Entropías

Introducción

La noción de entropía tiene su origen (como muchas otras nociones en matemática) en la física, específicamente en la termodinámica de fines del siglo XIX (vea [15]). Gracias al desarrollo de las probabilidades y la teoría de la información, poco a poco fue adentrándose en diversas áreas de la matemática, al punto que hoy en día se puede afirmar que, de entre los conceptos relativamente modernos de la física, uno de los que ha sido mejor asimilado por las matemáticas es precisamente el de la entropía.

La presentación que haremos dista mucho de ser una revisión de todos los aspectos ligados a la entropía. En efecto, a lo largo de estas notas adoptaremos en general una perspectiva "dinámica" del concepto. De manera más precisa, nos interesaremos en sistemas evolutivos que presentan un comportamiento difícil de predecir. En este contexto, la *entropía* corresponde a un parámetro que puede ser asociado de manera natural a una amplia gama de sistemas, permitiendo "medir" el grado de caoticidad de ellos: a sistemas más complejos se les asocia una mayor entropía, y los sistemas "equivalentes" tienen la misma entropía. Los sistemas de entropía nula corresponden así a sistemas relativamente simples.

1 La función $s\log(s)$

En esta sección explicaremos la naturalidad del uso de la función $\mathcal{H}(s)=-s\log(s)$ para medir la predicibilidad de un fenómeno.

Supongamos que un experimento pueda dar n resultados diferentes $A_1,\ldots,A_n$. A dicho experimento nos gustaría asociarle un parámetro (a saber, una *entropía*) que permita medir el *grado de incerteza* de sus posibles resultados. Si el experiemento está modelado en un espacio de probabilidad del que los A_i constituyen una partición $\mathcal{P}$, entonces quisiéramos que dicho parámetro $H(\mathcal{P}) = H(p_1,\ldots,p_n)$ dependiese sólo de las probabilidades p_i de cada resultado A_i, y no del experimento en sí. Otras propiedades naturales se imponen. Por ejemplo, si la probabilidad de un evento A_i es total, entonces no existe incerteza. Cuando no exista suceso de probabilidad igual a 1, nos gustaría que el grado de impredecibilidad fuese positivo. En resumen,

$$H(p_1,\ldots,p_n)=0 \quad \text{si y sólo si} \quad p_i=1 \quad \text{para algún} \quad p_i. \tag{1}$$

Puesto que H depende sólo de las probabilidades asociadas a los resultados del experimento, ella debiese ser una función *simétrica*, es decir, invariante bajo permutación de las coordenadas:

$$H(p_1,\ldots,p_i,\ldots,p_j,\ldots,p_n) = H(p_1,\ldots,p_j,\ldots,p_i,\ldots,p_n). \tag{2}$$

Si un resultado tiene mayor probabilidad de ocurrir que otro, entonces la incerteza no debiera ser maximal, pues dicho resultado debiese aparecer con mayor frecuencia (es decir, contamos *a priori* con cierta información

relevante sobre el experimento). Por lo tanto, el grado de incerteza debe ser máximo cuando los resultados están equidistribuidos, lo que justifica la siguiente propiedad:

$$H(p_1,\dots,p_n) \quad \text{asume su valor máximo en} \quad (p_1,\dots,p_n)=(1/n,\dots,1/n). \tag{3}$$

Si uno de los resultados del experimento tiene probabilidad nula de producirse, entonces podemos "descartar" dicho resultado, lo que se traduce en que

$$H(p_1,\dots,p_n,0)=H(p_1,\dots,p_n). \tag{4}$$

Las propiedades anteriores consideran sólo un experimento, pero si realizamos dos experiencias simultáneas, nos gustaría relacionar de alguna manera sus grados de incerteza. Claro está, si dichas experiencias son independientes, entonces la entropía de la "experiencia conjunta" debiese ser la suma de las originales. Supongamos ahora que tenemos dos experiencias no necesariamente independientes modeladas en un mismo espacio de probabilidad y cuyos posibles resultados quedan plasmados en particiones $\mathcal{P}_1=\{A_1,\dots,A_n\}$ y $\mathcal{P}_2=\{B_1,\dots,B_m\}$. La partición asociada al experiemento conjunto, denotada por $\mathcal{P}_1\vee\mathcal{P}_2$, es aquélla cuyos elementos son de la forma $A_i\cap B_j$. Quisiéramos entonces que fuese válida una fórmula del tipo

$$H(\mathcal{P}_1\vee\mathcal{P}_2)=H(\mathcal{P}_1)+H(\mathcal{P}_2/\mathcal{P}_1), \tag{5}$$

donde $H(\mathcal{P}_2/\mathcal{P}_1)$ correspondiera a una *entropía condicional* de $\mathcal{P}_2$ dada $\mathcal{P}_1$. En términos probabilísticos, la igualdad precedente se interpretaría diciendo que la entropía de un experimento conjunto es la suma de la entropía del primero con la entropía del segundo conociendo el resultado del primero. La definición más natural de entropía condicional correspondería así a la media de las entropías asociadas a las distribuciones condicionadas, es decir

$$H(\mathcal{P}_2/\mathcal{P}_1)=\sum_{i=1}^{n} p(A_i)\,H\left(\frac{p(B_1\cap A_i)}{p(A_i)},\dots,\frac{p(B_m\cap A_i)}{p(A_i)}\right).$$

Para resumir la discusión anterior, para cada $n\in\mathbb{N}$ consideremos $\Delta_n=\{(p_1,\dots,p_n): p_i\geq 0, \sum p_i=1\}$, y definamos $\Delta=\cup_n\Delta_n$. Buscamos entonces una función $H:\Delta\to\mathbb{R}$ tal que:

(i) $H(p_1,\dots,p_n)\geq 0$, con la igualdad si y sólo si algún p_i es igual a 1;

(ii) la restricción de H a cada Δ_n es invariante por cambios de orden de las coordenadas;

(iii) sobre cada Δ_n, la función H se maximiza en el punto $(1/n,\dots,1/n)$;

(iv) se tiene la igualdad $H(p_1,\dots,p_n,0)=H(p_1,\dots,p_n)$ para todo $(p_1,\dots,p_n)$;

(v) si $\mathcal{P}_1=\{A_1,\dots,A_n\}$ y $\mathcal{P}_2=\{B_1,\dots,B_m\}$ son dos particiones de un mismo espacio de probabilidades, entonces

$$H(\mathcal{P}_1\vee\mathcal{P}_2)=H(\mathcal{P}_1)+H(\mathcal{P}_2/\mathcal{P}_1).$$

La solución a este problema es esencialmente única, de acuerdo al siguiente teorema fundamental, cuya demostración hemos tomado de [14]. Para evitar confusiones, convengamos en la igualdad $0\log(0)=0$ (el lector notará rápidamente que, de acuerdo a (4), ésta es la convención natural).

Teorema 1.1. *Sea H una función que satisface las propiedades anteriores. Si la restricción de H a cada Δ_n es continua, entonces existe una constante $c > 0$ tal que para todo $(p_1, \ldots, p_n) \in \Delta$ se tiene*

$$H(p_1, \ldots, p_n) = c \sum_{i=1}^{n} p_i \log(1/p_i).$$

Demostración. Consideremos la función $U : \mathbb{N} \to \mathbb{R}$ definida por $U(n) = H(1/n, \ldots, 1/n)$. Comenzaremos probando que existe una constante $c > 0$ tal que $U(n) = c \log(n)$ para todo $n \in \mathbb{N}$. El lector reconocerá la similitud entre el argumento presentado a continuación y aquél que permite probar que toda función definida en el conjunto de los números naturales y que es positiva, multiplicativa y creciente, es necesariamente de la forma $n \mapsto n^c$ para algún $c \in \mathbb{R}$.

Por (iii) y (iv) tenemos

$$U(n) = H(1/n, \ldots, 1/n, 0) \leq H(1/(n+1), \ldots, 1/(n+1)) = U(n+1),$$

es decir, U es una función no decreciente. Afirmamos que para todo k, n en $\mathbb{N}$ se cumple

$$U(n^k) = k\, U(n). \tag{6}$$

Para verificar esto, consideremos k experimentos independientes entre sí que tengan resultados equidistribuidos $(1/n, \ldots, 1/n)$. La distribución del experimento conjunto es entonces $(1/n^k, \ldots, 1/n^k)$, y la propiedad (iv) implica que

$$U(n^k) = H(1/n^k, \ldots, 1/n^k) = \sum_{i=1}^{k} H(1/n, \ldots, 1/n) = k\, H(1/n, \ldots, 1/n) = k\, U(n).$$

Consideremos ahora dos enteros positvos arbitrarios m y n. Para cada $k \in \mathbb{N}$ existe un único $k' = k'(k) \in \mathbb{N}$ tal que

$$m^{k'} < n^k \leq m^{k'+1}.$$

Observe que estas desigualdades implican

$$\frac{k'}{k} \leq \frac{\log(n)}{\log(m)} \leq \frac{k'}{k} + \frac{1}{k},$$

y por lo tanto

$$\lim_{k \to \infty} \frac{k'}{k} = \frac{\log(n)}{\log(m)}.$$

Por otra parte, la monotonicidad de la función U y la desigualdad (6) implican que

$$k'\, U(m) \leq k\, U(n) \leq (k'+1)\, U(m),$$

por lo que

$$\left| \frac{U(n)}{U(m)} - \frac{\log(n)}{\log(m)} \right| \leq \frac{1}{k}.$$

Pasando al límite cuando k tiende al infinito obtenemos entonces

$$\frac{U(n)}{\log(n)} = \frac{U(m)}{\log(m)}.$$

Como m y n son enteros positivos arbitrarios, la expresión anterior es constante (y positiva), es decir que existe $c > 0$ tal que $U(n) = c\log(n)$ para todo $n \in \mathbb{N}$, como queríamos probar.

Fijemos ahora n números racionales $p_i = q_i/q$ tales que $\sum p_i = 1$ (cada q_i es un entero positivo). Queremos probar que

$$H(p_1, \ldots, p_n) = c\sum_{i=1}^{n} p_i \log(1/p_i). \tag{7}$$

Para ello, consideremos un experimento a n resultados A_i de probabilidad p_i. Imaginemos ahora otro experimento dependiente del primero consistente de q eventos $B_1, \ldots, B_q$ tales que:

- dichos eventos pueden ser distribuidos en n grupos conteniendo respectivamente $q_1, q_2, \ldots, q_n$ elementos;

- si el evento A_i tuvo lugar en el primer experimento, entonces en el segundo sólo pueden ocurrir los eventos del i-ésimo grupo, cada uno con la misma probabilidad $1/q_i$.

