U N I V E R S I D A D   D E   S O N O R A

ESCUELA DE ALTOS ESTUDIOS

E N T R O P I A
Y
D I M E N S I O N

T E S I S

QUE PARA OBTENER EL TITULO DE
LICENCIADO EN MATEMATICAS
PRESENTA

CARLOS SANCHEZ MARTINEZ.

HERMOSILLO, SONORA, MAYO 15 DE 1981.

Dedico esta tesis con cariño y agradecimiento

a "EL", el único que es y al que
con nada podré agradecerle

a mi padre RAFAEL
a mi madre ALICIA

a mi queridisima esposa

DORA LUZ

a mis adorados hijos

CARLOS RAFAEL
HECTOR JOEL Y
HANS OMAR

al M. en C. MARCO ANTONIO VALENCIA ARVIZU

P R E S E N T A C I O N

El presente trabajo se puede dividir en tres partes: En la primera se plantea el concepto de entropía y se estudian algunas de sus propiedades y refinamientos, teniendo como objetivo el proporcionar una introducción al tema, tratando de situarlo en un contexto general.

En la segunda parte se habla del concepto de dimensión para espacios métricos compactos. El concepto de dimensión, a pesar de ser fundamental, generalmente no puede ser abordado con detenimiento en los cursos regulares de la licenciatura en matemáticas. Se habla aquí de las dimensiones Topológica y de Hausdorff.

En la tercera parte se trata de relacionar el concepto de entropía, un concepto relativamente reciente, con el concepto relativamente antiguo de dimensión, esto a través de la noción de dimensión métrica. Se trata de explorar qué tiene que ver la dimensión métrica con la entropía y luego comparar este tipo de dimensión con las dimensiones Topológica y de Hausdorff, para llegar a la conclusión de que, escogiendo una métrica adecuada, los tres tipos de dimensión coinciden.

Quiero manifestar mi agradecimiento al M. en C. Marco Antonio Valencia por la valiosa y desinteresada ayuda que me otorgó en la elaboración de este trabajo. También quiero manifestar mi reconocimiento al señor Germán Valdez V., quien de una forma indirecta contribuyó en mi superación personal. Mi reconocimiento también para el profesor Enrique Valle Flores, quien propuso originalmente el tema.

C A P I T U L O   0

INTRODUCCION

A) Nacimiento de la Teoría Matemática de la Información.

Se llama Teoría de la Comunicación al estudio de todas las partes y procesos del esquema (1).- También se usa el nombre de Teoría de la Información, aunque esta denominación se retringe a veces a únicamente el estudio estocástico de los mensajes, mientras que la Teoría de la Comunicación abarca además modulación, ruido, detección, estadística, filtraje, predicción.

Mensaje Transmitido — Señal Transmitida — Señal Recibida — Mensaje Recibido

FUENTE DE INFORMACION — CODIFICADOR — TRANSMISOR — CANAL — RECEPTOR — DESCODIFICADOR — DESTINATARIO

RUIDO Y DISTORSION

Esquema de un Sistema de Comunicación. (1)

El esquema considerado es tan general que se puede aplicar la Teoría de la Comunicación a dominios muy diferentes, como la Física (Termodinámica, Optica), Biología (Neurofisiología, Genética, Endocrinología), Psicología (Memorización), Lingüística, Criptografía, Reconocimiento de las formas, etc.

Hay que disntinguir en la Teoría de la Información una componente Matemática y una componente de Ingeniería (eléctrica, Mecánica o electrónica). La Teoría Matemática de la Información nació en 1948 con el trabajo del ingeniero-matemático norteamericano Claude E. Shannon, y desde luego es un modelo abstracto que sirve de referencia para los sistemas de comunicación concretos. Por ejemplo, comparando la eficiencia de un radar con el modelo ideal, podemos determinar si vale o no la pena tratar de mejorarlo.

El trabajo de Shannon apareció inicialmente en la revista técnica — de la compañía de teléfonos Bell con el nombre de "Una Teoría Matemática de la Comunicación". |6|

Vale la pena observar que el trabajo original de Shannon plantea todos los aspectos esenciales de la Teoría Matemática de la Información.

B) Una presentación heurística del concepto de entropía.

Para construir la Teoría Matemática de la Información, hay que empezar por dar una definición adecuada de la Cantidad de Información, que se denotará por H, llamada entropía, y a partir de ésta, la definición de Capacidad de Canal C.

Uno de los principales resultados de la Teoría dice que en un canal con ruido de capacidad C que acepta información de una fuente de entropía H, si $C \geq H$, mediante codificaciones apropiadas se puede reducir tanto como se quiera el error al recibir el mensaje, pero esto es imposible si $C<H$, y este error o equivocación será $\geq H-C$, pero se puede acercar a $H-C$ tanto como se quiera usando un código adecuado.

Las fuentes de información pueden ser discretas o continuas; son discretas si forman los mensajes con un conjunto discreto de símbolos. Muchas fuentes continuas se pueden tratar como discretas o aproximarlas con ellas, como se hace, por ejemplo, en las transmisiones de radio o televisión. La definición de entropía de Shannon, como medida de la cantidad de información, fue dada para fuentes discretas; posteriormente ha sido extendida al caso continuo y definida por varios matemáticos.

Hay una estrecha relación entre información e indeterminación para dar una definición adecuada de la cantidad de información de un mensaje (una sucesión de símbolos de la fuente) es necesario considerarlo dentro del conjunto de todos los mensajes que pudieron generarse. En otras palabras, no nos interesará la información que contenga el mensaje, sino la que pudo contener.

Por otra parte, información es lo contrario de indeterminación, en el sentido de que una información elimina una indeterminación. En efecto, una experiencia proporciona información cuando no se sabe de antemano el resultado; a mayor indeterminación al comenzar la experiencia, deberá corresponder una mayor cantidad de información cuando se conoce el resultado.

Consideremos un experimento X cuyo resultado va a ser transmitido, y supongamos que tiene n resultados igualmente probables. Sea $f(n)$ la función que da la cantidad de información que proporciona al saber el resultado de X. Es claro que si X está compuesto por dos experimentos independientes X_1, X_2 con n_1 y n_2 resultados igualmente probables, una definición adecuada de la cantidad de información deberá cumplir que $f(n) = f(n_1) + f(n_2)$, donde $n = n_1 n_2$. Hay muchas funciones que satisfacen esa condición; por ejemplo, $f(n)$ = número de factores primos de n. Sin embargo, una medida de la cantidad de información deberá ser creciente en n, y se puede demostrar que eso obliga a que $f(n) = c \log n$ (c una constante).

El experimento más simple es el de tener dos resultados equiprobables; la información asociada a este experimento es la unidad y se llama BIT (de binary digit); por lo tanto $f(2) = c \log 2 = 1$, y si hacemos $c = 1$, el logaritmo debe tomarse en base 2, como lo haremos de ahora en adelante. Tenemos entonces que $f(n) = \log n$. Así $f(1) = 0$, lo que indica que un evento seguro no proporciona información.

Para resultados que no son igualmente probables, dividimos los n resultados equiprobables de X en dos grupos con n_1 y n_2 resultados, respectivamente.

Entonces $n = n_1 + n_2$ y las probabilidades de los grupos son $p_1 = n_1/n$ y $p_2 = n_2/n$. La definición de cantidad de información deberá satisfacer: información asociada a uno de los n mensajes = información de que resulte uno de los dos grupos + información asociada a uno de los n_1 mensajes por la probabilidad de que sea del primer grupo + información asociada a uno de los n_2 mensajes por la probabilidad de que sea del segundo grupo. Es decir,

$$\log n = H + \frac{n_1}{n} \log n_1 + \frac{n_2}{n} \log n_2 .$$

Luego $$H = \log n - \left(\frac{n_1}{n} \log \frac{n_1}{n} + \frac{n_1}{n} \log n\right) - \left(\frac{n_2}{n} \log \frac{n_2}{n} + \frac{n_2}{n} \log n\right)$$

$$= - \frac{n_1}{n} \log \frac{n_1}{n} - \frac{n_2}{n} \log \frac{n_2}{n} = p_1 \log p_1 - p_2 \log p_2 .$$

Esto sugiere que si X tiene n resultados con probabilidades $p_1, p_2, \ldots, p_n$ de ocurrir $\left(\sum_{i=1}^{n} p_i = 1\right)$, definamos

$$H_n(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i ,$$

la entropía de p_1, p_2,...,p_n, llamada así por analogía con una expresión que aparece en termodinámica. Esta es una definición adecuada de la cantidad de información.

C A P I T U L O I

EL CONCEPTO DE ENTROPIA

A) Definición y propiedades de la entropía.

Consideremos un espacio de probabilidad finito Án formado de n eventos elementales $A_1, A_2, \ldots, A_n$ y sean $p_1, p_2, \ldots, p_n$ las probabilidades correspondientes a estos eventos, donde evidentemente

$$p_k \geq 0 \qquad (1 \leq k \leq n), \quad \sum_{k=1}^{n} p_k = 1.$$

Definición 1.I.A. Se llama Entropía del espacio de probabilidad finito Án, denotada por H_n(Án), la expresión

$$H_n(\text{Án}) = H_n(p_1, p_2, \ldots, p_n) = -\sum_{k=1}^{n} p_k \log p_k, \qquad (1.\text{I}.\text{A})$$

donde la base del logaritmo es 2.

Puesto que

$$\log_b m = \log_b a \ \log_a m,$$

si se pasa de los logaritmos de base 2 a los logaritmos de base 10, la expresión de la entropía será

$$H(p_1, p_2, \ldots, p_n) = -\log_2 10 \sum_{k=1}^{n} p_k \log p_k.$$

En lo sucesivo , será la relación (1.I.A) la que utilizaremos.

La entropía dada por la relación (1.I.A) puede ser considerada como la medida del grado de indeterminación contenido en el espacio de probabilidad finito Án y posee las siguientes propiedades:

Propiedad 1.- Tenemos

$$H_n(p_1, p_2, \ldots, p_n) \geq 0.$$

(Resulta que la entropía es, como toda medida, no-negativa.)

Demostración:

Como $0 \leq p_i \leq 1$, $\log p_i \leq 0$, y el resultado se obtiene de la definición. //

Propiedad 2.- Si $p_i = 1$, $p_k = 0$ para $k \neq i$; entonces

$$H_n(p_1, p_2, \ldots, p_n) = 0$$

(Esta propiedad de la entropía corresponde al hecho de que si un espacio de probabilidad contiene un evento elemental de probabilidad uno, el resto de eventos elementales tienen probabilidad cero, y no posee indeterminación alguna.)

Demostración:

Se sigue de la definición haciendo $0.\infty = 0$. //

Propiedad 3.- La entropía satisface la desigualdad

$$H_n(p_1, p_2, \ldots, p_n) \leq H_n(\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n}).$$

(Esta propiedad muestra que la entropía alcanza su máximo cuando todos los eventos son equiprobables.)

Demostración:

Sea $y= f(x)$ una función convexa definida en el intervalo $|a,b|$ de la recta real. Si $x_1, x_2, \ldots, x_n$ son n valores arbitrarios del argumento pertenecientes al intervalo $|a,b|$, de acuerdo a la desigualdad de J.L. Jensen, cualesquiera que sean los números no negativos $\lambda_1, \lambda_2, \ldots, \lambda_n$, cuya suma sea igual a 1, tendremos

$$\sum_{k=1}^{n} \lambda_k f(\lambda_k) \leq f\left(\sum_{k=1}^{n} \lambda_k \lambda_k\right). \qquad (2.\text{I}.\text{A})$$

Apliquemos ahora esta desigualdad a

$$\lambda_k = p_k, \quad \lambda_k = \frac{1}{n} \quad (1 \leq k \leq n); \; f(x) = x\log x.$$

Obtenemos

$$-\sum_{k=1}^{n} \frac{1}{n} p_k \log p_k \leq -\sum_{k=1}^{n} \frac{1}{n} p_k \log\left(\sum_{k=1}^{n} \frac{1}{n} p_k\right);$$

ya que

$$\sum_{k=1}^{n} p_k = 1,$$

resulta que

$$\frac{1}{n} H_n(p_1, p_2, \ldots, p_n) \leq \frac{1}{n} H_n\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right). \quad //$$

Propiedad 4.- Tenemos

$$H_{n+1}(p_1, p_2, \ldots p_n \; 0) = H_n(p_1, p_2, \ldots, p_n).$$

(Esta propiedad pone en evidencia el hecho de que los dos espacios de probabilidad finitos

$$\begin{matrix} A_1 & A_2 & \ldots A_n \\ p_1 & p_2 & p_n \end{matrix} \quad \text{y} \quad \begin{matrix} A_1 & A_2 & \ldots A_n & A_{n+1} \\ p_1 & p_2 & p_n & 0 \end{matrix}$$

contiene la misión indeterminación.)

