Realizado el acto público de la Defensa y Mantenimiento de esta Tesis Doctoral el día 23 de Marzo de 1985, en la Universidad de Granada, ante el tribunal formado por:

Presidente: Dr. Don Enrique Trillas Ruíz

Vocales: Drª Dª Maria Amparo Vila Miranda

Dr. Don Elie Sanchez

Dr. Don Pedro Gil Alvarez

Secretario: Dr. Don José Luis Verdegay Galdeano

obtuvo la calificación de

APTO CUM LAUDE

R.56.545

"INFORMACION DIFUSA, RELACIONES ENTRE PROBABILIDAD Y POSIBILIDAD ".

Memoria que, para optar al grado de Doctor, presenta el licenciado en Ciencias Matemáticas - Serafín Moral Callejón.

Director de Tesis:
Profesor Dr. D. Miguel Delgado Calvo-Flores.

V° B°

UNIVERSIDAD DE GRANADA, FACULTAD DE CIENCIAS, Febrero de 1.985

Quiero expresar mi agradecimiento a los profesores Dr. D. Miguel Delgado Calvo-Flores, Dr. D. J.L. Verdegay Galdeano y Drª. Dª. Amparo Vila Miranda, por su labor de dirección y estímulo constante. Así mismo, quiero hacer extensivo este agradecimiento a todos los miembros del "grupo difuso" de Granada, a los del departamento de Estadística de esta Universidad y a todos aquellos que, de alguna forma, hayan contribuído a la realización de esta memoria.

INDICE

- INTRODUCCION GENERAL -

Según la filosofía bayesiana, la probabilidad de un suceso - representa el grado subjetivo de incertidumbre del experimentador,sobre la realización de dicho suceso. Debido al carácter -- subjetivo de este modelo, los investigadores bayesianos afirman que es suficiente para describir todos los casos de incertidumbre e, incluso, algunos de ellos llegan a afirmar que la probabilidad es la única representación válida de la incertidumbre. Discutible es ya la primera afirmación (muchos autores mantienen el carácter subjetivo de la probabilidad. Por ejemplo, Gnedenko (1.969) afirma: "La diversificada experiencia resultante de aplicar la probabilidad a una amplia gama de campos, nos enseña que muchos problemas de dar una estimación cuantitativa - de la probabilidad de un suceso tienen un significado intuitivo razonable solo bajo unas condiciones bastante definidas.... De la naturaleza no determinística de un suceso, no se sigue - que tenga sentido hablar de su probabilidad como un número definido, aún cuando éste no sea conocido"). Mucho más difícil - de aceptar nos resulta, entonces, la afirmación de que no existen otros modelos que el probabilístico (bayesiano), para describir los numerosos casos distintos de experimentos con incertidumbre.

Creemos que la razón histórica de la Teoría Bayesiana de la Probabilidad es la necesidad de aprovechar las potentes y bien desarrolladas herramientas que ofrecía la Teoría de la Probabi lidad clásica, para situaciones en las que ésta no era directamente aplicable.

Sin embargo, ultimamente se están desarrollando formas alternativas de representación de la incertidumbre y, dentro de este campo, podemos incluir el tema de la presente memoria. Desde -- nuestro punto de vista, son claves las siguientes referencias:

- Zadeh (1.965) inició el estudio de la incertidumbre derivada de la utilización de predicados difusas (predicados poco precisos). Esta vaguedad es inherente a todo sistema en el cual se emplee el lenguage humano. Un modelo para tratar esta información, tan poco precisa,lo proporciona la Teoría de Conjuntos Difusos.

- Sugeno (1.974) generalizó el concepto de medida de probabilidad, introduciendo las llamadas medidas difusas. Realmente, - Sugeno las introduce como medidas monótonas en subconjuntos ordinarios.

- Shafer (1.975), independientemente también de la Teoría de Conjuntos Difusos, introdujo un modelo de representación de la incertidumbre, la Teoría de la Evidencia, que generaliza la Teoría de la Probabilidad y utiliza medidas difusas en el sentido de Sugeno.

- Zadeh (1.978) introdujo las medidas de posibilidad como representación de informaciones dadas mediante predicados difusos. Estas medidas resultan ser evidencias del tipo analizado por -- Shafer, si bien, presentadas desde un punto de vista completamente distinto.

La Teoría de la Evidencia, en general, y la Teoría de la Posibilidad, en particular, han experimentado ultimamente un importante desarrollo, pero queda mucho por hacer, ya que, a nuestro criterio, no existen aún todas las herramientas necesarias para una amplia utilización de estos modelos.

El objetivo fundamental de este trabajo es el estudio de la incertidumbre representada por medidas de posibilidad y su rela ción con la incertidumbre probabilística.

El motivo que nos ha llevado a considerar este problema surge de un artículo de Y. Leung (1.980). En él y con base en el principio de máxima entropía, Leung propone un procedimiento para asignar una distribución de probabilidad "a priori" en un experimento, supuesto que se conoce una distribución de posibilidad sobre sus resultados. En concreto, se considera la distribución "a priori" solución de

$$\text{Max} \quad H = -k \sum_{a \in U} p(a) L(p(a))$$

$$\text{s.a.} \quad \sum_{a \in U} p(a) = 1$$

$$\sum_{a \in U} p(a)\pi(a) = \alpha$$

La restricción $\sum_{a \in U} p(a)\pi(a) = \alpha$, indica que la "consistencia" entre p y π (medida de acuerdo con el criterio introducido por Zadeh (1.978)) ha de ser α (que, en general, deberá estar próximo a uno).

Al tratar de generalizar este modelo (a las formas a las que posteriormente nos referiremos), encontramos que se necesitaba una mejor comprensión del término $\sum_{a \in U} p(a)\pi(a)$.

Zadeh (1.978) no define la consistencia de una forma precisa; solo se introduce, de forma intuitiva, que la cantidad $\sum_{a \in U} (a)\pi(a)$ es "una formalización aproximada de la observación heurística de que una disminución de la posibilidad lleva consigo una disminución de la probabilidad, pero no viceversa". Es decir, que existe la siguiente relación entre posibilidad y probabilidad: lo probable es posible, pero no reciprocamente.

Sobre el tema de la consistencia y relaciones entre posibilidad y probabilidad han trabajado, además de Zadeh, autores como Dubois y Prade (1.982, 1.983), Hohle (1.982), Natvig (1.982) y Lindley (1.980). En estos trabajos, hemos encontrado que el ---

principio de consistencia no ha sido definido con precisión y solo se han realizado algunas consideraciones intuitivas sobre él. Sin embargo, creemos que es un concepto particularmente interesante. Si en un sistema se trabaja con informaciones diversas, aunque estas sean de naturaleza distinta, es fundamental estudiar la coherencia, no contradicción o consistencia entre dichas informaciones. Por lo tanto, nos hemos propuesto profundizar en este concepto para definirlo rigurosamente.

Cuando originalmente nos planteamos este problema, nos encontramos que necesitábamos conceptos y relaciones que no habían sido consideradas con anterioridad. Por tanto, nos hemos visto obligados a profundizar en dos direcciones, cuyos resultados han sido fundamentales en la consecución de nuestros objetivos. Estas direcciones son:

- Los fundamentos de la Teoría de la Posibilidad. Así como la probabilidad está suficientemente bien estudiada, en la Teoría de la Posibilidad (debido a su corta existencia) existen muchos conceptos de los cuales no hay definiciones universalmente aceptadas, así como, parcelas de la misma que no han sido exploradas.