Las condiciones impuestas determinan únicamente la entropía condicional:

$$H(\mathcal{P}_2/\mathcal{P}_1) = \sum_{i=1}^{n} p_i\, H(1/q_i, \ldots, 1/q_i) = c\sum_{i=1}^{n} p_i \log(q_i) = c\sum_{i=1}^{n} p_i \log(p_i) + c\log(q).$$

Por otra parte, para $i \in \{1, \ldots, n\}$ cada evento de la forma $A_i \cap B_j$ se produce con probabilidad nula o igual a $p_i/q_i = 1/q$, y ésta última situación se produce en exactamente q_i ocasiones. Esto implica que $H(\mathcal{P}_1 \vee \mathcal{P}_2) = c\log(q)$, por lo que la condición (v) nos da

$$H(\mathcal{P}_1) + c\sum_{i=1}^{n} p_i \log(p_i) + c\log(q) = c\log(q),$$

de donde (7) se deduce inmediatamente.

Finalemente, puesto que hemos asumido que la (restricción a cada Δ_n de la) función H es continua, la validez de (7) para entradas racionales implica su validez para todo $(p_1, \ldots, p_n) \in \Delta_n$. □

En lo que sigue consideraremos siempre la normalización según la cual $c = 1$. Una partición (finita) $\mathcal{P} = \{A_1, \ldots, A_n\}$ de un epacio de probabilidad en conjuntos A_i de probabilidad p_i tiene entonces asociada una entropía igual a

$$H(\mathcal{P}) = \sum_{i=1}^{n} p_i \log(1/p_i).$$

La función $\mathcal{H}(s) = -s\log(s)$ satisface propiedades muy interesantes. Sin duda que la más utilizada es su concavidad, así como el hecho que $\mathcal{H}(0) = \mathcal{H}(1) = 0$. Ellas permiten probar (rigurosamente) muchas desigualdades frecuentemente utilizadas en la teoría, como por ejemplo

$$H(\mathcal{P}_2) \leq H(\mathcal{P}_2/\mathcal{P}_1), \tag{8}$$

que conjuntamente con (v) implica

$$H(\mathcal{P}_1 \vee \mathcal{P}_2) \leq H(\mathcal{P}_1) + H(\mathcal{P}_2). \tag{9}$$

Ahora bien, la desigualdad (8) es intuitivamente obvia: al considerar $H(\mathcal{P}_2/\mathcal{P}_1)$ estamos asumiendo el conocimiento del experimento asociado a $\mathcal{P}_1$, mientras que en $H(\mathcal{P}_2)$ no disponemos de tal información. En otras palabras, cada vez que disponemos *a priori* de cierta información, la incerteza será menor. Dejamos a cargo del lector la tarea de "asimilar" mediante argumentos heurísticos análogos (o probar usando desigualdades de convexidad) las siguientes propiedades:

(i) $H(\mathcal{P}_1 \vee \mathcal{P}_2/\mathcal{P}_3) = H(\mathcal{P}_1/\mathcal{P}_3) + H(\mathcal{P}_2/\mathcal{P}_1 \vee \mathcal{P}_3)$;

(ii) $H(\mathcal{P}_1 \vee \mathcal{P}_2/\mathcal{P}_3) \leq H(\mathcal{P}_1/\mathcal{P}_3) + H(\mathcal{P}_2/\mathcal{P}_3)$;

(iii) si cada elemento de $\mathcal{P}_1$ es unión de elementos de $\mathcal{P}_2$, entonces $H(\mathcal{P}_1) \leq H(\mathcal{P}_2)$;

(iv) si cada elemento de $\mathcal{P}_1$ es unión de elementos de $\mathcal{P}_2$, entonces $H(\mathcal{P}/\mathcal{P}_1) \geq H(\mathcal{P}/\mathcal{P}_2)$;

Para mayor información en relación con lo anterior, recomendamos la lectura de [4]. Para cerrar esta sección, presentamos un lema técnico que es constantemente utilizado en teoría ergódica.

Lema 1.2. *Sea $(a_n)_{n\in\mathbb{N}}$ una sucesión de números reales. Suponga que existe una constante $C \geq 0$ tal que*

$$|a_{m+n} - a_m - a_n| \leq C \tag{10}$$

para todo m, n en $\mathbb{N}$. Entonces existe un único $\rho \in \mathbb{R}$ tal que la sucesión $(|a_n - n\rho|)_{n\in\mathbb{Z}}$ es acotada. Dicho valor es igual al límite de la sucesión (a_n/n) cuando n tiende al infinito (en particular, este límite existe).

Demostración. Para cada $n \in \mathbb{N}$ consideramos el intervalo $I_n = \big[(a_n - C)/n, (a_n + C)/n\big]$. Afirmamos que I_{mn} está contenido en I_n para todo m, n en $\mathbb{N}$. En efecto, de (10) se obtiene $|a_{mn} - ma_n| \leq (m-1)C$, de donde se concluye que $a_{mn} + C \leq ma_n + mC$, y por lo tanto

$$\frac{a_{mn} + C}{mn} \leq \frac{a_n + C}{n}.$$

Análogamente se prueba que

$$\frac{a_{mn} - C}{mn} \geq \frac{a_n - C}{n},$$

y estas dos últimas desigualdades implican que $I_{mn} \subset I_n$.

Una aplicación simple de la propiedad de intersección finita muestra que la intersección $I = \cap_{n\in\mathbb{N}} I_n$ es no vacía. Si $\rho \in I$ entonces ρ pertenece a I_n para todo n, de donde se concluye fácilmente que

$$|a_n - \ n\rho| \leq C. \tag{11}$$

Luego, ρ satisface la afirmación del lema. Si $\rho' \neq \rho$ entonces

$$|a_n - n\rho'| = |(a_n - n\rho) + n(\rho - \rho')| \geq n|\rho - \rho'| - C,$$

por lo que $|a_n - n\rho'|$ tiende al infinito. Finalmente, de (11) se deduce que $|\rho - a_n/n| \leq C/n$ para todo n, por lo que $\rho = \lim(a_n/n)$. □

En lo que sigue, nosotros utilizaremos sólo la existencia del límite de la sucesión (a_n/n). Hemos querido sin embargo presentar la versión completa del lema, pues ella es muy importante en el estudio de invariantes de tipo "algebraico" de ciertos sistemas dinámicos (vea por ejemplo [10]).

Ejercicio 1.3. Sea $(a_n)_{n\in\mathbb{N}}$ una *sucesión casi subaditiva*, es decir una sucesión para la cual existe una constante $C \geq 0$ tal que para todo m, n en $\mathbb{N}$ se verifica

$$a_{m+n} \leq a_m + a_n + C.$$

Pruebe directamente la existencia del límite (en $\mathbb{R} \cup \{-\infty\}$) de la sucesión $(a_n/n)_{n\in\mathbb{N}}$. Pruebe además que si (a_n) es subaditiva, es decir si $C = 0$, entonces la sucesión (a_n/n) es decreciente.

2 Entropía medible

Para el estudio de la dinámica de una transformación, es fundamental trabajar con medidas invariantes. Es por ello que el teorema siguiente, debido a Bogoliubov y Krilov, es de vital importancia. Recuerde que si $T : X \to X$ es una transformación medible, la imagen de μ por T es denotada por $T_*(\mu)$ y definida por $T_*(\mu)(A) = \mu(T^{-1}(A))$ para cada subconjunto medible A de X. La medida μ es *invariante* (por T) si se cumple $T_*(\mu) = \mu$.

Teorema 2.1. *Si T es una tranformación continua de un espacio métrico compacto X, entonces existe al menos una medida de probabilidad sobre los boreleanos de X invariante por T.*

Demostración. Recuerde que el espacio de las medidas de probabilidad sobre los boreleanos de un espacio métrico compacto se identifica a la intersección de la esfera unidad con el cono positivo del espacio dual de las funciones continuas: cada $\mu \in Prob(X)$ induce un funcional lineal positivo $L_\mu : C(X) \to \mathbb{R}$, a saber

$$L_\mu(\varphi) = \int_X \varphi(x)\, d\mu(x).$$

Dotado de la topología débil estrella, $Prob(X)$ es entonces un espacio (métrico) compacto. Fijemos uno de sus elementos μ. Para cada $n \in \mathbb{N}$ consideramos la medida de probabilidad μ_n definida por

$$\mu_n = \frac{1}{n}\big(\mu + T_*(\mu) + (T^2)_*(\mu) + \ldots + (T^{n-1})_*(\mu)\big).$$

Por la compacidad de $Prob(X)$, la sucesión (μ_n) contiene una subsucesión convergente (μ_{n_k}). Sea ν el límite de esta subsucesión. Afirmamos que ν es invariante por T. En efecto, la igualdad $T_*(\nu) = \nu$ es equivalente a que, para toda función continua $\varphi : X \to \mathbb{R}$,

$$\int_X \varphi(x)\, d\nu(x) = \int_X \varphi(T(x))\, d\nu(x).$$

Ahora bien, a partir de $T_*(\mu_n) = \mu_n + ((T^n)_*(\mu) - \mu)/n$ se deduce que

$$\begin{aligned}
\int_X \varphi(T(x))\, d\nu(x) &= \int_X \varphi(x)\, dT_*(\nu)(x) \\
&= \lim_{k\to\infty} \int_X \varphi(x)\, dT_*(\mu_{n_k})(x) \\
&= \lim_{k\to\infty} \int_X \varphi(x)\, d\mu_{n_k}(x) + \lim_{k\to\infty} \frac{1}{n_k} \int_X \varphi(x)\, d(T^{n_k})_*(\mu)(x) - \lim_{k\to\infty} \frac{1}{n_k} \int_X \varphi(x)\, d\mu(x) \\
&= \int_X \varphi(x)\, d\nu(x),
\end{aligned}$$

que es la igualdad que queríamos demostrar. □

Denotaremos por $Prob_T(X)$ al espacio de las medidas de probabilidad de un espacio X que son invariantes por una transformación T.

Ejercicio 2.2. Considere la transformación $T : [0,1] \to [0,1]$ definida por $T(0) = 1$ y $T(x) = x/2$ para $x \in]0,1]$.