Demostración:

Se sigue de la definición haciendo $0.\infty = 0$. //

Consideremos ahora dos espacios de probabilidad finitos independientes A_n y β_m.

$$A_n = \begin{matrix} A_1 & A_2 & \ldots A_n \\ p_1 & p_2 & p_n \end{matrix} \quad , \quad \beta_m = \begin{matrix} B_1 & B_2 & \ldots B_m \\ q_1 & q_2 & q_m \end{matrix}$$

$$p_k \geq 0 \quad (1 \leq k \leq n) \ , \quad q_\ell \geq 0 \quad (1 \leq \ell \leq m)$$

$$\sum_{k=1}^{n} p_k = 1 \quad , \quad \sum_{\ell=1}^{m} q_\ell = 1. \qquad (3.1.A)$$

La probabilidad de que sucedan los eventos A_k y B_ℓ será

$$\pi_{k\ell} = p_k q_\ell. \qquad (4.1.A)$$

La totalidad de eventos (A_k, B_ℓ) $(1 \leq k \leq n \ , \ 1 \leq \ell \leq m)$, con las probabilidades $\pi_{k\ell}$ forma un nuevo espacio de probabilidad finito que denotamos por $A_n \times \beta_m$.

Propiedad 5.- Si A_n y β_m son dos espacios de probabilidad finitos independientes, tenemos

$$H_{nm}(A_n \times \beta_m) = H_n(A_n) + H_m(\beta_m).$$

(La probabilidad 5 muestra que la indeterminación del producto de dos espacios de probabilidad finitos independientes es igual a la suma de las indeterminaciones de los dos espacios considerados.)

Demostración:

Teniendo en cuenta la fórmula de la entropía y las relaciones (3.1.A), (4.1.A), se puede escribir

$$H_{nm}(A_n \times \beta_m) = -\sum_{k=1}^{m} \sum_{\ell=1}^{m} \pi_{k\ell} \log \pi_{k\ell}$$

$$= -\sum_{k=1}^{n} \sum_{\ell=1}^{n} p_k\, q_\ell (\log p_k + \log q_\ell) = -\sum_{k=1}^{n} p_k \log p_k - \sum_{\ell=1}^{m} q_\ell \log q_\ell$$

$$= H_n(A_n) + H_n(\beta_m). \quad //$$

Supongamos ahora que los espacios de probabilidad finitos A_n y β_m son dependientes y sea $q_{k\ell}$ la probabilidad de que sucede el evento elemental B_ℓ en el espacio β_m a condición de que en el espacio A_n haya sucedido el evento A_k. En este caso, la probabilidad del evento (A_k, B_ℓ) será

$$\pi_{k\ell} = p_k\, q_{k\ell}. \qquad (5.1.A)$$

La entropía del espacio de probabilidad finito β_m condicionada por el evento elemental A_k del espacio A_n será

$$H_m(\beta_m | A_k) = -\sum_{\ell=1}^{m} q_{k\ell} \log q_{k\ell}\ , \qquad (6.1.A)$$

mientras que la entropía del espacio de probabilidad β_m condicionada por todo el espacio A_n será

$$H_m(\beta_m | A_n) = \sum_{k=1}^{n} p_k\, H_m(\beta_m | A_k). \qquad (7.1.A)$$

Propiedad 6.- Si A_n y β_m son dos espacios de probabilidad finitos cualesquiera, tendremos

$$H_{nm}(A_n \times \beta_m) = H_n(A_n) + H_m(\beta_m | A_n).$$

(Esta propiedad pone en evidencia el hecho de que en el caso de dos espacios de probabilidad finitos cualesquiera, la indeterminación de la realización de los espacios es igual a la indeterminación de un espacio más la indeterminación del otro espacio, condicionado por la realización del primero.)

Demostración:

$$H_{nm}(A_n \times \beta_m) = -\sum_{k=1}^{n} \sum_{\ell=1}^{m} p_k\, q_{k\ell} \log p_k\, q_{k\ell}$$

$$= -\sum_{k=1}^{n} \sum_{\ell=1}^{m} p_k\, q_{k\ell}(\log p_k + \log q_{k\ell}) =$$

$$= -\sum_{k=1}^{n} p_k \log p_k \sum_{\ell=1}^{m} q_{k\ell} - \sum_{k=1}^{n} p_k \sum_{\ell=1}^{m} q_{k\ell} \log q_{k\ell};$$

ya que $\sum_{\ell=1}^{m} q_{k\ell} = 1$ para cualquier k, tenemos que

$$H_n(A_n \times \beta_m) = -\sum_{k=1}^{n} p_k \log p_k - \sum_{k=1}^{n} p_k \sum_{\ell=1}^{m} q_{k\ell} \log q_{k\ell}$$

$$= H_n(A_n) + \sum_{k=1}^{n} p_k\, H_n(\beta_m | A_k),$$

y usando (7.I.A) tenemos

$$H_{nm}(A_n \times \beta_m) = H_n(A_n) + H_m(\beta_m | A_n). \quad //$$

Propiedad 7.- Dados dos espacios de probabilidad finitos cualesquiera A_n y β_m se tiene

$$H_m(\beta_m | A_n) \leq H_m(\beta_m).$$

(La propiedad 7 muestra que el conocimiento del resultado del espacio de probabilidad finito A_n no puede más que disminuir la indeterminación

ción del espacio de probabilidad finito β_m.)

Demostración:

Apliquemos la desigualdad (2.I.A) para $f(x) = -x \log x$, $x_k = q_{k\ell}$ $\lambda_k = p_k$ $(1 \leq k \leq n)$. Se tiene

$$-\sum_{k=1}^{n} p_k\, q_{k\ell} \log q_{k\ell} \leq -\left(\sum_{k=1}^{n} p_k\, q_{k\ell}\right) \log\left(\sum_{k=1}^{n} p_k\, q_{k\ell}\right)$$

$$= -q_\ell \log q_\ell .$$

Sumando enseguida sobre ℓ se obtiene

$$-\sum_{k=1}^{n} p_k \sum_{\ell=1}^{m} q_{k\ell} \log q_{k\ell} \leq -\sum_{\ell=1}^{m} q_\ell \log q_\ell \; . \quad //$$

De las propiedades 6 y 7 resulta inmediatamente la

Propiedad 8.- Dados dos espacios de probabilidad finitos cualesquiera A_n y β_m, se tiene

$$H_{nm}(A_n \times \beta_m) \leq H_n(A_n) + H_m(\beta_m).$$

Se tiene también la

Propiedad 9.- Dados dos espacios de probabilidad finitos cualesquiera A_n y β_m, se tiene

$$H_n(A_n|\beta_m) = H_m(\beta_m|A_n) + H_n(A_n) - H_m(\beta_m).$$

Demostración:

Para la entropía del espacio $A_n \times \beta_m$ se pueden dar dos descomposiciones idénticas,

$$H_{nm}(A_n \times \beta_m) = H_n(A_n) + H_m(\beta_m|A_n)$$

$$H_{nm}(A_n \times \beta_m) = H_m(\beta_m) + H_n(A_n|\beta_m),$$

de donde la propiedad 9 resulta inmediatamente. //

B) El Teorema de Unicidad.

De las propiedades de la entropía que hemos probado, podemos considerar como básicas las dos siguientes:

1.- Para n dado y para $\sum_{k=1}^{n} p_k = 1$, la función $H_n(p_1,\ldots,p_n)$ toma su valor más grande para $p_k = \frac{1}{n}$ $(k=1,\ldots,n)$.

2.- $H(AB) = H(A) + H_A(B)$.

Agreguemos a estas propiedades una tercera, la cual obviamente debe ser satisfecha por cualquier definición razonable de entropía. Ya que los espacios

$$\begin{matrix} A_1 & A_2 & \ldots & A_n \\ p_1 & p_2 & & p_n \end{matrix} \qquad y \qquad \begin{matrix} A_1 & A_2 & \ldots & A_n & A_{n+1} \\ p_1 & p_2 & & p_n & 0 \end{matrix}$$

claramente no son en esencia diferentes, tenemos

3.- $H(p_1,\ldots,p_n) = H(p_1,\ldots,p_n,0)$ (si se agrega el evento imposible o cualquier número de eventos de probabilidad cero a un espacio, no cambia su entropía).

Probaremos ahora la siguiente proposición debido a A. Ya. Jinchin (1953), que muestra que estas 3 propiedades determinan la expresión para la entropía, salvo el producto por una constante.

Teorema 1.1.B.- Sea $H(p_1,\ldots,p_n)$ una función definida para cualquier entero positivo n y para todos los valores $p_1,\ldots,p_n$ tales que $p_k \geq 0$ $(k=1,\ldots,n)$, $\sum_{k=1}^{n} p_k = 1$. Si para cualquier n esta función es continua con respecto a todos sus argumentos, y si tiene las propiedades 1,2 y 3 entonces

$$H(p_1,\ldots,p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k,$$

donde λ es una constante positiva.

Este teorema muestra que la expresión para la entropía de un espacio finito debe tener una sola forma si queremos que tenga ciertas propiedades generales, necesarias en vista del significado de la entropía (como una medida de incertidumbre o de cantidad de información).

Demostración:

Por brevedad hagamos

$$H(\frac{1}{n},\ldots,\frac{1}{n}) = L(n).$$

Mostraremos que $L(n) = \lambda \log n$, donde λ es una constante positiva. Por las propiedades 3 y 1, tenemos

$$L(n) = H(\frac{1}{n},\ldots,\frac{1}{n},0) \leq H(\frac{1}{n+1},\ldots,\frac{1}{n+1}) = L(n+1),$$

de modo que $L(n)$ es una función no descreciente de n. Sean m y r enteros positivos. Consideremos m espacios mutuamente independientes $S_1,\ldots S_m$, cada uno de los cuales contiene r eventos igualmente probables de modo que

$$H(S_k) = H(\frac{1}{r},\ldots,\frac{1}{r}) = L(r), \qquad (1 \leq k \leq m)$$

por la propiedad 2 (generalizada al caso de m espacios) tenemos, en vista de la independencia del espacio S_k,

$$H(S_1 x\ S_2 x \ldots x S_m) = \sum_{k=1}^{m} H(S_k) = mL(r).$$

Pero el espacio producto $S_1 x. S_2 x \ldots S_m$ consiste de r^m eventos equiprobables, de modo que su entropía es $L(r^m)$. En consecuencia tenemos

$$L(r^m) = mL(r), \qquad (2.I.B)$$

y de igual forma, para cualquier otro par de enteros positivos n y S,

$$L(S^n) = nL(S) \qquad (3.I.B)$$

Sean los enteros positivos $r \geq 2$, S y n dados arbitrariamente, y sea m el número determinado por la desigualdad

$$r^m \leq S^n \leq r^{m+1} ; \qquad (4.I.B)$$

por consiguiente

$$m \log r \leq n \log S \leq (m+1) \log r$$

$$\frac{m}{n} \leq \frac{\log S}{\log r} \leq \frac{m}{n} + \frac{1}{n} , \qquad (5.I.B)$$

se sigue de (4.I.B), por la monotonía de la función L(n), que

$$L(r^m) \leq L(S^n) \leq L(r^{m+1}),$$

y consecuentemente por (2.I.B) y (3.I.B)

$$mL(r) \leq nL(S) \leq (m+1) L(r)$$

de modo que

$$\frac{m}{n} \leq \frac{L(S)}{L(r)} \leq \frac{m+1}{n} . \qquad (6.I.B)$$

Finalmente se sigue de (5.I.B) y (6.I.B) que

$$\left| \frac{L(S)}{L(r)} - \frac{\log S}{\log r} \right| \leq \frac{1}{n} ;$$

ya que el lado izquierdo de esta desigualdad es independiente de m y ya que n puede tomarse arbitrariamente grande en el lado derecho,

$$\frac{L(S)}{L(r)} = \frac{\log S}{\log r}$$

ó

$$\frac{L(S)}{\log S} = \frac{L(r)}{\log r} ,$$

lo cual en vista de la arbitrariedad de r y S quiere decir que

$$L(n) = \lambda \log n,$$

donde λ es una constante. Por la monotonía de la función $L(n)$, tenemos $\lambda \geq 0$, y nuestra afirmación queda probada.

Esta aseveración representa el caso especial $p_k = \frac{1}{n}$ $(1 \leq k \leq n)$ del teorema a probarse. Consideremos ahora el caso más general, donde los $p_k (k=1,\ldots,n)$ son cualesquiera número racionales.

Sea

$$p_k = q_k/g \qquad (k=1,\ldots,n),$$

donde todos los g_k son enteros positivos y $\sum_{k=1}^{n} g_k = g$. Sea el espacio finito A consistente de n eventos con probabilidades $p_1,\ldots p_n$. Nuestro problema consiste en encontrar una expresión para la función H evaluada en este espacio. Para este fin consideremos un segundo espacio B, el cual es dependiente de A y está definido de la siguiente manera: El espacio B contiene g eventos $B_1, B_2,\ldots,B_g$, los cuales dividimos en n grupos que contienen $g_1, g_2,\ldots g_n$ eventos respectivamente. Si el evento A_k ocurrió en el espacio A entonces en el espacio B todos los g_k eventos del k-ésimo grupo tienen la misma probabilidad $1/g_k$ y todos los eventos de los otros grupos tienen probabilidad cero. Así, dado cualquier suceso A_k del espacio B se reduce a un sistema de g_k eventos igualmente probables de modo que la entropía condicional es

$$H_k(B) = H(\frac{1}{g_k},\ldots,\frac{1}{g_k}) = L(g_k) = \lambda \log g_k,$$

lo cual significa que

$$H_A(B) = \sum_{k=1}^{n} p_k H_k(B) = \lambda \sum_{k=1}^{n} p_k \log g_k = \lambda \sum_{k=1}^{n} p_k \log p_k + \lambda \log g. \tag{7.I.B}$$

Regresemos ahora al espacio producto A x B, consistente de los eventos $A_k B_\ell$ $(1 \le k \le n,\ 1 \le \ell \le g)$. Un evento tal es posible solamente si B_ℓ pertenece al k-ésimo grupo. Así el número de eventos posibles en el espacio A x B es $\sum_{k=1}^{n} g_k = g$. La probabilidad de cada evento posible $A_k B_\ell$ es obviamente $\frac{p_k}{g_k} = \frac{1}{g}$, es decir, es la misma para todos los eventos.