-La Teoría de la Evidencia de Shafer. Esta teoría tiene, como casos particulares a la Posibilidad y la Probabilidad; por lo tanto, a la hora de estudiar la Teoría de la Posibilidad y las relaciones de ésta con la Probabilidad, siempre será interesante considerar el punto de vista global que proporciona la Teoría de la Evidencia.

En el primer capítulo, comenzamos haciendo unas definiciones generales sobre medidas difusas. En el segundo apartado, estudiamos la integral difusa introducida por Sugeno (1.974), junto con las generalizaciones (integrales seminormada y semiconormada) de F. Suárez (1.983, 1.984). Estas integrales serán un instrumento eficaz para medir la consistencia, tal y como veremos en el tercer capítulo de esta memoria.

El resto del Capítulo I está dedicado a estudiar la Teoría de la Evidencia, considerándose aquellos aspectos que utilizaremos en desarrollos posteriores. Hay que indicar que la Teoría de la Evidencia proporciona, en algunos casos, una interpretación intuitiva de resultados obtenidos teoricamente. Por otra parte, justifica algunos conceptos que introduciremos en capítulos posteriores.

Los problemas fundamentales que nos hemos planteado son el de la combinación de la evidencia (cuarto apartado) y el de medir la cantidad de información asociada a una evidencia (quinto apartado).

Para combinar la evidencia, solo se ha desarrollado un método, que es la regla de Demster y que proporciona buenos resultados en numerosas ocasiones (probabilidad condicionada clásica, probabilidad condicionada de Zadeh). Sin embargo no es un procedimiento universal. Shafer (1.976) reconoce que debe de existir cierta relación especial entre dos evidencias para que estas se puedan combinar según la regla de Dempster. Sería interesante disponer de un método que sirviese para combinar evidencias cualesquiera. Nosotros hemos definido una relación de inclusión entre evidencias (definición 1.16),que puede servir de base para obtener una regla de este tipo. De hecho, nos servirá para resolver el problema en el caso particular de evidencias posibilísticas.

Sobre medidas de la cantidad de incertidumbre para evidencias, solo existen dos trabajos: Yager (1.983) y Dubois y Prade (1.984). Yager propone dos medidas: entropía (mide la disonancia de la evidencia) y especificidad (mide la precisión de la evidencia). Nosotros introducimos las medidas de entropía superior y entropía inferior. Cada una de las cuales intenta recoger en una única cantidad la incertidumbre total (imprecisión y disonancia) de una evidencia.

En el capítulo segundo consideramos, en primer lugar, el problema de asignar una distribución de posibilidad a una información difusa. La diferencia esencial con el método propuesto en Zadeh (1.978) es que nosotros solo utilizamos distribuciones de posibilidad normalizadas. También resolvemos (apartado 2) el problema de la combinación de distribuciones de posibilidad y el de definir la compatibilidad de dos distribuciones de posibilidad.

A continuación, estudiamos cómo se puede medir la cantidad de información asociada a una distribución de posibilidad. Las medidas de entropía de De Lucca y Términi (1.972 , 1.979) valoran cuan difuso es un conjunto, de modo que si las trasladamos a las distribuciones de posibilidad no miden la incertidumbre asociada a la misma. Ha habido algunos intentos de medir esta incertidumbre directamente: Yager (1.981, 1.983), Higashi et al. (1.983) y Czogala et al. (1.982). Nosotros recogemos, en este punto, la axiomática propuesta en S. Moral (1.984), que está basada en las axiomáticas sobre medidas de información de Kampé de Feriet (1.970, 1.974). Presentamos, también, numerosos funcionales que pueden ser considerados como medidas de información (incertidumbre) y que serán de gran utilidad en el capítulo tercero, cuando planteemos el problema de asignar una distribución de posibilidad a una de probabilidad dada.

El último apartado de este segundo capítulo está dedicado a estudiar un tipo particular de informaciones difusas. En todo lo anterior, tanto las informaciones difusas como las evidencias y las distribuciones de posibilidad, se refieren a un elemento, x, desconocido, perteneciente a un referencial U. En este apartado, consideramos que lo desconocido es un subconjunto $A \subset U$, y disponemos de una información difusa sobre todos los elementos de dicho conjunto (afirmamos una propiedad difusa que han de cumplir todos sus elementos).

Consideramos como se puede asignar a estas informaciones, distribuciones de posibilidad que las hagan más manejables matemáticamente, lo que será esencial para nuestra formulación del principio de consistencia posibilidad-probabilidad.

Con las herramientas que hemos desarrollado en los capítulos primero y segundo, en el capítulo tercero nos planteamos el objetivo fundamental de esta memoria: estudiar las relaciones entre Posibilidad y Probabilidad, concretadas en tres problemas fundamentales:

1) Estudiar la consistencia entre posibilidad y probabilidad.

2) Asignar una distribución de probabilidad a una de posibilidad.

3) Asignar una distribución de posibilidad a una de probabilidad.

Comenzamos el capítulo haciendo un resumen de los trabajos más significativos que tratan sobre el principio de consistencia posibilidad-probabilidad. Además de Zadeh (1.978), la contribución más importante sobre dicho principio la realizan Dubois y Prade (1.982, 1.983). Dichos autores, recogiendo la idea intuitiva de Zadeh de que "lo probable es posible", establecen que una medida de probabilidad P y una de posibilidad Π son consistentes si y solo si

$$\Pi(A) \geq P(A), \quad \forall A \subset U$$

Con ayuda de esta condición, Dubois y Prade proponen una solución para los problemas 2) y 3). De hecho, el problema de asignar una distribución de posibilidad a una de probabilidad dad, se encuentra planteado, por primera vez, en Dubois y Prade (1.982).

Nuestro concepto de consistencia lo presentamos en el segundo apartado del tercer capítulo. Definimos la consistencia como una propiedad difusa relativa al grado de coherencia (ausencia de contradicción) existente entre una distribución de posibilidad y una de probabilidad.

Hay que destacar que, dadas una distribución de probabilidad y otra de posibilidad sobre un conjunto, no siempre tiene sentido hablar de consistencia. Si la distribución de posibilidad se refiere a un resultado particular de un experimento aleatorio, no cabe cuestionarse nada acerca de la coherencia (tal y como nosotros la definimos) de ambas distribuciones. Por el contrario, tendrá pleno sentido hablar de consistencia cuando la distribución de posibilidad se refiera a todos los resultados de un experimento aleatorio que tenga, como distribución de proba bilidad, la dada.

El caso de informaciones difusas y distribuciones de posibilidad sobre un resultado de un experimento aleatorio, ha sido considerado por Takeda et al. (1.975, 1.976). Sus resultados son coherentes con la regla de Demster de combinar la evidencia.

Como la consistencia es una propiedad difusa, deberá medirse en términos de grado de cumplimiento. Para este fin, en este mismo apartado, damos una axiomática que debe de cumplir un funcional,para que pueda considerarse como una medida de consistencia, y presentamos varios casos: la medida de Zadeh, la condición de Dubois y Prade (que da lugar a una medida "crisp") y una medida basada en las integrales seminormadas y semiconormadas, entre otras.

El problema de asignar una distribución de probabilidad a una de posibilidad se considera en el apartado tres. Nosotros proponemos que, siguiendo el principio de máxima entropía, la distribución de probabilidad que hay que considerar es la solución de

$$\begin{array}{ll} \text{Max} & H = -k \sum_{a \in U} p(a) L(p(a)) \\ \text{s.a.} & \sum_{a \in U} p(a) = 1 \\ & \text{"p es consistente con } \pi\text{"} \end{array}$$

que es un problema de programación difuso, cuya forma concreta dependerá de la medida de consistencia empleada. Resolveremos dicho problema por el método de J.L. Verdegay (1.980, 1.982), obteniendo la solución explícita para los casos más importantes de medidas de consistencia.