(i) Pruebe que no existe ninguna medida de probabilidad sobre $[0,1]$ invariante por T.

(ii) Dé un ejemplo de una métrica sobre $[0,1]$ para la cual T sea una transformación continua.

(iii) Concluya que X no puede ser compacto respecto a una métrica que satisface la propiedad en (ii).

Ejercicio 2.3. Sean $(X, \mathcal{A})$ e $(Y, \mathcal{B})$ dos espacios provistos de σ-álgebras. Suponga que existe una biyección medible $\Phi : X \to Y$ cuya inversa es medible. Dada una transformación medible $S : X \to X$, denote por $T : Y \to Y$ la transformación $\Phi \circ S \circ \Phi^{-1}$.

(i) Pruebe que si μ es una medida de probabilidad sobre $(X, \mathcal{A})$ invariante por S, entonces $\nu = \Phi_*(\mu)$ es una medida de probabilidad en $(Y, \mathcal{B})$ invariante por T (recuerde que $\Phi_*(\mu)(B) = \mu(\Phi^{-1}(B))$).

(ii) Pruebe que la aplicación $S : [0,1] \to [0,1]$ definida por

$$S(x) = 2x \quad \text{si} \quad x \in [0, 1/2], \qquad S(x) = 2(1-x) \quad \text{si} \quad x \in [1/2, 1],$$

preserva la medida de Lebesgue (esta transformación es conocida como la "transformación de la carpa", debido a su gráfica).

(iii) Considere el homeomorfismo $\Phi : [0,1] \to [0,1]$ dado por $\Phi(x) = \text{sen}^2(\pi x/2)$. Verifique que se tiene la igualdad

$$T(x) = \Phi \circ S \circ \Phi^{-1}(x) = 4x(1-x).$$

(iv) Concluya que la medida de probabilidad ν dada por

$$d\nu = \frac{dx}{\pi\sqrt{x(1-x)}}$$

es invariante por la transformación (logística) $T : [0,1] \to [0,1]$ definida por $T(x) = 4x(1-x)$.

Ejercicio 2.4. Considere la *aplicación de Gauss* $T :]0,1] \to]0,1]$ definida por $T(x) = \{\frac{1}{x}\}$. Pruebe que la medida μ dada por

$$d\mu = \frac{1}{\log(2)} \cdot \frac{dx}{1+x}$$

es una medida de probabilidad invariante por T.

Sea X un espacio provisto de una medida de probabilidad μ. Dada una partición $\mathcal{P}$ de X en una familia finita de conjuntos medibles, denotamos por $H(\mathcal{P}) = H(\mathcal{P}, \mu)$ su entropía respecto a μ, es decir

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

Si T es una transformación de X que preserva μ, para cada $n \in \mathbb{N}$ denotamos $\mathcal{P}_n = \mathcal{P} \vee \ldots \vee T^{-(n-1)}(\mathcal{P})$ la partición cuyos elementos son conjuntos de la forma $A_{i_1} \cap T^{-1}(A_{i_2}) \cap \ldots \cap T^{-(n-1)}(A_{i_{n-1}})$, donde los A_{i_j}

son elementos de la partición original $\mathcal{P}$. Afirmamos que la sucesión $(H(\mathcal{P}_n))_{n\in\mathbb{N}}$ es subaditiva. En efecto, por (9) tenemos

$$\begin{aligned} H(\mathcal{P}_{m+n}) &= H(\mathcal{P}\vee\ldots\vee T^{-(m+n-1)}(\mathcal{P})) \\ &\leq H(\mathcal{P}\vee\ldots\vee T^{-(m-1)}(\mathcal{P})) + H(T^{-m}(\mathcal{P})\vee\ldots\vee T^{-(m+n-1)}(\mathcal{P})) \\ &= H(\mathcal{P}\vee\ldots\vee T^{-(m-1)}(\mathcal{P})) + H(\mathcal{P}\vee\ldots\vee T^{-(n-1)}(\mathcal{P})) \\ &= H(\mathcal{P}_m) + H(\mathcal{P}_n). \end{aligned}$$

La subaditividad de la sucesión $(H(\mathcal{P}_n))_{n\in\mathbb{N}}$ permite entonces definir la entropía de T respecto a $\mathcal{P}$ por

$$h(T,\mu,\mathcal{P}) = \lim_{n\to\infty}\frac{H(\mathcal{P}_n)}{n} = \inf_{n>0}\frac{H(\mathcal{P}_n)}{n}.$$

La *entropía* (medible) de T respecto a μ es el supremo de los valores de la entropía respecto a diferentes particiones finitas, es decir

$$h(T,\mu) = \sup_{\mathcal{P}} h(T,\mu,\mathcal{P}).$$

Ejercicio 2.5. Pruebe que a lo largo de la definición precedente, es posible reemplazar las particiones finitas por particiones de entropía finita, sin alterar el valor final de la entropía de la transformación.

De la definición se concluye inmediatamente que la entropía respecto a una medida es invariante por conjugaciones medibles: si $S : X \to X$ preserva $\mu \in Prob(X)$ y $\Phi : X \to Y$ es una bijección medible con inversa medible, entonces para la transformación $T = \Phi\circ S\circ\Phi^{-1} : Y \to Y$ se cumple

$$h(T,\Phi_*(\mu)) = h(S,\mu).$$

Ejercicio 2.6. Generalizando lo anterior, pruebe que si S y T son *semiconjugadas*, es decir si existe $\Phi : X \to Y$ medible (no necesariamente invertible) tal que $\Phi\circ S = T\circ\Phi$, entonces se tiene la desigualdad de monotonicidad

$$h(T,\Phi_*(\mu)) \leq h(S,\mu). \tag{12}$$

Decir que $A_{i_1}\cap T^{-1}(A_{i_2})\cap\ldots\cap T^{-(n-1)}(A_{i_{n-1}})$ es el elemento de $\mathcal{P}_n$ que contiene a $x \in X$ es equivalente a prescribir (hasta el orden $n-1$) el itinerario de x respecto a $\mathcal{P}$:

$$x \in A_{i_1}, \quad T(x) \in A_{i_2}, \quad \ldots \quad T^{n-1}(x) \in A_{i_{n-1}}.$$

Consideremos ahora la aplicación T como un "experimento" cuyos resultados son leídos usando sólo los elementos de la partición prescrita $\mathcal{P}$. Si bien los experimentos repetidos $T,\ldots,T^n$ no son independientes entre sí, ellos pueden comportarse con cierta independencia cuando sus resultados son leídos respecto a la misma partición $\mathcal{P}$. Si esto ocurre, $H(\mathcal{P}_n)$ crece linealmente, y por lo tanto $h(T,\mathcal{P}) > 0$.

El cálculo de la entropía no siempre es sencillo, pues de acuerdo a la definición debemos tener simultáneamente en consideración todas las posibles particiones finitas medibles del espacio X. Sin embargo, las cosas se simplifican cuando existe una partición *generadora*, es decir una partición $\mathcal{P}$ tal que la σ-álgebra generada por la reunión de las $\mathcal{P}_n$ coincide con la σ-álgebra original de X. El importantísimo resultado siguiente fue obtenido a Kolmogorov y Sinai. Una demostración puede ser hallada en [19] o en [24].

Teorema 2.7. *Si $\mathcal{P}$ es una partición generadora de entropía finita, entonces $h(T,\mu) = h(T,\mu,\mathcal{P})$.*

Ejemplo 2.8. Sobre el espacio de sucesiones infinitas $X = \{0,1\}^{\mathbb{N}}$, considere la medida de Bernoulli equilibrada μ invariante por el desplazamiento $T : X \to X$ definido por $T(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$ (recuerde que la medida de Bernoulli equilibrada es aquélla que asigna a cada cilindro de longitud n una masa igual a $1/2^n$). La partición $\mathcal{P} = \{X_0, X_1\}$ de X dada por

$$X_i = \{(x_1, x_2, \ldots) : x_1 = i\}, \qquad i \in \{0,1\},$$

es generadora. En efecto, los elementos de $\mathcal{P}_n$ corresponden exactamente a los cilindros de longitud n, y la σ-álgebra sobre X es justamente aquélla generada por los cilindros (de longitud finita). Del teorema 2.7 concluimos que

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

Ahora bien, cada elemento de $\mathcal{P}_n$ tiene masa igual a $1/2^n$, y como $\mathcal{P}_n$ contiene 2^n cilindros, su entropía es igual a $-\log(1/2^n) = n\log(2)$. A partir de la definición obtenemos $h(T,\mu,\mathcal{P}) = \log(2)$, por lo que la entropía de T respecto a μ es igual a $\log(2)$.

Ejercicio 2.9. De manera más general, verifique que la entropía del desplazamiento en el espacio de sucesiones a k símbolos respecto a la medida de Bernoulli es igual a $\log(k)$.

Dadas una transformación medible $T : X \to X$ y una partición de entropía finita $\mathcal{P}$ de X, para cada $x \in X$ y cada $n \in \mathbb{N}$ designamos por $\mathcal{P}_n(x)$ el elemento de $\mathcal{P}_n$ que contiene a x. La entropía local de T en x (respecto a $\mathcal{P}$ y $\mu \in Prob_T(X)$) se define por

$$h_x(T,\mu,\mathcal{P}) = \lim_{n\to\infty} -\frac{\log(\mu(\mathcal{P}_n(x)))}{n}. \tag{13}$$

Obviamente, esta definición es pertinente sólo si el límite correspondiente existe. Sin embargo, esto ocurre para casi todo punto, como queda estipulado en la primera parte del importantísimo teorema siguiente, debido a Shannon, Mc Millan y Breiman.

Teorema 2.10. *La convergencia* (13) *se produce en casi todo punto $x \in X$ y en $L^1(X,\mu)$, y se tiene la igualdad*

$$h(T,\mu,\mathcal{P}) = \int_X h_x(T,\mu,\mathcal{P}). \tag{14}$$

Observe que la aplicación $x \mapsto h_x(T,\mu,\mathcal{P})$ es invariante por T. Luego, si T es ergódica respecto a μ, entonces ella es constante c.t.p., siendo el valor de dicha constante igual a $h(T,\mu,\mathcal{P})$.