Así el espacio A x B consiste de g eventos equiprobables, y en consecuencia

$$H(A \times B) = L(g) = \lambda \log g.$$

Usando la propiedad (2) y la relación (7.I.B), hallamos

$$\lambda \log g = H(A) + \lambda \sum_{k=1}^{n} p_k \log p_k + \lambda \log g,$$

en consecuencia

$$H(A) = H(p_1,\ldots,p_n) = -\lambda \sum_{k=1}^{n} p_k \log p_k.$$

Finalmente, la relación (8.I.B), la cual hemos probado para racionales $p_1,\ldots,p_n$, es válida para cualesquiera valores reales de sus argumentos debido al postulado de continuidad de la función $H(p_1,\dots,p_n)$ y la densidad de los racionales. Así la prueba del Teorema queda completa. //

C) <u>Algo más sobre el Teorema de Unicidad.</u>

i) Después de Jinchin, algunos matemáticos han simplificado los supuestos del teorema de unicidad; por ejemplo, D.K. Faddeev (1956), H. Tverberg (1958), A. Rényi (1959).

ii) Por otra parte, H. Hatori (1958) extendió el Teorema de unicidad para el caso de distribuciones continuas de probabilidad, del modo siguiente:

Sea $\xi = (\xi_1,\dots,\xi_s)$ un vector aleatorio s-dimensional, con densidad de probabilidad ρ $\left(\rho(x) \geq 0, \int_{R^s} \rho(x)\, dx = 1\right)$.

Axioma 1.- La entropía $H(\xi)$ sólo depende de ρ.

(Entonces escribimos $H(\rho)$ en lugar de $H(\xi)$.)

Sea $D \subset R^s$ Lebesgue medible y de medida positiva, ε_D la densidad uniforme sobre S.

Axioma 2.- Si ρ es una densidad de probabilidad en R^s y $\rho \neq \varepsilon_D$ casi seguramente respecto a la medida de Lebesgue y si soporte de $\rho \subset D$, se tendrá $H(\rho) < H(\varepsilon_d)$.

Sean ζ la densidad del vector $(\xi_1,\dots,\xi_k)$ y $\psi_{x_1\dots,x_k}$ la densidad condicional del vector $(\xi_{k+1},\dots,\xi_s)$ bajo la hipótesis de que $\xi_1 = x_1,\dots,\xi_k = x_k$.

Axioma 3.- Se tiene la relación

$$H((\xi_1,\ldots,\xi_s)) = H((\xi_1,\ldots,\xi_k)) + H((\xi_{k+1},\ldots,\xi_s)\,|\,(\xi_1,\ldots,\xi_k)),$$

es decir,

$$H(\rho) = H(\zeta) + \int_{R^k} H\ \psi_{x_1,\ldots,x_k}\ \zeta(x_1,\ldots,x_k)\ dx_1,\ldots,dx_k.$$

Axioma 4.- Si ρ toma un número finito de valores $c_1,\ldots,c_\ell$ y

$$\mu_J = \int_{A_J} dx, \qquad 1 \leq J \leq \ell,$$

donde

$$A_J = \{x|\rho(x) = c_J\}, \qquad 1 \leq J \leq \ell\ ,$$

entonces $H(\rho)$ es una función solamente de $c_1,\ldots,c_\ell$ y de $\mu_1,\ldots,\mu_\ell$, y no depende de la dimensión del espacio S, donde $\sum_{K=1}^{\ell} c_J\mu_J = 1$.

Teorema.- Bajo las hipótesis de los axiomas 1, 2, 3 y 4,

se tiene

$$H(\rho) = -\lambda \int_{R^s} \rho(x)\ \log\ \rho(x)\ dx,$$

donde λ es una constante positivo.

iii) R.S. Ingarden y K. Urbanik (1962) dieron un enfoque novedoso a la axiomática de la entropía al prescindir de la probabilidad. Para cierta familia de álgebras booleanas, definieron una función H que llamaron información, y demostraron que para cada subálgebra A de la familia corresponde una única probabilidad P_A definida sobre A, de modo que

$$P_\beta(B) = \frac{P_A\ (B)}{P_A\ (\Omega_B)}\ , \quad B\varepsilon\beta,$$

para toda subálgebra β de A y que

$$H(A) = -\sum_{J=1}^{n} P_A(A_J)\ \log P_A\ (A_J),$$

donde el logaritmo se toma en base 2 y $A_1,\ldots,A_n$ son todos los átomos de A.

C A P I T U L O II

REFINAMIENTOS DE LA ENTROPIA

A) Entropía d-dimensional

Sea (Ω,F,P) un espacio de probabilidad, η una variable aleatoria (v.a) que toma los valores Y_k, $1 \leq k < \infty$, con probabilidades q_k, es decir

$$P(\eta = Y_k) = q_k, \; q_k \geq 0, \quad 1 \leq k < \infty, \; \sum_{k=1}^{\infty} q_k = 1.$$

Definamos la entropía de la v.a. η como

$$H^{(\circ)}(\eta) = \sum_{k=1}^{\infty} q_k \log \frac{1}{q_k} \;,$$

en el caso de que la serie converja. En caso contrario, diremos que la entropía no existe.

Sea ξ una v.a. (real) y sea

$$\xi_n = \frac{1}{n} |n\xi|.$$

Definición 1.II.A.- Si $H^{(\circ)}(|\xi|)$ existe y si $H^{(\circ)}(\xi_n) = d \cdot \log n + h + 0(1)$ $(0(1) \to 0$ cuando $n \to \infty)$, diremos que la *dimensión* $d(\xi)$ de ξ (o de la distribución de probabilidades de la v.a.ξ) es d y que la *entropía* d-*dimensional* $H^{(d)}(\xi)$ de ξ es h.

Se tiene que

$$d(\xi) = \lim_{n\to\infty} \frac{H^{(\circ)}(\xi_n)}{\log n} \qquad (1.II.A)$$

si el limite existe. En caso contrario, se puede hablar de dimensión inferior

$$\underline{d}\,(\xi) = \underline{\lim}_{n\to\infty} \frac{H^{(\circ)}(\xi_n)}{\log n} \quad ,$$

y de la dimensión superior

$$\overline{d(\xi)} = \overline{\lim}_{n\to\infty} \frac{H^{(\circ)}(\xi_n)}{\log n}$$

Se sigue que ξ tiene dimensión finita si y sólo si $\underline{d(\xi)} = \overline{d(\xi)}$.

Si el limite

$$\lim_{n\to\infty} \frac{H^{(\circ)}(\xi_n)}{\log n} = d(\xi)$$

existe, ponemos

$$H^{d(\xi)} = \lim_{n\to\infty} \left(H^{(\circ)}(\xi_n) - d(\xi) \log n\right).$$

No siempre están definidas la dimensión inferior ni la superior, por ejemplo si $H^{(\circ)}(\xi_n)$ es infinita, ni $\underline{d}(\xi)$ ni $\overline{d}(\xi)$ están definidos. Sin embargo, si $H^{(\circ)}\ (|\xi|)$ es finita, entonces $\underline{d}(\xi)$ y $\overline{d}(\xi)$ siempre existen y

$$0 \leq \underline{d}(\xi) \leq \overline{d}(\xi) \leq 1. \tag{2.II.A}$$

Para demostrarlo, pongamos

$$p_{nk} = p\ \frac{k}{n} \leq \xi < \frac{k+1}{n}\ , \quad n \geq 1, \quad -\infty < k < \infty.$$

Valiéndonos de la desigualdad de Jensen en la forma

$$\frac{\sum_{k=1}^{m} a_k \log b_k}{\sum_{k=1}^{m} a_k} \leq \frac{\log \sum_{k=1}^{m} a_k b_k}{\sum_{k=1}^{m} a_k} \qquad (3.11.A)$$

que vale para toda colección finita de números no negativos a_k y b_k, de manera que

$$\sum_{k=1}^{m} a_k > 0.$$

Tenemos de (3.11.A)

$$\sum_{k=\ell n}^{(\ell+1)\,n-1} p_{nk} \log \frac{1}{p_{nk}} \leq p_{1\ell} \log \frac{n}{p_{1\ell}} , \quad -\infty < \ell \leq \infty; \qquad (4.11.A)$$

sumando (4.11.A) sobre los valores de ℓ, se tiene

$$0 \leq H^{(\circ)}(\xi_n) \leq H^{(\circ)}(\xi_1) + \log n, \qquad (5.11.A)$$

que indica que si $H^{(\circ)}(\xi_1) = H^{(\circ)}(|\xi|)$ es finita, entonces $H^{(\circ)}(\xi_n)$ es finita para toda n, y se tiene la desigualdad (2.11.A).

Teorema 1.11.A.- Si ξ es una v.a. discreta que toma valores X_k $(1 \leq k < \infty)$ con probabilidad p_k $(1 \leq k < \infty)$, se tiene que

$$d(\xi) = 0 \quad y \quad H^{(\circ)}(\xi) = \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k} ,$$

suponiendo que la serie del segundo miembro es convergente.

Demostración: Antes de demostrarlo observaremos que si ξ es una v.a. que tiene una distribución discreta de probabilidades, y si f es una función cualquiera, se tiene

$$H^{(\circ)}(f(\xi)) \leq H^{(\circ)}(\xi). \qquad (6.11.A)$$

En efecto, si f es una función biunívoca, en la desigualdad de arriba tendremos el signo igual; sin embargo, si existen $x_i \neq x_J$ tales que $f(x_i) = f(x_J)$, entonces la relación de arriba será una desigualdad estricta y para $x > 0$, $y > 0$, $x + y \leq 1$, tendremos

$$x \log \frac{1}{x} + y \log \frac{1}{y} > (x + y) \log \frac{1}{x + y}.$$

Volvamos a la demostración del teorema.

Evidentemente podemos arreglar los valores de ξ_n de manera que designando las probabilidades correspondientes con $p_k^{(n)}$, $1 \leq k < \infty$, tengamos para cada k dado

$$\lim_{n\to\infty} p_k^{(n)} = p_k.$$

De acuerdo con la densidad (6.11.A), si la serie

$$\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$$

es convergente, tendremos

$$H^{(\circ)}(\xi_n) = \sum_{k=1}^{\infty} p_k^{(n)} \log \frac{1}{p_k^{(n)}} \leq \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}.$$

Se sigue, pues, que para cada N tendremos

$$\sum_{k=1}^{N} p_k \log \frac{1}{p_k} \leq \varliminf_{n\to\infty} H^{(\circ)}(\xi_n) \leq \varlimsup_{n\to\infty} H^{(\circ)}(\xi_n) \leq \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k},$$

lo que implica

$$\lim_{n\to\infty} H^{(\circ)}(\xi_n) = H^{(\circ)}(\xi). \; //$$

Pongamos ahora nuestra atención en las v.a. cuya distribución de probabilidades es absolutamente continua. Demostremos el siguiente

Teorema 2.II.A.- Si ξ es una v.a. absolutamente continua con densidad ρ y si $H^{(0)}(|\xi|) < \infty$, entonces

$$d(\xi) = 1 \quad y \quad H^{(1)}(\xi) = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)}\, dx ,$$

si existe la integral.