La asignación de una distribución de posibilidad, π, a una de probabilidad, p, lo planteamos de forma "dual" al anterior:

$$\begin{array}{ll} \text{Max} & I(\pi) \\ \text{s.a.} & \pi \in \prod(U) \\ & \text{"}\pi\text{ es consistente con p"} \end{array}$$

donde $\prod(U)$ es el conjunto de las distribuciones de posibilidad normalizadas, sobre U, e I una medida de información en él.

Este problema lo resolvemos por el mismo método que el anterior. Obtenemos la solución explícita considerando varios casos de medidas de información y de consistencia.

Por último, hay que hacer notar que, a lo largo de toda la memoria, trabajamos con conjuntos finitos. Esto lo hacemos para poder centrarnos mejor en nuestros objetivos, evitando las perturbaciones que, por cuestiones de convergencia, medibilidad, etc. (ajenas al problema de la evaluación de la información), puede acarrear el uso de conjuntos infinitos.

Los problemas abiertos, por los que pensamos continuar el trabajo de esta memoria son:

- Profundizar en la cuestión de la combinación de la evidencia.

- Sistematizar las medidas de entropía asociadas a las evidencias, introduciendo una axiomática para las mismas.

- Extender el concepto de esperanza matemática a evidencias y distribuciones de posibilidad.

- Aplicar los resultados del tercer capítulo a la Estadística Matemática: test de hipótesis, estimación puntual y por intervalos de parámetros, etc..

- CAPITULO I -

Medidas Difusas. Integral de Sugeno. Teoría de la Evidencia.

0.-INTRODUCCION

En este capítulo vamos a presentar conceptos previos y herramientas necesarias, para el desarrollo de los dos capítulos posteriores; en los que estudiaremos informaciones difusas y posibilísticas y su relación con la incertidumbre probabilística.

Comenzamos con algunas definiciones generales sobre medidas difusas, siguiendo a Sugeno(1.972) .Consideramos especialmente los casos particulares de medidas de posibilidad y de probabilidad.Todos los conceptos están dados para conjuntos finitos ya que, como indicamos en la Introducción General no vamos a considerar conjuntos infinitos para evitar complicaciones no inherentes a nuestro tema.

En el segundo apartado estudiamos, en primer lugar, la integral de Sugeno(1.972) junto con las generalizaciones (integrales seminormadas y coseminormadas) de F.Suárez(1.983 y 1.984).Estas integrales serán utilizadas en el Capítulo III para obtener medidas de consistencia posibilidad-probabilidad.Dentro de este apartado hemos obtenido (Proposición 1.2) una relación de dualidad entre las integrales seminormada y coseminormada.Aunque señalamos algunas de las posibles aplicaciones de dicha relación, no las desarrollaremos ya que se salen del tema y objetivos de esta memoria.

Los puntos tercero, cuarto y quinto se dedican a estudiar diversos aspectos de la Teoría de la Evidencia de Shafer(1.976).Esta teoría ha sido desarrollada inicialmente fuera del marco de la Teoría de Conjuntos Difusos y está basada en los trabajos de

Dempster(1.967).Shafer considera una representación de la incertidumbre (evidencia) que generaliza la representación que proporciona la Teoría de la Probabilidad.Muchos probabilistas bayesianos afirman que la única representación "racional" de la incertidumbre (ver p.e. De Finetti(1.974),Lindley(1.982)) es la proporcionada por la Teoría de la Probabilidad; racionalidad que supone la aceptación y seguimiento de un conjunto de axiomas.La Teoría de la Evidencia no impone la aceptación de estos axiomas, -- por tanto, podemos considerarla como más general y, por otra parte, proporciona unos resultados y reglas de actuación que son -- bastante intuitivos. Como consecuencia de esto ha tenido un gran desarrollo en los últimos años, especialmente dentro de el campo de los Conjuntos Difusos . Este hecho está justificado por los - puntos de contacto existentes entre ambas teorías y que podemos concretar en:

-Las medidas de creencia y plausibilidad utilizadas por Shafer como representaciones de la evidencia son casos particulares de medidas difusas.

-Zadeh (1.978) propuso asignar a una información difusa una medida de posibilidad y estas medidas son un caso particular de -- plausibilidad anteriormente mencionadas.

El estudio de la Evidencia es particularmente importante para nosotros ya que, al contener como caso particular a la Posibilidad y a la Probabilidad, siempre será interesante considerar el punto de vista que proporciona esta teoría a la hora de estudiar las medidas de posibilidad y sus relaciones con las de probabilidad.

En el cuarto apartado se dan las relaciones y definiciones -- fundamentales que se encuentran en el libro de Shafer, además de una caracterización que obtenemos para las funciones de comunalidad.

En el punto cinco tratamos el problema de la combinación de - la evidencia. Existe una primera solución a este problema que -- viene proporcionada por la regla de Dempster; sin embargo, mostramos como dicha regla no tiene una validez universal para combinar todo tipo de evidencias. Nosotros no vamos a dar un método para efectuar esta combinación en cualquier situación, pero estableceremos las bases que nos van a permitir resolver este problema en el caso particular de evidencias posibilísticas, que es --

nuestro objetivo fundamental.

Yager (1.983) ha introducido dos medidas asociadas a una evidencia: entropía (mide la "disonancia" de la evidencia) y especificidad (mide la "precisión" de la evidencia). Yager afirma que estas medidas, conjuntamente, miden la cantidad de incertidumbre asociada a una evidencia: habrá más incertidumbre cuanto más entropía ("disonancia") y menos especificidad ("precisión") haya.

Dubois, Prade (1.984) han obtenido otra medida análoga a la de especificidad, generalizando las que Higashi, Klir (1.983) llaman "entropías posibilísticas".

El apartado quinto, y último de este capítulo, está dedicado a la presentación de estos conceptos, así como a la definición y análisis de las medidas de incertidumbre que hemos denominado entropía superior y entropía inferior. Con ellas intentamos resumir en un valor único la entropía y especificidad de una evidencia.

1.-DEFINICIONES GENERALES SOBRE MEDIDAS DIFUSAS

En este apartado vamos a recoger algunos de los conceptos fundamentales sobre medidas difusas. Nos vamos a centrar solamente en aquellos aspectos que sean de utilidad en nuestro posterio estudio de las medidas de posibilidad y su relación con las de probabilidad. Un estudio más amplio y detallado de estas cuestiones puede encontrarse en Dubois, Prade (1.980).

Sea U un conjunto que, por comodidad, consideraremos finito a lo largo de esta memoria. Sea $\mathcal{A} \subset \mathcal{P}(U)$ un álgebra. Vamos a comenzar dando la definición de medida difusa de Sugeno (1.972) en conjuntos finitos.

Definición 1.1.- Sea $h:\mathcal{A} \longrightarrow [0,1]$, se dice que h es una medida difusa en $(U,\mathcal{A})$ si y solo si cumple

i) $h(\phi)=0$, $h(U)=1$

ii) $A \subset B \Longrightarrow h(A) \leq h(B)$

Nota.- Para evitar problemas de medibilidad y dado el carácter finito de U, consideraremos siempre que $\mathcal{A} \equiv \mathcal{P}(U)$; y, por tanto las medidas difusas se definen en $(U,\mathcal{P}(U))$.