Ejercicio 2.11. Considere el desplazamiento en el espacio de sucesiones infinitas a k símbolos dotado de una medida de Bernoulli no necesariamente equilibrada (es decir, las probabilidades correspondientes a símbolos diferentes pueden ser distintas). Usando la *ley de los grandes números* de Borel, pruebe que la igualdad $h(T,\mu) = h_x(T,\mu,\mathcal{P})$ se verifica en casi todo punto $x \in X$ (considerando la partición inicial $\mathcal{P}$ en cilindros de longitud 1).

3 Entropía topológica

Dada una transformación T de un espacio métrico compacto X, para cada $\varepsilon > 0$ y cada $n \in \mathbb{N}$ denotemos por $H(T, n, \varepsilon)$ el número máximo de puntos del espacio X cuyas órbitas por T se ε-separan antes de n iteraciones. De manera más precisa, $H(T, n, \varepsilon)$ es el máximo entero positivo k para el cual existen puntos distintos $x_1, \ldots, x_k$ en X tales que, para cada $i \neq j$, se tiene $dist(T^m(x_i), T^m(x_j)) > \varepsilon$ para algún $m < n$.

Para espacios y transformaciones razonables, el valor de $H(T, n, \varepsilon)$ es finito. Definimos entonces

$$h(T, \varepsilon) = \limsup_{n \in \mathbb{N}} \frac{\log(H(T, n, \varepsilon))}{n}.$$

Observe que el valor de $h(T, \varepsilon)$ crece cuando $\varepsilon > 0$ decrece, lo que permite definir la *entropía topológica* de T por

$$h_{top}(T) = \lim_{\varepsilon \to 0} h(T, \varepsilon) = \sup_{\varepsilon > 0} h(T, \varepsilon).$$

Ejemplo 3.1. En el espacio $X = \{0, 1\}^{\mathbb{N}}$ consideramos la distancia

$$dist(x, y) = \sum_{i \geq 1} \frac{|x_i - y_i|}{2^i}.$$

Dados dos enteros positivos r y n, es fácil verificar que dos puntos $x = (x_1, x_2, \ldots)$ e $y = (y_1, y_2, \ldots)$ están $(1/2^r, n)$-separados por el desplazamiento T si y sólo si existe algún $m < n+r$ para el cual $x_m \neq y_m$. De esto se concluye que $h(T, n, 1/2^r) = 2^{n+r-1}$, por lo que

$$h(T, 1/2^r) = \limsup_{n \to \infty} \frac{\log(2^{n+r-1})}{n} = \log(2),$$

y por lo tanto $h_{top}(T) = \log(2)$.

Ejercicio 3.2. Pruebe que si T es un homeomorfismo de un espacio métrico compacto X, entonces se tiene la igualdad $h_{top}(T) = h_{top}(T^{-1})$.

La entropía topológica mide el grado de divergencia exponencial de las órbitas de una transformación. Su cálculo preciso no siempre es sencillo. Una de las herramientas usadas para él viene dada por el importantísimo *principio variacional*, contenido en el siguiente teorema.

Teorema 3.3. *Si T es una transformación continua de un espacio métrico compacto, entonces*

$$h_{top}(T) = \sup\{h(T, \mu),\ \mu \in Prob_T(X)\}. \tag{15}$$

La demostración de este resultado es técnicamente elaborada y puede ser hallada en [24]. Observe que en él no se estipula la existencia de una medida de probabilidad μ en X para la cual se tenga $h_{top}(T) = h(T, \mu)$. En efecto, un tal *estado de equilibrio* no siempre existe. A continuación presentamos un ejemplo sencillo pero "artificial" ilustrando esta situación.

Ejemplo 3.4. Sea $T_n : X_n \to X_n$ una sucesión de transformaciones continuas entre espacios métricos compactos tales que $h_{top}(T_n) < h_{top}(T_{n+1})$ para todo $n \in \mathbb{N}$ y $\sup_{n\in\mathbb{N}} h_{top}(T_n) < \infty$. Considere el espacio $X = \oplus X_n \cup \{\omega\}$ obtenido como la compactificación de Alexandrof de la reunión disjunta de los X_n. Defina $T : X \to X$ por $T(x) = T_n(x)$ si $x \in X_n$ y $T(\omega) = \omega$. La aplicación T es continua y no admite estado de equilibrio. En efecto, en caso contrario existiría una medida de probabilidad ergódica μ que realiza la igualdad en (15). Ahora bien, si $\mu \in Prob_T(X)$ es ergódica, entonces $\mu(X_n) = 1$ para algún $n \in \mathbb{N}$, o bien $\mu(\{\omega\}) = 1$. El segundo caso implica obviamente que $h(T, \mu) = 0$. El primer caso implica por su parte que $h(T, \mu) = h(T_n, \mu) \leq h_{top}(T_n) < h_{top}(T_{n+1}) \leq h_{top}(T)$, y por lo tanto μ no puede ser un estado de equilibrio. Dejamos la verificación de los detalles a cargo del lector.

No es difícil probar que la no existencia de estado de equilibrio implica que ninguna partición (de entropía finita) puede ser generadora para la transformación (vea [24] o [13]).

Ejercicio 3.5. Usando el principio variacional y la desigualdad (3), reobtenga el resultado del ejemplo 3.1. De manera más general, pruebe que la entropía topológica del desplazamiento en el espacio de las sucesiones a k símbolos es igual a $\log(k)$.

Ejercicio 3.6. Dé dos demostraciones diferentes de la nulidad de la entropía topológica de todo homeomorfismo del círculo.

Observación 3.7. Análogamente al caso de la entropía medible, de la definición se concluye inmediatamente que h_{top} es invariante por conjugaciones topológicas: si $S : X \to X$ es una aplicación continua y $\Phi : X \to Y$ es un homeomorfismo, entonces para la transformación $T = \Phi \circ S \circ \Phi^{-1} : Y \to Y$ se cumple $h_{top}(T) = h_{top}(S)$. Resulta interesante remarcar que el principio variacional permite refinar esta propiedad. En efecto, la relación (15) implica que si una biyección medible (no necesariamente continua) conjuga dos aplicaciones continuas, entonces éstas poseen la misma entropía topológica.

A continuación presentaremos una aplicación sencilla de la igualdad (15), ilustrando así la filosofía de un principio general de la teoría ergódica: para obtener resultados de naturaleza puramente topológica, a veces es necesario utilizar las medidas invariantes. Recordemos que, para una transformación continua T definida sobre un espacio métrico compacto X, el *conjunto no errante* $\Omega(T)$ es aquél constituido por los puntos $x \in X$ tales que, para toda vecindad V de x en X, existe $n \in \mathbb{N}$ tal que $T^n(V) \cap V \neq \emptyset$.

Proposición 3.8. *Si T es una transformación continua de un espacio métrico compacto, entonces la entropía topológica de la restricción de T a su conjunto no errante es igual a la entropía topológica de T.*

Demostración. De la definición se concluye inmediatamente que $\Omega(T)$ es un conjunto cerrado. Luego, por el principio variacional, para la entropía topológica $h(T|_{\Omega(T)})$ de la restricción de T a dicho conjunto se cumple

$$h(T|_{\Omega(T)}) = \sup\{h(T|_{\Omega(T)}, \mu),\ \mu \in Prob_T(\Omega(T)).\}$$

Para probar la proposición, debemos establecer entonces una correspondencia entre las medidas de probabilidad sobre X invariantes por T y aquéllas sobre $\Omega(T)$ invariantes por la restricción de T. Dicha correspondencia viene dada por la afirmación siguiente: si μ pertenece a $Prob_T(X)$, entonces el soporte de μ está contenido en $\Omega(T)$. Para probar esta afirmación, fijemos un punto arbitrario x en el complemento de $\Omega(T)$. Por definición, existe una vecindad V de x tal que $T^n(V) \cap V = \emptyset$ para todo $n \in \mathbb{N}$. De ello se concluye

que los conjuntos $T^{-n}(V)$ son dos a dos disjuntos para $n \geq 0$ (pues si $n > m \geq 0$ y $T^{-n}(V) \cap T^{-m}(V) \neq \emptyset$, entonces $T^{n-m}(V) \cap V \neq \emptyset$). Por la invariancia de μ concluimos que, para todo $n \in \mathbb{N}$,

$$1 \geq \mu\big(\cup_{k=0}^{n-1} T^{-k}(V)\big) = \sum_{k=0}^{n-1} \mu(T^{-k}(V)) = n\,\mu(V),$$

lo cual implica evidentemente que $\mu(V) = 0$. □

De entre los numerosos resultados de dinámica topológica que pueden ser obtenidos mediante métodos ergódicos, uno de los más notables es el que presentamos a continuación, debido a Katok [17].

Teorema 3.9. *Sea T un difeomorfismo de una superficie compacta de clase $C^{1+\tau}$ para algún $\tau > 0$. Si la entropía topológica de T es positiva, entonces T posee una infinidad de puntos periódicos.*

En efecto, bajo las hipótesis del teorema se puede concluir la existencia de una *herradura de Smale* [21] para algún iterado de T.

Ejercicio 3.10. Dada una transformación continua $T : X \to X$, denotemos por $P_n(T)$ al conjunto de los puntos periódicos de T cuyo periodo es igual a n. Uno de los problemas más interesantes de la teoría ergódica es la búsqueda de condiciones que garanticen la igualdad (conocida como "fórmula de Mañé")

$$h_{top}(T) = \lim_{n\to\infty} \frac{\log(|P_n(T)|)}{n}.$$

Verifique que esta fórmula es válida para el desplazamiento en el espacio de sucesiones infinitas de un número finito de símbolos.

4 La entropía en teoría de grupos

Diversas nociones de entropía han sido propuestas para el estudio general de acciones de grupos. Como veremos en las secciones que siguen, ellas deben tener en consideración simultáneamente la dinámica de la acción y la estructura algebraica del grupo. Los tópicos que presentamos son la ventana de entrada a un área de la matemática de gran desenvolvimiento en la actualidad.

4.1 Entropías topológicas

Una acción de un grupo Γ por homeomorfismos de un espacio métrico X es una correspondencia entre cada elemento $g \in \Gamma$ con una transformación continua $\varrho(g)$ de X, de modo que para todo g, h en Γ se cumple

$$\varrho(gh) = \varrho(g) \circ \varrho(h), \qquad \varrho(g^{-1}) = \varrho(g)^{-1}.$$

En otras palabras, $\varrho : \Gamma \to \mathrm{Homeo}(X)$ es un homomorfismo de grupos. Para simplificar, supondremos en lo que sigue que este homomorfismo es inyectivo, y denotaremos a $\varrho(g)$ simplemente por g.

