UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

OTRA TEORÍA MATEMÁTICA DE LA COMUNICACIÓN

Autor: Lcda. Yadira Matute
Tutor: Dra. María Luisa Capodieci

Barquisimeto, Mayo 2.007

UNIVERSIDAD CENTROCCIDENTAL "LISANDRO ALVARADO"
DECANATO DE CIENCIAS Y TECNOLOGIA
MAESTRIA EN CIENCIAS

OTRA TEORÍA MATEMÁTICA DE LA COMUNICACIÓN

Trabajo presentado para optar al Grado de
Magister Scientiarum. Mención Matemática

Autor: Lcda. Yadira Matute
Tutor: Dra. María Luisa Capodieci

Barquisimeto, Mayo 2.007

OTRA TEORÍA MATEMÁTICA DE LA COMUNICACIÓN

Autor: Lcda. Yadira Matute
Tutor: Dra. María Luisa Capodieci

Trabajo de grado aprobado

Dra. María Luisa Capodieci
Tutor

Lcda. Franca Laveglia

Lcdo. Iván Vásquez

Barquisimeto, 16 de mayo de 2.007

Dedicatoria

Dedico este triunfo primeramente a Dios, pues sin él no estaría aquí, ni sería nada de lo que soy.

A mis queridos padres y a mi hermano quienes siempre me han apoyado en todos los proyectos que he emprendido

Agradecimiento

A la profesora María Luisa Capodieci, quien no solo me guiado en el desarrollo de esta investigación, sino que también me ha brindado apoyo moral y su amistad.

A Miguel Ángel Monsalve por su valiosa colaboración en la traducción de ingles a español de algunos documentos usados para el desarrollo del trabajo de investigación.

A mis amigas Yackelin y Carolina quienes compartieron mis estudios y me han brindado su apoyo y su amistad.

A mis amigas Lorena y .... quienes siempre me han incentivado para seguir adelante.

A todas las personas y profesores que han contribuido con mi formación académica.

Indice

Indice de Cuadros

Resumen

En el presente trabajo de investigación se plantea la inquietud de realizar el análisis de la teoría matemática de la comunicación bajo la óptica de otras medidas de Incertidumbre las cuales fueron escogidas de acuerdo a los elementos exigidos en la mencionada teoría, estudiando ventajas y desventajas que ofrece cada medida escogida.

Una vez analizada las medidas de Incertidumbre escogidas, se concluye que la medida de Shannon parece ser la única medida que puede ser utilizada en la teoría matemática de la comunicación.

Palabras claves: Comunicación, incertidumbre, información, codificación.

UNIVERSIDAD CENTROCCIDENTAL "LISANDRO ALVARADO"
DECANATO DE CIENCIAS Y TECNOLOGIA
MAESTRIA EN CIENCIAS

OTRA TEORÍA MATEMÁTICA DE LA COMUNICACIÓN

Autor: Lic. Yadira Matute
Tutor: Dra. María Luisa Capodieci

RESUMEN

En el presente trabajo de investigación se plantea la inquietud de realizar el análisis de la teoría matemática de la comunicación bajo la óptica de otras medidas de Incertidumbre las cuales fueron escogidas de acuerdo a los elementos exigidos en la mencionada teoría, estudiando ventajas y desventajas que ofrece cada medida escogida.

Una vez analizada las medidas de Incertidumbre escogidas, se concluye que la medida de Shannon parece ser la única medida que puede ser utilizada en la teoría matemática de la comunicación.

Palabras claves: Comunicación, incertidumbre, información, codificación.

Introducción

El proceso de la comunicación es una teoría que ha sido estudiada desde hace bastante tiempo. Sin embargo es en 1.948 cuando Claude Shannon, en su trabajo "A mathematical theory of communication", plantea el problema fundamental de la comunicación dado por la reproducción exacta o aproximada de un mensaje emitido en un punto por un emisor y recibido en otro punto por un receptor.

Es así como nace una nueva rama de la Matemática Aplicada llamada Teoría de la Información y la codificación. Dicha rama se ha desarrollado en estos últimos años y pretendemos seguir haciéndolo sobre todo cuando hoy hablamos de una estructura difusa de la información, Capodieci [5]

En el estudio de la Teoría Matemática de la Comunicación se ha utilizado desde sus inicios como medida de información la entropía de Shannon

$$\mathbf{H}(\xi) = -\sum_{i=1}^{n} p(\chi_i) \log p(\chi_i)$$

definiendo a través de ella la cantidad de información de Shannon, especificándola tanto en el caso discreto como el continuo. Entendiéndose que la Transmisión de Mensajes consiste en hacer llegar cierta información contenida en los mensajes desde la fuente al receptor, cuyo objetivo esta centrado en el estudio de la circulación de la información en los elementos directamente relacionados a través del canal; el Codificador y el Traductor.

Las Medidas de Incertidumbre han tenido un desarrollo importante desde 1.948, año de nacimiento de la Teoría de la Información. Iniciándose sus estudios en el contexto de la teoría de la comunicación primero en forma implícita por Hartley, Nyquist después en forma explícita, separada y en dos contextos diferentes, por Wiener y Shannon. [1], [11] y [14].

Es importante destacar que hasta el comienzo de los años '50 el concepto de la probabilidad ha sido prácticamente el único instrumento usado para medir y manejar la incertidumbre. A partir de este periodo se ha puesto de manifiesto que existen situaciones que no pueden describirse, o no pueden describirse por completo, únicamente a través de la probabilidad.

La teoría axiomática propuesta por Forte y Kampé de Fériet, [7] y [8] intenta medir la Información y la Incertidumbre sin utilizar o sin utilizar únicamente las probabilidades. El enfoque proporcionado le da relevancia al "efecto sorpresa" relacionado con los acontecimientos.

Capodieci 1.986 y 1.999 ([4], [5]) y Vásquez 2.003 [15] han trabajado sobre la construcción y clasificación de medidas de incertidumbre, en dichos trabajos se han analizado una cantidad de medidas de incertidumbre verificando en algunos casos las propiedades que satisfacen dichas medidas.

Sin embargo hasta ahora solo se ha visto o se ha dado un intento de volver a la teoría de la comunicación utilizando alguna otra de las medidas estudiadas, analizando la posibilidad de que alguna otra medida de mejores resultados en la comunicación. Dicho trabajo lo esta realizando Capodieci en la actualidad.

En esta tesis consideraremos el trabajo presentado por Khinchin en su libro “Mathematical Foundations of Information Theory“ en el cual hace un análisis completo de la Teoría Matemática de la Comunicación de Shannon. En dicho trabajo Khinchin incorporó a la teoría de Shannon las ideas de Mc. Millan y Feinstein.

Planteamiento del problema.

Tomando como referencia el trabajo realizado hasta ahora por Capodieci en lo que respecta al análisis de la teoría Matemática de la Comunicación bajo la óptica de otras medidas de Incertidumbre, las cuales serán escogidas de acuerdo a los elementos exigidos en la mencionada teoría y sabiendo que uno de los elementos fundamentales en que se basa la entropía de Shannon es la probabilidad, ¿qué pasaría si utilizamos una medida que no dependa de la probabilidad?, ¿se podría hablar de una teoría matemática de la comunicación con otra medida de Incertidumbre?.

Con base en estas interrogantes en este trabajo nos planteamos los objetivos siguientes:

Objetivo General

Estudiar la teoría matemática de la comunicación utilizando una medida de Incertidumbre distinta a la de Shannon.

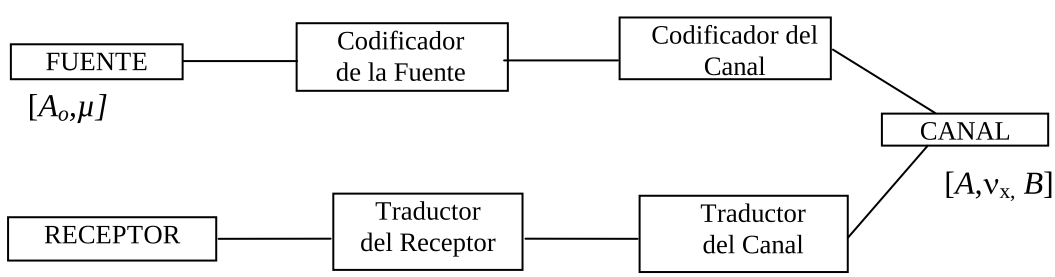
Objetivos Específicos

1. Estudiar la posibilidad de utilizar otra medida de Incertidumbre distinta a la de Shannon en la teoría matemática de la comunicación.
2. Verificar si la medida a utilizar debe depender de las probabilidades.
3. Analizar que sucede si se utiliza una medida que no depende de las probabilidades.

Capitulo I:

Antecedentes

La Transmisión de Mensajes consiste en hacer llegar cierta información contenida en los mensajes desde la fuente al receptor, el esquema estudiado se presenta a continuación.



En dicho esquema se incluyen los elementos esenciales en la transmisión de mensajes, el Codificador y el Traductor; cuyo objetivo esta centrado en el estudio de la circulación de la información a través del canal .

Para iniciar el trabajo que pretendemos desarrollar se introducen algunas definiciones indispensables para entender la teoría que se estudiará.

Definición 1.1.

Se entiende por "**alfabeto**" *A*, a un conjunto de símbolos, los cuales son llamados dígitos.

En el desarrollo de este trabajo se verá que tanto la fuente como el canal tienen su alfabeto, que en algunos casos pueden coincidir. Denotaremos por A_0 el alfabeto de la fuente y por *A* el alfabeto de entrada del canal.

Definición 1.2.

Sea Ω un espacio, $\mathscr{F} = \{E_j \subset \Omega\}$, $j \in I$ contable $\mathscr{F} \neq \emptyset$, una colección $\mathscr{F}$ de subconjuntos de Ω se denomina "**Campo de Borel**" si y solo si se cumple:

i. $E \in \mathscr{F} \Rightarrow E^c \in \mathscr{F}$

ii. $E_j \in \mathscr{F}, 1 \leq j < \infty \Rightarrow \left(\bigcup_{j=1}^{\infty} E_j\right) \in \mathscr{F}$

Definición 1.3.

Sean $t_1, t_2, \ldots, t_n \in Z^+$ tal que $t_1 < t_2 < \ldots < t_n$ y tomemos n símbolos o dígitos del alfabeto A, $\alpha_1, \alpha_2, \ldots, \alpha_n$, los cuales no necesariamente son distintos. Entonces, se conoce como “**cilindro de rango n**” o “**conjunto cilindro**” al conjunto:

$$\mathbb{C}_A = \{x \in A^I \,/\, x_{t_i} = \alpha_i ,\, i\in\{1,2,\ldots n\}\},$$

donde $A^I = \{x = (\ldots,x_{-1}, x_o ,x_1,...)/\ x_i\in A$, para algún $i\in Z\}$.

Definición 1.4.

Dado el alfabeto *A* y la intersección de todos los campos de Borel, F_A , la cual contiene todos los cilindros del alfabeto *A*, se denomina “**Extensión de Borel**” a la intersección F_A .

Se considera una secuencia infinita de dígitos $x = (\ldots,x_{-1}, x_o ,x_1,...)$, $x_i\in A$, a esta secuencia, se denomina “**evento elemental**” y al “**conjunto de todos los eventos elementales**” se representa por A^I, cuyos dígitos son del alfabeto *A*.

A continuación se identifican los elementos esenciales en la transmisión de mensajes como son fuente y canal.

Dados A_0 el alfabeto de la fuente y μ la medida de probabilidad definida para todo $S\in F_A$, con $\mu(A^I)=1$. Denotaremos la “**Fuente**” por$[A_0,\mu]$ y el “**canal**” por $[A,\nu_x, B]$, donde, *A* es el alfabeto de entrada al canal, *B* es el alfabeto de salida del canal, y ν_x (S) es la probabilidad de que “y sea recibido cuando un x dado es transmitido” donde $y\in S$ (esta probabilidad será dada por algún $x\in A^I$ y algún $S\in F_B$). En su trabajo básico, Shannon considera solamente fuentes con el carácter de cadena de Markov estacionaria.

Definición 1.5.

Sea una secuencia $x = (\ldots, x_{-1}, x_o, x_1, \ldots)$, $x_i \in A$, denotamos por Tx la aplicación (secuencia), $T: A^I \to A^I$, definida mediante:

$$Tx = (\ldots, x'_{-1}, x'_o, x'_1, x'_2, \ldots) \text{ donde } x'_k = x_{k+1}, \ (-\infty < k < +\infty)$$

El operador T se denomina "**Shift**" o "**cambio**" en una unidad de tiempo.

Es importante destacar que el conjunto A^I es invariante bajo el operador T, ya que $TA^I = A^I$

Definición 1.6.

$[A,\mu]$ es una "**fuente estacionaria**" si $\mu(TS) = \mu(S)$ para algún conjunto $S \in F_A$.

De aquí en adelante, todas las fuentes serán asumidas estacionarias.

Definición 1.7.

$[A,\mu]$ es una "**fuente ergódica**", si la probabilidad $\mu(S)$ de cada conjunto invariante $S \in F_A$, es 0 o 1.

Definición 1.8.

[A, ν_x, B], es un "**canal estacionario**" si ν_{Tx} (T S) = ν_x(S) para todo x $\in$ A^I y S$\in F_B$.

Definición 1.9.

[A, ν_x, B], se denomina "**canal sin anticipación**" si ν_x (y_n = b) es igual para todos los mensajes x transmitidos, para los cuales las señales $\ldots, x_{n-1}$, x_n son idénticas, donde la probabilidad de que y_n coincida con un dígito b $\in$ B, ν_x (y_n = b), no depende de todos los dígitos del mensaje transmitido x = $\ldots, x_{-1}$, x_0 , x_1 … y la distribución de los y_n es independiente de las señales transmitidas después de x_n, es decir, es independiente de x_k para k > n.

Definición 1.10.

[A, ν_x, B], se denomina "**canal con memoria finita**" si ν_x (y_n = b) es igual para todos los mensajes x transmitidos, para los cuales $x_{n-m}\ldots, x_{n-1}$, x_n son idénticos, donde la probabilidad de que y_n coincida con un dígito b $\in$ B, ν_x (y_n = b), depende de un número finito de dígitos del mensaje transmitido x = …, x_{-1}, x_0, x_1 … digamos x_{n-1}, $x_{n-2}\ldots, x_{n-m}$.

El número m, más pequeño para el cual se cumple lo anterior se denomina la "**memoria del canal**". En particular la distribución de y_n para un canal sin memoria (m=0) depende solo de x_n.

Definición 1.11.

La “**capacidad de un canal**” se define como la máxima cantidad de información que puede circular por él. Se entiende que el máximo esta extendido a todas las posibles distribuciones de probabilidad sobre el alfabeto de entrada.

Capitulo II:

Teoría de Shannon

En este capítulo analizaremos el estudio realizado por Khinchin en su libro “Mathematical Foundations of Information Theory“ en el cual hace un análisis completo de la Teoría Matemática de la Comunicación de Shannon. De acuerdo a las necesidades generadas por el propósito del trabajo que se esta realizando se incorporaran los elementos necesarios para aclarar a los lectores el desarrollo del contenido.

La prueba dada esta basada en las ideas de Shannon, McMillan, y Feinstein.

Se trabaja con un canal con ruido. Esto significa que no podemos determinar la secuencia de símbolos enviados a la entrada del canal de la secuencia recibida en la salida del canal a causa del ruido, puede suceder que dos secuencias diferentes en la entrada del canal den origen a la misma secuencia de salida.

La situación es mejorada considerablemente si conocemos dos grupos de secuencias de n-términos, B_1 y B_2, en la salida del canal tal que

1. B_1 y B_2 no contienen ninguna secuencia en común, y
2. Existe una gran probabilidad de que, la primera de las secuencias enviada es transformada en una de las secuencias del grupo B_1 y la segunda en una de las secuencias del grupo B_2. En este caso, al recibir una secuencia del grupo B_1 en la salida del canal, casi podemos asegurar que el mensaje fue el primero de las dos posibles secuencias, mientras que si una secuencia del grupo B_2 es recibida en la salida, entonces es casi cierto que la segunda secuencia era el mensaje. Bajo las circunstancias que hemos descrito, estas dos secuencias transmitidas son distinguibles.