Definición 1.2.- Sea $h: \mathcal{P}(U) \longrightarrow [0,1]$ una medida difusa. Se llama medida difusa dual de h a la medida h^* dada por

$$h^*(A) = 1-h(\overline{A})$$

Ejemplo 1.1.- Una medida de probabilidad $P: \mathcal{P}(U) \longrightarrow [0,1]$ es una medida difusa (cumple i) y ii)) cuya dual viene dada por

$$P^*(A) = 1-P(\overline{A}) = 1-(1-P(A)) = P(A)$$

Es decir $P^* \equiv P$. Las medidas de probabilidad son las únicas medidas difusas que coinciden con sus duales.

Las t-normas y t-conormas son de gran utilidad en el estudio de las medidas difusas, por lo que recogemos las siguientes definiciones; que pueden encontrarse en Schweizer, Sklar (1.963).

Definición 1.3.- Una aplicación

$$\otimes : [0,1] \times [0,1] \longrightarrow [0,1]$$

se dice que es una conorma triangular (t-conorma) si y solo si cumple

1) Propiedades de frontera.- $1 \otimes 1 = 1$, $0 \otimes a = a \otimes 0 = a$, $\forall a\ [0,1]$
2) Monotonía.- Si $a \leqslant b$, $c \leqslant d$, entonces $a \otimes c \leqslant b \otimes d$; $\forall a,b,c,d\ [0,1]$
3) Conmutativa.- $a \otimes b = b \otimes a$; $\forall a,b\ [0,1]$
4) Asociativa.- $a \otimes (b \otimes c) = (a \otimes b) \otimes c$; $\forall a,b,c\ [0,1]$

Una aplicación

$$\perp : [0,1] \times [0,1] \longrightarrow [0,1]$$

se dice que es una norma triangular (t-norma) si y solo si cumple las propiedades 2), 3) y 4) anteriores y, además

1') $0 \perp 0 = 0$, $1 \perp a = a \perp 1 = a$, $\forall a \in [0,1]$

F. Suárez (1.983) ha debilitado los anteriores conceptos introduciendo las definiciones de seminorma triangular (t-seminorma) y semiconorma triangular (t-semiconorma).

Definición 1.4.- Una aplicación $S' : [0,1] \times [0,1] \longrightarrow [0.1]$ que cumpla las propiedades 1) y 2) anteriores se dice que es una t-semiconorma.

Una aplicación $S : [0,1] \times [0,1] \longrightarrow [0,1]$ que cumpla las propiedades 1') y 2) anteriores se dice que es una t-seminorma.

Definición 1.5.- Sea g una medida difusa en U, se dice que g está basada en la conorma $\otimes$ si y solo si cumple

Para cada $A,B \subset U$, $A \cap B = \phi \Longrightarrow g(A \cup B) = g(A) \otimes g(B)$

-Se dice que la medida difusa f está basada en la norma $\perp$ si y solo si cumple

Para cada $A,B \subset U$ $\quad A \cup B = U \Longrightarrow f(A \cap B) = f(A) \perp f(B)$

Propiedad.- Si g es una medida difusa basada en la conorma $\otimes$ entonces su dual, g^*, es una medida difusa basada en la norma $\otimes'$ definida por: $a \otimes' b = 1-(1-a) \otimes (1-b)$. A esta norma $\otimes'$ se le llama norma dual de $\otimes$.

-Si f es una medida difusa basada en la norma $\perp$, entonces su dual, f^*, es una medida difusa basada en la conorma $\perp'$, definida por: $a \perp' b = 1-(1-a) \perp' (1-b)$. A esta conorma $\perp'$, se le llama conorma dual de .

Ejemplo 1.3.- La aplicación v, definida por avb=Max{a,b} es - una conorma en $[0,1]$. Las medidas, Π, que están basadas en esta conorma se llaman medidas de posibilidad y cumplen

$$\Pi(A \cup B) = \Pi(A) \vee \Pi(B) = \text{Max}\{\Pi(A), \Pi(B)\}, \forall A,B \subset U.$$

La función $\quad \pi: U \longrightarrow [0,1]$

$\quad a \longmapsto \Pi(\{a\})$

se llama distribución de posibilidad asociada a Π, y permite expresar la medida Π como

$$\Pi(A) = \underset{x \in A}{\text{Max}} \{\pi(x)\}$$

La medida dual de Π : $N(A) = 1-\Pi(\bar{A})$, es una medida basada en la norma $a v' b = 1-(1-a)v(1-b) = \text{Min}\{a,b\} = a \wedge b$; y, por tanto, cumple

$$N(A \cap B) = \text{Min}\{N(A), N(B)\}, \text{ para cada } A, B \subset U.$$

Este tipo de medidas reciben el nombre de medidas de necesi-- dad.

Notación.- A partir de ahora Max {a,b} y Min {a, b} lo notaremos, a veces, como avb y $a \wedge b$, respectivamente.

Ejemplo 1.3.- Las medidas de probabilidad están basadas, a la vez, en la conorma:

$$S_p(a,b) = a + b$$

y en su norma dual:

$$S_p'(a,b) = a + b - 1 .$$

2.-INTEGRAL DE SUGENO

Sugeno (1.972) introdujo la siguiente definición de integral difusa

Definición 1.6.-Dado $A \in \mathcal{T}(U)$, $h:U \longrightarrow [0,1]$ una aplicación y g una medida difusa en $(U, \mathcal{T}(U))$, entonces la integral difusa de h, sobre A, respecto a g, se define como el valor

$$\int_A h(x) \circ g(.) = \operatorname*{Sup}_{F \in \mathcal{T}(U)} \left[\operatorname{Min}\{ \operatorname*{Inf}_{x \in F} h(x),\ g(A \cap F) \} \right] \quad (1)$$

-Esta integral difusa se puede expresar de la forma

$$\int_A h(x) \circ g(.) = \operatorname*{Sup}_{\alpha \in [0,1]} \left[\operatorname{Min}\{\alpha ,\ g(A \cap H_\alpha)\} \right]$$

donde $H_\alpha = \{ x \in U \ / \ h(x) \geq \alpha \}$

Definición 1.7.- La integral difusa de h sobre $A \in \mathcal{T}(U)$, respecto a g es el valor

$$\int_A h(x) \circ g(.) = \int_U (\mu_A(x) \wedge h(x)) \circ g(.)$$

donde $\mu_A(x)$ es la función de pertenencia de A.

Algunas de las propiedades más importantes de esta integral son:

(1) Si $a \in [0,1]$, entonces $\int_U a \circ g(.) = a$

(2) $h \leq h' \implies \int_U h \circ g(.) \leq \int_U h' \circ g(.)$

(3) $\int_U (h_1 \vee h_2) \circ g \geq \int_U h_1 \circ g(.) \vee \int_U h_2 \circ g(.)$

(4) $\int_U (h_1 \wedge h_2) \circ g \leq \int_U h_1 \circ g \wedge \int_U h_2 \circ g$

(5) $A \subset B \implies \int_A h \circ g \leq \int_B h \circ g$

(6) Si g es una medida de probabilidad

$$\left| \int_U h \circ g - \int_U h . dg \right| \leq 1/4$$

(7) Si $U = \{u_1, \ldots, u_n\}$ con $h(u_i) \leq h(u_{i+1})$, $1 \leq i \leq n-1$ entonces

$$\oint_A h(u)\circ g(\cdot) = \operatorname*{Sup}_{1\leq i\leq n} \{ h(u_i) \wedge g(A \cap K_i)\}$$

donde $K_i = \{u_i, u_{i+1}, \ldots, u_n\}$

F. Suárez (1.983, 1.984) ha extendido este concepto de integral difusa de la siguiente forma.

<u>Definición 1.8.-</u> Sea S una t-seminorma y U un conjunto, entonces la integral seminormada de $h:U \longrightarrow [0,1]$, sobre $A \subset U$, respecto a la medida difusa g es el valor

$$N\text{-}\int_{S\ A} h(x)\circ g(\cdot) = \operatorname*{Sup}_{\alpha\in[0,1]} S(\alpha, g(A\cap H_\alpha))$$

donde $H_\alpha = \{x \ / \ h(x) \geq \alpha\}$

La integral de Sugeno es un caso particular de esta integral seminormada cuando $S(x,y) = x \wedge y$

<u>Definición 1.9.-</u> Sea S' una t-semiconorma y U un conjunto, entonces la integral semiconormada de $h:U \longrightarrow [0,1]$, sobre $A \subset U$, respecto a la medida difusa g, es el valor

$$C\text{-}\int_{S'\ A} h(x)\circ g(\cdot) = \operatorname*{Inf}_{\alpha\in[0,1]} S'(\alpha, g(A\cap H'_\alpha))$$

donde $H'_\alpha = \{x \ / \ h(x) > \alpha\}$

La integral de Sugeno es un caso particular de la integral semiconormada cuando $S'(x,y) = x \vee y$.