Supongamos que Γ sea finitamente generado y fijemos en él un sistema finito y simétrico de generadores $\mathcal{G}_1 = \{g_1, \ldots, g_m\}$ (el término *simétrico* significa que si $g \in \mathcal{G}_1$ entonces $g^{-1} \in \mathcal{G}_1$). Para cada $g \in \Gamma$

definimos la longitud de g (o la distancia de g al elemento neutro) como el número mínimo de elementos (no necesariamente distintos) de $\mathcal{G}_1$ que son necesarios para representar a g. De manera más precisa,

$$\text{long}(g) = dist(e, g) = \min\{n \in \mathbb{N} : g = g_{i_1} g_{i_2} \, g_{i_n}, \quad g_{i_j} \in \mathcal{G}_1\}.$$

La distancia entre dos elementos arbitrarios del grupo queda homogéneamente definida mediante la igualdad $dist(g, h) = dist(e, h^{-1}g)$. Los elementos del grupo Γ pueden así ser pensados como los vértices de un grafo, conocido como *grafo de Cayley* (asociado a $\mathcal{G}_1$). La inclusión de Γ (provisto de la métrica $dist$) en el grafo de Cayley (dotado de la métrica simplicial) es una isometría. Denotaremos por $B(e, n)$ al conjunto de elementos de Γ de longitud menor o igual a n (es decir, al conjunto de los vértices contenidos en la bola cerrada cuyo centro es el elemento neutro y cuyo radio en igual a n en el grafo de Cayley). La esfera de radio n correpondiente será denotada simplemente por S_n.

Inspirándose en la definición de la sección (3), dados $\varepsilon > 0$ y $n \in \mathbb{N}$ denotemos por $H(\Gamma, n, \varepsilon)$ la cantidad máxima de puntos de X que están ε-separados por algún elemento de $B(e, n-1)$. Definimos entonces

$$h(\Gamma, \varepsilon) = \limsup_{n\to\infty} \frac{\log(H(\Gamma, n, \varepsilon))}{2n} \tag{16}$$

El valor de $h(\Gamma, \varepsilon)$ aumenta cuando $\varepsilon > 0$ decrece. Definimos entonces la *entropía topológica* (de la acción de Γ) por

$$h(\Gamma) = \lim_{\varepsilon\to 0} h(\Gamma, \varepsilon) = \sup_{\varepsilon>0} h(\Gamma, \varepsilon).$$

Debe tenerse siempre en cuenta que el valor de $h(\Gamma)$ no es inherente a la estructura algebraica del grupo, sino que depende de la acción considerada. Ella depende también del sistema de generadores elegido. Sin embargo, el lector verificará sin dificultad que si la entropía de la acción es positiva para un sistema (finito) de generadores, entonces ella es positiva respecto a cualquier otro sistema (finito) de generadores.

Ejercicio 4.1. Pruebe que para acciones del grupo de los enteros, la definición precedente coincide con aquélla dada en la sección 3 (gracias a este ejercicio el lector comprenderá la naturalidad del factor 2 en el miembro a derecha de (16)).

Ejemplo 4.2. Considere dos transformaciones de Möbius hiperbólicas h_1 y h_2 cuyos conjuntos de puntos fijos sean disjuntos. Designemos por p (resp. p') el punto fijo repulsor (resp. atractor) de h_1. Cambiando h_2 por h_2^{-1} si es necesario, podemos suponer que el punto $q = h_2(p)$ pertenece al intervalo $]p, p'[$ (respecto a la orientación canónica del círculo). Fijemos $\delta > 0$ tal que $p + \delta < q$ y $h_2([p, p + \delta]) \subset]q, p'[$, y consideremos un entero positivo suficientemente grande k de modo que $h_1^{-k}(h_2(p + \delta)) < q$. Si en Γ fijamos un sistema de generadores que contiene a $g_1 = h_1^{-k}$ y $g_2 = h_2$ (así como a sus inversos), entonces para todo $n \in \mathbb{N}$ y todo $\varepsilon > 0$ menor que $\min\{\delta, dist(p + \delta, q), dist(q, g_2(p + \delta))\}$ se verifica

$$H(\Gamma, 2n, \varepsilon) \geq 2^{n+2} \tag{17}$$

(en particular, la entropía topológica de la acción es positiva, pues está minorada por $\log(2)/2$). En efecto, si denotamos $I_1 = [p, p + \delta]$ y $I_2 = g_2(I)$, entonces podemos definir

$$I_{1,1} = g_1(I_1), \qquad I_{1,2} = g_1(I_2),$$

$$I_{2,1} = g_2(I_{1,1}), \qquad I_{2,2} = g_2(I_{1,2}).$$

De manera más general, si los intervalos $I_{i_1,\dots,i_{n-1}}$ han sido definidos para cada $(i_1,\dots,i_{n-1}) \in \{1,2\}^{n-1}$, hacemos

$$I_{1,i_1,\dots,i_{n-1}} = g_1(I_{i_1,\dots,i_{n-1}}), \qquad I_{2,i_1,\dots,i_{n-1}} = g_2(I_{1,i_1,\dots,i_{n-1}}).$$

Los intervalos $I_{i_1,\dots,i_n}$ corresponden a aquéllos aparecen en la n-ésima etapa de la construcción de un conjunto de Cantor (observe sin embargo que nosotros necesitamos de $2(n-1)$ transformaciones para generar cada uno de ellos a partir de I_1 e I_2). Considerando los inversos de los elementos que originan dichos intervalos, se comprueba fácilmente que sus extremos son puntos que están $(2n,\varepsilon)$-separados por la acción. La desigualdad (17) se deduce entonces del hecho que existen 2^n intervalos en la n-ésima generación (y cada uno de ellos posee dos extremidades).

Resulta evidente de la definición y del ejemplo precedente que si el grupo Γ es "grande", entonces es más fácil obtener acciones de entropía positiva. Las nociones a continuación resultan por lo tanto naturales; ellas fueron introducidas por Milnor en [18].

Definición 4.3. Dados un grupo Γ y un sistema (finito y simétrico) de generadores $\mathcal{G}_1 = \{g_1,\dots,g_m\}$, la *función de crecimiento* $L = L_{\mathcal{G}_1} : \mathbb{N} \to \mathbb{N}$ es aquélla que asocia a cada $n \in \mathbb{N}$ el cardinal $L(n)$ de la bola $B(e,n)$ de centro e y radio n en el grafo de Cayley correspondiente. El grupo Γ tiene crecimiento polinomial si existe un polinomio P tal que $L(n) \leq P(n)$ para todo $n \in \mathbb{N}$. Si tal es el caso, el grado mínimo de un polinomio verificando dicha propiedad es llamado el *grado de crecimiento* de Γ. Se dice que Γ tiene crecimiento exponencial (resp. subexponencial) si $L(n)^{1/n}$ converge a un límite mayor que (resp. igual a) 1 (observe que por el lema 1.2, la expresión $L(n)^{1/n}$ es convergente; denotaremos por $c(\Gamma) = c(\Gamma,\mathcal{G}_1)$ el límite correspondiente).

Observe que las nociones de crecimiento exponencial, subexponencial o polinomial, son invariantes bajo cambio de sistema (finito) de generadores (a pesar de que el valor de $c(\Gamma,\mathcal{G}_1)$ depende de $\mathcal{G}_1$).

Ejercicio 4.4. Pruebe que si un grupo tiene crecimiento polinomial de grado k respecto a un sistema (finito) de generadores, entonces ocurre lo mismo respecto a cualquier otro sistema (finito) de generadores.

Observación. Un célebre teorema de Gromov estipula que un grupo finitamente generado es de crecimiento polinomial si y solamente si él contiene un subgrupo nilpotente de índice finito (vea [12]).

Como hemos visto en el ejemplo 4.2, si un grupo de homeomorfismos de S^1 posee un elemento con un punto fijo topológicamente contractivo cuyo dominio de atracción contiene al menos un elemento de la órbita del punto fijo, entonces la entropía topológica de la acción del grupo en cuestión es positiva (para el lector interesado no en la entropía sino que en el crecimiento de grupos de difeomorfismos, recomendamos la lectura de [20]). La recíproca de esta última afirmación es válida para grupos de diffeomorfismos de clase $C^{1+\tau}$ de S^1. Esto es una consecuencia de un resultado mucho más general (demostrado originalmente en clase C^2), debido a Ghys, Langevin y Walzac [11]. Una presentación de fácil acceso del teorema a continuación aparece en [5].

Teorema 4.5. *Si una foliación de codimensión 1 y transversalmente de clase $C^{1+\tau}$ (con $\tau > 0$) tiene entropía positiva, entonces ella posee hojas de tipo resorte hiperbólico.*

Algunas aclaraciones son necesarias. Primeramente, recuerde que en una foliación no corresponde precisamente a una acción de grupo. Sin embargo, si fijamos un sistema (completo) de transversales, entonces

las aplicaciones de holonomía constituyen un *pseudo-grupo* (es decir que las aplicaciones en cuestión sólo están definidas localmente). La noción de entropía se generaliza directamente para pseudo-grupos, y esta definición generalizada es la que consideramos en el teorema citado más arriba. Una hoja es *hiperbólica* si la holonomía asociada a un *lazo no homotópicamente trivial* en ella fija un punto de la transversal determinando un elemento del pseudo-grupo de holonomía con un punto fijo hiperbólico. Si la hoja en cuestión intersecta a la transversal dentro del dominio de atracción de dicho punto, entonces estamos en presencia de una hoja que se acumula sobre sí misma, o más precisamente de una *hoja resorte*

4.2 El problema de la medida invariante

El teorema de Bogoliubov y Krylov, el cual implica en particular que para todo homeomorfismo de un espacio métrico compacto existe (al menos) una medida de probabilidad invariante, no es válido para el caso general de acciones de grupos. Presentamos a continuación dos ejemplos sencillos; el segundo de ellos (que es una generalización del primero) fue ideado por Furstenberg y reviste especial interés (vea [25] para una discusión más detallada).

Ejemplo 4.6. El grupo de homeomorfismos del círculo del ejemplo 4.2 no preserva ninguna medida de probabilidad sobre S^1. Dejamos la verificación de esto a cargo del lector.