Así mismo grupos de tres, cuatro y más secuencias distinguibles (hasta ahora no tocamos sobre la pregunta de existencia de tales grupos) pueden ser determinadas en la entrada del canal de la misma forma.

Supongamos que pudimos encontrar un grupo constituido de un número grande k de tales secuencias distinguibles de n-términos en la entrada del canal. Si pudiéramos limitarnos a enviar solamente secuencias que pertenecen a este grupo, entonces de la secuencia recibida en la salida del canal, podríamos determinar casi sin error la secuencia transmitida.

¿Podemos hacer esto?, a fin de hacerlo así, es necesario que el número L de secuencias de n-términos distintos en la salida de la fuente dada, la cual es requerida para transmitir a través del canal, no debe exceder k; ya que bajo esta y solo esta condición será posible codificar todo el grupo de L secuencias, el cual tendría que ser transmitido dentro de nuestro grupo “distinguible” de k secuencias en la entrada del canal. Así, la desigualdad $L< k$ sirve como un criterio de la posibilidad de transmisión casi sin error, y nuestro esfuerzo debe dirigirse hacia el hecho de hacer L tan pequeño como sea posible y k tan grande como sea posible.

La primera meta se logra al usar el teorema de Mcmillan (pag. 20). Al dejar de lado el grupo de “baja probabilidad”, con probabilidad total tan pequeña como se desee, y restringiéndonos a transmitir secuencias del grupo de “alta probabilidad”. La reducción del número de esas secuencias es tal que lo hacemos aproximarse a 2^{nH_o}, donde H_o es la entropía de la fuente dada.

El segundo problema se resuelve usando el lema fundamental de Feinstein (lema 2.1, pag. 23), el cual acepta la existencia de grupos distinguibles para los cuales el número de términos es $k> 2^{n (C - \varepsilon)}$, donde C es la capacidad ergódica del canal, y ε es un número positivo tan pequeño como desee. Si $H_o < C$, entonces vemos de inmediato que es $L< k$ para una selección conveniente de los grupos distinguibles, y nuestro problema se resuelve.

Esta es la idea central de la prueba. Todo el resto es meramente técnico, aunque en esta técnica algunas veces están inmersas dificultades que requieren gran ingenio para resolverlas.

En todas las formulaciones de los teoremas de Shannon que existen en la literatura, estos teoremas están acompañados por la proposición reciproca: “Si $H_o > C$, entonces codificar con el efecto requerido es imposible”. Todos los autores consideran esta declaración como casi obvia y asignan a su prueba solo algunas líneas.

En la versión de la prueba, que aquí se presenta, no se ve la posibilidad de probar esta proposición recíproca. El problema aquí es la definición de C, una definición que usualmente se da sin importancia. Se dice que la capacidad C de un canal dado es el límite más pequeño de la tasa de transmisión sobre la salida del canal de todas las fuentes posibles, el cual coincide con el alfabeto de entrada del canal. Aquí las palabras “todas las fuentes posibles” deben ser reemplazadas por las palabras “todas las fuentes ergódicas posibles” en consecuencia la capacidad que se define se llama “ergódica”. Sin estos cambios, las pruebas tanto de Mc Millan como Feinstein se quiebran en su punto central. Pero si se entiende que C es la capacidad ergódica del canal dado entonces la proposición recíproca mencionada arriba no es solamente obvia, sino que aparentemente requiere nuevas ideas esenciales para su prueba. Esta dificultad puede ser superada utilizando limitaciones conocidas en sistemas de códigos admisibles.

Para iniciar el trabajo consideremos A_o y A los alfabetos de la fuente y de entrada al canal, respectivamente. Estos alfabetos en casi todos los estudios son considerados idénticos, sin embargo en el desarrollo de este trabajo, salvo se indique lo contrario, se trabajará el caso donde los alfabetos A_o y A son diferentes, es decir, el caso general para lo cual se supone que la salida de la fuente estacionaria $[A_o,\mu]$ está siendo transmitida por medio del canal estacionario $[A,\nu_x,B]$.

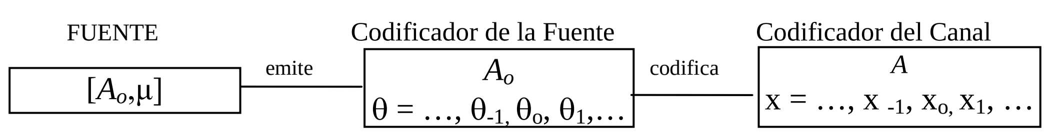
Antes de transmitir, es necesario transformar (codificar) la secuencia de dígitos del alfabeto A_o, emitidos por la fuente, en una secuencia de dígitos del alfabeto A. Usualmente se supone que un mensaje emitido por la fuente $[A_o,\mu]$ esta en la forma de una secuencia

(2.1) $\theta = \ldots, \theta_{-1}, \theta_o, \theta_1,\ldots,$

donde todos los θ_i son dígitos del alfabeto A_o. Luego transformamos (codificamos) cada θ en alguna secuencia x,

(2.2) $x = \ldots, x_{-1}, x_o, x_1, \ldots,$

donde todos los x_i son dígitos del alfabeto A. *La regla de esta transformación constituye el código usado*.



Así, desde el punto de vista matemático, un código es simplemente una función

$$x: A_o^I \to A^I \text{ tal que } \theta \to x(\theta), \text{ es decir, } x = x(\theta),$$

donde A_o^I y A^I es el conjunto de todas las secuencias cuyos dígitos están en el alfabeto A_o y A, respectivamente, es decir, el conjunto de todos los eventos elementales del espacio dado, cuyos dígitos están en el alfabeto A_o y A, respectivamente.

Es claro que formalmente *cada código puede ser considerado como un tipo de canal con alfabeto de entrada A_o y alfabeto de salida A*. Tales canales se conocen como canales sin ruido, es decir, a una secuencia θ en la entrada del canal corresponde una única secuencia $x = x(\theta)$ en la salida. A ese canal se denota por $[A_o, \rho_\theta, A]$, donde la forma de la función $\rho_\theta(M)$ ($M \in A^I$) es

$$\rho_\theta(M) = \begin{cases} 1 & \text{si } x(\theta) \in M \\ 0 & \text{si } x(\theta) \notin M \end{cases}$$

es decir, en los canales sin ruido se cumple que $\forall \theta \in A_o^I, \exists! x \,/\, x = x(\theta)$.

No todos los códigos tienen valor práctico. Veamos el siguiente ejemplo: En general podemos conocer la secuencia completa θ, es decir, el mensaje infinito completo de la fuente $[A_o, \mu]$ para determinar algún dígito x_k del texto codificado. En la práctica, nunca podremos determinar tal dígito. En consecuencia, los códigos importantes para la aplicación, son aquellos tal que, basta conocer alguna secuencia finita de dígitos $\theta_k \in A_o$ para determinar cada dígito $x_k \in A$ (y también cada secuencia finita de tales dígitos).

En particular, en la teoría de la información "*el código secuencial*", usado predominantemente, consiste de lo siguiente: Las secuencias (2.1) y (2.2) están divididas en secuencias finitas de cualquier longitud que son numeradas de izquierda a derecha, tal como los dígitos de esa secuencia. Se establece una regla la cual determina de forma única, la k-ésima subsecuencia de la secuencia x en términos de la k-ésima subsecuencia de la secuencia θ.

Retomando el problema de transmisión de la salida de la fuente $[A_o,\mu]$ a través del canal $[A,\nu_x,B]$, codificamos cada mensaje θ de la fuente $[A_o,\mu]$ en algún mensaje específico x, compuesto de dígitos del alfabeto A. Luego pasamos este x a través del canal $[A,\nu_x,B]$ y obtenemos algún mensaje $y \in B^I$ en su salida.

El resultado de relacionar el código seleccionado al canal $[A,\nu_x,B]$, puede ser considerado como un nuevo canal $[A_0, \lambda_\theta, B]$. Determinar la función de probabilidad $\lambda_\theta(Q)$ $(Q \in F_B)$ es fácil. En efecto como el código transforma el mensaje $\theta \in \mathbf{A}_0^I$ en el mensaje $x(\theta) \in A^I$, la probabilidad $\lambda_\theta(Q)$ de obtener $y \in Q$ para un θ dado es la probabilidad de obtener $y \in Q$ si el mensaje $x = x(\theta)$ entra en el canal $[A,\nu_{x,}B]$, es decir,

$$\lambda_\theta (Q) = \nu_{x(\theta)}(Q),$$

en consecuencia el canal $[A_0, \lambda_\theta, B]$ se puede escribir como $[A_0, \nu_{x(\theta)}, B]$.

El proceso que se esta examinando consiste en relacionar la fuente $[A_o,\mu]$ a ese nuevo canal $[A_0, \lambda_\theta, B]$; aquí el alfabeto de la fuente coincide con el alfabeto de entrada del canal, así que tenemos una situación con la cual estamos familiarizados.

En particular al relacionar la fuente al canal dado, la fuente "compuesta" $[C,\omega]$ con $C= A \text{ X } B$, donde la distribución de probabilidad ω es tal que para $S_1 \in F_{A_o}$, $S_2 \in F_B$ se tiene que

$$(2.3) \qquad \omega (S_1 \text{ X } S_2) = \int_{S_1} \lambda_\theta (S_2) d\mu(\theta) = \int_{S_1} \nu_{x(\theta)} (S_2) d\mu(\theta) .$$

2.1. Primer Teorema de Shannon

Dado lo complicado de la demostración se procede a hacer un proceso inductivo donde se van dando explicaciones en la medida que avanzamos, estableciendo definiciones, proposiciones, lemas, notaciones y otros elementos necesarios para lograr la demostración de este Teorema, el cual será enunciado al final en la pagina 29.

Supongamos que se tiene una fuente ergódica $[A_o,\mu]$ y un canal estacionario, sin anticipación $[A, \nu_x, B]$ con memoria finita m. Sea H_o la entropía de la fuente $[A_o,\mu]$, y sea C la capacidad ergódica del canal $[A, \nu_x, B]$. Se asume que $H_o < C$, y se escoge un número positivo $\lambda < \frac{1}{2}(C - H_o)$. Entonces, se selecciona un código particular que transforma la salida de la fuente $[A_o, \mu]$ en el alfabeto A del canal dado.

Primero que todo, notemos que la fuente $[A_o, \mu]$, siendo ergódica, tiene la propiedad *E* (Teorema de Mc. Millan), la cual indica que para un n suficientemente grande, la secuencia α de dígitos del alfabeto A_o puede ser dividida en dos grupos, un grupo de "alta probabilidad", para el cual cada secuencia α cumple

$$\frac{\log \mu(\alpha)}{n} + H_o > -\lambda$$

o equivalentemente

$$\mu(\alpha) > 2^{-n(H_0+\lambda)}$$

y un grupo de "baja probabilidad", cuya probabilidad total es tan pequeña como se desee.

Proposición 2.1.

$\frac{\log \mu(\alpha)}{n} + H_o > -\lambda$ es equivalente a $\mu(\alpha) > 2^{-n(H_0+\lambda)}$.

Demostración:

En efecto,

$$\frac{\log \mu(\alpha)}{n} + \mathrm{H_o} > -\lambda \Leftrightarrow \frac{\log \mu(\alpha)}{n} > -\lambda - \mathrm{H_o}$$

$$\Leftrightarrow \log \mu(\alpha) > -n(\lambda + \mathrm{H_o})$$

$$\Leftrightarrow \mu(\alpha) > 2^{-n(H_0+\lambda)}.$$

Proposición 2.2.

El número de secuencias en el grupo de "alta probabilidad" es menor que $2^{n(H_0+\lambda)} < 2^{n(C-\lambda)}$.

Demostración:

Supongamos que r es el número de secuencias en el grupo de alta probabilidad. Luego,

$$1 \geq \sum_{i=1}^{r} \mu_i > \sum_{i=1}^{r} 2^{-n(H_0+\lambda)} = r\,2^{-n(H_0+\lambda)} \quad \Rightarrow \; r < 2^{n(H_0+\lambda)}$$

Por otra parte,

$$\lambda < \frac{C - H_0}{2} \Rightarrow \quad 2\lambda < (C - H_0)$$

$$\Rightarrow \quad \lambda+\lambda < (C - H_0)$$

$$\Rightarrow \quad \lambda + H_0 < C - \lambda$$

$$\Rightarrow \quad n(H_o+\lambda) < n(C-\lambda) \qquad (n>0)$$

$$\Rightarrow \quad 2^{n(H_0+\lambda)} < 2^{n(C-\lambda)}$$

así tenemos $r < 2^{n(H_0+\lambda)} < 2^{n(C-\lambda)}$.

A efectos de facilitar la comprensión de la demostración, en lo que sigue utilizaremos las notaciones siguientes:

	Notación
secuencias del grupo de alta probabilidad	$\alpha_1, \alpha_2, \ldots,$
conjunto de todas las secuencias en el grupo de baja probabilidad	$\alpha_o,$
elementos de los conjuntos A^I	$x = \ldots, x_{-1}, x_o, x_1, \ldots$
elementos de los conjuntos B^I	$y = \ldots, y_{-1}, y_o, y_1, \ldots$
la secuencia (cilindro) $x_{-m}, \ldots, x_{-1}, x_o, x_1, \ldots, x_{n-1}$ de longitud n+m, n fijo, donde todos los x_i toman un valor definido, dígitos del alfabeto *A*	$\mathcal{U}$
número de secuencias distintas u	a^{m+n}
número de dígitos del alfabeto *A*	A
secuencia (cilindro) $y_o, y_1, \ldots, y_{n-1}$ de longitud n, donde todos los y_i toman un valor definido (dígito del alfabeto *B*)	$\mathbf{Y}$
número de secuencias distintas $\mathbf{\upsilon}$	b^n
número de dígitos del alfabeto *B*	B

Además como la memoria del canal es m, entonces la probabilidad, $\nu_x(\mathfrak{v})$, de recibir $y \in \mathfrak{v}$ en la salida del canal cuando x es transmitido, es la misma para todo x perteneciente a la misma secuencia u. Así esta probabilidad depende solo de las secuencias u y $\mathfrak{v}$ escogidas y es natural denotarla por $\nu_u(\mathfrak{v})$.

Similarmente, si V es la unión de varias secuencias $\mathfrak{v}$, podemos denotar por $\nu_u(V)$ la probabilidad del evento $y \in V$ cuando algún $x \in u$ es transmitido.

Definición 2.1.

Sea λ alguna constante ($0 < \lambda < ½$). Llamaremos a $\{u_i\}$ ($1 \leq i \leq N$) un "**grupo distinguible**" de u-secuencias si existe un grupo $\{V_i\}$ ($1 \leq i \leq N$) de conjuntos $V_i \subset B^I$, donde cada V_i es un conjunto de varias secuencias $\mathfrak{v}$ tal que :

i. V_i y V_k no tienen secuencias comunes para $i \neq k$ y

ii. $\nu_{u_i}(V_i) > 1 - \lambda$ ($1 \leq i \leq N$).

Lema 2.1. (Lema Fundamental de Feinstein)

Si un canal dado es estacionario, sin anticipación y con memoria finita m, entonces para $\lambda > 0$, suficientemente pequeño y n suficientemente grande, existe un grupo distinguible $\{u_i\}$ $(1 \leq i \leq N)$ de u–secuencias, con $N > 2^{n(C-\lambda)}$ miembros, donde C es la capacidad ergódica del canal.

