

*Luego* $\displaystyle \frac{1}{n} H_\mu\left(\alpha^n \mid \delta^-\right) = \frac{1}{n}\sum_{k=1}^{n} H_\mu\left(T^k\alpha \mid \alpha^{k-1} \bigvee \delta^-\right) \to_{n\to\infty} H_\mu\left(\alpha \mid \alpha^-\right).$

*b) Supongamos que $\alpha \preceq \delta$. Entonces*

$$\frac{1}{n} H_\mu\left(\alpha^n \mid \delta^-\right) \leq \frac{1}{n} H_\mu\left(\alpha^n \mid \alpha^-\right) \to_{n\to\infty} H_\mu\left(\alpha \mid \alpha^-\right)$$

*tomando $\delta = \alpha$ en (a).*

*Recíprocamente $H_\mu(\delta^n \mid \delta^-) = H_\mu(\delta^n \bigvee \alpha^n \mid \delta^-)$, de donde*

$$\frac{1}{n} H_\mu\left(\alpha^n \mid \delta^-\right) = \frac{1}{n} H_\mu\left(\delta^n \mid \delta^-\right) - \frac{1}{n} H_\mu\left(\delta^n \mid \alpha^n \bigvee \delta^-\right).$$

*Luego*

$$\liminf_{n\to\infty} \frac{1}{n} H_\mu\left(\alpha^n \mid \delta^-\right) \geq H_\mu\left(\delta \mid \delta^-\right) \geq \limsup_{n\to\infty} \frac{1}{n} H_\mu\left(\delta^n \mid \alpha^n \bigvee \alpha^-\right).$$

*Por (a) como $\alpha \preceq \delta$ se tiene*

$$\begin{array}{rl} \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\delta^n \mid \alpha^-\right) - & \\ \limsup_{n\to\infty}\frac{1}{n}H_\mu\left(\delta^n \mid \alpha^n \bigvee \alpha^-\right) & = \liminf_{n\to\infty}\frac{1}{n}(H_\mu(\delta^n|\alpha^-) - H_\mu(\delta^n|\alpha^n \bigvee \alpha^-)) \\ & = \liminf_{n\to\infty}\frac{1}{n}H_\mu\left(\alpha^n \mid \alpha^-\right) \end{array}$$

*pues* $H_\mu\left(\delta^n \mid \alpha^-\right) = H_\mu\left(\delta^n \bigvee \alpha^n \mid \alpha^-\right) = H_\mu\left(\alpha^n \mid \alpha^-\right) + H_\mu\left(\delta^n \mid \alpha^n \bigvee \alpha^-\right).$
♠

***Proposición.*** Sean $\alpha, \delta$ particiones de entropía finita entonces

$$h_\mu\left(T, \alpha \bigvee \delta\right) = h_\mu\left(T, \delta\right) + H_\mu\left(\alpha \mid \alpha^- \bigvee \delta_T\right)$$

$\left(\text{donde } \alpha^- = \bigvee_{i\geq 1} T^{-i}\alpha, \delta_T = \bigvee_{i\in\mathbb{Z}} T^{-i}\delta\right).$

***Demostración:*** *Recordemos la notación* $\alpha^n = \bigvee_{i=0}^{n} T^i\alpha$. *De acuerdo al lema anterior teníamos:*

$$h_\mu\left(T, \alpha \vee \delta\right) =_{n\to\infty}^{lim} \frac{1}{n}H_\mu\left(\alpha^n \vee \delta^n \mid \alpha^- \vee \delta^-\right),$$

*pues* $\alpha^n \vee \delta^n = \left(\alpha \vee \delta\right)^n, \alpha^- \vee \delta^- = \left(\alpha \vee \delta\right)^-$. *Por otra parte*

$$\frac{1}{n}H_\mu\left(\alpha^n \vee \delta^n \mid \alpha^- \vee \delta^-\right) = \frac{1}{n}H_\mu\left(\delta|\alpha^- \vee \delta^-\right) + \frac{1}{n}H_\mu\left(\alpha^n \mid \alpha^- \vee \delta^n\right)$$