Ejemplo 4.7. La acción natural del grupo $\mathrm{SL}(n,\mathbb{R}) = \{M \in \mathrm{M}(n\times n,\mathbb{R}) : \det(M) = 1\}$ sobre el espacio proyectivo $\mathbb{RP}^{n-1}$ no admite medida invariante (donde $n \geq 2$).

Para demostrar esta afirmación, consideremos una sucesión (g_k) de elementos de $\mathrm{SL}(n,\mathbb{R})$ que escape de todo compacto. Cada g_k puede ser representado por una matriz M_k tal que $\|M_k\| = 1$, donde $\|\cdot\|$ es una norma completa en el espacio de matrices $n \times n$. Pasando a una subsucesión podemos suponer que M_k converge a cierta matriz M, la cual será necesariamente no invertible (pues en caso contrario (g_k) no escaparía de los compactos). Sea E la imagen de $\mathbb{R}^{n-1}$ por la aplicación lineal correspondiente a M.

Supongamos que μ sea una medida de probabilidad sobre $\mathbb{RP}^{n-1}$ invariante por $\mathrm{SL}(n,\mathbb{R})$. Dejamos al lector verificar, a partir de la igualdad $M_k(\mu) = \mu$ (válida para todo $k \in \mathbb{N}$), que $M(\mu) = \mu$. Esto último implica que el soporte de μ está contenido en E (identificamos un espacio vectorial a su imagen en $\mathbb{RP}^{n-1}$).

Sea $h \in \mathrm{SL}(n,\mathbb{R})$ tal que $\dim(E \cap h(E)) < \dim(E)$. De la igualdad $h(\mu) = \mu$ se concluye que $\mathrm{sop}(\mu)$ queda incluido en $E \cap h(E)$. Procediendo inductivamente concluimos que μ es una medida cuyo soporte es un espacio vectorial de dimensión uno, es decir, un punto en $\mathbb{RP}^{n-1}$. Sin embargo, esto es imposible, ya que la acción de $\mathrm{SL}(n,\mathbb{R})$ en $\mathbb{RP}^{n-1}$ es transitiva.

La importancia de la existencia de medidas invariantes para acciones de grupos amerita la siguiente definición.

Definición 4.8. Un grupo Γ es *promediable* si toda acción de Γ por homeomorfismos de un espacio métrico compacto admite (al menos) una medida de probabilidad invariante.

Existen muchas caracterizaciones de los grupos promediables. Nosotros nos concentraremos en el caso de grupos finitamente generados. Para ilustrar nuestro problema, tratemos de utilizar la estrategia de la prueba del teorema de Bogoliubov y Krylov para obtener una medida invariante para una acción arbitraria

de un grupo Γ sobre un espacio X. Fijemos un sistema (finito y simétrico) $\mathcal{G}_1 = \{g_1, \ldots, g_m\}$ de generadores de Γ, así como una medida de probabilidad μ sobre X. Para cada $n \in \mathbb{N}$ consideremos la medida

$$\mu_n = \frac{1}{L_{\mathcal{G}_1}(n-1)} \sum_{g \in B(e,n-1)} g_*(\mu).$$

Pasando a una subsucesión tenemos que $\lim_{k\to\infty} \mu_{n_k} = \nu$ para cierta medida de probabilidad ν. El problema que se presenta es que ν no es necesariamente una medida invariante. En efecto, si tratásemos de repetir el argumento de la demostración del teorema 2.1, entonces deberíamos estimar una expresión del tipo

$$\frac{1}{L_{\mathcal{G}_1}(n_k-1)} \sum_{\substack{g \in B(e,n_k-1),\\ g_i \in \mathcal{G}_1}} (g_i g)_*(\mu).$$

Sin embargo, esta expresión no converge necesariamente a cero, pues puede darse el caso en que la cantidad de elementos del conjunto $B(e, n_k) \setminus B(e, n_k - 1)$ sea grande en comparación a $L_{\mathcal{G}_1}(n_k - 1)$. Las siguientes definiciones resultan entonces naturales.

Definición 4.9. Dado un subconjunto $A \subset \Gamma$, definimos el *borde geométrico* de A como el conjunto

$$\partial A = \bigcup_{g \in \mathcal{G}_1} \big(A \Delta gA\big),$$

donde Δ denota la diferencia simétrica de los conjuntos respectivos.

Definición 4.10. Una sucesión de Følner para Γ es una sucesión (A_n) de subconjuntos finitos de Γ tales que

$$\lim_{n\to\infty} \frac{|\partial A_n|}{|A_n|} = 0.$$

Utilizando el argumento de Bogoliubov y Krylov, no es difícil verificar que si Γ admite una sucesión de Følner entonces toda acción de Γ por homeomorfismos de un espacio compacto admite una medida invariante. El panorama completo queda aclarado entonces por el siguiente teorema, debido a Følner [8].

Teorema 4.11. *Un grupo finitamente generado es promediable si y sólo si admite una sucesión de Følner.*

Es importante remarcar que esta caracterización es independiente del sistema de generadores. En efecto, no es difícil verificar que el cuociente de las funciones longitud de un elemento $g \in \Gamma$ con respecto a dos sistemas finitos de generadores está acotado por una constante que depende de ambos sistemas y es independiente de g. Por lo tanto una sucesión de Følner respecto a un sistema origina una sucesión de Følner respecto al otro sistema.

Ejercicio 4.12. Pruebe que si un grupo discreto posee un subgrupo libre a dos generadores, entonces dicho grupo no es promediable.

Ejercicio 4.13. Pruebe que todo grupo abeliano es promediable.

Ejercicio 4.14. Pruebe que la propiedad de promediabilidad es estable por *operaciones elementales*, es decir:

(i) todo subgrupo de un grupo promediable es promediable;

(ii) todo grupo que es límite directo de grupos promediables es promediable;

(iii) el cuociente de un grupo promediable es promediable;

(iv) la extensión de un grupo promediable por otro grupo promediable es promediable.

Concluya que todo grupo virtualmente soluble es promediable. Si tiene problemas para probar estas afirmaciones, vea el capítulo 3 de [25].

Ejercicio 4.15. Pruebe de dos maneras diferentes que todo grupo de crecimiento polinomial es promediable.

Ejercicio 4.16. Generalizando el ejercicio anterior, pruebe que todo grupo finitamente generado y de crecimiento subexponencial es promediable.

La discusión de este capítulo muestra que, para desarrollar una "teoría medible de grupos" que sea interesante, deben ser introducidas nuevas ideas. Una de ellas consiste en considerar acciones (generalmente libres) que *a priori* preservan una medida de probabilidad. Este punto de vista, desarrollado por Connes, Weiss, Popa, Gaboriau y otros, se ha revelado muy fecundo y ha estrechado la teoría de grupos con otras áreas de la matemática, como el álgebra de operadores [6]. Otra perspectiva interesante, y que nosostros introduciremos en las secciones que siguen, consiste en considerar el grupo como un sistema dinámico en sí mismo: la estructura del grupo puede ser leída directamente de su acción por traslaciones (teorema de Cayley), y si asignamos "parámetros de frecuencia" a dichas traslaciones (equivalentemente, para las transiciones entre los elementos), entonces es esperable recuperar parte de esa información algebraica mediante argumentos probabilísticos.

4.3 Caminatas aleatorias y entropía asintótica

Consideremos un grupo (enumerable) Γ provisto de una medida de probabilidad μ. La *caminata aleatoria* sobre Γ siguiendo la ley dada por μ consiste en el movimiento de un elemento a otro en el grupo, de modo que en cada paso la probabilidad $p(g \to h)$ de ir desde g a h es igual a $\mu(g^{-1}h)$. Evidentemente, el "comportamiento estadístico" de la caminata depende en gran medida de la estructura del grupo, pues cada camino cerrado de la caminata corresponde a una relación algebraica en Γ.

Para estudiar la evolución de la caminata, en el espacio de las sucesiones finitas de largo n de elementos de Γ podemos considerar la medida producto $\mu \times \cdots \times \mu$ (n factores). Bajo la acción de la aplicación $\Gamma^n \to \Gamma$ dada por

$$(g_1, g_2, \ldots, g_n) \mapsto g_1 g_2 \cdots g_n,$$

la imagen de dicha medida es denominada la n-ésima *convolución* de μ consigo misma, y denotada por $\mu^{*(n)}$. De manera un poco más general, si μ y ν son medidas de probabilidad sobre Γ, entonces la convolución $\mu * \nu$ es una nueva medida de probabilidad sobre Γ, definida por

$$\mu * \nu(h) = \sum_{fg=h} \mu(f)\, \nu(g).$$

Cuando hablemos de la probabilidad de que cierto suceso que depende de n pasos ocurra en la caminata inducida por μ, estaremos tácitamente pensando en la probabilidad correspondiente dada por $\mu^{*(n)}$. Por

ejemplo, la probabilidad de transición entre g y h en n pasos es $p^{(n)}(g \to h) = \mu^{*(n)}(g^{-1}h)$. Observe que si μ es *simétrica*, es decir si $\mu(g) = \mu(g^{-1})$ para todo $g \in \Gamma$, entonces lo mismo ocurre para la medida $\mu^{*(k)}$ para todo $k \in \mathbb{N}$. En dicho caso se tiene la igualdad $p^{(k)}(g \to h) = p^{(k)}(h \to g)$ para todo g, h en Γ y todo $k \in \mathbb{N}$.

La noción de entropía (asintótica) para caminatas aleatorias sobre grupos fue introducida por Avez en [2]. La idea es considerar la caminata aleatoria como "sistema dinámico", interpretando los posibles estados de esta caminata (tras el número correspondiente de pasos) como los "iterados" del sistema. De manera más precisa, consideremos la entropía $H(\mu)$ de la partición $\mathcal{P}$ de Γ en sus elementos, es decir

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

Para cada $n \geq 1$ definamos

$$H_n(\mu) = H(\mu^{*(n)}) = -\sum_{g \in \Gamma} \mu^{*(n)}(g) \log(\mu^{*(n)}(g)).$$

Observe que $H_n(\mu)$ puede ser igual a infinito si el soporte de μ no es finito. Sin embargo, es fácil verificar que si $H(\mu) = H_1(\mu)$ es finito, entonces $H_n(\mu)$ es finito para todo $n \geq 1$.