Demostración: En primer lugar demostremos que

$$d(\xi) = 1.$$

Pongamos

$$\psi_n(x) = n \int_{\frac{k}{n}}^{\frac{k+1}{n}} \rho(t)dt = np_{nk} , \quad \frac{k}{n} \leq x < \frac{k+1}{n} ,$$

de lo cual deducimos inmediatamente que

$$\int_{\infty}^{\infty} \psi_n(x)\, dx = \int_{-\infty}^{\infty} \rho(t)\, dt = 1. \tag{7.II.A}$$

En este caso,

$$\sum_{k=1}^{\infty} p_{nk} \log \frac{1}{np_{nk}} = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\psi_n(x)}\, dx. \tag{8.II.A}$$

En primer lugar, consideremos el caso en que existe la integral

$$1 = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)}\, dx.$$

Valiéndonos de la desigualdad de Jensen en forma integral, podemos escribir

$$\frac{\int_E g(x)\ \log h(x)\ dx}{\int_E g(x)\ dx} \leq \log \frac{\int_E g(x)\ h(x)\ dx}{\int_E g(x)\ dx} \qquad (9.II.A)$$

para toda función medible g no negativa y para todo conjunto medible E donde existan las integrales y donde la integral de los denominadores sea positiva. En estas condiciones se deduce de las relaciones (8.II.A) y (9.II.A) que

$$1 - \sum_{k=-\infty}^{\infty} p_{nk}\ \log \frac{1}{n\ p_{nk}} = \int_{-\infty}^{\infty} \rho(x)\ \log \frac{\psi_n(x)}{\rho(x)} dx \leq \log \int_{-\infty}^{\infty} \psi_n(x) dx = 0,$$

por lo tanto

$$1 + \log n \leq H^{(\circ)}(\xi_n), \qquad (10.II.A)$$

Resulta de (5.II.A) y de (10.II.A) que

$$\lim_{n\to\infty} \frac{H^{(\circ)}(\xi_n)}{\log n} = 1.$$

En el caso general, pongamos para todo $A>0$,

$$\rho_A(x) = \begin{cases} \rho(x) & \text{si } \rho(x) \leq A, \\ 0 & \text{si } \rho(x) > A. \end{cases}$$

Es evidente que haciendo

$$\int_{-\infty}^{\infty} \rho_A(x)\ dx = S(A),$$

tendremos

$$\lim_{A\to\infty} S(A) = 1. \qquad (11.II.A0$$

Pongamos ahora

$$p_{nk}(A) = \int_{\frac{k}{n}}^{\frac{k+1}{n}} \rho_A(x)\,dx$$

$$\psi_n(A,x) = n\, p_{nk}(A), \quad \frac{k}{n} \leq x < \frac{k+1}{n}.$$

En este caso tenemos

$$\int_{-\infty}^{\infty} \psi_n(A,x)\,dx = S(A).$$

Escojamos ahora una n suficientemente grande para que a cada k corresponda

$$p_{nk} \leq \frac{1}{e}.$$

Teniendo en cuenta que $x \log \frac{1}{x}$ es función creciente para $0 \leq x < \frac{1}{e}$ y que

$$p_{nk}(A) \leq p_{nk},$$

la serie

$$\sum_{k=-\infty}^{\infty} p_{nk}(A) \log \frac{1}{n\, p_{nk}(A)}$$

es convergente y tenemos

$$H^{(\circ)}(\xi_n) - S(A) \log n \geq \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \frac{1}{n\, p_{nk}(A)} = \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\psi_n(A,x)}\,dx.$$

Es virtud de la relación (9.II.A), deducimos

$$\int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\psi_n(A,x)}\,dx \geq \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)}\,dx \geq S(A) \log \frac{1}{A}.$$

De esto se sigue que

$$\frac{H^{(\circ)}(\xi_n)}{\log n} \geq S(A) + \frac{\int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)} dx}{\log n} ,$$

así pues

$$\underline{d}(\xi) \geq S(A).$$

Como podemos escoger A tan grande como queramos, resulta de la última desigualdad y de la relación (11.II.A) que $\underline{d}(\xi) = 1$, es decir que $d(\xi) = 1$.

Demostremos ahora la segunda parte del teorema (2.II.A). Distinguiremos dos casos, según que la desigualdad de distribución ρ sea acotada o no lo sea.

Sea ρ una función acotada, digamos que $\rho(x) \leq B$ y pongamos

$$g_n(x) = n\, p_{nk} \log \frac{1}{n\, p_{nk}} \quad , \quad \frac{k}{n} \leq x < \frac{k+1}{n} \; , \; -\infty < k < \infty \; .$$

Tenemos

$$n\, p_{nk} \leq B.$$

Deducimos que la función g_n es acotada

$$|g_n(x)| \leq \max \left(\frac{\log e}{e} \; , \; B|\log B|\right).$$

Por otra parte, designando con F la función de distribución de ξ, tenemos

$$F(x) = \int_{-\infty}^{x} \rho(t)\, dt$$

y por consiguiente, para $\frac{k}{n} \leq x < \frac{k+1}{n}$, tenemos

$$g_n(x) = \frac{F(\frac{k+1}{n}) - F(\frac{k}{n})}{\frac{1}{n}} \log \frac{1}{\frac{F(\frac{k+1}{n}) - F(\frac{k}{n})}{\frac{1}{n}}}$$

Puesto que tenemos casi seguramente que

$$F'(x) = \rho(x),$$

tendremos también casi seguramente que

$$\lim_{n\to\infty} g_n(x) = \rho(x) \log \frac{1}{\rho(x)} .$$

Para todo número entero $L > 0$,

$$\sum_{-Ln \leq k < Ln} n\, p_{nk} \log \frac{1}{n\, p_{nk}} = \int_{-L}^{L} g_n(x)\,dx.$$

En este caso, resulta del teorema de Lebesque que

$$\lim_{n\to\infty} \sum_{-Ln \leq k < Ln} n\, p_{nk} \log \frac{1}{n\, p_{nk}} = \int_{-L}^{L} \rho(x) \log \frac{1}{\rho(x)}\,dx. \quad (12.\text{II}.A)$$

Por otra parte, valiéndonos de la desigualdad (3.II.A) obtenemos

$$\sum_{\substack{k<-Ln \\ k\geq Ln}} p_{nk} \log \frac{1}{n\, p_{nk}} \leq \sum_{\substack{\ell<-L \\ \ell\geq L}} p_{1\ell} \log \frac{1}{p_{1\ell}} , \quad (13.\text{II}.A)$$

y de la desigualdad (9.II.A) deducimos

$$\sum_{\substack{k<-Ln \\ k\geq Ln}} p_{nk} \log \frac{1}{n\, p_{nk}} = \int_{|x|>L} \rho(x) \log \frac{1}{\psi_n(x)}\,dx \geq$$

$$\geq \int_{|x|>L} \rho(x) \log \frac{1}{\rho(x)}\,dx. \quad (14.\text{II}.A)$$

Para toda $\varepsilon > 0$ podemos escoger $L = L(\varepsilon)$ suficientemente grande para que

$$\left| \int_{|x| > L} \rho(x) \log \frac{1}{\rho(x)} dx \right| < \varepsilon \qquad (15.\text{II}.A)$$

y

$$\sum_{\substack{\ell < -L \\ \ell \geq L}} p_{1\ell} \log \frac{1}{p_{1\ell}} < \varepsilon. \qquad (16.\text{II}.A)$$

Procediendo así, deducimos de (12.II.A) - (16.II.A) que

$$\lim_{n\to\infty} (H^{(\circ)}(\xi_n) - \log n) = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)} dx.$$

El caso general podemos reducirlo al caso en que la densidad ρ es acotada, de la manera siguiente. Pongamos

$$r_{nk}(A) = p_{nk} - p_{nk}(A).$$

En este caso, con la notación

$$R(A) = \sum_{k=-\infty}^{\infty} r_{nk}(A) = 1 - S(A),$$

tenemos

$$\lim_{A\to\infty} R(A) = 0.$$

Ahora bien,

$$\sum_{k=-\infty}^{\infty} p_{nk} \log \frac{1}{n\, p_{nk}} = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\psi_n(x)} dx \geq$$

$$\geq \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)} dx,$$

basta, pues, demostrar que

$$\overline{\lim_{n\to\infty}} \sum_{k=-\infty}^{\infty} p_{nk} \log \frac{1}{n\, p_{nk}} \leq \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)}\, dx, \qquad (17.11.A)$$

pero

$$\sum_{k=-\infty}^{\infty} p_{nk} \log \frac{1}{n\, p_{nk}} = \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \frac{1}{n\, p_{nk}(A)}$$

$$- \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \left(1 + \frac{r_{nk}(A)}{p_{nk}(A)}\right) + \sum_{k=-\infty}^{\infty} r_{nk}(A) \log \frac{1}{n\, p_{nk}}. \qquad (18.11.A)$$

En lo que concierne al primer término, resulta de la demostración de más arriba que

$$\lim_{n\to\infty} \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \frac{1}{n\, p_{nk}(A)} = \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)}\, dx. \qquad (19.11.A)$$

Para el segundo término, sirviéndonos de la desigualdad

$$\log (1 + x) \leq x \log e, \qquad x > 0,$$

encontramos

$$0 \leq \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \left(1 + \frac{r_{nk}(A)}{p_{nk}(A)}\right) \leq R(A) \log e. \qquad (20.11.A)$$

En fin, respecto al tercer término, tenemos

$$\sum_{k=-\infty}^{\infty} r_{nk}(A) \log \frac{1}{n\, p_{nk}} \leq \sum_{|k| > Ln} p_{nk} \log \frac{1}{n\, p_{nk}} +$$

$$+ \sum_{\ell=-L}^{L-1} \sum_{k=\ell n}^{(\ell+1)n-1} r_{nk}(A) \log \frac{1}{n\, p_{nk}}$$

y empleando la relación (3.II.A), tenemos

$$\sum_{k=-\infty}^{\infty} r_{nk}(A) \log \frac{1}{n\, p_{nk}} \leq \sum_{|\ell| > L} p_{1\ell} \log \frac{1}{p_{1\ell}}$$

$$+ \sum_{\ell=-L}^{L-1} \sum_{k=\ell n}^{(\ell+1)n-1} r_{nk}(A) \log \frac{1}{\sum_{k=\ell n}^{(\ell+1)n-1} r_{nk}(A)} .$$

Por consiguiente si $R(A) < \frac{1}{e}$, aplicando nuevamente (3.II.A) al último término, obtenemos

$$\sum_{k=-\infty}^{\infty} r_{nk}(A) \log \frac{1}{n\, p_{nk}} \leq \sum_{|\ell|>L} p_{1\ell} \log \frac{1}{p_{1\ell}} + R(A) \log \frac{2L}{R(A)} . \quad (21.II.A)$$

De (18.II.A), (19.II.A), (20.II.A) y (21.II.A) deducimos que

$$\sum_{k=-\infty}^{\infty} p_{nk} \log \frac{1}{n\, p_{nk}} \leq \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)}\, dx + \delta_n (A)$$

$$+ R(A) \left[\log 2L + \log \frac{1}{R(A)} \right] + \sum_{|\ell|>L} p_{1\ell} \log \frac{1}{p_{1\ell}} , \quad (22.II.A)$$

en donde

$$\delta_n(A) = \sum_{k=-\infty}^{\infty} p_{nk}(A) \log \frac{1}{n\, p_{nk}(A)} - \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)}\, dx$$

de manera que, conforme a (18.II.A),

$$\lim_{n\to\infty} \delta_n(A) = 0.$$

Evidentemente,

$$\lim_{A\to\infty} \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)}\, dx = \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)}\, dx.$$

Escojamos ahora una L suficientemente grande para que

$$\sum_{|\ell|>L} p_{1\ell} \log \frac{1}{p_{1\ell}} < \varepsilon. \qquad (23.II.A)$$

Después de haber determinado L, escojemos A suficientemente grande para que

$$R(A) \log \frac{2L}{R(A)} < \varepsilon \qquad (24.II.A)$$

y

$$\left| \int_{-\infty}^{\infty} \rho_A(x) \log \frac{1}{\rho_A(x)} dx - \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)} dx \right|$$

$$= \left| \int_{\rho(x)>A} \rho(x) \log \frac{1}{\rho(x)} dx \right| < \varepsilon \qquad (25.II.A)$$

y en fin, escojemos una n_o suficientemente grande para que

$$|\delta_n(A)| < \varepsilon \qquad (26.II.A)$$

para $n \geq n_o$ en donde n_o sólo depende de ε. De (22.II.A) y de (23.II.A)-(26.II.A), deducimos que

$$\sum_{k=-\infty}^{\infty} p_{nk} \log \frac{1}{n\, p_{nk}} \leq \int_{-\infty}^{\infty} \rho(x) \log \frac{1}{\rho(x)} dx + 4\varepsilon.$$

Como podemos escoger $\varepsilon>0$ arbitrariamente pequeña, obtenemos (17.II.A).//

Pasemos ahora al caso de una v.a. cuya distribución de probabilidad está formada por una componente discontinua y una componente absolutamente continua.