Las propiedades más importantes de estas integrales son:

i) Si $a \in [0,1]$, entonces $N\text{-}\int_{S\ U} a\circ g(\cdot) = a$, $C\text{-}\int_{S'\ U} a\circ g(\cdot) = a$

ii) $h_1 \leq h_2 \implies N\text{-}\int_{S\ A} h_1(x)\circ g(\cdot) \leq N\text{-}\int_{S\ A} h_2(x)\circ g(\cdot)$ y

$$C\text{-}\int_{S'\ A} h_1(x)\circ g(\cdot) \leq C\text{-}\int_{S'\ A} h_2(\cdot)\circ g(\cdot)$$

iii) $N\text{-}\int_{S\ U} \mu_A(x)\circ g(\cdot) = C\text{-}\int_{S'\ U} \mu_A(x)\circ g(\cdot) = g(A)$, para cada $A \subset U$

iv) $N\text{-}\int_{S\ A} (h_1 \vee h_2)(x)\circ g(\cdot) \geq N\text{-}\int_{S\ A} h_1(x)\circ g(\cdot) \vee N\text{-}\int_{S\ A} h_2(x)\circ g(\cdot)$

$$C\text{-}\int_{S',A}(h_1 \vee h_2)(x)\circ g(.) \geq C\text{-}\int_{S',A} h_1(x)\circ g(.) \vee C\text{-}\int_{S',A} h_2(x)\circ g(.)$$

v) $A_1 \subset A_2 \implies$ $$N\text{-}\int_{S\ A_1} h(x)\circ g(.) \leq N\text{-}\int_{S\ A_2} h(x)\circ g(.)$$

$$C\text{-}\int_{S',A_1} h(x)\circ g(.) \leq C\text{-}\int_{S',A_2} h(x)\circ g(.)$$

vi)-Sea $M = N\text{-}\int_{S\ U} h(x)\circ g(.)$ y sea h' la función

$$h'(x) = \begin{cases} M & \text{si } h(x) < M \\ h(x) & \text{en otro caso} \end{cases}$$

se verifica, entonces, que

$$N\text{-}\int_{S\ U} h(x)\circ g(.) = N\text{-}\int_{S\ U} h'(x)\circ g(.)$$

-Sea $M' = C\text{-}\int_{S',U} h(x)\circ g(.)$ y la función h^*, definida por

$$h^*(x) = \begin{cases} M' & \text{si } h(x) > M' \\ h(x) & \text{en otro caso} \end{cases}$$

se verifica, entonces que

$$C\text{-}\int_{S',U} h(x)\circ g(.) = C\text{-}\int_{S',U} h^*(x)\circ g(.)$$

vii) Si S es una t-seminorma y S' una t-semiconorma. Entonces para cada $A \subset U$ se cumple

$$N\text{-}\int_{S\ A} h(x)\circ g(.) \leq C\text{-}\int_{S',A} h(x)\circ g(.)$$

viii) $$N\text{-}\int_{S\ A} h(x)\circ g(.) = \sup_{H\in\mathcal{T}(U)} S(\inf_{x\in H} h(x),\ g(A\cap H)) \qquad (2)$$

Estas propiedades pueden encontrarse demostradas en F. Suárez (1.983).

La siguiente proposición proporciona un resultado paralelo a (2) para el caso particular de integrales semiconormadas

<u>Proposición 1.1.-</u> Sea g una medida difusa en U, entonces

$$C\text{-}\int_{S',A} h(x)\circ g(.) = \operatorname*{Inf}_{H\in\mathfrak{I}(U)} S'(\operatorname*{Sup}_{x\in H} h(x),\ g(A\cap H)) \qquad (3)$$

<u>Demostración.-</u>

Sabemos que $C\text{-}\int_{S',A} h(x)\circ g(.) = \operatorname*{Inf}_{\alpha\in[0,1]} S'(\alpha,\ g(A\cap H'_\alpha))$

donde $H'_\alpha = \{x\in U \ / \ h(x) > \alpha\}$

Ahora bien, para todo $\alpha\in[0,1]$ se tiene

$\operatorname*{Sup}_{x\in H'_\alpha} h(x) \leqslant \alpha$ y, por tanto

$$S'(\operatorname*{Sup}_{x\in H'_\alpha} h(x),\ g(A\cap H'_\alpha)) \leqslant S'(\alpha,\ g(A\cap H'_\alpha))$$

de donde

$$\operatorname*{Inf}_{H\in\mathfrak{I}(U)} S'(\operatorname*{Sup}_{x\in H} h(x),\ g(A\cap H)) \leqslant$$

$$\operatorname*{Inf}_{\alpha\in[0,1]} S'(\operatorname*{Sup}_{x\in H'_\alpha} h(x),\ g(A\cap H'_\alpha)) \leqslant$$

$$\operatorname*{Inf}_{\alpha\in[0,1]} S'(\alpha,\ g(A\cap H'_\alpha)) = C\text{-}\int_{S',A} h(x)\circ g(.) \qquad (4)$$

Por otra parte, si $H\in\mathfrak{I}(U)$ podemos considerar $\alpha_H = \operatorname*{Sup}_{x\in H} h(x)$ con lo que se obtiene $H = H'_{\alpha_H}$, $\alpha_H\in[0,1]$. Así pues, para cada H en (U), existe $\alpha_H\in[0,1]$ tal que

$$S'(\operatorname*{Sup}_{x\in H} h(x),\ g(A\cap H)) = S'(\alpha_H,\ g(A\cap H'_{\alpha_H}))$$

y, por tanto

$$\operatorname*{Inf}_{H\in\mathfrak{I}(U)} S'(\operatorname*{Sup}_{x\in H} h(x),\ g(A\cap H)) \quad \operatorname*{Inf}_{\alpha\in[0,1]} S'(\alpha,\ g(A\cap H'_\alpha)),$$

que, junto con (4), completa la demostración. #

La siguiente proposición relaciona las integrales seminormadas y semiconormadas.