Como el canal $[A,\nu_x,B]$ cumple las hipótesis del lema fundamental de Feinstein, tenemos que para n suficientemente grande existe un grupo distinguible $\{u_i\}$ $(1 \leq i \leq N)$ de secuencias u de longitud n + m (consistiendo de dígitos del alfabeto A), con $N > 2^{n(C-\lambda)}$ miembros. Siendo N mas grande que el número de secuencias α_i en el grupo de alta probabilidad considerado, podemos asociar a cada secuencia α_i una secuencia u_i tal que a diferente u_i le corresponde una α_i diferente. Haciendo esto, una secuencia mas pequeña del grupo distinguible no ha sido usada, luego asociamos a esta última secuencia con todas las secuencias α_o del grupo de baja probabilidad. Después de hacer esto, a cada secuencia α, de n-términos de dígitos del alfabeto A_o corresponde una secuencia específica u de longitud n + m del alfabeto A, la cual pertenece al grupo distinguible $\{u_i\}$.

Ahora dividimos la secuencia θ de símbolos del alfabeto A_o, la cual ha sido codificada, en subsecuencias de longitud n, y la secuencia x, en la cual θ es codificada, en subsecuencias de longitud n + m. Enumeramos ambos conjuntos de subsecuencias de izquierda a derecha, como es usual. La k-ésima subsecuencia en el mensaje θ será una de las secuencias α; entonces seleccionamos como la k-ésima subsecuencia en el mensaje x, la secuencia u_i de nuestro grupo distinguible el cual corresponde a esa secuencia α. Haciendo esto para todo k ($-\infty < k < +\infty$), es claro que se determina, en forma única, $x = x(\theta)$ en términos del θ dado. En este sentido, se establece un código específico, que se mantendrá en lo que sigue. Obviamente, este código es un ejemplo del "código secuencial" discutido en la sección anterior.

Si se relaciona el código seleccionado, al canal $[A,\nu_x,B]$ obtenemos, como vimos anteriormente, un nuevo canal $[A_0, \lambda_\theta, B]$, donde $\lambda_\theta(Q) = \nu_{x(\theta)}(Q)$ para $Q \in F_B$.

Siendo la fuente $[A_o,\mu]$ la que alimenta este nuevo canal, llegamos, de acuerdo a lo planteado, a la fuente "compuesta" $[C,\omega]$, donde $C = A_o$ X B y $\omega(S_1 \text{ x } S_2)$ (donde $S_1 \in F_{A_0}$, $S_2 \in F_B$) está dada por la fórmula (2.3). Denotaremos las diferentes secuencias de longitud n, consistiendo de dígitos del alfabeto B, por $ß_k$ (k=1,2,…). Entonces por la definición de la distribución ω, nosotros tenemos para algún $k \geq 1$ e $i \in \{0,1,\ldots,N\}$

$$\omega\,(\alpha_i \text{ x } ß_k) = \int_{\alpha_i} \lambda_\theta(\beta_k)d\mu(\theta) = \int_{\alpha_i} \nu_{x(\theta)}(\beta_k)d\mu(\theta) \quad .$$

Pero decir que $\theta \in \alpha_i$ es equivalente a decir que $x(\theta) \in u_i$, como $\nu_x(\beta_k)$ asume el mismo valor para todo $x \in u_i$, se designa por $\nu_{u_i}(\beta_k)$, entonces tenemos que

$\nu_{x(\theta)}(\beta_k) = \nu_{u_i}(\beta_k)$ para todo $\theta \in \alpha_i$. Así, encontramos

(2.4) $\omega(\alpha_i \times ß_k) = \mu(\alpha_i).\nu_{u_i}(\beta_k)$.

Es importante destacar que u_0 es la secuencia del grupo distinguible $\{u_i\}$ en la cual todo "el grupo de baja probabilidad" es codificado.

A fin de evitar alguna mala interpretación, recordaremos, el significado del evento $\alpha_i x \beta_k$, cuya probabilidad está dada por (2.4); $\alpha_i \times \beta_k$ denota el evento que:

1). La secuencia de n-términos transmitida por la fuente $[A_o,\mu]$ coincide con α_i (si i>0) o pertenece al grupo de baja probabilidad (si i=0), y

2). después de la trasmisión a través del canal $[A_o,\lambda_\theta,\beta]$, la secuencia de n-términos transmitida por la fuente $[A_o,\mu]$ da una secuencia de n+m-términos de dígitos del alfabeto B, donde los últimos n dígitos son la secuencia β_k.

Sea i_k el valor de i ($0\leq i \leq N$) para el cual la probabilidad $\omega(\alpha_i , \beta_k) = \omega(\alpha_i \times \beta_k)$ tiene su valor más grande (si hay varios de tales valores de i, entonces tomaremos a alguno de ellos como i_k). Como la probabilidad condicional de la secuencia α_i para una secuencia β_k es igual a

$$\frac{\omega(\alpha_i,\beta_k)}{\omega(A_0^I,\beta_k)},$$

donde el denominador es independiente de i, entonces también podemos decir que α_{i_k} es la secuencia α_i, más probable para una secuencia dada β_k.

Sea el conjunto

$$\sum_{k}\sum_{i\neq i_k}\omega(\alpha_i, B_k) = \mathcal{P}.$$

se puede considerar $\mathcal{P}$ como la probabilidad de que la secuencia α_i en la entrada del canal $[A_o, \lambda_\theta, \beta]$ no sea la más probable para una secuencia dada β_k en la salida del canal.

La secuencia u_i, en la cual codificamos todas las secuencias α_i forma un grupo distinguible, eso implica la existencia de un grupo $\{B_i\}$ $(1 \leq i \leq n)$ (donde cada B_i es la unión de varias secuencias β_k tal que

1) $\nu_{u_i}(B_i) > 1 - \lambda$, y

2) B_i y B_j no tienen secuencias en común para $i\neq j$.

Pero sumando (2.4) sobre todo $\beta_k \subset B_i$, encontramos

$$\omega\left(\alpha_i x B_i\right) = \mu(\alpha_i).\nu_{u_i}(B_i) > (1-\lambda).\mu(\alpha_i) \qquad (0 \leq i \leq N).$$

Por lo tanto, la probabilidad condicional de obtener una secuencia $ß_k \subset B_i$, en la salida del canal $[A_o, \lambda_\theta, ß]$, dado que tenemos una secuencia α_i en la entrada del canal, es:

$$\mathcal{P}_{\alpha_i}(B_i) = \frac{\omega(\alpha_i, B_i)}{\mu(\alpha_i)} > 1 - \lambda \qquad (0 \le i \le N).$$

Así vemos que dos sistemas completos de eventos α_i $(0 \le i \le N)$ y $ß_k$ $(k=1,2,\ldots)$ satisfacen todas las premisas del lema siguiente:

Lema 2.2.

Si para cada ε $(0< \varepsilon <1)$, un conjunto Δ_i de eventos B_k puede ser asociado con cada A_i $(1 \le i \le N)$ tal que

1) $p(\Delta_i \Delta_j) = 0 \qquad (i \neq j)$ y

2) $p_{\alpha_i}(\Delta_i) > 1 - \varepsilon \qquad 0 \le i \le N;$

entonces $P < \varepsilon$.

<u>Demostración:</u>

Sean:

$$P = 1 - \sum_{k=1}^{n} p(A_{i_k} B_k); \qquad 1 - P = \sum_{k=1}^{n} p(A_{i_k} B_k),$$

Δ_o el conjunto de eventos B_k, si existe, que no están en algunos de los conjuntos Δ_i $(1 \le i \le n)$; entonces, el rango de sumatoria en la última suma puede ser expandido a $\Delta_o, \Delta_1, \ldots, \Delta_n$ y por lo tanto,

$$1 - P = \sum_{k=1}^{n} \sum_{B_k \in \Delta_i} p(A_{i_k} B_k) + \sum_{B_k \in \Delta_0} p(A_{i_k} B_k) \ge \sum_{k=1}^{n} \sum_{B_k \in \Delta_i} p(A_{i_k} B_k)$$

De acuerdo a la definición del subíndice i_k, el lado derecho de esta desigualdad puede solamente disminuir si reemplazamos el subíndice i_k, en cada término de la suma por algún subíndice de 1 a n. Así, en particular

$$1 - P \geq \sum_{i=1}^{n} \sum_{B_k \in \Delta_i} p(A_i B_k) = \sum_{i=1}^{n} p(A_i \Delta_i) = \sum_{i=1}^{n} p(A_i) p_{A_i}(\Delta_i) > (1-\varepsilon) \sum_{i=1}^{n} p(A_i) = 1 - \varepsilon$$

por lo tanto, $P \leq \varepsilon$

En consecuencia, usando este lema para $\Delta_i = B_i$, $B_k = \beta_k$, $A_i = \alpha_i$ y $\varepsilon = \lambda$ encontramos que es $\wp \leq \lambda$.

Este importante resultado se puede enunciar como sigue:

Teorema 2.1: (Primer Teorema de Shannon)

Sean:

a) un canal estacionario, sin anticipación $[A,\nu_x,B]$ con capacidad ergódica C y memoria finita m,

b) una fuente ergódica $[A_o,\mu]$ con entropía $H_o < C$ y

c) $\varepsilon>0$.

Entonces, para n suficientemente grande, la salida de la fuente $[A_o,\mu]$ puede ser codificada en el alfabeto A de tal forma que cada secuencia α_i de n dígitos del alfabeto A_o es transformada en una secuencia u_i de (n + m) dígitos del alfabeto A, tal que si la secuencia u_i es trasmitida a través del canal dado, se puede determinar la secuencia transmitida α_i con probabilidad, de la secuencia recibida en el canal de salida, mayor que $1-\varepsilon$.

La definición de la secuencia α_i, ocasiona la selección de la secuencia más probable como α_i, dado los últimos n dígitos mas recibidos en el canal de salida. El primer teorema de Shannon es a menudo formulado con menos exactitud diciendo que para $H_o < C$ es siempre posible encontrar un código tal que la secuencia transmitida puede suponerse como la secuencia recibida con un error que no excede un número ε arbitrariamente pequeño.

2.2.Segundo Teorema de Shannon.

Al igual que en el Primer Teorema, se procede a hacer un proceso inductivo donde se van dando explicaciones en la medida que avanzamos, estableciendo definiciones, proposiciones, lemas, notaciones y otros elementos necesarios para lograr la demostración de este Teorema.

En esta sesión se utilizará la fuente $[A_o,\mu]$, el canal $[A,\nu_x,B]$ y el código de la sección 2.1, así como toda la notación introducida allí. Por supuesto se considerará un nuevo problema pertinente a todo este proceso, a saber, la evaluación de su razón de trasmisión, es decir, la cantidad de información promedio dada por un dígito en el canal de salida. Para esto, se examinará primero el espacio de probabilidad finita $\alpha_i\ \beta_k$ ($0 \leq i \leq N$, $k = 1, 2, \ldots$), con la distribución $\omega(\alpha_i, \beta_k)$, considerada en 2.1.

Recordemos que α_i es uno de los n-términos de la secuencia del grupo de alta probabilidad en la salida de la fuente dada para $i \geq 1$, que α_o es el conjunto del grupo de baja probabilidad de cada secuencia y que β_k es un n-término de la secuencia de dígitos del alfabeto *B* (en el canal de salida). Se denota por $\wp$ la cantidad

$$\wp = \sum_{k} \sum_{i \neq i_k} \omega(\alpha_i, \beta_k)$$

donde i_k es el valor del subíndice i para el cual la probabilidad $\omega(\alpha_i, \beta_k)$ es la más grande, y se prueba que $\wp < \lambda$.

Definición 2.2. [A]

Una función real Ψ es una función **cóncava diferenciable** (f.c.d) en el intervalo (a,b), si está definida y es dos veces diferenciable en (a,b) y además $\Psi''(X) \leq 0$ $\forall X \in (a, b)$.

Si Ψ es cóncava diferenciable en (a,b) entonces $-\Psi$, se dice que es **convexa diferenciable** .

Lema 2.3. (Desigualdad de Jensen) [A]

Si Ψ es una función cóncava diferenciable en (a,b), entonces $\forall X_K \in (a,b)$ $K = 1,2,..,n$ y $\forall (q_1, q_2, ..., q_n) \in \Gamma_n$ $(n = 2,3,...)$, se cumple la desigualdad:

$$\Psi\left(\sum_{K=1}^{n} q_K X_K\right) \geq \sum_{K=1}^{n} q_K \Psi(X_K)\,. \tag{2.5}$$

Si $\varphi = -\Psi$, entonces es convexa diferenciable. Luego se cumple,

$$\varphi\left(\sum_{K=1}^{n} q_K X_K\right) \leq \sum_{K=1}^{n} q_K \varphi(X_K) \tag{2.6}$$

Lema 2.4.

Sean A y B dos espacios, n y m el número de eventos elementales A_i y B_k de los espacio A y B respectivamente. Para n >1 se cumple que

$$H_B(A) \leq P \log (n-1) - P \log P - (1-P) \log (1-P).$$

Demostración:

Sea $f(x) = x \log x$, entonces:

$$H_B(A) = -\sum_k p(B_k) \sum_i p_{B_k}(A_i) \log(p_{B_k}(A_i))$$

$$= -\sum_k p(B_k) \sum_i f(p_{B_k}(A_i))$$

$$= H_1 + H_2$$

donde

$$H_1 = -\sum_k p(B_k) f\left[p_{B_k}(A_{i_k})\right]$$

$$H_2 = -\sum_k p(B_k) \sum_{i \neq i_k} f\left[p_{B_k}(A_i)\right],$$

tomando $\lambda_k = p(B_k)$, $x_k = p_{B_k}(A_{i_k})$ y aplicando la desigualdad (2.6) en H_1 ,por

ser f convexa, se deduce que:

$$\begin{aligned} H_1 &= -\sum_k \lambda_k f(x_k) \leq -f\left(\sum_k \lambda_k x_k\right) = -f\left(\sum_k p(B_k) p_{B_k}(A_{i_k})\right) \\ &= -f\left(\sum_k p(B_k A_{i_k})\right) = -f(1-P) = -(1-P)\log(1-P) \end{aligned}$$

es decir,

(2.7) $\quad H_1 \leq -(1-P) \log (1-P)$.