Lema 4.17. *Si $H_n(\mu) < \infty$ para todo $n \in \mathbb{N}$ entonces la sucesión $(H_n(\mu))_{n \in \mathbb{N}}$ es subaditiva.*

Demostración. Basta probar la desigualdad

$$H(\mu_1 * \mu_2) \leq H(\mu_1) + H(\mu_2)$$

para dos medidas de probabilidad cualesquiera sobre el mismo grupo Γ. Esta desigualdad puede ser verificada directamente, utilizando para ello la propiedad de concavidad de la función $\mathcal{H}(s) = -s \log(s)$. Para una prueba más conceptual, observemos primeramente que, por definición, la aplicación $\Gamma \times \Gamma \to \Gamma$ dada por $(g, h) \mapsto gh$ envía la medida producto $\mu_1 \times \mu_2$ sobre $\mu_1 * \mu_2$. Por la monotonicidad de la entropía, es decir por (12), tenemos

$$H(\mu_1 * \mu_2) \leq H(\mu_1 \times \mu_2).$$

Por otra parte, se cumple la relación

$$H(\mu_1 \times \mu_2) = H(\mu_1) + H(\mu_2).$$

Las dos relaciones anteriores concluyen la prueba del lema. □

Observación 4.18. El mismo argumento de la demostración precedente permite probar que si $H_1(\mu) < \infty$ entonces $H_n(\mu) \leq nH(\mu) < \infty$ para todo $n \geq 1$. Recíprocamente, si $H(\mu) = \infty$ entonces $H_n(\mu) = \infty$ para todo $n \geq 1$.

Definición 4.19. Si $H(\mu) < \infty$ definimos la entropía de la caminata aleatoria correspondiente a μ por

$$h(\Gamma, \mu) = \lim_{n \to \infty} \frac{H_n(\mu)}{n},$$

y extendemos esta definición por $h(\Gamma, \mu) = \infty$ si $H(\mu) = \infty$.

En términos probabilísticos, $h(\Gamma,\mu)$ es la información promedio de uno de los factores de un producto $h_n = g_1 \cdots g_n$ de n variables aleatorias independientes $g_i \in \Gamma$ con distribución μ.

Ejemplo 4.20. Si $\mu = \tau\mu_1 + (1-\tau)\mu_2$ entonces

$$\begin{aligned}
H(\mu^{*(n)}) &= \sum_{g\in\Gamma} \mathcal{H}\left((\tau\mu_1 + (1-\tau)\mu_2)^{*(n)}(g)\right) \\
&= \sum_{g\in\Gamma} \mathcal{H}\left(\sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} \sum_{h\in\Gamma} \mu_1^{*(i)}(h)\mu_2^{*(n-i)}(h^{-1}g)\right) \\
&\geq \sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} \sum_{g\in\Gamma} \mathcal{H}\left(\sum_{h\in\Gamma} \mu_1^{*(i)}(h)\mu_2^{*(n-i)}(h^{-1}g)\right) \\
&\geq \sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} \sum_{g,h\in\Gamma} \mathcal{H}\left(\mu_1^{*(i)}(h)\mu_2^{*(n-i)}(h^{-1}g)\right) \\
&= \sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} \sum_{g,h\in\Gamma} \left(\mu_1^{*(i)}(h)\mathcal{H}(\mu_2^{*(n-i)}(h^{-1}g)) + \mu_2^{*(n-i)}(h^{-1}g)\mathcal{H}(\mu_1^{*(i)}(g))\right) \\
&= \sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} (H(\mu_1^{*(i)}) + H(\mu_2^{*(n-i)})) \\
&\geq \sum_{i=0}^{n} \binom{n}{i} \tau^i (1-\tau)^{n-i} (ih(\Gamma,\mu_1) + (n-i)h(\Gamma,\mu_2)) \\
&= n(\tau h(\Gamma,\mu_1) + (1-\tau)h(\Gamma,\mu_2)),
\end{aligned}$$

por lo que

$$h(\Gamma,\mu) = \lim_{n\to\infty} \frac{H(\mu^{*(n)})}{n} \geq \tau h(\Gamma,\mu_1) + (1-\tau)h(\Gamma,\mu_2).$$

Luego, si la entropía de Γ respecto a una combinación convexa de dos medidas de probabilidad μ_1 y μ_2 es positiva, entonces su entropía respecto a al menos una de las medidas originales es positiva.

Ejercicio 4.21. Sobre el grupo libre a dos generadores L_2 considere la medida de probabilidad μ simétrica y equidistribuida sobre dichos generadores. Calcule el valor de $h(L_2,\mu)$.

La entropía permite obtener información sobre un invariante probabilístico asociado a un grupo, a saber su borde de Poisson-Furstenberg [9]. De esta manera, ella aparece relacionada con el espacio de las funciones *harmónicas* sobre Γ: si $H(\mu) < \infty$ entonces $h(\mu) > 0$ si y solamente si Γ admite funciones harmónicas acotadas no triviales. Recordemos que una función $\varphi : \Gamma \to \mathbb{R}$ es harmónica (respecto a μ) si ella verifica la *igualdad de la media*, es decir si para todo $g \in \Gamma$ se cumple

$$\varphi(g) = \sum_{h\in\Gamma} \varphi(gh)\mu(h).$$

Lamentablemente, no tenemos espacio suficiente para desarrollar en mayor profundidad este apasionante tema. Nos limitaremos entonces a probar sólo un resultado concreto, a saber la desigualdad que relaciona

el crecimiento, la razón de escape al infinito y la entropía de un grupo. El lector notará cierta similitud entre dicho resultado y el principio variacional estudiado en la sección 3. Para el estudio de otros tipos de "equilibrio" ligados a la entropía, esta vez en geometría diferencial, vea [3].

Definición 4.22. Fijemos un sistema finito y simétrico de generadores en un grupo Γ. Respecto a dicho sistema viene asociada una función distancia *dist*. Si μ es una medida de probabilidad sobre Γ, entonces se define el k-ésimo momento de μ por

$$M_k(\mu) = \sum_{g\in\Gamma} dist(e,g)^k\, \mu(g)$$

Para analizar la relación entre la entropía y el primer momento de una medida, necesitamos del siguiente lema elemental, cuya demostración ha sido tomada de [19].

Lema 4.23. *Si una sucesión de números $\mu_n \in]0,1[$ satisface $\sum_{n\in\mathbb{N}} n\,\mu_n < \infty$, entonces se cumple*

$$\sum_{n\in\mathbb{N}} \mu_n \log(1/\mu_n) < \infty.$$

Demostración. Si I denota el conjunto de los enteros positivos tales que $\log(1/\mu_n) < n$, entonces se tiene

$$\begin{aligned}\sum_{n\in\mathbb{N}} \mu_n \log(1/\mu_n) &= \sum_{n\in I} \mu_n \log(1/\mu_n) + \sum_{n\notin I} \mu_n \log(1/\mu_n)\\ &\leq \sum_{n\in I} n\mu_n + \sum_{n\notin I} \mu_n \log(1/\mu_n).\end{aligned}$$

Observe que si n no pertenece a I entonces $e^{-n} \geq \mu_n$. Usando la desigualdad (válida para todo $t \in]0,1[$)

$$\sqrt{t}\log(1/t) \leq \frac{2}{e},$$

concluimos que

$$\sum_{n\notin I} \mu_n \log(1/\mu_n) \leq \frac{2}{e}\sum_{n\notin I} \sqrt{\mu_n} \leq \frac{2}{e}\sum_{n\notin I} e^{-n/2} < \infty,$$

lo cual finaliza la demostración. □

Proposición 4.24. *Si μ es una medida de primer momento finito, entonces su entropía también es finita.*

Demostración. Sea d la cardinalidad del conjunto prescrito de generadores de Γ. La cardinalidad de cada esfera S_k está superiormente acotada por d^k, por lo que

$$\begin{aligned}-\sum_{g\in S_k} \mu(g)\log(\mu(g)) &= -\mu(S_k)\sum_{g\in S_k} \frac{\mu(g)}{\mu(S_k)} \log\left(\frac{\mu(g)}{\mu(S_k)}\right) - \sum_{g\in S_k} \mu(g)\log(\mu(S_k))\\ &\leq \mu(S_k)\log(d^k) + \mu(S_k)\log(1/\mu(S_k)).\end{aligned}$$

Por la hipótesis $M_1(\mu) < \infty$ y el lema anterior tenemos

$$\sum_{k\in\mathbb{N}} \mu(S_k)\log(1/\mu(S_k)) = M < \infty,$$

de donde concluimos que

$$H(\mu) = -\sum_{k\in\mathbb{N}}\sum_{g\in S_k} \mu(g)\log(\mu(g)) \leq \log(d)\, M_1(\mu) + M. \quad \square$$

Queremos introducir ahora un parámetro, al que llamaremos *razón de escape al infinito*, que permita medir el comportamiento estadístico de la distancia al elemento neutro de un elemento representado por una palabra que es el producto de n generadores. La idea es que, dado que pueden producirse simplificaciones a lo largo de esta palabra, es esperable que dicha distancia sea menor que n.

Definición 4.25. Sea Γ un grupo enumerable provisto de una medida de probabilidad μ de primer momento finito. La *razón de escape al infinito* de Γ con respecto a μ se define por

$$I = I(\mu) = \lim_{n\to\infty} \frac{\mathbb{E}(I_n)}{n}, \tag{18}$$

donde $\mathbb{E}(I_n)$ designa la esperanza de la función $I_n(\omega) = dist(g_1\cdots g_n, e)$ respecto a la medida $\mu^{*(n)}$ (donde $\omega = (g_1, g_2, \ldots) \in \Gamma^{\mathbb{N}}$).

Obviamente, debemos verificar que esta definición es pertinente, es decir que el límite correspondiente siempre existe. Ello es una consecuencia casi directa del teorema ergódico subaditivo. En efecto, denotando por T el desplazamiento en $\Gamma^{\mathbb{N}}$, vemos que la sucesión de funciones I_n satisface de manera evidente la desigualdad

$$I_{m+n}(w) = I_n(w) + I_m(T^n(w)).$$

Por el teoreme ergódico subaditivo, tenemos la convergencia c.t.p. de I_n/n hacia cierta función límite I. Puesto que el cambiar un segmento finito de w no altera el valor de dicho límite, la ley 0-1 de Kolmogorov implica que I es constante c.t.p. Finalmente, siendo siempre I_n/n menor o igual a 1, el teorema de convergencia dominada implica que dicho valor constante es igual a la razón de escape.