<u>Proposición 1.2.-</u> Si g^* es la medida difusa dual de g y S y S' son duales $(S(x,y) = 1 - S'(1-x,\ 1-y))$ entonces

$$N\text{-}\int_{S\ U} h(x)\circ g(.) \ + \ C\text{-}\int_{S',U} (1 - h(x))\circ g^*(.) = 1 \qquad (5)$$

Demostración.-

$$1 - N\text{-}\int_{S\,U} h(x)\circ g(.) = 1 - \sup_{H\in\mathfrak{J}(U)} S(\inf_{x\in H} h(x),\ g(H)) =$$

$$\inf_{H\in\mathfrak{J}(U)} \{1 - S(\inf_{x\in H} h(x),\ g(H)) =$$

$$\inf_{H\in\mathfrak{J}(U)} S'(1 - \inf_{x\in H} h(x),\ 1 - g(H)\ ) =$$

$$\inf_{H\in\mathfrak{J}(U)} S'(\sup_{x\in H} (1- h(x)),\ g^{*}(\overline{H})) = \inf_{H\in\mathfrak{J}(U)} S'(\sup_{x\in\overline{H}} (1-h(x)),\ g^{*}(H))$$

$= C\text{-}\int_{S'\,U} (1 - h(x))\circ g(.)$, de acuerdo con la proposición anterior. #

Corolario.- Para todo $A \subset U$

$$N\text{-}\int_{S\,A} h(x)\circ g(.) + C\text{-}\int_{S'\,A} (1 - h(x))\circ g_A^{*}(.) = g(A)$$

donde $g_A^{*}(H) = g(A) - g(A\cap\overline{H})$

Nota 1.- La proposición y corolario anteriores permiten hacer una demostración casi inmediata de muchas de las propiedades de las integrales semiconormadas a partir de las de las seminormadas; y es un buen instrumento para obtener nuevas propiedades de ambas integrales. Por ejemplo, sabemos que si g es una medida de posibilidad se tiene

$$N\text{-}\int_{S\,A} (h_1 \vee h_2)\circ g(.) = N\text{-}\int_{S\,A} h_1(x)\circ g(.) \vee N\text{-}\int_{S\,A} h_2(x)\circ g(.)$$

$$C\text{-}\int_{S'\,A} (h_1 \vee h_2)\circ g(.) = C\text{-}\int_{S'\,A} h_1(x)\circ g(.) \vee C\text{-}\int_{S'\,A} h_2(x)\circ g(.) \quad (6)$$

Se obtiene, entonces como aplicación inmediata del corolario, que si g es una medida de necesidad

$$N\text{-}\int_{S\,A}(h_1 \wedge h_2)\circ g(.) = N\text{-}\int_{S\,A} h_1(x)\circ g(.) \wedge N\text{-}\int_{S\,A} h_2(x)\circ g(.) \tag{7}$$

$$C\text{-}\int_{S',A}(h_1 \wedge h_2)\circ g(.) = C\text{-}\int_{S',A} h_1(x)\circ g(.) \wedge C\text{-}\int_{S',A} h_2(x)\circ g(.) \tag{8}$$

No nos extenderemos más en esta dirección, para no salirnos de los objetivos generales de esta memoria; pero consideramos que es un tema de estudio interesante.

Nota 2.- En F. Suárez (1.983) se indica que una forma de extender medidas difusas de $\mathcal{P}(U)$ a $\tilde{\mathcal{P}}(U)$ es considerar las medidas definidas por

$A \in \tilde{\mathcal{P}}(U)$ $$\tilde{g}_S(A) = N\text{-}\int_{S\,U} \mu_A(x)\circ g(.)$$

$$\tilde{g}_{S'}(A) = C\text{-}\int_{S',U} \mu_A(x)\circ g(.)$$

Se obtiene, entonces que si S y S' son duales y g^* es la medida dual de g

$$\tilde{g}_S(A) + \tilde{g}^*_{S'}(\bar{A}) = 1 \tag{9}$$

donde $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$, para todo $x \in U$.

La expresión (9) se puede considerar como la definición de dualidad de medidas difusas en $\mathcal{P}(U)$; con lo cual, podemos afirmar que esta extensión, bajo ciertas condiciones, conserva la dualidad.

Por último cabe señalar que, como

$$⨍_U h(x)\circ g(.) = N\text{-}\int_{\wedge\,U} h(x)\circ g(.) = C\text{-}\int_{\vee\,U} h(x)\circ g(.)$$

entonces, notando $\tilde{g}(A) = ⨍_U \mu_A(x)\circ g(.)$, se tiene que

$$\tilde{g}(A) + \tilde{g}^*(\bar{A}) = 1$$

para todo $A \in \mathcal{P}(U)$ y cualesquiera que sean g y g^* duales.

3.-TEORIA DE LA EVIDENCIA DE SHAFER

En este apartado nos vamos a limitar a exponer algunas de las definiciones y propiedades básicas de esta teoría, además de unos resultados que hemos obtenido para las funciones de comunalidad. No demostraremos aquellos teoremas que no sean originales remitiéndonos al trabajo original de Shafer (1.976), donde se puede encontrar una exposición más amplia y detallada de este tema.

Definición 1.10.- Sea Bel una aplicación $Bel: \mathcal{P}(U) \longrightarrow [0,1]$; se dice que Bel es una medida de creencia si y solo si cumple

1) $Bel(\phi) = 0$

2) $Bel(U) = 1$

3) Si $n \in N$ y $A_1, A_2, \ldots, A_n \subset U$, entonces se tiene que

$$Bel(A_1 \cup \ldots \cup A_n) \geqslant \sum_{\substack{I \subset \{1,..,n\} \\ I \neq \phi}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i)$$

donde $|I|$ es el cardinal de I.

Proposición 1.3.- Las medidas de creencia son medidas difusas

Definición 1.11.- Una aplicación $m: \mathcal{P}(U) \longrightarrow [0,1]$, se dice que es una asignación básica de probabilidad (A.B.P.) si y solo si cumple

1) $m(\phi) = 0$

2) $\sum_{A \subset U} m(A) = 1$

-La cantidad $m(A)$ se llama número básico de probabilidad de A.

Teorema 1.1.- La aplicación $Bel: \mathcal{P}(U) \longrightarrow [0,1]$ es una medida de creencia si y solo si existe una asignación básica de probabilidad, m, tal que

$$Bel(A) = \sum_{B \subset A} m(B) \quad , \text{ para todo } A \subset U.$$

-Dicha asignación básica viene dada por la expresión

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} Bel(B) \quad , \text{ para toto } A \subset U.$$

Definición 1.12.- La medida dual de una medida de creencia

$$Pl(A) = 1 - Bel(\overline{A})$$

se dice que es una medida de plausibilidad.

Proposición 1.3.- Si m es la asignación básica de probabilidad asociada a la creencia Bel; entonces

$$Pl(A) = 1 - Bel(\overline{A}) = \sum_{A \cap B \neq \phi} m(B) \quad , \forall A \subset U$$

Nota.- A las medidas de plausibilidad Shafer les llama probabilidades superiores; pero nosotros hemos preferido llamarles de plausibilidad pues así vienen referenciadas en la bibliografía específica de conjuntos difusos.

Notemos como existe una correspondencia biunívoca entre las asi naciones básicas de probabilidad, las medidas de creencia y las medidas de plausibilidad definidas sobre un conjunto U. A nivel intuitivo, tanto cualquiera de ellas, como las funciones de comunalidad que veremos más adelante, representan algún tipo de información, de evidencia, sobre un elemento desconocido, $x \in U$. Cada una de estas representaciones está basada en distintos criterios : por ejemplo, Bel(A) mide la creencia total sobre que un cierto elemento, x, pertenezca a A, mientras que m(A) mide la creencia que se asigna exactamente a A, sin contar la que se asigna a las partes propias de A (ver Shafer (1.976) para una explicación más detallada). De acuerdo con estas ideas, cuando en un determinado contexto, nos refiramos a "la evidencia sobre un elemento desconocido de U...", estaremos haciendo alusión a cualquiera de sus representaciones.