Por otra parte, y usando la desigualdad (2.6)

$$-\sum_k p(B_k) f\left[1 - p_{B_k}(A_{i_k})\right] \leq -f\left(\sum_k p(B_k)\left[1 - p_{B_k}(A_{i_k})\right]\right)$$

$$= -f\left(1 - \sum_k p(B_k A_{i_k})\right) = -f(P) = -P\log P$$

así tenemos que:

(2.8) $$-\sum_k p(B_k) f\left[1 - p_{B_k}(A_{i_k})\right] \leq -P\log P$$

como

$$1 - p_{B_k}(A_{i_k}) = \sum_{i \neq i_k} p_{B_k}(A_i)$$

entonces para algún k fijo, $(1 \leq k \leq m)$

$$f\left[1 - p_{B_k}(A_{i_k})\right] = \sum_{i \neq i_k} p_{B_k}(A_i) \log \sum_{i \neq i_k} p_{B_k}(A_i)$$

$$= (n-1)\sum_{i\neq i_k}\frac{p_{B_k}(A_i)}{n-1}\log\left(\frac{n-1}{n-1}\sum_{i\neq i_k}p_{B_k}(A_i)\right)$$

$$= (n-1)\sum_{i\neq i_k}\frac{p_{B_k}(A_i)}{n-1}\log(\frac{1}{n-1}\sum_{i\neq i_k}p_{B_k}(A_i)) + \log(n-1)\sum_{i\neq i_k}p_{B_k}(A_i)$$

$$= (n-1)f\left(\frac{\sum_{i\neq i_k}p_{B_k}(A_i)}{n-1}\right) + \log(n-1)\sum_{i\neq i_k}p_{B_k}(A_i)$$

es decir,

(2.9) $$f\left[1-p_{B_k}(A_{i_k})\right] = \; = (n-1)f\left(\frac{\sum_{i\neq i_k}p_{B_k}(A_i)}{n-1}\right) + \log(n-1)\sum_{i\neq i_k}p_{B_k}(A_i),$$

como $\sum_{i=1}^{n-1}\frac{1}{n-1} = 1$ y $p_{B_k}(A_i) > 0$ $(1 \leq i \leq n,\ i\neq i_k)$, por la desigualdad (2.6),

tenemos:

(2.10) $$f\left(\frac{\sum_{i\neq i_k}p_{B_k}(A_i)}{n-1}\right) \leq \sum_{i\neq i_k}\frac{1}{n-1}f\left[p_{B_k}(A_i)\right]$$

luego, sustituyendo (2.10) en la desigualdad (2.9), obtenemos,

(2.11) $$f\left[1-p_{B_k}(A_{i_k})\right] \leq \sum_{i\neq i_k}f\left[p_{B_k}(A_i)\right] + \log(n-1)\sum_{i\neq i_k}p_{B_k}(A_i)$$

multiplicando a ambos lados de la desigualdad (2.11) por $p(B_k)$ y sumando sobre

k de 1 hasta m, y sabiendo que $P = \sum_{k=1}^{m}\sum_{\substack{i=1\\ i\neq i_k}}^{n}p(A_i, B_k) = \sum_{k=1}^{m}\sum_{\substack{i=1\\ i\neq i_k}}^{n}p(B_k)p_{B_k}(A_i)$,

tenemos

$$\sum_k p(B_k) f\left[1 - p_{B_k}(A_{i_k})\right] \le \sum_k p(B_k) \sum_{i \ne i_k} f\left[p_{B_k}(A_i)\right] + \sum_k p(B_k) \log(n-1) \sum_{i \ne i_k} p_{B_k}(A_i)$$

$$\le \sum_k p(B_k) \sum_{i \ne i_k} f\left[p_{B_k}(A_i)\right] + P\log(n-1) = -H_2 + P\log(n-1)$$

así tenemos

(2.12) $$\sum_k p(B_k) f\left[1 - p_{B_k}(A_{i_k})\right] \le -H_2 + P\log(n-1)$$

de (2.8) y (2.12), obtenemos,

$- P\log P - H_2 + P\log(n-1) \ge 0$

finalmente, usando (2.7) se concluye que

$$H_1 \le -(1-P)\log(1-P) - P\log P - H_2 + P\log(n-1)$$

de donde,

$$H_B(A) = H_1 + H_2 \le P\log(n-1) - P\log P - (1-P)\log(1-P) \,.$$

Siguiendo con el desarrollo de la teoría, a este punto podemos aplicar el lema 2.4, que acabamos de demostrar, donde los espacios $\{\alpha_i\}$ y $\{\beta_k\}$ juegan el rol del espacio A y B de este lema, y el número N+1 juega el rol del número n. Si $H_\beta(\alpha)$ es la entropía condicional del espacio $\{\alpha_i\}$ dado β_k, promediado sobre todo β_k, entonces tenemos:

$$H_\beta(\alpha) \le \mathcal{P} \log N - \mathcal{P} \log \mathcal{P} - (1 - \mathcal{P}) \log (1 - \mathcal{P})$$

Aquí, el número N es el número de secuencias en el grupo de alta probabilidad; como vimos en la proposición 2.2, $N < 2^{n(C-\lambda)} < 2^{nC}$ de donde $\log N < nC$. Tomando en cuenta que $\mathcal{P} < \lambda$ y, $-\mathcal{P}\log\mathcal{P} - (1-\mathcal{P})\log(1-\mathcal{P}) < 1$ para algún $\mathcal{P}$ $(0 < \mathcal{P} < 1)$, tenemos:

$$H_\beta(\alpha) < \lambda\, n\, C + 1.$$

En particular, como λ puede ser escogido arbitrariamente pequeña cuando n es suficientemente grande, tenemos que cuando $n \to \infty$:

$$H_\beta(\alpha) = O(n).$$

Aquí,

$H_\beta(\alpha) = H(\beta\alpha) - H(\beta)$

$$= -\sum_k \sum_{i=0}^{N} \omega(\beta_k, \alpha_i)\log(\omega(\beta_k, \alpha_i)) + \sum_k \eta(\beta_k)\log(\eta(\beta_k))$$

$$= -\sum_k \sum_{i=0}^{N} p_{\beta_k}(\alpha_i)\eta(\beta_k)\log(p_{\beta_k}(\alpha_i)\eta(\beta_k)) + \sum_k \eta(\beta_k)\log(\eta(\beta_k))$$

$$= -\sum_k \eta(\beta_k)\sum_{i=0}^{N} p_{\beta_k}(\alpha_i)(\log(p_{\beta_k}(\alpha_i)) + \log\eta(\beta_k)) + \sum_k \eta(\beta_k)\log(\eta(\beta_k))$$

$$= -\sum_k \eta(\beta_k)\sum_{i=0}^{N} p_{\beta_k}(\alpha_i)\log(p_{\beta_k}(\alpha_i)) - \sum_k \eta(\beta_k)\sum_{i=0}^{N} p_{\beta_k}(\alpha_i)\log\eta(\beta_k)$$
$$+ \sum_k \eta(\beta_k)\log(\eta(\beta_k))$$

$$= -\sum_k \eta(\beta_k) \sum_{i=0}^{N} p_{\beta_k}(\alpha_i) \log(p_{\beta_k}(\alpha_i)) - \sum_k \eta(\beta_k) \log \eta(\beta_k) \sum_{i=0}^{N} p_{\beta_k}(\alpha_i)$$
$$+ \sum_k \eta(\beta_k) \log(\eta(\beta_k))$$

$$= -\sum_k \eta(\beta_k) \sum_{i=0}^{N} p_{\beta_k}(\alpha_i) \log(p_{\beta_k}(\alpha_i)) - \sum_k \eta(\beta_k) \log \eta(\beta_k)$$
$$+ \sum_k \eta(\beta_k) \log(\eta(\beta_k))$$

$$= -\sum_k \eta(\beta_k) \sum_{i=0}^{N} p_{\beta_k}(\alpha_i) \log(p_{\beta_k}(\alpha_i)).$$

Así,

$$H_\beta(\alpha) = -\sum_k \eta(\beta_k) \sum_{i=0}^{N} f\left[p_{\beta_k}(\alpha_i)\right],$$

donde

$$f(x) = x \log x \,; \qquad \eta(\beta_k) = \omega\,(A^I,\beta_k); \quad y \quad p_{\beta_k}(\alpha_i) = \frac{\omega(\alpha_i,\beta_k)}{\omega(A^I,\beta_k)} = \frac{\omega(\alpha_i,\beta_k)}{\eta(\beta_k)}$$

En todo lo anterior α_o, distinto de α_i para i>0, no denota una secuencia individual sino todo el grupo de baja probabilidad de secuencia de n-términos formados arriba por dígitos del alfabeto *A*. Ahora dividimos éste grupo en secuencias separadas.

$\alpha_1', \alpha_2', \ldots, \alpha_q'$

y en lugar del espacio (α_i) ($0 \le i \le N$) consideramos el espacio (α_i, α'_j) ($1 \le i \le N$, $1 \le j \le q$) de todas las secuencias de n-términos consistiendo de dígitos del alfabeto A_o, con probabilidad adecuada $\mu(\alpha_i)$, $\mu(\alpha'_j)$. Dejamos el espacio (β_k) para más adelante. Sea $H_\beta(\alpha,\alpha')$ la entropía condicional del espacio (α_i,α'_j) para un β_k dado; promediado sobre β_k, encontramos

$$(2.13)\qquad H_\beta(\alpha,\alpha') = -\sum_k \eta(\beta_k)\left\{\sum_{i=1}^{N} f\left[p_{\beta_k}(\alpha_i)\right] + \sum_{j=1}^{q} f\left[p_{\beta_k}(\alpha_j')\right]\right\}$$

$$= -\sum_k \eta(\beta_k)\sum_{i=0}^{N} f\left[P_{\beta_k}(\alpha_i)\right] + \sum_k \eta(\beta_k) f\left[P_{\beta_k}(\alpha_0)\right] - \sum_k \eta(\beta_k)\sum_{j=1}^{q} f\left[P_{\beta_k}(\alpha_j')\right]$$

$$= H_\beta(\alpha) + \sum_k \eta(\beta_k) f\left[p_{\beta_k}(\alpha_o)\right] + R\ ,$$

donde

$$(2.14)\qquad R = -\sum_k \eta(\beta_k)\sum_{j=1}^{q} f\left[p_{\beta_k}(\alpha_j')\right].$$

Notemos que

$$P_{\beta_k}(\alpha_j') = \frac{\omega(\alpha_j',\beta_k)}{\eta(\beta_k)},$$

donde

$$\omega(\alpha_j',\beta_k) = \int_{\alpha_j'} \lambda_\theta(\beta_k)\, d\mu(\theta).$$

Como $\lambda_\theta(\beta_k)$ tiene el mismo valor $\lambda_{\alpha_0}(\beta_k)$ para todo $\theta \in \alpha_j' \subset \alpha_o$, entonces

$$p_{\beta_k}(\alpha_j') = \frac{\lambda_{\alpha_o}(\beta_k)\mu(\alpha_j')}{\eta(\beta_k)}$$

esta dada únicamente por los datos de nuestro problema.

El segundo término en el lado derecho de (2.13) es negativo, en efecto

$f\lfloor P_{\beta_k}(\alpha_0)\rfloor = P_{\beta_k}(\alpha_0)\ln\lfloor P_{\beta_k}(\alpha_0)\rfloor < 0$ ya que $0 < p_{\beta_k} < 1 \Rightarrow \ln\lfloor p_{\beta_k}(\alpha_0)\rfloor < 0$.

De allí que $\sum_k \eta(\beta_k) f\lfloor p_{\beta_k}(\alpha_0)\rfloor < 0$, entonces

(2.15) $$H_\beta(\alpha,\alpha') < H_\beta(\alpha) + R,$$

y como ya tenemos calculado $H_\beta(\alpha)$, falta solamente calcular la cantidad R definida por la suma (2.14).

Lema 2.5.

Sean A y B dos espacios finitos compuestos de eventos elementales A_i y B_k respectivamente, con probabilidades $p(A_i)$ ($1 \leq i \leq n$), $p(A_i) > 0$; $\sum_{i=1}^{n} p(A_i) = 1$ y $p(B_k)$ ($1 \leq k \leq m$), $p(B_k) > 0$; $\sum_{k=1}^{m} p(B_k) = 1$ entonces

(2.16) $$\sum_{i=1}^{n} p(A_i) \sum_k {}^{*} p_{A_i}(B_k) \log p_{A_i}(B_k) \geq \sum_k {}^{*} p(B_k) \log p(B_k),$$

donde $\sum_{K}^{*}$ denota la suma sobre ciertos valores del subíndice k (no necesariamente todos, pero si los mismos valores en ambos lados de la desigualdad).

Lema 2.6.

Para tres espacios finitos A, B y C se cumple

(2.17) $H_{AB}(C) \leq H_B(C)$.

Usando el Lema 2.5 encontramos que

$$(2.18) \qquad R \leq -\sum_{j=1}^{q} \mu(\alpha_j') \log \mu(\alpha_j') .$$

Aquí, por definición del grupo de baja probabilidad, tenemos que $\sum_{j=1}^{q} \mu(\alpha_j') = \mu(\alpha_o) < \lambda$. Fácilmente se ve que el valor más grande de la suma (2.18), bajo la condición suplementaria.

$$\sum_{j=1}^{q} \mu(\alpha_j') = \varepsilon ,$$

se obtiene para $\mu(\alpha_j') = \varepsilon/q$ $(1 \leq j \leq q)$ y es $\varepsilon [\log q + \log (1/\varepsilon)]$.

En efecto:

$$\sum_{j=1}^{q} \mu(\alpha_j') \log(\mu(\alpha_j')) = -\varepsilon \log\left(\frac{\varepsilon}{q}\right) = -\varepsilon(\log \varepsilon - \log q) = \varepsilon(\log q + \log \varepsilon^{-1})$$

$$= \varepsilon\,[\,\log q + \log (1/\varepsilon)],$$

por lo tanto, en nuestro caso

$$R < \lambda[\log q + \log (1/\lambda)].$$

Como q es el número de secuencias en el grupo de baja probabilidad y es menor que a^n, el número de todas las secuencias de n términos, (a es el número de dígitos en el alfabeto A_o), por lo tanto

$$R < \lambda n \log a + \lambda \log (1/\lambda) < \lambda n \log a + 1$$

Vemos que $R = O(n)$ cuando $n \to \infty$, y como $H_\beta(\alpha) = O(n)$, como probamos anteriormente, se sigue de (2.8) que:

(2.19) $$H_\beta(\alpha,\alpha') = O(n) \qquad (n \to \infty)$$

Ahora cambiamos nuestra notación un poco. El espacio $\alpha(\alpha_1, \alpha_2, \ldots)$ será el conjunto de todos las secuencias de n-términos de dígitos del alfabeto A_o, sea del grupo de alta o baja probabilidad, así que el número de todos los α_i es a^n, donde "a" es el número de dígitos en el alfabeto A_o. Como antes el espacio $\beta(\beta_1, \beta_2, \ldots)$ denota el conjunto de todas las secuencias de n-términos de dígitos del alfabeto B. El producto (α,β) de esos dos espacios tiene la distribución

$$\omega(\alpha_i, \beta_k) = \int_{\alpha_i} \lambda_\theta(\beta_k)\,d\mu(\theta) = \lambda_{\alpha_i}(\beta_k)\mu(\alpha_i),$$

y las distribuciones de los espacios α y β están dadas por la función $\mu(\alpha_i)$ y

$$\eta(\beta_k) = \omega\left(A_o{}^I, \beta_k\right) = \sum_i \mu(\alpha_i)\lambda_{\alpha_i}(\beta_k),$$

respectivamente. La entropía condicional del espacio α dado β_k, promediado sobre β_k, la cual denotamos por $H_\beta(\alpha)$ es obviamente la cantidad $H_\beta(\alpha,\alpha')$ que acabamos de calcular. Por lo tanto, por (2.19) tenemos:

(2.20) $$H_\beta(\alpha) = O(n) \qquad (n \to \infty).$$

El proceso de transmisión de información que consideramos es el siguiente. La salida de la fuente $[A_o,\mu]$ es dividida en secuencias de longitud n, y cada secuencia α_i es trasmitida a través del canal $[A_o,\lambda_\theta,B]$, dando como salida una secuencia de (n + m) dígitos del alfabeto *A*. Los últimos n-dígitos de esa secuencia forman la secuencia β_k “recibida”. La salida de nuestro proceso de transmisión consiste de tal secuencia. Nuestro problema es calcular la razón de la transmisión. Para hacer esto, consideremos una secuencia de longitud s = nt + r de la salida de la fuente $[A_o, \mu]$, donde t es algún entero positivo y $0 \leq r \leq n$. Denotamos a tal secuencia por X y el conjunto (espacio) de todas esas secuencias (para un *s* dado) por {X}. Sea la secuencia de *s*-términos Y (de dígitos del alfabeto *B*) recibida en el canal de salida, cuando la secuencia X es trasmitida a través del canal $[A_o, \lambda_\theta, B]$, y {Y} el conjunto (espacio) de tales secuencias Y. Denotamos por $H_Y(X)$ la entropía condicional del espacio {X}, promediado sobre Y.

Cada secuencia X de longitud s = nt + r puede ser descompuesta en t secuencias consecutivas $\alpha^{(1)}$, $\alpha^{(2)},\ldots,\alpha^{(t)}$ de longitud n y una secuencia "restante" α^{*} de longitud r < n. Claramente podemos considerar el espacio {X} como el producto de los (t +1) espacios $\{\alpha^{(j)}\}$ $(1 \leq j \leq. t)$ y $\{\alpha^{*}\}$, donde cada uno de los primeros t espacios tiene la estructura del espacio que se considero arriba. Es obvio que en general los (t + 1) espacios serán mutuamente dependiente. De la propiedad básica de la entropía de un espacio producto, tenemos:

$$H_{Y_O}(X) \leq \sum_{j=1}^{t} H_{Y_O}\left[\alpha^{(j)}\right] + H_{Y_o}(\alpha^{*}),$$

donde Y_o es alguna secuencia fijada Y. Así, promediando sobre Y_o, encontramos:

$$H_{Y}(X) \leq \sum_{j=1}^{t} H_{Y}\left[\alpha^{(j)}\right] + H_{Y}(\alpha^{*}). \qquad (2.21)$$

Se descompuso la secuencia X en t secuencias $\alpha^{(j)}$ y la secuencia residual α^{*}, luego se puede descomponer de forma equivalente la secuencia Y en t secuencias $\beta^{(j)}$ $(1 \leq j \leq t)$ de longitud n y una secuencia residual β^{*} de longitud r < n. Entonces el espacio {Y} será el producto de los (t+1) espacios $\{\beta^{(j)}\}$ $(1 \leq j \leq t)$ y $\{\beta^{*}\}$. La secuencia $\beta^{(j)}$ $(1\leq j \leq t)$ es la secuencia, en el canal de salida, correspondiente a la secuencia $\alpha^{(j)}$, en su entrada. Así, el espacio producto $\{\alpha^{(j)},\beta^{(j)}\}$ $(1 \leq j \leq t)$ tiene la distribución $\omega(\alpha_j,\beta_k)$ considerada arriba, y el espacio $\{\beta_k^{(j)}\}$ tiene la distribución $\eta(\beta_k)$.