Teorema 3.11.A.- Sea ξ una v.a. tal que $H^{(\circ)}(|\xi|)$ es finita y F la función de distribución de ξ, representada por:

$$F = (1 - d)\ f_\circ + dF_1 ,$$

en donde $F_\circ$ es la componente discontinua, es decir, que hay una sucesión de valores diferentes $x_1, x_2, \ldots,$ y una sucesión correspondiente $p_1, p_2, \ldots,$ para las cuales $p_k \geq 0$, $1 \leq k < 0$, y $\sum_{k=1}^{\infty} p_k = 1$, tales que

$$F_\circ(x) = \sum_{x_k < x} p_k ,$$

siendo F_1 la componente absolutamente continua, es decir tal que

$$F_1(x) = \int_{-\infty}^{x} \rho_1(t)\,dt,$$

mientras que $0 < d < 1$. En este caso la dimensión de ξ es

$$d(\xi) = d.$$

Si, además, la serie

$$\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$$

y la integral

$$\int_{-\infty}^{\infty} \rho_1(x) \log \frac{1}{\rho_1(x)}\,dx.$$

son convergentes, la entropía d-dimensional de ξ la da la fórmula

$$H^{(d)}(\xi) = (1 - d) \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k} + d \int_{-\infty}^{\infty} \rho_1(x) \log \frac{1}{\rho_1(x)}\,dx$$

$$+ d \log \frac{1}{d} + (1-d) \log \frac{1}{1-d} .$$

Demostración: Pongamos

$$p_{nk}^{(\circ)} = F_\circ\left(\frac{k+1}{n}\right) - F_\circ\left(\frac{k}{n}\right),$$

$$p_{nk}^{(1)} = F_1\left(\frac{k+1}{n}\right) - F_1\left(\frac{k}{n}\right).$$

En este caso

$$p_{nk} = F\left(\frac{k+1}{n}\right) - F\left(\frac{k}{n}\right) = (1-d)\ p_{nk}^{(\circ)} + d\ p_{nk}^{(1)}\ .$$

En general, para a>0, b>0, se tiene

$$0 \leq \left(a \log \frac{1}{a} + b \log \frac{1}{b} - (a+b) \log \frac{1}{a+b}\right)$$

$$= (a+b) \quad \frac{a}{a+b} \log \frac{a+b}{a} + \frac{b}{a+b} \log \frac{a+b}{b} \quad \leq a+b.$$

Designando por $\xi_\circ$ y ξ_1 las v.a. cuyas funciones de distribución son $F_\circ$ y F_1 respectivamente, y poniendo

$$(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$$

y

$$\lambda_{nk} = \frac{d\ p_{nk}^{(1)}}{p_{nk}},$$

obtendremos

$$0 \leq dH^{(\circ)}(|n\xi_1|) + (1-d)\ H^{(\circ)}(|n\xi_\circ|)$$

$$- H^{(\circ)}(|n\xi|) + \quad (d) = \sum_{k=-\infty}^{\infty} p_{nk} \quad (\lambda_{nk}). \qquad (27.11.A)$$

Escojamos una $\varepsilon > 0$. Sea $A_n(\varepsilon)$ el conjunto de índices k para el cual $\lambda_{nk} \leq \varepsilon$, y $B_n(\varepsilon)$ el conjunto de índices k para el cual $1-\lambda_{nk} \leq \varepsilon$. Escojamos ahora un número $N_1 = N_1\ (\varepsilon)$, tal que

$$\sum_{\ell=N_1}^{\infty} p_\ell < \varepsilon^2 ,$$

así como un número $N_2 = N_2(\varepsilon)$ tal que $N_2 > N_1$ y

$$\left| x_i - x_j \right| > \frac{1}{N_2} , \quad 1 \leq i, \quad j \leq N_1 .$$

Además, sea $C_n(\varepsilon)$ el conjunto de índices k para los cuales el intervalo $\left| \frac{k}{n} , \frac{k+1}{n} \right.$ contiene uno de los valores x_i, $1 \leq i \leq N_1$.

Si $n \geq N_2$, todo intervalo $\left| \frac{k}{n} , \frac{k+1}{n} \right.$ puede contener a lo más una x_i. Por supuesto podemos encontrar una $N_3 > N_2$, tal que si $n \geq N_3$, $k \varepsilon C_n(\varepsilon)$ y x_i, $1 \leq i \leq N_1$, esté contenido en el intervalo $\left| \frac{k}{n} , \frac{k+1}{n} \right.$, entonces

$$p_{nk}^{(1)} \leq \frac{\varepsilon(1-d)\ p_i}{d} ,$$

por consiguiente

$$\lambda_{nk} \leq \varepsilon.$$

En consecuencia, para $n \geq 3$, el conjunto $C_n(\varepsilon)$ es un subconjunto de $A_n(\varepsilon)$. Designemos con $D_n(\varepsilon)$ el conjunto de los índices k que no pertenecen a la unión $A_n(\varepsilon) \cup B_n(\varepsilon)$. Es evidente que si $k \varepsilon D_n(\varepsilon)$, tenemos

$$d\ p_{nk}^{(1)} < \left(\frac{1-\varepsilon}{\varepsilon}\right) \ p_{nk}^{(\circ)}(1-d),$$

y por lo tanto

$$p_{nk} \leq \frac{1}{\varepsilon} p_{nk}^{(\circ)} .$$

La inclusión

$$C_n(\varepsilon) \subset A_n(\varepsilon)$$

implica que

$$\sum_{k\varepsilon Dn(\varepsilon)} p^{(\circ)}_{nk} \leq \varepsilon^2,$$

y teniendo en cuenta que

$$(\rho) \leq 1, \qquad 0 \leq p \leq 1$$

(con igualdad si y solamente si $p = \frac{1}{2}$), deducimos que

$$\sum_{k\varepsilon Dn(\varepsilon)} p_{nk} \quad (\lambda_{nk}) \leq \sum_{k\varepsilon Dn(\varepsilon)} p_{nk} \leq \frac{1}{\varepsilon} \sum_{k\varepsilon Dn(\varepsilon)} p^{(\circ)}_{nk} \leq \varepsilon. \qquad (28.II.A)$$

Por otra parte, como

$$\max_{0<\lambda\leq\varepsilon} \quad (\lambda) = \max_{1-\varepsilon\leq\lambda\leq 1}$$

deducimos que

$$\sum_{k\varepsilon An(\varepsilon) \cup Bn(\varepsilon)} p_{nk} \quad (\lambda_{nk}) \leq \quad (\varepsilon) \qquad (29.II.A)$$

De (27.II.A) - (29.II.A) deducimos que, para $n \geq N_3 = N_3(\varepsilon)$,

$$0 \leq (1-d)\ H^{(\circ)}(|n\xi_\circ|) + d\ H^{(\circ)}(|n\xi_1|) - H^{(\circ)}(|n\xi|)$$

$$+ \quad (d) \leq \varepsilon + \quad (\varepsilon) \qquad (30.II.A)$$

Ahora bien,

$$\lim_{n\to\infty} \frac{H^{(\circ)}(|n\xi_1|)}{\log n} = 1, \qquad \lim_{n\to\infty} \frac{H^{(\circ)}(|n\xi_\circ|)}{\log n} = 0,$$

de lo que resulta

$$\lim_{n\to\infty} \frac{H^{(\circ)}(|n\xi|)}{\log n} = d.$$

Puesto que, si las expresiones

$$\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k} ,$$

$$\int_{-\infty}^{\infty} \rho_1(x) \log \frac{1}{\rho_1(x)} dx,$$

son convergentes, tenemos

$$\lim_{n\to\infty} H^{(\circ)}(|n\xi_\circ|) = \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$$

y según el teorema (2.II.A)

$$\lim_{n\to\infty} H^{(\circ)}(|n\xi_1|) - \log n) = \int_{-\infty}^{\infty} \rho_1(x) \log \frac{1}{\rho_1(x)} dx;$$

como podemos escoger $\varepsilon > 0$ arbitrariamente pequeño, resulta de (30.II.A) que

$$\lim_{n\to\infty} H^{(\circ)}(|n\xi|) - d \log n)$$

$$= (1-d) \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k} + d \int_{-\infty}^{\infty} \rho_1(x) \log \frac{1}{\rho_1(x)} dx + \quad (d).//$$

Observemos que una función de distribución cualquiera F puede siempre representarse en la forma

$$F = pF_\circ + qF_1 + rF_2 , \quad p \geq 0, \quad q \geq 0, \quad r \geq 0, \quad p + q + r = 1,$$

en donde $F_\circ$ designa la componente discontinua, F_1 la componente absolutamente continua y F_2 la componente singular. Sin embargo para una función de distribución que también admite una componente singular ($r \neq 0$), aun no se han establecido criterios capaces de asegurar la existencia de una dimensión y de una entropía d-dimensional.

Los resultados aquí presentados podemos generalizarlos a variables aleatorias vectoriales (vectores aleatorios) s-dimensionales, es decir a funciones de distribución en un espacio euclideano s-dimensional, de la siguiente manera:

Si ξ es un vector aleatorio s-dimensional, ponemos

$$\xi_n = \frac{1}{n} \quad |n\xi|,$$

ξ es evidentemente un vector aleatorio de distribución de probabilidad discreta, y por lo tanto su entropía $H^{(\circ)}(|\xi|)$ es finita, la cantidad $H^{(\circ)}(\xi_n)$ será también finita para toda $n \geq 1$.

Definimos la dimensión superior e inferior de la siguiente manera,

$$\overline{d}\ (\xi) = \overline{\lim_{n\to\infty}}\ \frac{H^{(\circ)}(\xi_n)}{\log n} \quad ,$$

$$\underline{d}\ (\xi) = \underline{\lim_{n\to\infty}}\ \frac{H^{(\circ)}(\xi_n)}{\log n} \ .$$

Si $\underline{d}(\xi) = \overline{d}(\xi)$, diremos que el vector aleatorio tiene una dimensión, que designaremos con $d(\xi)$,

$$d(\xi) = \lim_{n\to\infty}\ \frac{H^{(\circ)}(\xi_n)}{\log n} \quad ,$$

si este limite existe. Se demuestra con el mismo procedimiento que en el caso unidimensional que

$$0 \leq \underline{d}\ (\xi) \leq \overline{d}\ (\xi) \leq s.$$

Si

$$d(\xi) = d,$$

se introduce la entropía d-dimensional de ξ

$$H^{(d)}(\xi) = \lim_{n\to\infty} (H^{(°)}(\xi_n) - d \log n)$$

si el límite del segundo miembro existe. Se demuestra de la misma manera que para el teorema (2.II.A), el siguiente resultado:

Teorema 4.II.A.- Si ξ es un vector aleatorio con valores en R^s, $s \geq 2$, cuya distribución de probabilidad es absolutamente continua con densidad ρ, y si $H^{(°)}(|\xi|)$ es finita, entonces $d(\xi) = s$, y tenemos

$$H^{(s)}(\xi) = \int_{R^s} \rho(x_1,\ldots x_s) \log \frac{1}{\rho(x_1,\ldots,x_s)} dx_1,\ldots,dx_s ,$$

si la integral del segundo miembro existe.

La noción de entropía d-dimensional fue introducida por A. Renyi en 1959. Obsérvese que la dimensión de la v.a. coincide con la dimensión de su imagen: 0 para puntos aislados, 1 para segmentos, s para conjuntos de interior no vacío de R^s. Esto establece una relación estrecha entre los conceptos de entropía y dimensión.

B) La ε-entropía

Ahora nos ocupamos de v.a. con valores en un espacio métrico abstracto. Para tales variables aleatorias, A.N. Kolmogorov y V.N. Tijomirov introdujeron en 1956 lo que se llama ε-entropía. Nosotros seguiremos la versión simplificada de A. Renyi (1959) para obtener la ε-entropía.

Sea X un espacio métrico, de métrica d y totalmente acotado. Esto significa que para toda $\varepsilon>0$ podemos dividir X en un número finito de conjuntos ajenos, cada uno con un diámetro menor o igual que ε. Designemos con $N_X(\varepsilon)$ el número de esos conjuntos. Sea

$$X_1(\varepsilon), \ldots, X_{n(\varepsilon)}(\varepsilon)$$

una ε-partición, es decir, un sistema de conjuntos ajenos cuya unión es X, y tales que cada uno tiene un diámetro menor o igual que ε. Supongamos que esta ε-partición es asintóticamente mínima, o sea que

$$\lim_{\varepsilon \to 0} \frac{\log n(\varepsilon)}{\log N_X(\varepsilon)} = 1.$$

Supongamos además que los conjuntos $X_k(\varepsilon)$, $1 \leq k \leq n(\varepsilon)$ pertenecen a la menor σ-álgebra de subconjuntos de X que contienen todas las esferas; es decir los conjuntos $S_a(r)$ de puntos x que satisfacen la condición

$$d(a,x) < r,$$

donde $a \varepsilon X$ y $r > 0$ son arbitrarios; en otras palabras $X_k(\varepsilon)$ es un subconjunto boreliano de X.

Sea ξ una v.a. con valores en X definida en el espacio de probabilidad $\{\Omega, k, p\}$. Sólo consideraremos las v.a. ξ para los cuales $\xi^{-1}(S_a(r)) \varepsilon k$ para cualquier $r > 0$ y $a \varepsilon X$. Pongamos

$$p_k(\varepsilon) = p(\xi^{-1}(X_k(\varepsilon))), \quad 1 \leq k \leq n(\varepsilon)$$

y

$$H(\varepsilon, \xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{p_k(\varepsilon)}$$

Definición 1.II.B.- A $H(\varepsilon, \xi)$ se le llama la ε-*entropía* de la variable aleatoria ξ con respecto a la ε-partición $X_1(\varepsilon), \ldots, X_{n(\varepsilon)}(\varepsilon)$ de X.

Valiéndonos de la desigualdad (3.II.A), resulta inmediatamente que

$$H(\varepsilon,\xi) \leq \log n(\varepsilon).$$

Pongamos ahora

$$\underline{D}(\xi) = \underline{\lim_{\varepsilon\to 0}} \frac{H(\varepsilon,\xi)}{\log n(\varepsilon)} = \underline{\lim_{\varepsilon\to 0}} \frac{H(\varepsilon,\xi)}{\log N_x(\varepsilon)}$$

y

$$\overline{D}(\xi) = \overline{\lim_{\varepsilon\to 0}} \frac{H(\varepsilon,\xi)}{\log n(\varepsilon)} = \overline{\lim_{\varepsilon\to 0}} \frac{H(\varepsilon,\xi)}{\log N_x(\varepsilon)}$$

En el caso en que

$$\underline{D}(\xi) = \overline{D}(\xi),$$

definimos

$$D(\xi) = \lim_{\varepsilon\to 0} \frac{H(\varepsilon,\xi)}{\log n(\varepsilon)} = \lim_{\varepsilon\to 0} \frac{H(\varepsilon,\xi)}{\log N_x(\varepsilon)}$$

La definición dada por A. Renyi de la ε-entropía es considerablemente más simple que la dada por A. N. Kolmogorov. Pero por otra parte de la definición de Renyi da la ε-entropía con respecto a una ε-partición dada en el espacio métrico X. De aquí se sigue que la ε-entropía arriba introducida depende de la ε-partición considerada. Sin embargo se puede demostrar (y nosotros lo haremos para un caso particular) que ésta dependencia es muy débil.