Definición 1.13.- Sea m una asignación básica de probabilidad en U. Se dice que $A \subset U$ es un elemento focal de m si y solo si $m(A) > 0$.

Los siguientes ejemplos nos ayudarán a comprender la naturaleza de las distintas representaciones de la evidencia que hemos introducido.

Ejemplo 1.4.- Una medida de probabilidad P en U es, a su vez, una medida de creencia y de plausibilidad. Los únicos elementos focales de su asignación básica de probabilidad son los conjuntos unitarios, y se tiene

$$m(\{a\}) = P(\{a\}) = p(a) \text{ , para todo } a \in U$$

donde p es la distribución de probabilidad asociada a P.

Recíprocamente, toda evidencia tal que sus únicos elementos focales sean los conjuntos unitarios tiene medidas de creencia y plausibilidad asociadas que coinciden con una misma medida de probabilidad. En base a esto, llamaremos a este tipo de evidencias evidencias probabilísticas.

Ejemplo 1.5.- Las medidas de posibilidad y necesidad son medidas de plausibilidad y creencia, respectivamente. Su asignación básica de probabilidad, m, es tal que sus elementos focales están anidados, es decir, existen $A_i \in \mathcal{P}(U)$, i = 1,..,n tales que

$$A_1 \subsetneq A_2 \subsetneq \ldots \subsetneq A_n \subseteq U \qquad y \qquad \sum_{i=1}^{n} m(A_i) = 1 .$$

Si $U = \{x_1, x_2, \ldots, x_n\}$ y Π es una medida de posibilidad en U con distribución asociada π de tal forma que

$$1 = \pi(x_1) \geq \pi(x_2) \geq \ldots \geq \pi(x_n)$$

entonces la asignación básica asociada a Π, m_π, es

$$m_\pi(A) = \begin{cases} \pi(x_i) - \pi(x_{i+1}) & \text{si } A=\{x_1,..,x_i\}, i=1,..,n-1 \\ \pi(x_n) & \text{si } A=U \\ 0 & \text{en otro caso} \end{cases}$$

Es inmediato comprobar que sus elementos focales están anidados.

Recíprocamente, toda asignación básica tal que sus elementos focales estén anidados, tiene medidas de plausibilidad y creencia asociadas que son medidas de posibilidad y necesidad, respectivamente. A este tipo de evidencias las llamaremos evidencias posibilísticas.

Ejemplo 1.6.- La asignación básica de probabilidad

$$m(A) = \begin{cases} 1 & \text{si } A = U \\ 0 & \text{si } A \neq U \end{cases} \qquad (10)$$

es apropiada cuando no sabemos nada sobre x, excepto que pertene ce a U. Shafer resalta que esta "representación de la ignorancia" difiere de la que se realiza desde el punto de vista bayesiano; donde la ignorancia se representa como la distribución de probabilidad uniforme en U.

Las medidas de creencia y plausibilidad asociadas a esta m -- son

$$Bel(A) = \begin{cases} 0 & \text{si } A \neq U \\ 1 & \text{si } A = U \end{cases}$$

$$Pl(A) = \begin{cases} 0 & \text{si } A = \phi \\ 1 & \text{si } A \neq \phi \end{cases}$$

Pl es una medida de posibilidad con distribución, π , asociada

$$\pi(a) = 1 \ , \quad \text{para todo} \quad a \in U.$$

Definición 1.14.- Dada una asignación básica de probabilidad, m, en $\mathcal{P}(U)$, se llama función de comunalidad asociada a m a la aplicación $Q: \mathcal{P}(U) \longrightarrow [0,1]$, definida por

$$Q(A) = \sum_{\substack{B \subset U \\ A \subset B}} m(B)$$

A la cantidad Q(A) se le llama número de comunalidad de A.

Los dos siguientes teoremas que permiten expresar una medida de creencia (plausibilidad) a través de su función de comunalidad y reciprocamente, pueden encontrarse en Shafer (1.976).

Teorema 1.2.- Supongamos que Bel y Q son la medida de creencia y función de comunalidad asociadas a una asignación básica - de probabilidad m; entonces

$$Bel(A) = \sum_{B \subset \overline{A}} (-1)^{|B|} Q(B)$$

$$Q(A) = \sum_{B \subset A} (-1)^{|B|} Bel(\overline{B}) \quad ,\text{para todo } A \subset U$$

donde $|B|$ es el cardinal de B.

Teorema 1.3.- Sean Pl y Q las medidas de plausibilidad y función de comunalidad asociadas a una asignación básica de probabilidad m; entonces

$$Pl(A) = \sum_{\substack{B \subset A \\ B \neq \phi}} (-1)^{|B|+1} Q(B)$$

$$Q(A) = \sum_{B \subset A} (-1)^{|B|+1} Pl(B) \quad , \text{ para todo } A \subset U, A \neq \phi.$$

Las funciones de comunalidad constituyen una herramienta importante, ampliamente utilizada, de la Teoría de la Evidencia. Existen dos cuestiones que no hemos encontrado resueltas y que permiten una utilización más eficaz de esta función de comunalidad. Estas son

1) Expresar la asignación básica de probabilidad, m, en función de Q.

2) Caracterizar las funciones Q: $\mathcal{P}(U) \longrightarrow [0,1]$ que son función de comunalidad de alguna asignación básica m.

Estos dos problemas quedan resueltos mediante los dos teoremas siguientes, para cuya demostración necesitamos de un lema previo que, a su vez, constituye un resultado bien conocido de la Teoría de Conjuntos.

Lema 1.- Si U es un conjunto no vacío, entonces tiene el mismo número de subconjuntos con un número par de elementos que con un número impar, es decir:

$$\sum_{A \in \mathcal{P}(U)} (-1)^{|A|} = \begin{cases} 1 & \text{si } U = \phi \\ 0 & \text{si } U \neq \phi \end{cases} \qquad (11)$$

Teorema 1.4.- Si m es una A.B.P. y Q su función de comunalidad asociada, entonces

$$m(A) = \sum_{A \subset B} Q(B)(-1)^{|B-A|} \quad , \text{ para todo } A \subset U.$$

Demostración

Por la definición de Q

$$\sum_{A \subset B} Q(B)(-1)^{|B-A|} = \sum_{A \subset B} \Big[\sum_{B \subset C} m(C))(-1)^{|B-A|} \Big] =$$

$$\sum_{A \subset C} \Big[m(C) \; (\sum_{A \subset B \subset C} (-1)^{|B-A|}) \Big] ,$$

y esto último es igual a m(A) si comprobamos que

$$\sum_{A \subset B \subset C} (-1)^{|B-A|} = \begin{cases} 1 & \text{si } C = A \\ 0 & \text{si } A \neq C \end{cases} \qquad (12)$$

lo cual es equivalente a

$$\sum_{D \subset (C-A)} (-1)^{|D|} = \begin{cases} 1 & \text{si } D = \phi \\ 0 & \text{si } D \neq \phi \end{cases} \qquad (13)$$

Ahora bien, (13) es una consecuencia inmediata del lema 1. #

Teorema 1.5.- Sea una función Q: $\mathcal{P}(U) \longrightarrow R$. Entonces, una condición necesaria y suficiente para que Q sea la función de comunalidad de una A.S.B.,m , es que se cumplan las peopiedades:

1) $Q(\phi) = 1 \; , \qquad \sum_{A \neq \phi} Q(A)(-1)^{|A|+1} = 1$ (14)

2) Para todos $n \in N$; $A_1, A_2, \ldots, A_n \subset U$, se tiene

$$Q(A_1 \cap A_2 \cap \ldots \cap A_n) \geq \sum_{I \subset \{1,..,n\}} (-1)^{|I|+1} Q(\bigcup_{i \in I} A_i) \qquad (15)$$