Sea $B^{(j)}$ $(1 \leq j \leq t)$ el conjunto de todas las secuencias $\beta^{(l)}$ $(1\leq l\leq t)$ y β^* que conforman Y, con la excepción de $\beta^{(j)}$. Entonces con el objeto de fijar una secuencia específica Y, podemos fijar $\beta^{(j)}$ y $B^{(j)}$. En otras palabras, podemos considerar {Y} como el producto de los espacios $\{\beta^{(j)}\}$ y $B^{(j)}$. (Aquí j es alguno de los números 1,2,…, t). Así, por el Lema 2.6, encontramos para algún j $(1\leq j\leq t)$.

$$H_Y\left[\alpha^{(j)}\right] = H_{\beta^{(j)}B^{(j)}}\left[\alpha^{(j)}\right] \leq H_{\beta^{(j)}}\left[\alpha^{(j)}\right] = H_\beta(\alpha) \tag{2.22}$$

Por otro lado, el espacio $\{\alpha^*\}$ obviamente contiene a^r eventos, donde a es el número de dígitos en el alfabeto A_o. Pero por una de las propiedades fundamentales de la entropía, la entropía de un espacio finito no excede al logaritmo del número de eventos en el espacio. Así, para alguna secuencia Y_o escogida

$$H_{Y_o}(\alpha^*) \leq r \log(a) < n \log(a)$$

y, en consecuencia, el promedio de la entropía condicional $H_Y(\alpha^*)$ satisface

$$H_Y(\alpha^*) < n\log(a). \tag{2.23}$$

Entonces por (2.22) y (2.23), la desigualdad (2.21) da

$$H_Y(X) \leq t\, H_\beta(\alpha) + n \log(a),$$

de la cual sigue por (2.20), que para un $\lambda>0$ arbitrariamente pequeño, para n suficientemente grande, y para algún $t \geq 1$

$$H_Y(X) < \lambda tn + n\log(a) \leq \lambda s + n\log(a).$$

Se ha caracterizado la cantidad $H_Y(X)$, “entropía residual” de la secuencia X transmitida a través del canal, como la cantidad de información que permanece en la secuencia después de la transmisión, es decir, la información perdida durante la transmisión.

Como la cantidad de información contenida en la secuencia X antes de la transmisión es sH_o, la cantidad de información transmitida es $sH_o - H_Y(X)$.

Pero transmitir cada secuencia $\alpha^{(j)}$, de longitud n, requiere el pasar n + m símbolos a través del canal, y otros tantos como símbolos sean necesarios para transmitir la secuencia α^*. Por lo tanto el número total de símbolos que pasan a través del canal durante la transmisión de la secuencia X es (t +1)(n + m). Así, un símbolo en el canal de salida lleva el promedio de cantidad de información

$$\frac{sH_o - H_Y(X)}{(t+1)(n+m)} \geq \frac{sH_o - \lambda s - nlog\ a}{n(t+1)(1+\frac{m}{n})} \geq \frac{sH_o - \lambda s - nlog\ a}{(s+n)(1+\frac{m}{n})} = \frac{H_o - \lambda - \frac{nlog\ a}{s}}{(1+\frac{n}{s})(1+\frac{m}{n})}.$$

Si seleccionamos n y t lo suficientemente grande tal que $m/n < \varepsilon$ y $n/s \leq 1/t < \varepsilon$, entonces el lado derecho de la desigualdad anterior será más grande que

$$\frac{H_o - \lambda - \varepsilon log(a)}{(1+\varepsilon)^2} < H_o - 2\lambda \qquad \text{para } \varepsilon \text{ suficientemente pequeño.}$$

Este importante resultado significa que con el código seleccionado cada dígito recibido, en la salida del canal, aporta en promedio una cantidad de información, tan cercana como se desea, promedio este que depende del rendimiento de la salida de la fuente. En otras palabras, la transmisión de información ocurre a una razón arbitrariamente cercana a aquella con que es emitida la información de acuerdo al rendimiento de la fuente. Por supuesto, todo esto bajo la condición que $H_o < C$, y para códigos de secuencias suficientemente grandes.

En esta relación, notamos que si n puede ser tomada también grande, entonces el valor práctico del método del código descrito es nulo, ya que se debería esperar demasiado para decodificar (descifrar) el texto recibido en la salida del canal. Así, desde el punto de vista práctico, podría ser considerado interesante investigar la relación entre n y λ. Feinstein obtuvo un resultado interesante en ese sentido, pero no hablaremos de ello aquí. Desde el punto de vista práctico, podríamos notar que en ambos métodos, el de Feinstein y el de Shannon, la construcción de un código con la característica requerida no esta dada; la existencia de tal código se prueba, pero no ofrece indicación de cómo encontrarlo actualmente.

El resultado que hemos obtenido es el segundo teorema de Shannon, enunciado como sigue:

Teorema 2.2. (Segundo teorema de Shannon)

Bajo la condición del Teorema 2.1 existe un código tal que la razón de transmisión es tan cercano a H_o como se desea.

Capítulo 3

Medidas de Información y de Incertidumbre

En este capítulo presentaremos definiciones y teoremas concernientes a la teoría de la "**Información** y de la **Incertidumbre**". Es una sección introductoria a dicha teoría cuyo contenido es necesario para poder desarrollar y establecer en el capítulo siguiente el fundamento de este trabajo. Por esta razón la teoría se presenta de manera esencial, esto es sin demostrar los teoremas y sin mayores detalles en las definiciones y en resultados ya establecidos.

3.1. Definiciones y Observaciones Generales

Definición 3.1.

Una "**Estructura de Información**" es una terna I = (**Ω, *S,*** *K*) donde:

- **Ω** es el conjunto de los eventos elementales.
- ***S*** es una subálgebra de ***P* (Ω)** (álgebra de los eventos observables).
- *K* es una subfamilia de K_m, clase de las colecciones finitas de álgebras ***m***-independientes. [1] [10])

Definición 3.2. (B. Forte - J. Kampé de Fériet [8])

Una "**Medida de Información**" (de evento) sobre una estructura I es una aplicación

$$J: \boldsymbol{S} \rightarrow \mathbf{R}^{+} \qquad \text{tal que:}$$

(3.1) $J(\boldsymbol{\Omega}) = 0\ ,\ J(\varnothing) = +\infty$

(3.2) $A \subset B \ \Rightarrow\ J(A) \ \geq\ J(B)$

(3.3) $\left\{\mathcal{A}^{i},\ i = 1,\dots,n\right\} \in K, A_{j_i} \in \mathcal{A}^{i} \ \Rightarrow\ J\left(\bigcap_{i=1}^{n} A_{j_i}\right) = \sum_{i=1}^{n} J\left(A_{j_i}\right).$

[1] Sea $\mathcal{A}^1, \dots, \mathcal{A}^n$ una colección de subálgebras de *S* diremos que estas álgebras son ***m***-independientes si para cada $A_{j_i} \in \mathcal{A}^{i}, (i = 1,\dots,n)\, A_{j_i} \neq \varnothing$ se cumple $\bigcap_{i=1}^{n} A_{j_i} \neq \varnothing$. Si las álgebras son generadas por las particiones $\pi^1,\dots,\pi^n$, diremos también que las particiones son ***m*** -independientes ($(\pi^1,\dots,\pi^n) \in K$).

La expresión (3.3) se puede generalizar sustituyendo la ley aditiva por la ley *

quedando

(3.3)' $\left\{\mathcal{A}^{i},\ i = 1,\ldots,m\right\} \in \mathcal{K}, A_{j_i} \in \mathcal{A}^{i} \Rightarrow$

$$J\left(\bigcap_{i=1}^{n} A_{j_i}\right) = \mathop{*}_{i=1}^{n} J\left(A_{j_i}\right) \qquad (\forall\ n \leq m).$$

Propiedades Elementales

(3.4) $\mathbf{J}(A \cap B) \geq \sup[\mathbf{J}(A), \mathbf{J}(B)]$

(3.5) $\mathbf{J}(A \cup B) \leq \inf[\mathbf{J}(A), \mathbf{J}(B)]$

Definición 3.3.

A la estructura $\mathbb{I}$ = (**Ω,** ***S, K,*** J) se le denomina "**Espacio de Información**" (de eventos).

Definición 3.4.

Una "**familia de experiencias**" es una familia de particiones finitas $A=\{A_1,\ldots,A_n\}$ de conjuntos de ***S,*** subalgebra ***P*** **(Ω),** de cuyos elementos componentes A_i son también elementos de ***S***.

Definición 3.5.

Al conjunto $\mathcal{A} = \bigcup_{i=1}^{n} A_i$ se le denomina " **soporte de la experiencia** " . Si $\mathcal{A} = \mathbf{\Omega}$

decimos que la experiencia es " **completa** "; en caso contrario, "**incompleta** ".

Observación 3.1.

En la familia de experiencias ξ tomaremos en consideración:

a. La relación de orden "< " , refinamiento de particiones. Definida por:

(3.6) 1. $A < B \Leftrightarrow [\ \mathcal{A} = \mathcal{B}$ y $\forall A_i \in A, \exists B_j \in B / A_i \subset B_j]$ $\veebar$

2. $A < B \Leftrightarrow [A = \{\ \mathcal{A}\ \}, B = \{\ \mathcal{B}\ \}$ y $\mathcal{A} \subset \mathcal{B}]$.

Decimos entonces que A es "**más fina**" que B. En el caso 1 (experiencias con el mismo soporte) se habla de refinamiento propio, en el caso 2 (experiencias impropias) de refinamiento impropio.

b. Las operaciones "$\wedge$", producto de particiones y "$\vee$", unión de particiones definidas por:

(3.7) $A \wedge B = \{\ A_i \cap B_j / A_i \in A, B_j \in B\}$

(3.8) $A \vee B = \{\ A_1 , ..., A_n, B_1, ..., B_m / A_i \in A, B_j \in B\ \}$ solo si $\mathcal{A} \cap \mathcal{B} = \varnothing$.

Observación 3.2.

El producto "$\wedge$" es asociativo, conmutativo, distributivo respecto a "$\vee$" y monótono respecto a la relación de orden " < "; vale decir,

(3.9) $(A < A') \wedge (B < B') \Rightarrow A \wedge B < A' \wedge B'$.

Definición 3.6.

Una “**Estructura de Incertidumbre**” es un par $\mathbb{H} = (\mathbb{I}, \xi)$ donde:

- $\mathbb{I} = (\boldsymbol{\Omega}, \boldsymbol{S}, \boldsymbol{K}, J)$ es un espacio de información según la definición 3.3.
- ξ es una familia de experiencias.

Definición 3.7. (B. Forte [7])

La “**Medida de Incertidumbre**” contenida en una experiencia $\pi \in \xi$ sobre una estructura $\mathbb{H}$, es una aplicación

$H: \xi \rightarrow \mathbf{R}^+$ con las siguientes propiedades:

(3.10) $\quad H(\{\mathcal{A}\}) = J(\mathcal{A})$

(3.11) $\quad$ Valores universales: $H(\boldsymbol{\Omega}) = 0$, $H(\varnothing) = +\infty$

(3.12) $\quad$ Monotonía: $A \in \xi, B \in \xi, A < B \Rightarrow H(A) \geq H(B)$

(3.13) $\quad H\left(\bigwedge_{\mathbf{i}=1}^{\mathbf{n}} \pi_{\mathbf{i}}\right) = \mathop{*}_{i=1}^{n} H(\pi_{\mathbf{i}}) \quad (\forall\, n \leq m)$

donde π_i es una partición compuesta por subconjuntos del álgebra $\mathcal{A}_i$.

Definición 3.8.

Un “**Espacio de Incertidumbre**” es una estructura $\mathbf{H} = (\mathbb{H}, H)$ donde:

- $\mathbb{H}$ es una estructura de incertidumbre.
- $H: \xi \rightarrow \mathbf{R}^+$ es una medida de incertidumbre sobre $\mathbb{H}$.

3.2. Propiedades

Las siguientes propiedades, (3.14) - (3.18), son consecuencia inmediata de la definición 3.7.

(3.14) $\mathbf{H}(\{\mathbf{\Omega}\}) = \mathbf{J}(\mathbf{\Omega}) = 0$

$\mathbf{H}(A) \geq \mathbf{H}(\{\mathcal{A}\}) = \mathbf{J}(\mathcal{A})$

(3.15) $\mathbf{H}(A \vee B) \geq \mathbf{H}(A \vee \{\mathcal{B}\}) \geq \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\}) \geq \mathbf{J}(\mathcal{A} \cup \mathcal{B})$

$\mathbf{H}(A \vee B) \geq \mathbf{H}(\{\mathcal{A}\} \vee B) \geq \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\}) \geq \mathbf{J}(\mathcal{A} \cup \mathcal{B})$

(3.16) $\mathbf{H}(A \wedge B) \geq \mathbf{H}(A \wedge \{\mathcal{B}\}) \geq \mathbf{H}(\{\mathcal{A}\} \wedge \{\mathcal{B}\}) \geq \mathbf{J}(\mathcal{A} \cap \mathcal{B})$

$\mathbf{H}(A \wedge B) \geq \mathbf{H}(\{\mathcal{A}\} \wedge B) \geq \mathbf{H}(\{\mathcal{A}\} \wedge \{\mathcal{B}\}) \geq \mathbf{J}(\mathcal{A} \cap \mathcal{B})$

(3.17) $\mathbf{H}(A \vee \{\emptyset\}) = \mathbf{H}(A)$ (siendo $A \vee \{\emptyset\} = A$)

(3.18) $\mathbf{H}(\{\emptyset\}) = +\infty$

(3.19) **Simetría** (sigue inmediatamente de la definición de partición). Sea p una permutación de {1,2,...,n} entonces se tiene

$\mathbf{H}(\{A_1,...,A_n\}) = \mathbf{H}(\{A_{p(1)},...,A_{p(n)}\})$.

La definición 3.7. es muy débil. La clase de las aplicaciones **H** que cumplen con las propiedades (3.14), (3.15), (3.16) contiene funciones muy diferentes cuantitativa y cualitativamente, de allí la necesidad de estudiar otras propiedades que definen familias de incertidumbre homogéneas entre sí, es así como aparecen las propiedades siguientes.

Idempotencia [7]

(3.20) $\mathcal{A} \cap \mathcal{B} = \varnothing$; $\mathbf{H}(A) = \mathbf{H}(B) = h \Rightarrow \mathbf{H}(A \vee B) = h$

Compositividad [3]

(3.21) $\mathbf{H}(A \vee B) = M[\mathcal{A}, \mathcal{B}, \mathbf{H}(A), \mathbf{H}(B)]$

donde a la función M se le denomina ley de composición.