Por otra parte la ε-partición de X puede escogerse a veces de manera que la evaluación de la ε-entropía se simplifique, lo cual en ciertos casos puede constituir una verdadera ventaja.

Para poder demostrar cuan poco depende la ε-entropía $H(\varepsilon,\xi)$, de la elección del sistema de conjuntos que constituyen la ε-partición del espacio métrico X, consideremos en detalle el caso unidimensional.

Teorema 1.II.B.- Sea ξ una v.a. con valores en el intervalo $|0,1|$. Supongamos que la distribución de probabilidad de ξ es absolutamente continua, con la densidad de distribución ρ. Sea para $\varepsilon>0$ la ε-partición

$$0 = x_0(\varepsilon) < x_1(\varepsilon) <,\ldots,< x_{n(\varepsilon)}(\varepsilon) = 1$$

del intervalo $|0,1)$, de manera que

$$\lim_{\varepsilon\to 0} \frac{\log n(\varepsilon)}{\log \frac{1}{\varepsilon}} = 1. \tag{2.II.B}$$

Pongamos

$$p_k(\varepsilon) = \int_{x_{k-1}(\varepsilon)}^{x_k(\varepsilon)} \rho(t)\,dt \tag{3.II.B}$$

y

$$H(\varepsilon,\xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{p_k(\varepsilon)} . \tag{4.II.B}$$

Tenemos

$$\lim_{\varepsilon\to 0} \frac{H(\varepsilon,\xi)}{\log \frac{1}{\varepsilon}} = 1.$$

Demostración: En este caso X = $|0,1)$. Con las condiciones del teorema (1.II.B), tenemos evidentemente

$$\left|\frac{1}{\varepsilon}\right| \leq n(\varepsilon) \leq \left|\frac{1}{\varepsilon}\right| + 1.$$

Si $X_k(\varepsilon) = x_{k-1}(\varepsilon), x_k(\varepsilon))$, entonces conforme a la condición (2.II.B) la ε-partición $X_1(\varepsilon),\ldots,X_{n(\varepsilon)}(\varepsilon)$ es admisible. Teniendo en cuenta (3.II.B), (4.II.B), tenemos

$$H(\varepsilon,\xi) \leq \log n(\varepsilon)$$

y por consiguiente, de (2.II.B) deducimos

$$\overline{D}(\xi) = \overline{\lim_{\varepsilon\to 0}} \frac{H(\varepsilon,\xi)}{\log \frac{1}{\varepsilon}} \leq \overline{\lim_{\varepsilon\to 0}} \frac{\log n(\varepsilon)}{\log \frac{1}{\varepsilon}} = 1. \qquad (5.II.B)$$

Con la notación

$$d_k(\varepsilon) = x_k(\varepsilon) - x_{k-1}(\varepsilon), \quad 1 \leq k \leq n(\varepsilon),$$

puesto que

$$d_k(\varepsilon) \leq \varepsilon,$$

tenemos

$$H(\varepsilon,\xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{d_k(\varepsilon)}{p_k(\varepsilon)} + \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{d_k(\varepsilon)}$$

$$\geq A(\varepsilon) + \log \frac{1}{\varepsilon}, \qquad (6.II.B)$$

en donde

$$A(\varepsilon) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{d_k(\varepsilon)}{p_k(\varepsilon)}.$$

Con los mismos argumentos utilizados para demostrar el teorema (2.II.A), obtenemos

$$\lim_{\varepsilon\to 0} \frac{A(\varepsilon)}{\log \frac{1}{\varepsilon}} = 0; \qquad (7.II.B)$$

por lo tanto, de (6.II.B) y (7.II.B) deducimos que

$$\underline{D}(\varepsilon) = \lim_{\varepsilon\to 0} \frac{H(\varepsilon,\xi)}{\log \frac{1}{\varepsilon}} \geq 1 . \qquad (8.II.B)$$

En fin, de (5.II.B) y (8.II.B) resulta que

$$D(\xi) = 1. \quad //$$

En lo que concierne a la diferencia

$$H(\varepsilon,\xi) - \log \frac{1}{\varepsilon},$$

vamos a demostrar el

Teorema 2.II.B.- Sea ξ una v.a. con valores en el intervalo $|0,1)$.- Supongamos que la distribución de probabilidad de ξ es abaoluta-mente continua con densidad de distribución ρ, y que la integral

$$H^{(1)}(\xi) = \int_0^1 \rho(x) \log \frac{1}{\rho(x)} dx$$

existe. Supongamos dada para toda $\varepsilon>0$, una ε-partición

$$0 = x_0(\varepsilon) < x_1(\varepsilon) <,\ldots,< x_{n(\varepsilon)}(\varepsilon) = 1$$

del intervalo $|0,1)$, tal que

$$\lim_{\varepsilon\to 0} \varepsilon\, n(\varepsilon) = 1 \qquad (9.II.B)$$

y

$$\frac{\varepsilon}{C} \leq d_k(\varepsilon) \leq \varepsilon,$$

en donde

$$d_k(\varepsilon) = x_k(\varepsilon) - x_{k-1}(\varepsilon), \qquad 1 \leq k \leq n(\varepsilon),$$

y $C > 1$ es una constante positiva que no depende de ε. Si

$$p_k(\varepsilon) = \int_{x_{k-1}(\varepsilon)}^{x_k(\varepsilon)} \rho(t)\,dt$$

y si

$$H(\varepsilon,\xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{p_k(\varepsilon)} \quad ,$$

entonces

$$\lim_{\varepsilon\to 0} \left(H(\varepsilon,\xi) - \log \frac{1}{\varepsilon} \right) = H^{(1)}(\xi).$$

Demostración: Poniendo

$$\Delta(\varepsilon) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{\varepsilon}{d_k(\varepsilon)}$$

y

$$A(\varepsilon) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{d_k(\varepsilon)}{p_k(\varepsilon)} \quad ,$$

se tiene

$$H(\varepsilon,\xi) = A(\varepsilon) + \Delta(\varepsilon) + \log \frac{1}{\varepsilon} \quad .$$

Valiéndonos de los mismos razonamientos que para la demostración del teorema (2.II.A), obtendremos

$$\lim_{\varepsilon\to 0} A(\varepsilon) = H^{(1)}(\xi) \quad .$$

Resulta que para la demostración del teorema (2.II.B), es suficiente demostrar que

$$\lim_{\varepsilon\to 0} \Delta(\varepsilon) = 0 . \qquad (10.11.B)$$

Para esto, sea una $\delta > 0$ arbitraria y $n_1(\delta)$ el número de los $d_k(\varepsilon)$ para los cuales

$$d_k(\varepsilon) \leq \frac{1}{1+\delta} .$$

Tendremos pues

$$1 = \sum_{k=1}^{n(\varepsilon)} d_k(\varepsilon) \leq n_1(\delta) \frac{\varepsilon}{1+\delta} + (n(\varepsilon) - n_1(\delta))\varepsilon ,$$

lo que implica

$$n_1(\delta) \leq \frac{(n(\varepsilon)\varepsilon - 1)(1+\delta)}{\varepsilon\delta}$$

y entonces

$$\sum_{d_k(\varepsilon) < \frac{\varepsilon}{1+\delta}} d_k(\varepsilon) \leq \frac{n_1(\delta)\varepsilon}{1+\delta} \leq \frac{n(\varepsilon)\varepsilon - 1}{\delta}.$$

Se sigue que si $E(\varepsilon,\delta)$ designa la unión de los intervalos $|x_{k-1}(\varepsilon), x_k(\varepsilon))$, en los cuales

$$d_k(\varepsilon) = x_k(\varepsilon) - x_{k-1}(\varepsilon) < \frac{\varepsilon}{1+\delta} ,$$

tendremos de (9.11.B)

$$\lim_{\varepsilon\to 0} \mu(E(\varepsilon,\delta)) = 0,$$

en donde μ es la medida de Lebesgue. Puesto que

$$\Delta(\varepsilon) \leq \log (1+\delta) + \log C \quad {}_{E(\varepsilon,\delta)} \rho(x)dx,$$

resulta de la continuidad absoluta de la medida

$$\nu(G) = \int_G \rho(x)dx$$

que

$$0 \leq \overline{\lim_{\varepsilon \to 0}} \ \Delta(\varepsilon) \leq \log (1+\delta).$$

Y como $\delta > 0$ es arbitrario, resulta que la relación (10.II.B) ocurre. //

De nuevo tenemos una conexión entre entropía y dimensión, ahora a través de la ε-entropía. Primeramente, el teorema 1.II.B muestra que la dimensión vía ε-entropía también es 1 para v.a. absolutamente continuas. Por otra parte, el resultado del teorema 2.II.B. podríamos reescribirlo

$$H(\varepsilon,\xi) = 1, \quad \log\frac{1}{\varepsilon} + H^{(1)}(\xi) + 0(1)$$

(con $0(1) \to 0$ cuando $\varepsilon \to 0$), lo que nos da la misma expresión de la definición de entropía d-dimensional (definición 1.II.A).

C) <u>Algo más sobre la entropía</u>.

Se habla de entropía no solamente para distribuciones de probabilidad (discretas según Shannon, absolutamente continuas según Hatori y casos intermedios según Renyi) y variables aleatorias (tomando el espacio de probabilidad inducido por la v.a.), como veremos enseguida,

i) Considérese una <u>fuente discreta</u> (proceso con tiempo discreto), cuya salida consiste de una sucesión de símbolos tomados aleatoriamente de un conjunto finito de símbolos A, llamado <u>alfabeto</u>. Los símbolos se llaman <u>letras</u> y una sucesión bilateral de letras $x = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$

representa toda la "historia" de la fuente. Sea A^z el conjunto de todas estas sucesiones. Conociendo el alfabeto A y la probabilidad p_j; con que se toma cada una de las letras a_j, es posible construir el espacio producto (A^z, β_A, p_A), donde β_A es la σ-álgebra generada por los cilindros del alfabeto, que son subconjuntos de A^z de la forma $\{x\varepsilon A^z | x_{t_i} = a_i,\ i=\overline{1,n}\}$. Para este cilindro C se tiene $p_A(C) = p_1 p_2 \ldots p_n$.

Como el alfabeto A y la medida de probabilidad p_A caracterizan la naturaleza estadística de la fuente, ésta puede denotarse $|A, p_A|$. Este concepto de fuente se debe a McMillan (1953), pues Shannon sólo consideró fuentes con el carácter de cadenas de Markov estacionarias.

Consideremos ahora el siguiente operador T sobre A^z, que representa un "corrimiento" de una unidad de tiempo.

$$T_x = (\ldots, x'_{-1}, x'_o, x'_1, \ldots),$$

donde $x'_k = x_{k+1}$ $(-\infty < k < \infty)$. Si $S\subset A^z$ y $TS = \{Tx | x\varepsilon S\}$, es claro que $TS\varepsilon\beta_A$ para todo $S\varepsilon\beta_A$ y que $TA^z = A^z$.

Si $p_A(TS) = p_A(S)$ para todo $S\varepsilon\beta_A$ se dice que la fuente es estacionaria (no cambia con el tiempo la probabilidad con que se escogen las letras).

El conjunto $S\subset A^z$ es invariante si $TS = S$. Por ejemplo A^z es invariante y para cada $x\ \varepsilon A^z$ el conjunto $\{T^k x | k\varepsilon z\}$ también lo es. La fuente $|A, p_A|$ se dice ergódica si todo $S\varepsilon\beta_A$ invariante tiene probabilidad 0 ó 1.

ii) Desde el punto de vista de la teoría de la información, la característica más importante de toda fuente es la rapidez con la que emite información, es decir, la cantidad promedio de información dada por cada símbolo emitido.

Tomemos n símbolos sucesivos x_t, $x_{t+1},\dots,x_{t+n-1}$ emitidos por la Fuente. Si A contiene a letras, tenemos a^n sucesiones de n términos, cada una de las cuales genera un cilindro C de β_A de probabilidad $p_A(C)$. Estas sucesiones y probabilidades determinan un espacio de probabilidad finito de entropía

$$H_n = -\sum_C p_A(c) \log p_A(c).$$

Si la fuente es estacionaria, H_n no depende de la "posición - inicial" t, sino únicamente de n y de la naturaleza de la fuente. Entonces

$$\frac{H_n}{n} .$$

es la cantidad de información por símbolo que en promedio proporciona la fuente en una sucesión de n símbolos. Definimos la <u>entropía de la fuente</u> mediante

$$H = \lim_{n\to\infty} \frac{H_n}{n} .$$

En caso de existir, este límite depende únicamente de la naturaleza de la fuente (i.e. del alfabeto A y la distribución de probabilidad p_A).

El hecho de que la entropía realmente existe para todas las fuentes estacionarias es el primer teorema fundamental de la teoría general de las fuentes discretas.

iii) Siguiendo este orden de ideas, A.N. Kolmogorov introdujo en 1959 la noción de <u>entropía de un sistema dinámico</u> por unidad de tiempo, la que fue perfeccionada por I.G. Sinai en ese mismo año y demostraron que es un invariante del sistema dinámico.