*y por el lema* $\frac{1}{n}H_\mu\left(\delta^n \mid \left(\alpha \vee \delta\right)^-\right) = h_\mu\left(\delta, T\right)$. *Además se tiene*

$$\begin{array}{rl} \frac{1}{n}H_\mu\left(\alpha^n \mid \alpha^- \vee \delta^- \vee \delta^n\right) & = \frac{1}{n}\left[H_\mu\left(\alpha^{n-1} \mid \alpha \vee \delta^- \vee \delta^n\right) + H_\mu\left(T^n\alpha \mid \alpha^- \vee \delta^n \vee \alpha^{n-1}\right)\right] \\ & = \frac{1}{n}\sum_{k=0}^{n} H_\mu(T^k\alpha|\alpha^- \bigvee \delta^- \bigvee \delta^n \bigvee \alpha^{k-1}) \\ & = \frac{1}{n}\sum_{k=0}^{n} H_\mu(\alpha|\alpha^- \bigvee \delta^- \bigvee \delta^{n-k}) \end{array}$$

*ahora bien* $H_\mu\left(\alpha \mid \alpha^- \vee \delta^- \vee \delta^n\right) \to_{n\to\infty} H_\mu\left(\alpha \mid \alpha^- \vee \delta_T\right)$ *con lo que se obtiene el resultado.* ♠

***Demostración del Teorema:*** *Debemos mostrar que*

$$\mathcal{P}\left(T\right) = \bigvee_{\alpha: H_\mu(\alpha)<\infty} \bigcap_{n\geq 0} T^{-n}\alpha.$$

*Sea* $\alpha^- = \bigvee_{k\geq 1} T^{-k}\alpha$. *Entonces*

$$T^{-n}\alpha^- = \bigvee_{k\geq n+1} T^{-k}\alpha \searrow \bigcap_{n\geq 0} T^{-n}\alpha^-.$$

*Por separabilidad,* $\exists\,(\alpha_k)_{k\in\mathbb{N}}$ *sucesión creciente de particiones finitas tales que* $\bigvee_{k\geq 0} \alpha_k = \mathcal{P}(T)$, *pues* $\alpha_k \subseteq \mathcal{P}(T), \alpha_k \subseteq \alpha_k^-$.

*Lo anterior implica que* $T\alpha_k \subseteq T\alpha_k^- = \alpha_k \vee \alpha_k^- = \alpha_k^-$ *ya que* $\alpha_k \subseteq \alpha_k^-$. *Luego* $T\alpha_k \subseteq \alpha_k^- \wedge \alpha_k \subseteq T^{-1}\alpha_k^-$. *Análogamente se prueba que* $\alpha_k \subseteq T^{-n}\alpha_k^-, \forall n \geq 0$. *Por lo que*

$$\alpha_k \subseteq \bigcap_{n\geq 0} T^{-n}\alpha_k^- \subseteq \bigvee_{\alpha: H_\mu(\alpha)<\infty} \bigcap_{n\geq 0} T^{-n}\alpha^-.$$

*Como* $\alpha_k \nearrow \mathcal{P}(T)$ *se concluye que*

$$\mathcal{P}(T) \subseteq \bigvee_{\alpha: H_\mu(\alpha)<\infty} \bigcap_{n\geq 0} T^{-n}\alpha^-.$$

*Recíprocamente sea* $\alpha$ *de entropía finita. Sea* $\delta$ *partición de entropía finita tal que* $\delta \subseteq \bigcap_{n\geq 0} T^{-n}\alpha^-$. *Nos basta probar que* $\delta \subseteq \mathcal{P}(T)$. *Si* $\delta \subseteq \bigcap_{n\geq 0} T^{-n}\alpha^-$ *entonces* $\delta_T \subseteq \alpha^-$.