Ramificación [9]

(3.22) $\mathbf{H}(A \vee B) = \mathbf{H}(A \vee \{\mathcal{B}\}) + \mathbf{H}(\{\mathcal{A}\} \vee B) - \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\})$

$\forall\, A, B \in \xi,\ \mathcal{A} \cap \mathcal{B} = \varnothing$

o equivalentemente

(3.22)' $\mathbf{H}(\{A_1, A_2, A_3, \ldots, A_n\}) - \mathbf{H}(\{A_1 \cup A_2, A_3, \ldots, A_n\}) =$

$$= \Phi\left(\left\{\bigcup_{i=3}^{n} A_i, A_1, A_2\right\}\right)$$

Ramificación generalizada [6]

(3.23) $\mathbf{H}(A \vee B) = \Psi[\mathbf{H}(A \vee \{\mathcal{B}\}), \mathbf{H}(\{\mathcal{A}\} \vee B), \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\})]$

donde a ψ se le denomina ley de g-localidad .

Aditividad [7]

(3.24) $(A, B) \in \boldsymbol{K}, A \wedge B \in \xi \Rightarrow \mathbf{H}(A \wedge B) = \mathbf{H}(A)+\mathbf{H}(B).$

En la aditividad la ley de independencia es la clásica $*=+$ luego en éste caso la incertidumbre condicional es $\mathbf{H}(B \mid A)=\mathbf{H}(A \wedge B)-\mathbf{H}(A)$ [2]

Aditividad fuerte [13]

(3.25) $(A, B) \in \boldsymbol{K}, \mathbf{H}(A \wedge B) = \mathbf{H}(A) + \mathbf{H}(B \mid A)$

Expansibilidad

La restricción de $\mathbf{H} : \xi \to \mathbf{R}^+$ a $\xi_n \subset \xi$, con ξ_n conteniendo las particiones con n elementos, define una función $\mathbf{H}_\mathbf{n}:\xi_n \to \mathbf{R}^+$ sobre un oportuno subconjunto de $\boldsymbol{S}^\mathbf{n}$. La expansibilidad consiste en :

(3.26) $\mathbf{H}_\mathbf{n}(\{A_1,...,A_n\}) = \mathbf{H}_{\mathbf{n+1}}(\{A_1,...,A_k,\emptyset,A_{k+1},...,A_n\}),$

$\forall k=1,...n$

Proposición 3.1. Si una medida H cumple con la Ramificación entonces cumple con la Ramificación Generalizada.

2 $\mathbf{H}(B \mid A)$ es la incertidumbre condicional de la partición B con respecto a la partición A. Ver definición 3.9.

Demostración.

En efecto si cumple con la Ramificación se tiene que:

$$\mathbf{H}(A \vee B) = \mathbf{H}(A \vee \{\mathcal{B}\}) + \mathbf{H}(\{\mathcal{A}\} \vee B) - \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\})$$

Así basta definir $\Psi(x, y, z) = x + y - z$. Luego

$$\mathbf{H}(A \vee B) = \Psi[\mathbf{H}(A \vee \{\mathcal{B}\}), \mathbf{H}(\{\mathcal{A}\} \vee B), \mathbf{H}(\{\mathcal{A}\} \vee \{\mathcal{B}\})].$$

3.3. Incertidumbre Condicional.

Para cualquier partición $A=\{A_1, ..., A_n\}$ la cantidad H(A) mide la incertidumbre que se tiene sobre el resultado de una experiencia cuyos éxitos previstos son los eventos $A_1, ..., A_n$, sin saber nada de lo que puede acontecer.

Por otro lado, muchas veces antes de ejecutar la experiencia ya se conoce algo sobre los posibles éxitos y este conocimiento disminuye la incertidumbre. Típico es el caso donde antes de efectuar la experiencia ya se sabe que el resultado será un evento elemental perteneciente al conjunto $\mathcal{A} = \bigcup_{i=1}^{n} A_i$, por supuesto sin saber a cual de los éxitos A_1 o A_2 ... o A_n pertenecerá. En esta condición la incertidumbre sobre cuál será el éxito no será H (A) sino otra cantidad que indicaremos por K(A) y la llamaremos “incertidumbre condicional” de la partición A.

La primera definición dada con el objeto de medir esa nueva incertidumbre, ha sido propuesta por P. Benvenuti y B. Forte [2], quienes introdujeron la cantidad K(A) por medio de la relación

(3.27) $$K(A) = H(A) - H(\{\mathcal{A}\}) = H(A) - J(\mathcal{A}).$$

Esta definición, aunque es aceptable en el contexto clásico, no pone en evidencia directamente las características fundamentales de la información condicional y por otro lado no es fácilmente generalizable a espacios donde la incertidumbre del producto de dos experiencias independientes no se conforma con el axioma clásico $H(A \wedge B) = H(A) + H(B)$.

Resulta más conveniente introducir esta noción en forma axiomática de manera que, la definición misma subraye las características esenciales. Adecuados teoremas de representación permitirán en lo sucesivo explicitar las formas de K(A) en los distintos espacios de incertidumbre.

Definición 3.9.

Sea $\mathbf{H} = (\mathbb{I} = (\boldsymbol{\Omega}, \boldsymbol{S}, \boldsymbol{K}, J), \xi, H)$ un espacio de incertidumbre. Se da el nombre de "**incertidumbre condicional**" asociada a ese espacio a cada aplicación:

$$K: \xi \to \mathbf{R}^+$$

con las siguientes propiedades:

1. K: $\xi \to \mathbf{R}^+$ es una medida de incertidumbre sobre un espacio $\mathbb{K}$ cuyos elementos $\boldsymbol{\Omega}$, $\boldsymbol{S}$, K, *, ξ, son los mismos del espacio $\mathbb{H}$.
2. $K(\{\mathcal{A}\}) = 0$
3. $K(A) = f\, [\, H(A), J(\mathcal{A})]$.

Las propiedades 1. y 2. se pueden sustituir por la siguiente forma más compacta:

1'. K: $\xi \to \mathbf{R}^+$ es una medida de incertidumbre sobre la estructura $(\mathbb{I}^*, \xi)$ donde en el espacio $\mathbb{I}^* = (\boldsymbol{\Omega}, \boldsymbol{S}, K, J^*)$, J^* es la medida de información trivial, $J^*(\mathcal{A}) = 0$ $(\forall \mathcal{A} \in \boldsymbol{S})$.

Teorema 3.1.

Si en el espacio de información $\mathbb{I}$, la ley de independencia es la clásica ($x * y = x + y$) y si además se pide que la función f sea continua y universal, entonces

$$f(x, y) = c\,(x - y) \qquad c > 0 \tag{3.28}$$

y por consiguiente

$$K(A) = c\,[\, H(A) - J(\mathcal{A})] \qquad c > 0. \tag{3.29}$$

Otro caso de condicionamiento es aquel en el cual se mide la incertidumbre de una experiencia π después de efectuarse una experiencia previa π'. La incertidumbre sobre cuál será el éxito de π estará entonces condicionada por el éxito de la experiencia π'.

Aunque no sea estrictamente vinculante, en lo que sigue supondremos que la partición condicionante π' tenga soporte contenido en la partición condicionada π. Esta hipótesis no es demasiado limitativa porque los eventos elementales que no pertenecen a elementos de π pueden en muchos casos ser eliminados sin modificar la parte que realmente condiciona nuestro conocimiento. En otras palabras, si nosotros estamos interesados en la partición π podemos ignorar todos los eventos elementales que no le pertenecen.

Poniéndonos como observadores de ambas experiencias, nos daremos cuenta que cualquiera sea el éxito de la primera experiencia π', éste modifica nuestros conocimientos sobre la segunda experiencia π; por lo tanto, saber cómo está estructurada la experiencia π', influye sobre la evaluación que atribuiremos a la incertidumbre de π.

Hablaremos de incertidumbre condicional de la partición π respecto a la partición π', denotando esta cantidad por $H(\pi|\pi')$. Pues la incertidumbre condicional será una aplicación:

$$\mathbf{H}: \xi \times \xi \to \mathbf{R}^+$$

a la cual es natural imponer las siguientes condiciones:

a. La restricción de **H** a los pares del tipo $(A,\{\mathcal{A}\})$ determina una aplicación

$$K: \xi \rightarrow \mathbf{R}^{+}$$

que debe ser una medida de incertidumbre sobre $(\mathbf{I}, \xi)$.

b. La incertidumbre de A relativa a cualquiera de sus refinamientos tiene que ser nula.

Las condiciones a. y b. se traducen en las siguientes relaciones:

	a'	$A'' < A' \Rightarrow H(A''\mid\{\mathcal{A}\}) \geq H(A'\mid\{\mathcal{A}\})$
(3.30)	a"	$(A, B) \in K \Rightarrow H(A \wedge B \mid \{\mathcal{A} \cap \mathcal{B}\}) = H(A\mid\{\mathcal{A}\}) * H(B\mid\{\mathcal{B}\})$
	b	$A' < A \Rightarrow H(A \mid A') = 0.$

Si queremos definir $H(\pi|\pi')$ de manera que sea coherente con el espacio de incertidumbre $(\mathbf{I}, \xi, H)$ es conveniente que la aplicación $\mathbf{H}: \xi \times \xi \rightarrow \mathbf{R}^{+}$ dependa de sus argumentos π y π' sólo a través de medidas de incertidumbres absolutas asociadas de cualquier manera a esas particiones. En forma más precisa, resulta natural pensar que sea

$$H(\pi|\pi') = F[H(\pi\wedge\pi'), H(\pi')]. \tag{3.31}$$

En efecto, aunque a primera vista podría parecer natural pedir que $H(\pi|\pi')$ dependa de $H(\pi)$ y $H(\pi')$ no hay que olvidar el hecho que: si se conoce la estructura de la partición π' y de la condicionada π, también se conoce la partición $\pi \wedge \pi'$ y si ya se conoce un éxito A'_j de π', la incertidumbre que queda es la incertidumbre sobre la partición $A_1 \cap A'_j$, $A_2 \cap A'_j$, ... $A_n \cap A'_j$. Por esto el observador que pueda ver ambas particiones se da cuenta que la incertidumbre relativa depende de la partición condicionante y de su producto con la partición condicionada.

Por otro lado la hipótesis (3.5) es compatible con la propiedad de aditividad fuerte de la incertidumbre de Shannon, $H(A|B) = H(A \wedge B) - H(A)$, y además tiene analogía con la definición de probabilidad condicional, $P(A|B) = p(A \cap B) / p(B) = f[p(A \cap B), p(B)]$.

Como en la hipótesis el soporte de π' esta contenido en el soporte de π, entonces tenemos que $H(\pi' \wedge \pi) \geq H(\pi')$.

Luego F está definida en un subconjunto de $\Gamma_2 := \{(x, y) / x \geq y\}$.

3.4. Compositividad. Ley de composición

Definición 3.10.

Sea $\mathbf{I} = (\boldsymbol{\Omega}, \boldsymbol{S}, \boldsymbol{K}, *, \mathbf{J})$ un espacio de información, **J** es "**compositiva** " si para dos eventos A, B en ***S***, tales que $A \cup B \in \boldsymbol{S}$ y $A \cap B = \emptyset$ tenemos:

$$\mathbf{J}(A \cup B) = \mathbf{F}[\mathbf{J}(A), \mathbf{J}(B)] \tag{3.32}$$

A la función **F**(x,y) = x **T** y se le denomina "**ley de composición**" para la información.

La forma más general de las leyes de composición continuas es proporcionada por el siguiente teorema de C.H. Ling [12].

Teorema 3.2.

Sea $\{\Delta_i = (a_i,b_i) \,/\, i=1,...,I \leq +\infty\}$ una sucesión finita o numerable de intervalos disjuntos y $f_i : [a_i,b_i] \rightarrow \mathbf{R}^+$ una familia de funciones decrecientes (estrictamente) con $f_i(b_i) = 0$ y sea

$$f_i^{(-1)}(\mu) = f_i^{-1}\left[\,\inf\{\, f_i(a_i), \mu \,\}\,\right]$$

la seudoinversa de f_i .Cada función definida haciendo

(3.33) $$F(x,y) = \begin{cases} f_i^{(-1)}\left[f_i(x) + f_i(y)\right] & \text{si } (x,y) \in \Delta_i^2 \\ \inf(x,y) & \text{en otro caso} \end{cases},$$

representa una posible ley de composición y viceversa; cada posible ley de composición es de esta forma bajo una adecuada selección de $\{\Delta_i\}$ y $\{f_i\}$. Los elementos del conjunto $\boldsymbol{\Lambda} = \mathbf{R}^+ - \bigcup_{i \in I}(\alpha_i, \beta_i)$ se llaman **F**-idempotentes y cumplen con la condición **F** (x,x) = x Al igual que * también **F** es una ley de semigrupo, pero de tipo norma triangular (t-norma) .

Definición 3.11.

Una " **t-norma** " sobre [a,b] es una función $G:[a,b]^2 \rightarrow [a,b]$ que satisface las siguientes propiedades

a. $x \, G \, y = y \, G \, x$

b. $x \, G \, (y \, G \, z) = (x \, G \, y) \, G \, z$

(3.34) c. $x_1 < x_2 \Rightarrow x_1 G \, y \leq x_2 \, G \, y$

d. $x \, G \, b = x$.

Capítulo 4

Otras Medidas de Incertidumbre

En este capítulo tomaremos como punto de partida las medidas de incertidumbre analizadas por Capodieci 1.999 [5] y Vásquez 2.003 [15], de estas medidas serán escogidas aquellas que cumpliendo con las exigencias de la Teoría Matemática de la Comunicación de Shannon puedan ser utilizadas para obtener una Teoría Matemática de la Comunicación alternativa.

Para ello, primero presentaremos un cuadro resumen de las Medidas de Incertidumbre analizadas por Capodieci 1.999 [5] y Vásquez 2.003 [15], indicando las propiedades que se verifican para cada una de ellas donde la medida que encabeza la lista es la de Shannon.

Con el objeto de cubrir nuestro primer objetivo:

1. Estudiar la posibilidad de utilizar otra medida de Incertidumbre distinta a la de Shannon en la teoría matemática de la comunicación;

comenzamos por escoger, de las medidas presentadas desde el cuadro 1 al 5, aquellas medidas que tenían el mayor número de propiedades comunes con la entropía de Shannon.

Con las medidas escogidas intentamos repetir el procedimiento utilizado en el Capítulo 2, relativo al primero y segundo teorema de Shannon. Particularmente detectamos que los elementos mas relevantes del desarrollo de la teoría se encuentran en las demostraciones de los lemas 2.2 y 2.4.

El Lema 2.2 tiene hipótesis relacionadas con las propiedades que tiene la medida de probabilidad, por lo que llegamos a la primera conclusión: se detectó que se necesita una distribución de probabilidad o en su defecto una distribución de la medida de información escogido, lo que no se logró en el ámbito nítido luego se concluye que *se debía abordar el trabajo con medidas que dependieran de la probabilidad*, es decir llegamos al esquema desarrollado por Wiener, el cual plantea lo siguiente:

Dado un espacio de probabilidad ($\boldsymbol{\Omega}$, $\boldsymbol{S}$, p). A cada evento A $\in$ $\boldsymbol{S}$ esta asociado una función J(A) (medida de información de A) que satisface las siguientes propiedades:

Axioma 4.1 J(A) = F[p(A)] J depende de A solo a través de p(A).

Axioma 4.2 Si A y B son independientes $J(A \cap B) = J(A) + J(B)$

Esto es la información asociada a la realización a la vez, de dos eventos independientes, es la suma de las informaciones asociadas a la realización de los dos eventos.

Del axioma 4.1 tenemos:

$$F[\ p(A \cap B)\ ] = J(A \cap B) = J(A) + J(B)$$

$$= F[\ p(A)\ ] + F[\ p(B)\ ]$$

Axioma 4.3 Dos eventos independientes para la información deben ser independientes para la probabilidad.

En base al axioma 4.2 tenemos que

$$F[\ p(A)\ ] + F[\ p(B)\ ] = J(A) + J(B) = J(A \cap B)$$

$$= F[\ p(A \cap B)\ ] = F[\ p(A).\ p(B)\ ]$$

es decir

(4.1) $F[\ p(A).\ p(B)\ ] = F[\ p(A)\ ] + F[\ p(B)\ ].$

Axioma 4.4 La relación (4.1) debe ser verificada para cualquier valor $x = p(A)$, $y = p(B)$. Esto es debe ser

(4.2) $F(x.y) = F(x) + F(y) \qquad \forall\ x,y \ \ 0 < x,y < 1.$

Observación 4.1:

Las únicas funciones continuas que satisfacen estos axiomas son las funciones

$$F(x) = c \log x.$$

Si imponemos a las informaciones de eventos ser además una cantidad positiva tendremos en definitiva

$$J(A) = - c \log p(A) \qquad c > 0. \tag{4.3}$$

Con lo que verificamos lo planteado en los objetivos 2 y 3

2. Verificar si la medida a utilizar debe depender de las probabilidades.
3. Analizar que sucede si se utiliza una medida que no depende de las probabilidades.