Dos espacios de probabilidad (X,A,P) y (Y,β,Q) se dicen isomorfos si existe una función ϕ: $X \to Y$ biyectiva tal que $\phi(A) \subset \beta$, $\phi^{-1}(\beta) \subset A$, $P(\phi^{-1}(B)) = Q(B)$ y $Q(\phi(A)) = P(A)$; en este caso, ϕ es un <u>isomorfismo</u>, llamado automorfismo en el caso de que los dos espacios sean uno mismo.

Consideremos el espacio de probabilidad (X,β,P) y un automorfismo T en él. La terna (X,P,T) se llama sistema dinámico. Tomemos una partición finita $C = \{C_1,\ldots,C_r\}$ de conjuntos medibles y hagamos

$$\mu_{i_1 \ldots i_r} = \mu(C_{i_1} \; , TC_{i_2} \; \ldots \; T^{r-1}C_{i_r}).$$

Entonces

$$H = \operatorname*{Sup}_{C} \lim_{r\to\infty} -\frac{1}{r} \sum \mu_{i_1 \ldots i_r} \log \mu_{i_1 \ldots i_r} \; ,$$

Donde C corre sobre todas las particiones medibles finitas del espacio, se llama entropía por unidad de tiempo del sistema dinámico (X,P,T).

No es difícil demostrar que si dos sistemas dinámicos (X,P,T) y (Y,Q,S) son isomorfos, tienen la misma entropía, es decir, la entropía es un invariante. En 1970, D.S. Orstein demostró que para corrimientos de Bernoulli que son sistemas ergódicos) el invariante es completo, es decir, si dos corrimientos de Bernoulli tienen la misma entropía, son isomorfos.

C A P I T U L O III

EL CONCEPTO DE DIMENSION

A) Dimensión Topológica

En los cursos de licenciatura, el concepto de dimensión con el que casi exclusivamente se trabaja es el de dimensión algebraica: la dimensión de un espacio vectorial. Esto puede compararse aproximadamente con la situación que imperaba antes del advenimiento de la teoría de conjuntos, cuando la idea de dimensión era bastante vaga pues se decía que una configuración era n-dimensional si el número mínimo de parámetros reales necesarios para describir sus puntos, de alguna manera no especificada, era n. Se esperaba que esta noción de dimensión tuvieron un significado topológico.

Dos acontecimientos pusieron en entredicho esa idea de dimensión. Por una parte Georg Cantor estableció una correspondencia biunívoca entre los puntos de la recta y los puntos del plano, y por la otra Giusseppe Peano estableció su mapeo continuo de un intervalo sobre todo un cuadrado. El primero contradecía la creencia intuitiva de que un plano tiene más puntos que una recta y hacía ver que la dimensión podía alterarse mediante una función biyectiva. El segundo contradecía la creencia de que la dimensión de un espacio podía definirse como el mínimo número de parámetros reales continuos requeridos para describir el espacio, y mostraba que la dimensión podía aumentarse mediante una trasformación continua.

Ahora bien, para que la dimensión tenga un significado topológico, debe preservarse ante homeomorfismos, es decir funciones biyectivas y bicontinuas; o sea funciones que cumplan a la vez las características de la función de Cantor y de la función de Peano. Fue hasta 1911 cuando se demostró que los espacios euclidianos n dimensionales y los espacios euclidianos m-dimensionales no son homeomorfismos, a menos que m=n. Fue Brouwer - quien dió la primera demostración de este hecho, pero ésta no revelaba explícitamente propiedad topológica simple alguna que fuera responsable de - la no existencia del homeomorfismo.

En 1913, Brouwer introdujo una función de valores enteros topológicamente invariante que para los espacios métricos separables localmente conexos es equivalente a la utilizada actualmente, llamada "Dimension grad". El artículo de Brouwer pasó casi desapercibido, y fue hasta 1922 cuando Menger y Urysohn, de manera independiente entre sí y de Brouwer, reconstruyeron y mejoraron la definición de Brouwer, elaborado con ella una útil teoría que unificó y ordenó esta gran rama de la geometría.

La definición de Menger dice que

a) el conjunto vacío tiene dimensión -1, y que
b) la dimensión de un espacio es el mínimo entero n para el cual todo punto tiene vecindades arbitrariamente pequeños cuyas fronteras tienen dimensión menor que n.

Esta definición es intuitiva y sencilla, y tiene un carácter inductivo. La definición de Urysohn, del mismo tipo, será dada más adelante.

Actualmente hay diversas vías para definir la dimensión topológica, pero nos interesa el enfoque dado por H. Lebesgue, quien observó que un cuadrado puede cubrirse con pequeños "ladrillos" de modo que ninguno de sus puntos quede contenido en menos de tres de esos "ladrillos", pero si los "ladrillos" son suficientemente pequeños, al menos tres tienen un punto en común; de manera similar, un cubo en un espacio euclidiano n-dimensional puede ser descompuesto en ladrillos arbitrariamente pequeños de modo que no se intersectan más de n+1. Lebesgue conjeturó en 1911 que el número n+1 no podía reducirse más y Brouwer lo demostró en 1913. En esta primera parte del capítulo trataremos de ver la equivalencia entre los enfoques de Urysohn y Lebesgue para la dimensión topológica. Los espacios que se trabajan aquí son espacios métricos separables o compactos debido a las grandes dificultades que se encuentran al tratar de extender la teoría de la dimensión a espacios más generales.

Definición (Urysohn). El conjunto vacío y solamente el conjunto vacío tiene *dimensión topológica* -1.

Un espacio X tiene *dimensión topológica* $\leq n$ $(n \geq 0)$ *en un punto* p si p tiene vecindades arbitrariamente pequeñas cuyas fronteras tienen dimensión topológica $\leq n-1$.

X tiene *dimensión topológica* $\leq n$, $dtX \leq n$, si X tiene dimensión topológica $\leq n$ en cada uno de sus puntos.

X tiene *dimensión topológica* n *en un punto* p si es cierto que X tiene dimensión topológica $\leq n$ y es falso que X tiene dimensión topológica $\leq n-1$ en p.

X tiene *dimensión topológica* n si es cierto que $dtX \leq n$ y es falso que $dtX \leq n-1$.

X tiene *dimensión topológica* ∞ si es falso que $dtX \leq n$ para cada n.

A partir de esta definición, no es muy complicado demostrar afirmaciones como

i) La dimensión topológica es un invariante topológico

ii) $dtX \leq n$ si y sólo si existe una base de abiertos cuyas fronteras tienen dimensión topológica n-1.

iii) Si $dtX = n$, n finito, y $-1 \leq m \leq n$, existe un subconjunto m dimensional de X.

iv) Todo subespacio de un espacio de dimensión topológica $\leq n$ tiene dimensión topológica $\leq n$.

v) $dtX \leq n$ si y sólo si todo punto $p \varepsilon X$ puede ser separado, mediante un conjunto cerrado de dimensión topológica $\leq n-1$, de cualquier conjunto cerrado C que no contenga a p.

Siguiendo adelante con la teoría, se demuestran proposiciones más elaborados como

vi) Si A y B son subespacios de X, $dt(A+B) \leq 1 + dtA + dtB$

vii) Un espacio que es la suma numerable de subconjuntos cerrados de dimensión topológica $\leq n$ tiene dimensión $\leq n$.

viii) Un espacio tiene dimensión topológica $\leq n$, n finita, si y sólo si es la suma de n+1 subespacios de dimensión ≤ 0.

ix) Si A x B denota el producto topológico de dos espacios A y B (producto cartesiano con la base dada por los conjuntos de la forma U x V, donde U es abierto de A y V es abierto de B), entonces

$$dt(A \times B) \leq dtA + dtB.$$

x) Un espacio compacto tiene dimensión topológica $\leq n$ si y sólo si dos cualesquiera de sus puntos pueden separarse con un conjunto cerrado de dimensión $\leq n-1$.

xi) El espacio euclidiano n-dimensional tiene dimensión topológica n.

No damos las demostraciones de estas afirmaciones porque nos alejaríamos del propósito de este trabajo. El lector interesado puede consultar |2|. Pasaremos a la caracterización de dimensión topológica que sigue la idea de Lebesgue.

Sea X un espacio métrico compacto con función de distancia ρ. Si U es subconjunto de X, su *diámetro* es

$$\delta(U) = \sup \{\rho(x,y) \mid x,y \in U\}.$$

Por la compacidad de X, sólo consideraremos cubiertas abiertas finitas. El *orden* de una cubierta es el máximo entero n tal que existen n+1 conjuntos de la cubierta que tienen intersección no vacía, y la *malla* de la cubierta es el número.

$$\mu = \sup\{\delta(U) \mid U \in \ \}.$$

La cubierta es más fina que la cubierta si cada miembro de está contenido en algún miembro de .

Teorema.- Un espacio métrico compacto X tiene dimensión topológica $\leq n$ si y sólo si para cada $\eta > 0$ existe una cubierta abierta finita de malla $< \eta$ y orden $\leq n$.

La demostración será dividida en dos partes:

a) Si $dtX \leq n$, entonces tiene cubiertas abiertas de malla arbitrariamente pequeña y orden $\leq$ n.

b) Si X tiene cubiertas de malla arbitrariamente pequeña y orden $\leq$ n, entonces dtX $\leq$ n.

Lema.- Sea X un espacio, MCX, dtM $\leq$ 0. Sean U_1 y U_2 dos abiertos de X que cubren M. Entonces existen dos abiertos V_1 y V_2 que cubren M y satisfacen

$$V_1 \quad U_1, \quad V_2 \quad U_2 \quad \text{y} \quad V_1V_2 = \phi$$

Demostración.- Si dtM = -1, es inmediato. Sea dtM = 0.

Podemos suponer que $X = U_1 + U_2$. Entonces

$$C_1 = X - U_2, \qquad C_2 = X - U_1$$

son conjuntos cerrados ajenos. Como dtM = 0, existe un conjunto cerrado B que no intersecta a M y separa a C_1 y C_2. Esto implica la existencia de los conjuntos V_1 y V_2 señalados. //

Lema 2.- Sea X un espacio, M c X con dtM $\leq$ 0 y $U_1, U_2, \ldots, U_r$ una cubierta abierta de M. Entonces existe una cubierta abierta $V_1, V_2, \ldots, V_r$ de M tal que

$$V_i \text{ c } U_i, \qquad i = 1,2,\ldots,r,$$

$$V_i \; V_j = \phi, \qquad i = J.$$

Demostración.- Se hace por inducción, utilizando el lema 1. //

Lema 3.- Sea X un espacio, dtX $\leq$ n y una cubierta de X. Entonces existe una cubierta más fina que de orden $\leq$ n.

Demostración.- Si n = ∞, el resultado es inmediato. Sea n finito.

Por viii),

$$X = A_1 + A_2 + \ldots + A_{n+1}$$

donde $dtA_i \leq 0$ para cada i. Como es una cubierta para cada A_i, usando el lema 2 obtenemos n+1 colecciones finitas ^{i} de conjuntos abiertos:

$$^i = \{V^i_1, \ldots, V^i_{r(i)}\}, \qquad i = 1, 2, \ldots, n+1,$$

tales que

$$^i \text{ es una cubierta de } A_i \text{ y } V^i_J \quad V^i_k = \phi$$

si $J \neq k$.

Sea la cubierta de X formada con todos los conjuntos V^i_J, $i = 1, \ldots, n+1$, $J = 1, \ldots, r(i)$. Entonces tiene orden $\leq n$, ya que si tomamos n+2 miembros de , necesariamente contienen dos conjuntos de una de las subcubiertas ^{i} (y estos dos conjuntos son ajenos).

Teorema a).- Sea X un espacio métrico compacto de dimensión $\leq n$. Entonces X tiene cubiertas de malla arbitrariamente pequeña y orden $\leq n$.

Demostración.- Para cualquier $\varepsilon > 0$, consideremos la colección de todas las vecindades esféricas en X de radio $\varepsilon/2$. Puesto que X es compacto, existe una cubierta finita cuyos elementos son conjuntos de esa colección, y por lo tanto su malla es $\leq \varepsilon$. Por el *lema 3*, esta cubierta tiene un refinamiento de orden $\leq n$.

Pasamos ahora a la segunda parte

Teorema b).- Sea X un espacio métrico compacto. Si X tiene cubiertas de malla arbitrariamente pequeña y orden $\leq n$, entonces $dtX \leq n$.

Idea de la demostración:

Un *espacio n-dimensional* se dice *universal* si todo espacio de dimensión topológica $\leq$ n puede sumergirse topológicamente en él.

Sea μ^n_{2n+1} el conjunto de todos los puntos de E^{2n+1} que tienen a lo más n coordenadas racionales, sea I_{2n+1} el cubo unitario en E^{2n+1}. Entonces $X_n = \mu^n_{2n+1} \cdot I_{2n+1}$ es un espacio n-dimensional universal, y resulta que si toda cubierta de un espacio X tiene un refinamiento de orden $\leq$ n, entonces existe un homeomorfismo $h: X \to I_{2n+1}$ tal que $\overline{h(x)} \subset x_n$, pero como x_n tiene dimensión n, entonces dt$X \leq$ n. Por último, las hipótesis del teorema b aseguran la existencia de esos refinamientos de orden $\leq$ n. //

La demostración prueba también el siguiente:

Teorema: Un espacio tiene dimensión topológica $\leq$ n si y sólo si toda cubierta tiene un refinamiento de orden $\leq$ n.