*Ahora bien:*

$$\begin{aligned}
H_\mu\left(\delta \vee \alpha \mid \overbrace{\delta^- \vee \alpha^-}^{(\delta\vee\alpha)^-}\right) &= h_\mu(\delta \vee \alpha, T) \\
&= H_\mu\left(\alpha \mid \underbrace{\delta^- \vee \alpha^-}_{=\alpha^-}\right) + H_\mu(\delta|\alpha \vee \delta^- \vee \alpha^-) \\
&= h_\mu(T,\alpha) + H_\mu(\delta|\alpha \vee \delta^- \vee \alpha^-) \\
&= h_\mu(T,\delta) + H_\mu\left(\alpha \mid \underbrace{\delta_T \vee \alpha^-}_{=\alpha^- \textit{ pues } \delta_T \subseteq \alpha^-}\right) \\
&= h_\mu(T,\delta) + H_\mu(\alpha \mid \alpha^-) \\
&= h_\mu(T,\delta) + h_\mu(T,\alpha)
\end{aligned}$$

*Se concluye que* $h_\mu(T,\delta) = 0$, *lo que implica* $\delta \subseteq \mathcal{P}(T)$. ♠

***Observación:*** *Si* $(X, \mathcal{B}, \mu, T)$ *es un s.d.a. invertible, se tiene que* $\exists \mathcal{A}$, *sub*$\sigma-$ *álgebra de* $\mathcal{B}$ *tal que*

$$T^{-1}\mathcal{A} \subseteq \mathcal{A}, T^n\mathcal{A} \nearrow \mathcal{B}\,(mod\mu)\,, T^{-n}\mathcal{A} \searrow \mathcal{P}\,(T)\,(mod\mu)$$

***Definición***. Sea $(X, \mathcal{B}, \mu, T)$ un s.d.a. invertible. Se dice que es un $K-$sistema si existe una sub $\sigma-$ álgebra $\mathcal{A}$ de $\mathcal{B}$ tal que

$$T^{-1}\mathcal{A} \subseteq \mathcal{A}, T^n\mathcal{A} \nearrow \mathcal{B}\,(mod\mu)\,, T^{-n}\mathcal{A} \searrow \mathcal{N}\,(mod\mu)$$

donde $\mathcal{N}$ es la $\sigma-$ álgebra trivial.

***Observación:*** *En este caso se puede probar que* $\mathcal{P}\,(T) = \mathcal{N}$.

***Teorema***. *Sea* $\left(A^{\mathbb{Z}}, \mathcal{B}, \mu_P, \sigma\right)$ *un shift de Markov, con* $P$ *irreducible y aperiódica. Entonces es* $K-$ *sistema.*

***Demostración:*** *Recordemos que* $\mu_P\,\{x : x_l = i_0, .., x_{l+k} = i_k\} = \pi_{io}P_{i_0i_1}...P_{i_{k-1}i_k}$ *donde* $\pi$ *es el único vector de probabilidad tal que* $\pi^T = \pi^T P$. *Dado que* $P$ *es irreducible y aperiódica:* $P_{ij}^{(n)} \to_{n\to\infty} \pi_j$.

*Sea* $\delta_0$ *la partición de coordenada cero ,* $\delta_0 = \{C_a : a \in A\}$ *con* $C_a = \{x : x_0 = a\}$. *Sea* $\mathcal{A} = \bigvee_{n\geq 0} \sigma^{-n}\delta_0 = \delta_0 \vee \delta_0^-$. *Claramente* $\sigma^{-1}\mathcal{A} = \delta_0^- \subseteq \mathcal{A}$ *y* $\sigma^l\mathcal{A} = \bigvee\limits_{n\geq 0} \sigma^n\delta_0 \nearrow_{l\to\infty} \mathcal{B}\,(mod\mu)$, *donde* $\mathcal{B}$ *es la* $\sigma-$*álgebra producto.*

*La parte de* $K-$ *sistema, que es específica al shift de Markov irreducible y aperiódico es* $\sigma^{-n}\mathcal{A}\searrow_{n\to\infty}\mathcal{N}\,(mod\mu_P)$. *Para esto tomemos* $C = \{x : x_0 = i_0, ...., x_l = i_l\}$ *y* $D = \{x : x_0 = j_0, ..., x_m = j_m\}$ *tales que* $C \in \bigvee\limits_{n=-l}^{0} \sigma^n\delta_0, D \in \bigvee\limits_{n=-m}^{\infty} \sigma^n\delta_0$. *Entonces,*

$$\begin{aligned} \mu_P\,\{C \textstyle\bigcap \sigma^{-n}D\} &= \mu_P\,\{x \in C, \sigma^n x \in D\} \\ &= \mu_P\,\{x_0 = i_0, ..., x_l = i_l, (\sigma^n x)_0 = j_0, ..., (\sigma^n x)_m = j_n\} \\ &= \mu_P\,\{x_0 = i_0, ..., x_l = i_l, x_n = j_0, ..., x_{n+m} = j_m\} \end{aligned}$$