La medida debe ser probabilística y no se verifica el lema cuando no lo es, ya que la teoría se basa en una distribución de probabilidad.

En vista que los trabajos analizados, Capodieci [4] y Vásquez [15], han sido realizados desde el punto de vista axiomático, la medida J representa una medida de información que de acuerdo a lo visto hasta ahora, debe ser la medida de Shannon-Wiener, específicamente $J(A) = - \log p(A)$. Luego se hace necesario la escogencia de las medidas que siendo probabilísticas tengan el mayor número de propiedades comunes a la entropía de Shannon.

Es así como obtenemos el cuadro 6.

Al intentar reproducir la demostración del Lema 2.4 se nos presentaron inconvenientes que no pudimos saldar.

A continuación presentamos algunos de los cálculos hechos para determinar las propiedades de aditividad y aditividad fuerte en las medidas escogidas en el cuadro 6. En la demostración trabajaremos con Π_A = A.

ADITIVIDAD

En la demostración de la aditividad, A y B son esquemas independientes

$j_i = -c \log p_i$, c>0

$\pi_{kl} = p_k q_l$

p_k: es la probabilidad de que ocurra A_k

q_l: es la probabilidad de que ocurra B_l

A_k pertenece al esquema A

B_l pertenece al esquema B

Probemos la aditividad para las entropías 2-6 del cuadro 6:

2.- $H(A) = -\log\left(\sum_{i=1}^{n} e^{-j_i}\right)$

$$-\mathrm{H(AB)} = \log\left(\sum_k \sum_l e^{c \log \pi_{kl}}\right)$$

$$= \log\left(\sum_k \sum_l e^{c \log p_k q_l}\right)$$

$$= \log(\sum_k \sum_l e^{c\log p_k + c\log q_l})$$

$$= \log(\sum_k \sum_l e^{c\log p_k} e^{c\log q_l})$$

$$= \log(\sum_k e^{c\log p_k} \sum_l e^{c\log q_l})$$

$$= \log(\sum_k e^{c\log p_k}) + \log(\sum_l e^{c\log q_l})$$

$$= -H(A) - H(B)$$

Por lo tanto, $H(AB) = H(A) + H(B)$

3.- $H(A) = \left(\sum_{i=1}^{n} j_i^{\alpha} \right)^{1/\alpha}$

$$H(AB) = \left(\sum_k \sum_l (-c \log \pi_{kl})^{\alpha} \right)^{1/\alpha}$$

$$= \left(\sum_k \sum_l (-c \log p_k q_l)^{\alpha} \right)^{1/\alpha}$$

$$= \left(\sum_k \sum_l (-c \log p_k - c \log q_l)^{\alpha} \right)^{1/\alpha}$$

$$H(A) = \left(\sum_k (-c \log p_k)^{\alpha} \right)^{1/\alpha}$$

$$H(B) = \left(\sum_l (-c \log q_l)^{\alpha} \right)^{1/\alpha}$$

$$H(A) + H(B) = \left(\sum_k \left(-c\log p_k\right)^{\alpha}\right)^{1/\alpha} + \left(\sum_l \left(-c\log q_l\right)^{\alpha}\right)^{1/\alpha}$$

$\neq$ H(AB). Solo se cumple si $\alpha = 1$.

4.- $H(A) = J_A + \frac{k}{\alpha}\left[\sum_{i=1}^{n} j_i^{\alpha} - J_A^{\alpha}\right]$, $\alpha \neq 0$, $k > 0$.

$$H(AB) = J_{AB} + \frac{k}{\alpha}\left[\sum_i \sum_l j_{il}^{\alpha} - J_{AB}^{\alpha}\right]$$

$$= p(AB)\log p(AB) + \frac{k}{\alpha}\left[\sum_i \sum_l \left(p_i q_l \log p_i q_l\right)^{\alpha} - \left(p(AB)\log p(AB)\right)^{\alpha}\right]$$

$$= (p(A) + p(B))\log p(AB) +$$

$$\frac{k}{\alpha}\left[\sum_i p_i^{\alpha} \sum_l q_l^{\alpha}\left(\log p_i + \log q_l\right)^{\alpha} - \left(p(A)+p(B)\right)^{\alpha}\left(\log(p(A)+p(B))\right)^{\alpha}\right]$$

$$= p(A)\log p(AB) + p(B))\log p(AB) +$$

$$\frac{k}{\alpha}\left[\sum_i p_i^{\alpha} \sum_l q_l^{\alpha}\left(\log p_i + \log q_l\right)^{\alpha} - \left(p(A)+p(B)\right)^{\alpha}\left(\log(p(A)+p(B))\right)^{\alpha}\right]$$

Por otro lado,

$$H(A) = p(A)\log p(A) + \frac{k}{\alpha}\left[\sum_i \left(p_i \log p_i\right)^{\alpha} - \left(p(A)\log(p(A)\right)^{\alpha}\right]$$

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditividad	Aditividad Fuerte	Expansibilidad
1	$\mathbf{H}(\Pi_A) = \left[\dfrac{\sum_{i=1}^{n} \mathbf{J}_i e^{-\mathbf{J}_i}}{e^{-\mathbf{J}_A}}\right]$ (Shannon)	X	X	X	X	X J tipo Shannon	X J tipo Shannon	X
2	$\mathbf{H}(\Pi_A) = \left(\sum_{i=1}^{n} \mathbf{J}_i^{\alpha}\right)^{1/\alpha}$	No	X	X $\alpha = 1$	X	X $\alpha = 1$	No	No
3	$\mathbf{H}(\Pi_A) = \dfrac{1}{\alpha}\left[\log \sum_{i=1}^{n} e^{\alpha \mathbf{J}_i}\right]$ (Renyi)	No	X	No	X	X	X	No
4	$\mathbf{H}(\Pi_A) = \left[\dfrac{\sum_{i=1}^{n} \mathbf{J}_i^{\alpha} - \beta \mathbf{J}_A^{\alpha}}{1-\beta}\right]^{1-\beta}$	No	X	No	X	X $\alpha = 1$, $\beta=0$	X $\alpha = 1$, $\beta=0$	No
5	$\mathbf{H}(\Pi_A) = \dfrac{1}{1-\alpha} \log \left[\dfrac{\sum_{i=1}^{n} e^{\alpha \mathbf{J}_i} - \beta e^{\alpha \mathbf{J}_A}}{1-\beta}\right]$	No	X	No	X	X $\beta=0$	X	X $\alpha <0$, $\beta\neq1$
6	$\mathbf{H}(\Pi_A) = -\log\left(\sum_{i=1}^{n} e^{-\mathbf{J}_i}\right)$	No	X	No	No	X	X	X

Cuadro 1: Cuadro Resumen de propiedades de Medidas de Incertidumbre (1-6). Capodieci [3]

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditividad	Aditividad Fuerte	Expansibilidad
7	$\mathbf{H}(\Pi_A) = \left[\frac{\sum_{i=1}^{n} e^{-s J_i} [J_i]^k}{\left(\sum_{i=1}^{n} e^{-J_i} \right)^s} + (n-1)\, c \right]^{1/k}$ $(k \geq 0, s \leq 1)$	X s=1	No	No	No	X s = k =1 y c=0	X k=1	X
8	$\mathbf{H}(\Pi_A) = (1-t)^{-1} \log \left[\frac{\sum_{i=1}^{n} 2^{(1-t-s) J_i}}{\sum_{i=1}^{n} 2^{-s J_i}} \right]$ (Kapur) $(t \neq 1, t > 0, s \geq 1)$	X	No	No	No	X	X	X

Cuadro 2: Cuadro Resumen de propiedades de Medidas de Incertidumbre (7-8). Capodieci [3]

La clase de medidas escogidas es aquella que depende de $J_i = J(A_i)$. Estas medidas son compatibles con todas las informaciones de evento J que tienen la propiedad de compositividad. En particular son compatibles con las informaciones de tipo inf y con las de tipo M. En las medidas anteriores $\alpha > 0$, $k > 0$ y $0 \leq \beta < 1$. $A = \{A_1, A_2, A_3, ..., A_n\}$, $B = \{B_1, B_2, B_3, ..., B_m\}$, $\mathcal{A} = \bigcup_{i=1}^{n} A_i$, $\mathcal{B} = \bigcup_{j=1}^{m} B_j$. La propiedad de aditividad y aditividad fuerte están probadas para la ley de independencia clásica $x * y = x + y$.

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditividad	Aditividad Fuerte	Expansibilidad
1	$\mathbf{H}(\Pi_A) = \prod_{i=1}^{n}(\mathbf{J}_i + 1) - 1$	No	X	No	X	No	No	No
2	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \frac{K}{\alpha}\left[\sum_{i=1}^{n}\mathbf{J}_i^{\alpha} - \mathbf{J}_A^{\alpha}\right]$ $\alpha = 0$, $k>0$	No	X J compositiva	X	X	No	No	No
3	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \frac{K}{\alpha}\log\left[\frac{\prod_{i=1}^{n}(\mathbf{J}_i + 1)}{\mathbf{J}_A + 1}\right]$	No	X J compositiva	X	X	No	No	No
4	$\mathbf{H}(\Pi_A) = \left[\sum_{i=1}^{n}\frac{1}{\mathbf{J}_i}\right]^{-1}$	No	X	X	X	No	X	X
5	$\mathbf{H}(\Pi_A) = \mathbf{J}_A \cdot \frac{\left[\prod_{i=1}^{n}(\mathbf{J}_i + 1)\right]^k}{(\mathbf{J}_A + 1)^k}$	No	X J compositiva	No	X	No	No	No
6	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \left[\frac{\sum_{i=1}^{n} e^{-\mathbf{J}_i} - e^{\mathbf{J}_A}}{\alpha}\right]$	No	X J compositiva	X	X	No	No	X

Cuadro 3: Cuadro Resumen de propiedades de Medidas de Incertidumbre (1-6). Vásquez [8]

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditividad	Aditividad Fuerte	Expansibilidad
7	$\mathbf{H}(\Pi_A) = \mathbf{J}_A \cdot e^{\left[k\left(\sum_{i=1}^{n} \mathbf{J}_i^{\alpha} - \mathbf{J}_A^{\alpha}\right)\right]}$	No	X J compositiva	No	X	No	No	No
8	$\mathbf{H}(\Pi_A) = \log\left(\sum_{i=1}^{n} e^{\alpha \mathbf{J}_i}\right) + (1-\alpha)\mathbf{J}_A, 0 < \alpha \leq 1$	No	No	No	No	X J Shann.	X	No
9	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \log\left[1 - k\,\mathbf{J}_A^{\alpha} + k\sum_{i=1}^{n} \mathbf{J}_i^{\alpha}\right]$	No	X J compositiva	No	X	No	No	No
10	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \log\left[1 + k\left(\sum_{i=1}^{n} e^{\alpha \mathbf{J}_i} - e^{\alpha \mathbf{J}_A}\right)\right]$	No	X J compositiv	No	X	No	No	No
11	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + e^{\left[k\left(\sum_{i=1}^{n} \mathbf{J}_i^{\alpha} - \mathbf{J}_A^{\alpha}\right)\right]} - 1$	No	X J compositiva	No	No	No	No	No
12	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \left[\frac{\prod_{i=1}^{n}(\mathbf{J}_i+1)}{\mathbf{J}_A+1}\right] - 1$	No	X J compositiva	X	X	No	No	No

Cuadro 4: Cuadro Resumen de propiedades de Medidas de Incertidumbre (7-12). Vásquez [8]

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditividad	Aditividad Fuerte	Expansibilidad
13	$\mathbf{H}(\Pi_A) = \mathbf{J}_A + \frac{1}{\alpha}\left[\sum_{i=1}^{n}\frac{1}{\mathbf{J}_i} - \frac{1}{\mathbf{J}_A}\right]$	No	X J compositiva	X	X	No	No	X
14	$\mathbf{H}(\Pi_A) = \mathbf{J}_A \cdot \left[k\sum_{i=1}^{n} e^{\alpha \mathbf{J}_i} - k e^{\alpha \mathbf{J}_A} + 1\right]^{1/\alpha}$	No	X J compositiva	X Con α=1	X Con α=1	No	No	No
15	$H(\Pi_A) = J_A + \log\left[1 - e^{-J_A} + \sum_{i=1}^{n} e^{-J_i}\right]$	No	No	No	No	No	No	X

Cuadro 5: Cuadro Resumen de propiedades de Medidas de Incertidumbre (13-15). Vásquez [8]

	Entropía	Idempotente	Compositiva	Ramificación	Ramificación generalizada	Aditiva	Aditividad Fuerte	Expansibilidad
1.	$H(\prod_A) = \sum_{i=1}^{n} p_i \log p_i \Big/ \sum_{i=1}^{n} p_i$	X	X	X	X	X	X J tipo Shannon	X
2.	$H(\prod_A) = -\log\left(\sum_{i=1}^{n} e^{-j_i}\right)$	No	X	No	X	X	X	X
3.	$H(\prod_A) = \left(\sum_{i=1}^{n} j_i^{\alpha}\right)^{1/\alpha}$	No	X	X $\alpha = 1$	X	X $\alpha = 1$	No	No
4.	$H(\prod_A) = J_A + \frac{k}{\alpha}\left[\sum_{i=1}^{n} j_i^{\alpha} - J_A^{\alpha}\right]$ $\alpha \neq 0, k > 0$	No	X J compositiva	X	X	No	No	No
5.	$H(\prod_A) = \left[\sum_{i=1}^{n} \frac{1}{j_i}\right]^{-1}$	No	X	X	X	No	X	X
6.	$H(\prod_A) = \log\left(\sum_{i=1}^{n} e^{\alpha J_i}\right) + (1-\alpha) J_A$ $0 \leq \alpha \leq 1$	No	No	No	No	X	X	No

Cuadro 6: Medidas seleccionadas.

La medida 1 es la medida de Shannon.

$$H(B) = p(B)\log p(B) + \frac{k}{\alpha}\left[\sum_l (q_l \log q_l)^\alpha - (p(B)\log(p(B))^\alpha\right]$$

$$H(A) + H(B) = p(A)\log p(A) + p(B)\log P(B) +$$

$$\frac{k}{\alpha}\left[\sum_i (p_i \log p_i)^\alpha + \sum_l (q_l \log q_l)^\alpha - \left[(p(A)\log(p(A))^\alpha + (p(B)\log(p(B))^\alpha\right]\right]$$

por lo tanto,

$H(A) + H(B) \neq H(AB)$. Solo se cumple si $\alpha = 1$

5.- $$H(A) = \left[\sum_i \frac{1}{j_i}\right]^{-1}$$

$$= \left[\sum_i \frac{1}{-c\log p_i}\right]^{-1}$$

$$H(AB) = \left[\sum_k \sum_i \frac{1}{-c\log \Pi ki}\right]^{-1}$$

$$= \left[\sum_k \sum_i \frac{1}{-c\log p_k q_i}\right]^{-1}$$

$$= \left[\sum_k \frac{1}{p_k} \sum_i \frac{1}{-c(\log p_k + \log q_i)}\right]^{-1}$$

$$H(A) + H(B) = \left[\sum_k \frac{1}{-c\log p_k}\right]^{-1} + \left[\sum_i \frac{1}{-c\log q_i}\right]^{-1}$$

por lo tanto,

$H(AB) \neq H(A) + H(B)$

$$6.\text{- } H(A) = \log\left(\sum_{i=1}^{n} e^{\alpha j_i}\right) + (1-\alpha)J(A), \quad 0 < \alpha \le 1$$

$$H(AB) = \log\left(\sum_i \sum_k e^{\alpha J(A_i B_k)}\right) + (1-\alpha)J(AB)$$

$$= \log\left(\sum_i \sum_k e^{\alpha J(A_i)+\alpha J(B_k)}\right) + (1-\alpha)(J(A)+J(B))$$

$$= \log\left(\sum_i \sum_k e^{\alpha J(A_i)} e^{\alpha J(B_k)}\right) + (1-\alpha)(J(A)+(1-\alpha)J(B))$$

$$= \log\left(\sum_i e^{\alpha J(A_i)} \sum_k e^{\alpha J(B_k)}\right) + (1-\alpha)(J(A)+(1-\alpha)J(B))$$

$$= \log\left(\sum_i e^{\alpha J(A_i)}\right) + (1-\alpha)(J(A) + \log\left(\sum_k e^{\alpha J(B_k)}\right) + (1-\alpha)J(B))$$

$$= H(A) + H(B)$$

ADITIVIDAD FUERTE

En la demostración de la aditividad fuerte, A y B son esquemas mutuamente dependientes.