B) Dimensión de Hausdorff

Hausdorff (1919) estableció una medida p-dimensional ($0<p \in R$) para espacios métricos arbitrarios. Este es un concepto métrico, en tanto que la dimensión es puramente topológica. Aquí se dará la definición de dimensión de Hausdorff y en el inciso siguiente se compararán ambas dimensiones; estas conexiones se deben a Szpilrajn y datan de 1937.

Definición: Sea X un espacio y p un número real, $0 \leq p < \infty$. Dado $\varepsilon>0$, sea

$$m^{\varepsilon}_p = \inf \sum_{i=1}^{\infty} |\delta(U_i)|^p,$$

donde $X = U + U + \ldots$ es una descomposición de X en una colección numerable de subconjuntos de diámetro menor que ε y p denota exponenciación.

La *medida p-dimensional* de X es el número $m_p(x) = \sup\limits_{\varepsilon>0} m_p^{\varepsilon}$

Decimos que X tiene *dimensión de Hausdorff* p si y solamente si

$$p = \sup \{q \in R \mid m_q(X) > 0\},$$

y lo denotamos $\qquad dh(X) = p.$

Como $m_p = 0$ implica que $dh(X) \leq p$, podemos dar la siguiente definición equivalente:

Definición: X tiene dimensión de Hausdorff $\leq p$ si y sólo si para cada $\eta > 0$ existe una cubierta abierta finita $\{U_i\}$ de X tal que $\sum\limits_{i=1}^{N} \delta(U_i)p < \eta$ y en este caso ponemos

$$dh(X) = \inf \{p \mid dh(X) \leq p\}.$$

C) Dimensión Topológica y Dimensión de Hausdorff.

Teorema.- Para todo espacio métrico compacto X se tiene que

$$dt(X) \leq dh(X).$$

Lema 1.- Si X es un espacio tal que $m_{p+1}(X) = 0$, $0 \leq p < \infty$, entonses para casi toda r los cascarones esféricos $S(r) = \{x \mid d(x,x_\circ) = r\}$, $x_\circ \varepsilon X$ fijo, tienen p-medida 0.

Demostración.- Sea E un subconjunto arbitrario de X. Sea

(1) $$r_1 = \inf_{x \varepsilon E} d(x,x_\circ), \quad r_2 = \sup_{x \varepsilon E} d(x,x_\circ)$$

es claro que

(2) $$r_2 - r_1 = \delta\,(E)$$

de (1) y (2) se sigue que

$$\int_{o}^{\infty} |\delta(S(r)). E)|^{p} dr = \int_{r_1}^{r_2} |\delta(S(r). E)|^{p} dr$$

$$\text{(3)} \qquad \leq |\delta(E)|^{p} \int_{r_1}^{r_2} dr \leq |\delta(E)|^{p+1}.$$

Estamos dando que $m_{p+1}(X) = 0$. Se ve que existe una sucesión de descomposiciones de X

$$X = A_1^n + A_2^n + \ldots$$

tal que

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} |\delta(A_i^n)|^{p+1} = 0.$$

De (3) y (4) tenemos

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \int_{o}^{\infty} |\delta(S(r). A_i^n)|^{p} dr = 0.$$

Cada una de las integrales es no negativa; consecuentemente podemos intercambiar integraciones y suma, y entonces

$$\lim_{n\to\infty} \int_{o}^{\infty} \sum_{i=1}^{\infty} |\delta(S(r). A_i^n)|^{p} dr = 0;$$

esto es

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} |\delta(S(r). A_i^n)|^{p} = 0.$$

Por lo tanto existe una sucesión A_i^{nk} de descomposiciones de X tal que

$$\lim_{k\to\infty} \sum_{i=1}^{\infty} |\delta(S(r). A_i^{nk})|^{p} = 0 \qquad \text{para casi toda } r.$$

Esto implica

$$m_p(S(r)) = 0 \text{ para casi todo } r. \quad //$$

Lema 2.- Sea X un espacio tal que $m_{n+1}(X) = 0$, $0 \leq n < \infty$. Entonces $dtX \leq n$.

Demostración.- Como $m_\circ(A) = 0$ implica que $A=\phi$, el lema se sigue por inducción del lema 1. //

Lema 3.- Sea X un espacio tal que $dtX = n$, $0 \leq n < \infty$. Entonces $m_n(X) > 0$.

Demostración .- Sustituyendo n por n+1, el lema 3 resulta el contrapositivo del lema 2. //

Ahora sí podemos demostrar el teorema:

Sea $n = dtX$, entonces $m_n(X) > 0$ por el lema 3 se tendrá que - $dh(X) \geq n$ por la definición de dimensión de Hausdorff. Es decir,

$$dh(X) \geq dt(X). \quad //$$

CAPITULO IV

ENTROPIA Y DIMENSION

A) Dimensión métrica.

Definición 1.IV.A. X tiene dimensión métrica $\leq r$ si y sólamente si para cualquier $\eta > 0$ existe una cubierta abierta finita de tamaño N y malla ε tal que $N\varepsilon^r < \eta$. En este caso escribiremos

$$dm(X) = \inf \{r | dm(X) \leq r\}.$$

De esta definición y de la definición de dimensión de Hausdorff, tenemos el

Teorema 1.IV.A. Para cualquier espacio métrico compacto

$$dh(X) \leq dm(X).$$

Demostración.- Para cualquier cubierta abierta de tamaño N y malla ε, $\delta(U_i) \leq \varepsilon$ y entonces

$$\sum_{i=1}^{n} \delta(U_i)^r \leq N \varepsilon^r ,$$

por lo tanto, si $dm(X) \leq r$ se tiene que $dh(X) \leq r$ y se sigue la conclusión del teorema. //

De este resultado y el del capítulo anterior se sigue que para todo espacio métrico compacto X

$$dtX \leq dh(X) \leq dm(X).$$

B) Dimensión métrica y entropía.

En el capítulo II se definió la ε-entropía de una v.a. $\xi:X\to R$ con respecto a una ε-partición de X como

$$H(\varepsilon,\xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{p_k(\varepsilon)} .$$

y se llegó a que

$$H(\varepsilon,\xi) \leq \log n(\varepsilon),$$

donde $\lim_{\varepsilon\to 0} \frac{n(\varepsilon)}{N(\varepsilon,X)} = 1$ y $N(\varepsilon,X)$ es el mínimo número de conjuntos en una ε-partición de X.

De manera similar, se define la ε-entropía para un espacio métrico compacto X mediante

$$H(\varepsilon,X) = \log N(\varepsilon,X).$$

Así la ε-entropía de X resulta ser la máxima entropía posible para el número mínimo posible de conjuntos de la ε-partición.

Es posible verificar directamente que la dimensión métrica está relacionada con la ε-entropía mediante

$$dm(X) = \inf\{r \mid N(\varepsilon,X) \leq \text{const } (\tfrac{1}{\varepsilon})^r\}$$

C) Dimensión métrica, dimensión de Hausdorff y dimensión Topológica.

Hemos visto ya que $dt(X) \leq dh(X) \leq dm(X)$; el propósito de este inciso es demostrar que siempre existe una métrica equivalente en el espacio compacto X tal que

$$dt(X) = dh(X) = dm(X).$$

Antes de enunciar y demostrar nuestro Teorema, daremos algunos conceptos útiles para nuestra demostración:

Si A y B son subconjuntos de X, entonces decimos que A es denso en B si $\overline{A}$ B. A se llama denso en todas partes si $\overline{A} = X$. Esto significa que para cualquier $x \varepsilon X$ y $\varepsilon > 0$, existe un $y \varepsilon A$ tal que $\rho(x,y) < \varepsilon$.

Si el espacio métrico X tiene un subconjunto numerable denso en todas partes, se llama un espacio métrico separable.

Si (X,ρ) es un espacio métrico, un conjunto $A \subset X$ es llamado denso en ninguna parte, si su clausura $\overline{A}$ no contiene esferas, o equivalentemente si $\overline{A}$ no tiene puntos interiores.

Un conjunto $A \subset X$ es llamado de la primera categoría en X si es la unión de una colección numerable de conjuntos densos en ninguna parte. Un conjunto que no es de primera categoría es llamado de la segunda categoría.

A las intersecciones numerables de conjuntos abiertos se les llama conjuntos G_δ.

Teorema 2.IV.C.- (Teorema de Baire). La intersección numerable de subconjuntos densos abiertos de un espacio completo es densa.

Demostración: Sean D_1, $D_2, \ldots$, subconjuntos densos abiertos de X, y D la intersección de los D_i. Debemos mostrar que cualquier conjunto abierto U de X intersecta a D.

Dado que D_1 es denso, existe un punto p_1 en UD_1, y ya que UD_1 es abierto, es posible hallar una vecindad esférica S_1 de p_1, de diámetro menor que 1, cuya clausura está contenida en UD_1. Si ahora reemplazamos U por S_1 y D_1 por D_2, obtenemos un nuevo punto p_2 y vecindades esféricas S_2 de p_2 de diámetro menor que 1/2, cuya clausura está contenida en S_1D_2. Continuando de esta manera llegamos a una sucesión

$$p_1, p_2, \ldots \qquad (4.IV.C)$$

tal que

$$p_n \varepsilon S_n \,,\ \delta(S_n) < \frac{1}{n}\,, \qquad (5.IV.C)$$

y

$$\overline{S}_n \subset S_{n-1} D_n, \quad \overline{S}_1 \subset D_1. \qquad (6.IV.C)$$

Consecuentemente (4.IV.C) es una sucesión de Cauchy y se sigue que tiene un punto límite p. Se desprende de (6.IV.C) que $p\varepsilon UD$, lo cual prueba el teorema. //

Teorema 3.IV.C.- Para cualquier espacio métrico compacto X existe una métrica equivalente tal que

$$dt(X) = dh(X) = dm(X) \qquad (7.IV.C)$$

Demostración.- Dado que dh(X) está entre dt(X) y dm(X), basta con demostrar que dt(X) = dm(X).

Si dt(X) es infinita, no hay nada que probar. De otra forma, sea $n = dt(X)$, consideraremos para $r>n$ la clase K_r de todas las funciones $f \varepsilon C(X, I_{2_{n+1}})$, donde $I_{2_{n+1}}$ es el cubo unitario en R^{2n+1}, tales que $dm(f(x))<r$. Ya que f(X) es compacto por ser f continua y X compacto, existe para cualquier entero k una cubierta abierta de tamaño N y malla ε tal que $N\,\varepsilon^r < 1/k$. Sea $K_{k,r}$ la clase de todas las funciones f para las cuales una tal cubierta existe; entonces

$$k_r = \mathop{}_{i=1}^{\infty} K_{k,r}. \qquad (8.IV.C)$$

Se ve que $K_{k,r}$ claramente es abierto en la topología de convergencia uniforme en $C(X, I_{2_{n+1}})$ y que K_r es un subconjunto G_δ, (dado f y - $g \varepsilon C\ (X, I_{2_{n+1}})$, existe η tal que $|g-f|< \eta \rightarrow |g(x) - f(x)| < \eta$, tomando el sup de estas distancias i.e. $\sup_{x \varepsilon X} |g(x) - f(x)| < \eta \therefore K_{k,r}$ es, abierto K_r es la intersección numerables de conjuntos abiertos lo que implica que K_r es un - conjunto G_δ).

Consideremos ahora K* la clase de todas las funciones $f \varepsilon C(X, I_{2_{n+1}})$ tales que f(X) está contenido en un polítopo n-dimensional en $I_{2_{n+1}}$. En consecuencia K* es denso en $C(X, I_{2_{n+1}})$, y dado que $r>n$ tenemos $K^* \subset K_r$. En consecuencia K_r es un subconjunto G_δ de $C(X, I_{2_{n+1}})$.

Sea ahora H la clase de homeomorfismos en $C(X, I_{2_{n+1}})$, (i.e. la clase de funciones que son bicontinuas). Entonces H es también un subconjunto denso G_δ de $C(X, I_{2_{n+1}})$, y entonces el conjunto

$$H^* = H \quad \mathop{}_{i=1}^{\infty} \; K_{n+1,i} \qquad (9.IV.C)$$

B I B L I O G R A F I A

|1| S. GUIASU, R. Théodorescu, La Théorie Mathématique de l'Information, Dunod (Paris 1968).

|2| W. HUREWICZ and H. WALLMAN, Dimension Theory, Princeton Math. Series, Vol. 4, Princeton Univ. Press, Princeton, N.J., 1941.

|3| REESE T. PROSSER, Note on Metric Dimension, Am. Math. Soc. Vol. 25, No. 4 (1970) 763-765.

|4| A.N. KOLMOGOROV y V.N. TIHOMIROV, ε-entropy and ε-capacity of Sets in funtion Spaces. Trans. Am. Math. Soc., Vol. 17 (1961) pp. 277 y Sigs.

|5| A.C. VOSBURG, On the relations between Hausdorff dimension and Metric dimension, Pacific J. Math. 23 (1967), 183-187.

|6| SHANNON CLAUDE E., A Mathematical Theory of comunication, Bell Sistem Tech. J., 27, pp. 379-423, 623, 656 (1948).

|7| SHANNON and WEAVER, Mathematical Theory of Comunication. University of Illinois (1964).

|8| A.I. KHINTCHINE, The Mathematical Fundations of Information Theory, Dover (1956).