*Tomando* $n \geq l$ *se tiene que*

$$\mu_P\left\{C \bigcap \sigma^{-n}D\right\} = \pi_{io}P_{i_0,i_1}...P^{n-l}_{i_{l-1},i_l} \cdot P_{j_0,j_1}...P_{j_{m-1},j_m} \to_{n\to\infty} \mu_P\,(C) \cdot \mu_P\,(D)\,.$$

*Sea $C$ fijo. Sea $\epsilon > 0$. Entonces existe $n(\epsilon)$ que no depende de $D$ tal que para todo $n \geq n(\epsilon)$, todo $m \in \mathbb{N}$ y todo cilindro $D$*

$$(1-\epsilon)\,\mu_P(C)\cdot\mu_P(D) \leq \mu_P\left(C \bigcap \sigma^{-n} D\right) \leq (1+\epsilon)\,\mu_P(C)\cdot\mu_P(D)$$

*o equivalentemente*

$$(1-\epsilon)\,\mu_P(C) \leq \mu_P\left(C \mid \sigma^{-n} D\right) \leq (1+\epsilon)\,\mu_P(C).$$

*Como $\mathcal{A} = \bigvee_{n\leq 0} \sigma^n \delta_0 = lim_{m\to\infty} \bigvee_{n=-m}^{0} \sigma^n \delta_0$ se deduce que*

$$(1-\epsilon)\,\mu_P(C) \leq \mathbb{E}_{\mu_P}\left(1_C \mid \sigma^{-n}\left(\bigvee_{n=-m}^{0} \sigma^n \delta_0\right)\right) \leq (1+\epsilon)\,\mu_P(C)$$

*y $(1-\epsilon)\,\mu_P(C) \leq \mathbb{E}_{\mu_P}(1_C \mid \sigma^{-n}\mathcal{A}) \leq (1+\epsilon)\,\mu_P(C)$. Haciendo $\epsilon \searrow 0$ se deduce que $\mathbb{E}_{\mu_P}(1_C \mid \mathcal{A}^-) = \mu_P(C)$ en todo $C$. Concluimos $\mathbb{E}\mu_P(1_C \mid \mathcal{A}^-) = \mu_P(C), \forall C \in \mathcal{B}$. En particular, como $\mathcal{A}^- \subseteq \mathcal{B}$ para todo $C \in \mathcal{A}^-$se tiene*

$$1_C = \mathbb{E}_{\mu_P}\left(1 \mid \mathcal{A}^-\right) = \mu_P(C) \wedge \mu_p(C) = 0 \text{ ó } 1, \forall C \in \mathcal{A}^- \text{ de donde } \mathcal{A}^- \subseteq \mathcal{N}.$$

♠

***Observación:*** *Como el shift de Bernoulli $\mu_\pi$ con $\pi > 0$ es un caso particular del shift de Markov irreducible y aperiódico, concluimos que los shifts de Bernoulli invertibles son $K-$sistemas.*

*W. Parry, Topics in Ergodic Theory, Cambridge University Press 1981.*

# Bibliografía


*[1] K. Petersen. Ergodic Theory, Cambridge Studies in Advanced Mathematics 2, 1983.*

*[2] M. Denker, C. Grillenberger, K. Sigmund, Ergodic Theory on Compact Spaces, Lecture Notes in Mathematics 527, Springer 1976.*

*[3] D. Lind, B. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge University Press 1995.*

*[4] P. Walters, An Introduction to Ergodic Theory, Graduate Texts in Mathematics 79, Springer-Verlag 1982.*