$j_i = -c \log p_i$, c>0

$\pi_{kl} = p_k q_{kl}$ $\quad 1 \le k \le n,\ 1 \le l \le m$

q_{kl} es la probabilidad que el evento B_l del esquema B ocurra, dado que el evento A_k del esquema A ocurre, esto es:

p_k: es la probabilidad de que ocurra A_k

q_l: es la probabilidad de que ocurra B_l

A_k pertenece al esquema A

B_l pertenece al esquema B

La medida de incertidumbre de Shannon, es fuertemente aditiva, es decir, H(AB) = H(A) + H_A(B). Queremos ver si las medidas escogidas cumplen esta propiedad.

2.- $H(A) = -\log\left(\sum_{i=1}^{n} e^{-j_i}\right)$

$$-\mathrm{H(AB)} = \log(\sum_k \sum_l e^{c\log\pi_{kl}})$$

$$= \log(\sum_k \sum_l e^{c\log p_k q_{kl}})$$

$$= \log(\sum_k \sum_l e^{c(\log p_k + \log q_{kl})})$$

$$= \log(\sum_k \sum_l e^{c\log p_k}\, e^{c\log q_{kl}})$$

$$= \log(\sum_k e^{c\log p_k} \sum_l e^{c\log q_{kl}})$$

$$= \log(\sum_k e^{c\log p_k}) + \log(\sum_l e^{c\log q_{kl}})$$

$$= -H(A) - H_{A_k}(B)$$

Luego cumple la aditividad fuerte.

5.- $\mathrm{H(A)} = \left[\sum_i \frac{1}{j_i}\right]^{-1}$

$$\mathrm{H(A \wedge B) - H(A)} = \left[\sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{J_i + J_j}\right]^{-1} - \left[\sum_{i=1}^{n} \frac{1}{J_i}\right]^{-1}$$

$$= \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{J_i + J_j}} - \frac{1}{\sum_{i=1}^{n}\frac{1}{J_i}} = \frac{\sum_{i=1}^{n}\frac{1}{J_i} - \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{J_i + J_j}}{\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{J_i + J_j}\right)\left(\sum_{i=1}^{n}\frac{1}{J_i}\right)} =$$

$$= \frac{[H(A)]^{-1} - [H(A\wedge B)]^{-1}}{[H(A\wedge B)]^{-1}[H(A)]^{-1}}$$

$$= F[H(A\wedge B), H(A)].$$

Luego cumple la aditividad fuerte.

6.- $H(A) = \log\left(\sum_{i=1}^{n} e^{\alpha J_i}\right) + (1-\alpha)J_A$

Vásquez [15] probó que:

$H(AB) = H(A) + H_A(B)$

De las medidas escogidas en el Cuadro 6 solo 2.- y 6.- tienen ambas propiedades luego aparentemente son las únicas medidas de incertidumbre a las cuales les podemos deducir la entropía condicionada, la cual es necesaria para trabajar con el Lema 2.4.

ENTROPIA CONDICIONAL

2.- $H(A) = -\log\left(\sum_{i=1}^{n} e^{-j_i}\right)$

Como $H(AB) = H(A) + H_A(B)$, entonces

$$H_A(B) = H(AB) - H(A)$$

$$= -\log(\sum_k \sum_l e^{-j_{ki}}) + \log \sum_k e^{-j_k}$$

$$= -(\log(\sum_k \sum_l e^{-j_{kl}}) - \log \sum_k e^{-j_k})$$

$$= -\log(\sum_k \sum_l e^{-j_{ki}} \div \sum_k e^{-j_k})$$

$$= -\log\left(\frac{\sum_k \sum_l e^{c\log p_k} e^{c\log q_{ki}}}{\sum_k e^{c\log p_k}}\right)$$

$$= -\log\left(\frac{\sum_k \sum_l e^{c\log p_k + c\log q_{ki}}}{\sum_k e^{c\log p_k}}\right)$$

$$= -\log\left(\frac{\sum_k \sum_l e^{c\log p_k} e^{c\log q_{ki}}}{\sum_k e^{c\log p_k}}\right)$$

$$= -\log\left(\frac{\sum_k e^{c\log p_k} \sum_l e^{c\log q_{kl}}}{\sum_k e^{c\log p_k}}\right)$$

6.- $H(A) = \log\left(\sum_{i=1}^{n} e^{\alpha J_i}\right) + (1-\alpha) J_A$

$$H_A(B) = H(AB) - H(A)$$

$$= \log\left(\sum_k \sum_l e^{\alpha J_{kl}}\right) + (1-\alpha) J(AB) \; - \log\left(\sum_k e^{\alpha J_k}\right) + (1-\alpha) J(A)$$

$$= \log\left(\frac{\sum_k \sum_l e^{\alpha J_{kl}}}{\sum_k e^{\alpha J_k}}\right) + (1-\alpha)(J(AB) - J(A))$$

$$= \log\left(\frac{\sum_k \sum_l e^{-\alpha c \log p_k q_{kl}}}{\sum_k e^{-\alpha c \log p_k}} \right) + (1-\alpha)(J(AB) - J(A))$$

LEMA 2.4

2.- $H(A) = -\log\left(\sum_{i=1}^{n} e^{-j_i} \right)$

$$H_A(B) = -\log\left(\frac{\sum_k \sum_i e^{c \log p_k q_{ki}}}{\sum_k e^{c \log p_k}} \right)$$

Sea $P = \sum_k \sum_{\substack{i \\ i \neq i_k}} p(A_i B_k) = \sum_k \sum_{\substack{i \\ i \neq i_k}} p(A_i) p_{A_i}(B_k) = \sum_k \sum_{\substack{i \\ i \neq i_k}} p_k q_{ki}$

donde i_k es el valor del subíndice i para el cual la probabilidad $P(A_i B_k)$ asume su valor mas grande

p_k : es la probabilidad de que ocurra A_k

q_{kl} : es la probabilidad de que ocurra B_i dado que ocurrió A_k ($p_{A_k}(B_i)$)

$$2^{H_A(B)} = \left(\frac{\sum_k \sum_i e^{c \log p_k q_{ki}}}{\sum_k e^{c \log p_k}} \right)^{-1}$$

$$= \left(\frac{\sum_k \sum_{\substack{i \\ i \neq i_k}} e^{c \log p_k q_{ki}} + \sum_k e^{c \log p_k q_{k i_k}}}{\sum_k e^{c \log p_k}} \right)^{-1}$$

$$=\left(\frac{\sum_{k}\sum_{\substack{i\\ i\neq i_k}} e^{c\log p_k q_{ki}}}{\sum_{k} e^{c\log p_k}}+\frac{\sum_{k} e^{c\log p_k q_{ki_k}}}{\sum_{k} e^{c\log p_k}}\right)^{-1}$$

Si c=1 nos queda:

$$2^{H_A(B)}=\left(\frac{\sum_{k}\sum_{\substack{i\\ i\neq i_k}} e^{\log p_k q_{ki}}}{\sum_{k} e^{\log p_k}}+\frac{\sum_{k} e^{\log p_k q_{ki_k}}}{\sum_{k} e^{\log p_k}}\right)^{-1}$$

$$=\left(\frac{\sum_{k} e^{\ln p_k q_{ki} / \ln 2}}{\sum_{k} e^{\ln p_k / \ln 2}}+\frac{\sum_{k} e^{\ln p_k q_{ki_k} / \ln 2}}{\sum_{k} e^{\ln p_k / \ln 2}}\right)^{-1}$$

$$=\left(\frac{\sum_{k}\sum_{\substack{i\\ i\neq i_k}} e^{\ln p_k q_{ki}\,\ln 2}}{\sum_{k} e^{\ln p_k \ln 2}}+\frac{\sum_{k} e^{\ln p_k q_{ki_k}\,\ln 2}}{\sum_{k} e^{\ln p_k \ln 2}}\right)^{-1}$$

$$=\left(\frac{\sum_{k}\sum_{\substack{i\\ i\neq i_k}} (p_k q_{ki})^{\ln 2}+\sum_{k} (p_k q_{ki_k})^{\ln 2}}{\sum_{k} p_k{}^{\ln 2}}\right)^{-1}$$

$$H_A(B)=\log\left(\frac{\sum_{k}\sum_{\substack{i\\ i\neq i_k}} (p_k q_{ki})^{\ln 2}+\sum_{k} (p_k q_{ki_k})^{\ln 2}}{\sum_{k} p_k{}^{\ln 2}}\right)^{-1}$$

No es posible replicar el lema 2.4 para esta medida de incertidumbre.

6.- $H(A)=\log\left(\sum_{i=1}^{n} e^{\alpha J_i}\right)+(1-\alpha)J_A$

$$H_A(B)=\log\left(\frac{\sum_{k}\sum_{l} e^{-\alpha c \log p_k q_{kl}}}{\sum_{k} e^{-\alpha c \log p_k}}\right)+(1-\alpha)(J(AB)-J(A))$$

$$2^{H_A(B)}=\left(\frac{\sum_{k}\sum_{l} e^{-\alpha c \log p_k q_{kl}}}{\sum_{k} e^{-\alpha c \log p_k}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}$$

Si c = 1

$$2^{H_A(B)}=\left(\frac{\sum_{k}\sum_{l} e^{-\alpha \log p_k q_{kl}}}{\sum_{k} e^{-\alpha \log p_k}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}$$

$$=\left(\frac{\sum_{k=!}^{n}\sum_{\substack{l=! \\ l\neq l_k}}^{m} e^{-\alpha \log p_k q_{kl}}+\sum_{k=!}^{n} e^{-\alpha \log p_k q_{kl_k}}}{\sum_{k=!}^{n} e^{-\alpha \log p_k}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}$$

$$=\left(\frac{\sum_{k=!}^{n}\sum_{\substack{l=! \\ l\neq l_k}}^{m} e^{\log(p_k q_{kl})^{-\alpha}}}{\sum_{k=!}^{n} e^{-\alpha \log p_k}}+\frac{\sum_{k=!}^{n} e^{\log(p_k q_{kl_k})^{-\alpha}}}{\sum_{k=!}^{n} e^{-\alpha \log p_k}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}$$

$$=\left(\frac{\sum_{k=!}^{n}\sum_{\substack{l=! \\ l\neq l_k}}^{m} e^{\frac{\ln(p_k q_{kl})^{-\alpha}}{\ln 2}}}{\sum_{k=!}^{n} e^{\frac{\ln p_k^{-\alpha}}{\ln 2}}}+\frac{\sum_{k=!}^{n} e^{\frac{\ln(p_k q_{kl_k})^{-\alpha}}{\ln 2}}}{\sum_{k=!}^{n} e^{\frac{\ln p_k^{-\alpha}}{\ln 2}}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}$$

$$
=\left(\frac{\sum_{k=!}^{n}\sum_{\substack{l=!\\ l\neq l_k}}^{m} e^{\ln(p_k q_{kl})^{-\frac{\alpha}{\ln 2}}}}{\sum_{k=!}^{n} e^{\ln p_k^{-\frac{\alpha}{\ln 2}}}}+\frac{\sum_{k=!}^{n} e^{\ln(p_k q_{kl_k})^{-\frac{\alpha}{\ln 2}}}}{\sum_{k=!}^{n} e^{\ln p_k^{-\frac{\alpha}{\ln 2}}}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}
$$

$$
=\left(\frac{\sum_{k=!}^{n}\sum_{\substack{l=!\\ l\neq l_k}}^{m} (p_k q_{kl})^{-\frac{\alpha}{\ln 2}}}{\sum_{k=!}^{n} (p_k)^{-\frac{\alpha}{\ln 2}}}+\frac{\sum_{k=!}^{n} (p_k q_{kl_k})^{-\frac{\alpha}{\ln 2}}}{\sum_{k=!}^{n} (p_k)^{-\frac{\alpha}{\ln 2}}}\right)\bullet 2^{(1-\alpha)(J(AB)-J(A))}
$$

Tampoco es posible replicar el lema 2.4 para esta medida de incertidumbre.

De todo el trabajo desarrollado podemos concluir que la medida de incertidumbre de Shannon parece ser la única que puede ser utilizada en la Teoría Matemática de la Información.

Bibliografía

[1] Ash, R.B.: “Information theory”. Interscience J. Wiley. 1.965

[2] Benvenuti, P. - Forte, B. - Kampè de Fèriet, J.: "Su una classe di misure di informazione regolari a traccia Shannoniana". Ani. Sem. Mat. Univ. Modena. XVII, pp. 99-108. 1969.

[3] Bertoluzza, C. - Schneider, M.: "Informations Totalement composables". Lecture Notes in Math. m. 398. 1974.

[4] Capodieci, M.L.: “Clasificación y Construcción de las Medidas de Incertidumbre”. Trabajo de grado de M.I.M. UCLA - IUEPB - IUP. 1986.

[5] Capodieci, M.L.: “Sobre algunos problemas del Análisis del Incierto”. Tesis Doctoral. UCV. 1999.

[6] Divari, M. – Pandolfi, M.: "Su una legge compositiva dell'informaziones". Rend. Mat. Univ. Roma, serie 6(3), pp. 805-817. 1970.

[7] Forte, B.: “ Measures of Information: the general axiomatic theory”. R.I.R.O., 3° année, N° R-2, pp.63-90. 1.969.

[8] Kampé de Feriét, J - Forte, B.: “ Information et Probabilité”. C.R. Acad. Sci. Paris. Ser A 265.

[9] Forte, B. – Pintacuda, N.: "Sull'informazione associata alle esperienze incompleta". Annali di Matematica pura ed applicata. (IV), Vol, LXXX, pp.215-234.

[10] Kappos, D.A.: "Strukturtheorie der Wahrscheimlichkeits-raumeüd Felder". Springer Verlag. 1966.

[11] Khinchin, A.I.: “ Mathematical Foundations of Information Theory”. Dover. 1.957.

[12] C.H. Ling: "Representation of asociative functions". Publ. Math. V.12, pp.189-212. 1965.

[13] Poggi, C.: "Entropie G-diramative ed Entropie G-locali in spazio dell'informazione con legge di independenza di tipo generale". Statistica, ann XLIII, n.2.1983.

[14] Shannon, C.E.: " A Mathematical Theory of Communication" Bell System Tech. J.27. 1.948.

[15] Vásquez, I.: "Compositividad, Ramificación y Maximalidad en la Incertidumbre". Trabajo de Grado. Maestría. UCLA. Barquisimeto-Venezuela. 2.003.

Bibliografía de Consulta

[A] ACZEL, J. - DAROCZY, Z. " On measures of information and their characterizations". Academic Press. New York. MATHEMATICS IN SCIENCE AND ENGINEERING. Volume 115.

[B] BERTOLUZZA, C. - DE SIMONI, S. " Elementos de la Teoría de la Información".

[C] CAPODIECI, M.L. " Notas de Teoría de la Información y de la Incertidumbre". Notas de la asignatura electiva Elementos de la Teoría de la Información, dictada a los estudiantes de la Licenciatura en Ciencias Matemáticas del Decanato de Ciencias de la UCLA. 1988.