El hecho que $I_n(\omega)/n$ converja para casi toda trayectoria ω a la razón de escape al infinito puede ser pensado como la validez de una "ley de los grandes números" en el grupo Γ. Presentamos a continuación algunos ejemplos.

Ejemplo 4.26. Si en $\Gamma = \mathbb{Z}$ consideramos la probabilidad estándar entonces se comprueba fácilmente que la igualdad $I_{2n+1}(g_1\cdots g_{2n+1}) = I_{2n}(g_1\cdots g_{2n}) + 1$ se verifica con probabilidad $\frac{1}{2}\left[1 + \frac{1}{2}\binom{2n}{n}\right]$, mientras que $I_{2n+1}(g_1\cdots g_{2n+1}) = I_{2n}(g_1\cdots g_{2n}) - 1$ se cumple con probabilidad $\frac{1}{2}\left[1 - \frac{1}{2}\binom{2n}{n}\right]$. Mediante la aproximación de Stirling $n! \sim \sqrt{2\pi n}\,(n/e)^n$, se concluye que existen constantes positivas $\bar{c}, c$ tales que

$$\mathbb{E}(I_{2n+1}) = \sum_{j=0}^{n} \frac{1}{2^{2j}}\binom{2j}{j} \leq \sum_{j=0}^{n} \frac{\bar{c}}{\sqrt{j}} \leq c\sqrt{n},$$

por lo que

$$\lim_{n\to\infty} \frac{I_{2n+1}}{2n+1} \leq \lim_{n\to\infty} \frac{c\sqrt{n}}{2n+1} = 0,$$

de donde se obtiene $I = 0$.

Ejemplo 4.27. La razón de escape del grupo libre L_k respecto a la medida de probabilidad simétrica y equidistribuida sobre sus generadores es igual a $(k-1)/k$. Como ya analizamos el caso en que $k=1$, supondremos que k es al menos igual a 2. En tal caso, si $n \geq 1$ entonces con probabilidad $(2k-1)/2k$ se tiene $I(g_1 \cdots g_n g_{n+1}) = I(g_1 \cdots g_n) + 1$, mientras que $I(g_1 \cdots g_n g_{n+1}) = I(g_1 \cdots g_n) - 1$ se cumple con probabilidad $1/2k$. De ello se deduce que

$$\mathbb{E}(I_{n+1}) = \mathbb{E}(I_n) + \frac{2k-1}{2k} - \frac{1}{2k},$$

lo cual implica la igualdad $I = (k-1)/k$ vía un argumento sencillo de suma telescópica.

A continuación expondremos una desigualdad interesantísima que relaciona la razón de escape al infinito con el crecimiento y la entropía de un grupo (vea [23] para mayores detalles).

Teorema 4.28. *Si μ es una medida de probabilidad cuyo soporte es finito y genera a Γ como semigrupo, entonces se tiene la desigualdad*

$$h(\Gamma, \mu) \leq I(\mu)\, c(\Gamma, sop(\mu)).$$

Demostración. Fijemos $\varepsilon > 0$ y denotemos por $X_{\varepsilon,n}$ al conjunto de elementos de Γ a los que se llega en n pasos de la caminata pero cuya distancia al elemento neutro se sitúa en el intervalo $[(1-\varepsilon)nI, (1+\varepsilon)nI)]$. La discusión anterior muestra que $\mu^{*(n)}(X_{\varepsilon,n}) \geq 1-\varepsilon$ para n suficientemente grande. Escribamos la medida $\mu^{*(n)}$ como la combinación convexa de dos medidas de probabilidad $\mu_{1,n}$ y $\mu_{2,n}$, donde $\mu_{1,n}$ es la restricción normalizada de $\mu^{*(n)}$ a $X_{\varepsilon,n}$ y $\mu_{2,n}$ es la restricción normalizada de $\mu^{*(n)}$ al complemento de $X_{\varepsilon,n}$. Una aplicación sencilla de (3) muestra que

$$H(\Gamma, \mu^{*(n)}) \leq (1-\varepsilon) H(\mu_{1,n}) - \log(1-\varepsilon) + \varepsilon \log(|B_n|).$$

Dividiendo por n y pasando al límite obtenemos

$$h(\Gamma, \mu) \leq (1-\varepsilon) \liminf_{n\to\infty} \frac{H(\mu_{1,n})}{n} + \varepsilon\, c(\Gamma, sop(\mu)). \tag{19}$$

Observe ahora que, considerando como conjunto generador a $sop(\mu)$, el conjunto $X_{\varepsilon,n}$ está contenido en la bola $B(e, (1+\varepsilon)nI)$. Puesto que el valor de $H(\mu_{1,n})$ está acotado por $\log(|X_{\varepsilon,n}|)$, a partir de (19) obtenemos

$$h(\Gamma, \mu) \leq (1-\varepsilon) \liminf_{n\to\infty} \frac{\log(|B(e, (1+\varepsilon)nI)|)}{n} + \varepsilon\, c(\Gamma, sop(\mu)).$$

Como el valor de $\log(|B(e, (1+\varepsilon)nI|)/n$ tiende a $(1+\varepsilon)Ic(\Gamma, sop(\mu))$ cuando n tiende al infinito, tenemos

$$h(\Gamma, \mu) \leq (1-\varepsilon^2) Ic(\Gamma, sop(\mu)) + \varepsilon c\,(\Gamma, sop(\mu)).$$

Siendo esta desigualdad válida para todo $\varepsilon > 0$, concluimos finalmente que $h(\Gamma, \mu) \leq I(\mu)\, c(\Gamma, sop(\mu))$. $\square$

Ejercicio 4.29. Pruebe que para el caso de la medida simétrica y equidistribuida sobre los generadores del grupo libre L_k, la desigualdad anterior es una igualdad.

La desigualdad $h(\Gamma,\mu) \leq I(\mu)\, c(\Gamma, sop(\mu))$ implica que si $I(\mu) = 0$ entonces $h(\Gamma,\mu) = 0$. La recíproca de esta última afirmación también es válida, de acuerdo a un interesantísimo resultado de Varopoulos [1, 22].

Para concluir estas notas, enunciamos una versión del teorema de Shannon, Mc Millan y Breiman válido para caminatas aleatorias sobre grupos. El resultado siguiente fue obtenido por Kaimanovich y Vershik en [16] (vea también [7]).

Teorema 4.30. *Si μ es una medida de probabilidad de entropía finita sobre un grupo enumerable Γ, entonces la convergencia*

$$\lim_{n\to\infty} -\frac{\log(\mu^{*(n)}(g_n))}{n} = h(\Gamma,\mu)$$

tiene lugar tanto en casi todo punto $\omega = (g_1, g_2, \ldots) \in \Gamma^{\mathbb{N}}$ como en $L^1(\Gamma^{\mathbb{N}}, \mu^{\mathbb{N}})$

Referencias

[1] Alexopoulos, G. On the mean distance of random walks on groups. *Bull. Sci. Math.* **111** (1987), 189-199.

[2] Avez, A. Entropie des groupes de type fini. *C. R. Acad. Sci. Paris* **275** (1972), 1363-1366.

[3] Besson, G., Courtois, G. & Gallot, S. Entropies et rigidités des espaces localement symétriques de courbure strictement négative. *Geom. Funct. Anal.* **5** (1995), 731-799.

[4] Billingsley, P. *Ergodic theory and information.* Wiley Series in Prob. and Math. Statistics (1965).

[5] Candel, A. & Conlon, L. *Foliations I.* Graduate Studies in Mathematics **23**, American Mathematical Society, Providence (2000).

[6] Connes, A. Nombres de Betti L^2 et facteurs de type Π_1 (d'après D. Gaboriau et S. Popa). *Astérisque* **294** (2004), 321-333.

[7] Derriennic, Y. Quelques applications du théorème ergodique sous-additif. *Astérisque* **74** (1980), 183-201.

[8] Følner, E. On groups with full Banach mean value. *Math. Scand.* **3** (1955), 243-254.

[9] Furstenberg, H. Random walks and discrete subgroups of Lie groups. *Advances in Probability and Related Topics* vol **1** (1971), 1-63.

[10] Ghys, É. Groupes d'homéomorphismes du cercle et cohomologie bornée. *Cont. Math.* **58** (1987), 81-106.

[11] Ghys, É., Langevin, R. & Walczak, P. Entropie géométrique des feuilletages. *Acta Math.* **160** (1988), 105-142.

[12] Gromov, M. Groups of polynomial growth and expanding maps. *Publ. Math. de l'IHES* **53** (1981), 53-73.

[13] Keller, G. *Equilibrium states in ergodic theory.* London Mathematical Society Student Texts **42**, Cambridge University Press (1998).

[14] Khinchin, A. *Mathematical Foundations of Information Theory.* Dover Publ. Math. (1957).

[15] Khinchin, A. *Mathematical Foundations of Statistical Mechanics.* Dover Publ. Math. (1949).

[16] Kaimanovich, V. & Vershik, A. Random walks on discrete groups: boundary and entropy. *Ann. Probab.* **11** (1983), 457-490.

[17] Katok, A. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Publ. Math. de l'IHES* **51** (1980), 137-173.

[18] Milnor, J. A note on curvature and fundamental group. *J. Diff. Geometry* **2** (1968), 1-7.

[19] Mañé, R. *Introdução à teoria ergódica.* Projeto Euclides **14**, IMPA (1983).

[20] Navas, A. A group of diffeomorphisms of the interval with intermediate growth. Prepublicación (2004).

[21] Palis, J. & Takens, F. *Hyperbolicity and sensitive chaotic dynamics at homoclinic bifurcations. Fractal dimensions and infinitely many attractors.* Cambridge Studies in Advanced Mathematics **35**, Cambridge University Press (1993).

[22] Varopoulos, N. Long rate estimates for Markov chains. *Bull. Sci. Math.* **109** (1985), 225-252.

[23] Vershik, A. Dynamic theory of growth in groups: entropy, boundaries, examples. *Russian Math. Surveys* **55** (2000), 667-733.

[24] Walters, P. *An introduction to ergodic theory.* G.T.M. **79**, Springer Verlag (1982).

[25] Zimmer, R. *Ergodic theory of semisimple groups.* Monographs in Mathematics, Birkhäuser (1984).

Andrés Navas

IHES, 35 route de Chartres, 91440 Bures sur Yvette, France (anavas@ihes.fr)

Univ. de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile (andnavas@uchile.cl)