3) $0 \leq Q(U) \leq Q(U - \{a\})$, para todo $a \in U$ (16)

Demostración

En primer lugar demostraremos la condición necesaria; es decir que dada una A.B.P., m, entonces

$$Q(A) = \sum_{A \subset B} m(B)$$

cumple (14), (15), y (16). En efecto:

- $Q(\phi) = \sum_{\phi \subset B} m(B) = \sum_{B \subset U} m(B) = 1$

- $\sum_{A \subset U} Q(A)(-1)^{|A|} = m(\phi) = 0$, según el teorema anterior

y, por tanto,

$$\sum_{A \neq \phi} Q(A)(-1)^{|A|+1} = 1$$

- Si notamos $J(B) = \{i \;/\; A_i \subset B\}$, entonces se tiene

$$\sum_{\substack{I \subset \{1,..,n\} \\ I \neq \phi}} (-1)^{|I|+1} Q(\bigcup_{i \in I} A_i) =$$

$$\sum_{\substack{I \subset \{1,..,n\} \\ I \neq \phi}} (-1)^{|I|+1} \{ \sum_{\bigcup_{i \in I} A_i \subset B} m(B) \} =$$

$$\sum_{J(B) \neq \phi} \left[m(B) \sum_{\substack{I \subset J(B) \\ I \neq \phi}} (-1)^{|I|+1} \right] = \sum_{J(B) \neq \phi} m(B) \leqslant$$

$$\sum_{A_1 \cap \ldots \cap A_n \subset B} m(B) = Q(A_1 \cap \ldots \cap A_n)$$

con lo que queda probado (15).

- La propiedad (16) es un caso particular de

$$A \subset B \implies Q(A) \geqslant Q(B) \tag{17}$$

y cuyo cumplimiento se deduce inmediatamente de la definición de Q.

Para demostrar la condición suficiente construiremos para una

Q que cumpla (14), (15) y (16) una A.B.P., m, cuya función de co munalidad sea la Q dada. Definamos de acuerdo con el teorema anterior

$$m(A) = \sum_{A \subset B} Q(B)(-1)^{|B-A|} \quad , \text{ para todo } A \subset U$$

Para ver que m es una A.B.P., deberemos verificar que se cumple:

$$m(A) = \sum_{A \subset B} Q(B)(-1)^{|B-A|} \geqslant 0 \quad , \text{ para todo } A \subset U \tag{18}$$

$$\sum_{A \subset U} m(A) = 1 \tag{19}$$

$$m(\phi) = 0 \tag{20}$$

-La propiedad (18) para $A = U$ o $A = U - \{a\}$,$a \in U$, es una consecuencia inmediata de (16), ya que $m(U) = Q(U) \geqslant 0$ y $m(U-\{a\}) = Q(U-\{a\}) - Q(U)$.

Sea ahora $A \subset U$ con $|U - A| \geqslant 2$ y consideremos la familia $\{A_a\}_{a \in U-A}$, donde $A_a = A \cup \{a\}$. De acuerdo con (15) se tiene

$$Q(\bigcap_{a \in U-A} A_a) > \sum_{\substack{H \subset U-A \\ H \neq \phi}} Q(\bigcup_{a \in H} A_a)(-1)^{|H|+1}$$

Por otra parte, $\bigcap_{a \in U-A} A_a = A$, por ser $|U-A| \geqslant 2$; y es inmediato que,

$$\sum_{\substack{H \subset U-A \\ H \neq \phi}} Q(\bigcup_{a \in H} A_a)(-1)^{|H|+1} = \sum_{\substack{A \subset B \\ A \neq B}} Q(B)(-1)^{|B-A|+1}$$

de donde se deduce

$$Q(A) \geqslant \sum_{\substack{A \subset B \\ A \neq \phi}} Q(B)(-1)^{|B-A|+1}$$

y se tiene entonces que

$$\sum_{A \subset B} Q(B)(-1)^{|B-A|} \geqslant 0 \quad , \text{ tal como queríamos probar.}$$

- La propiedad (19) se demuestra teniendo en cuenta que:

$$\sum_{A \subset U} m(A) = \sum_{A \subset U} \left\{ \sum_{A \subset B} Q(B)(-1)^{|B-A|} \right\} =$$

$$\sum_{B \subset U} \left[Q(B) \left(\sum_{A \subset B} (-1)^{|B-A|} \right) \right] = \sum_{B \subset U} \left[Q(B)(-1)^{|B|} \left(\sum_{A \subset B} (-1)^{|A|} \right) \right] =$$

$= Q(\phi) = 1$, ya que, por el lema 1

$$\sum_{A \subset B} (-1)^{|A|} = \begin{cases} 1 & \text{si } B = \phi \\ 0 & \text{en otro caso} \end{cases}$$

de acuerdo con el lema 1 y la condición (14) de Q.

- La propiedad (20) es inmediata ya que

$$m(\phi) = \sum_{\phi \subset A} (-1)^{|A|} Q(A) = Q(\phi) - \sum_{\substack{A \subset U \\ A \neq \phi}} (-1)^{|A|+1} Q(A) =$$

$1 - 1 = 0$.

- Para terminar la demostración solo nos queda probar que la función de comunalidad asociada a esta m, que notaremos Q*, coincide con Q. En efecto,

$$Q^*(A) = \sum_{A \subset C} m(B) = \sum_{A \subset B} \left(\sum_{B \subset C} Q(C)(-1)^{|C-B|} \right) =$$

$$\sum_{A \subset C} \left[Q(C) \left(\sum_{A \subset B \subset C} (-1)^{|C-B|} \right) \right] =$$

$$\sum_{A \subset C} \left[Q(C)(-1)^{|C|} \left(\sum_{D \subset C-A} (-1)^{|D|} \right) \right] = Q(A); \text{ ya que}$$

$$\sum_{D \subset C-A} (-1)^{|D|} = \begin{cases} 1 & \text{si } C-A=\phi \ (A = C) \\ 0 & \text{en otro caso} \end{cases}$$

de acuerdo con el lema 1. #

4.-LA COMBINACION DE LA EVIDENCIA

En este apartado vamos a estudiar dos problemas que están estrechamente relacionados. Consideremos dos asignaciones básicas de probabilidad, m_1 y m_2, que suponemos que representan sendas -

evidencias sobre un x U desconocido. Nos proponemos, en primer lugar, medir la contradicción existente entre m_1 y m_2; y, en segundo lugar, caso de que estas no sean totalmente contradictorias, definir una nueva asignación, m, que recoja la combinación de las evidencias contenidas en m_1 y m_2.

La regla de Dempster (1.967) proporciona una solución para el segundo de los problemas anteriores. Dicha regla está estudiada con detalle en Shafer (1.976). Suponiendo que m_1 y m_2 tienen elementos focales $A_1, A_2, \ldots, A_k$ y $B_1, \ldots, B_l$, respectivamente, y que no existe contradicción total entre ellas, es decir si

$$\sum_{A_i \cap B_j = \phi} m_1(A_i).m_2(B_j) < 1$$

entonces la combinación de m_1 y m_2, según la regla de Dempster, proporciona

$$m(A) = \frac{\displaystyle\sum_{A_i \cap B_j = A} m_1(A_i).m_2(B_j)}{1 - \displaystyle\sum_{A_i \cap B_j = \phi} m_1(A_i).m_2(B_j)} \qquad (21)$$

donde $m(A) = 0$, si $A_i \cap B_j \neq A$, para todos i:1,..,k; j:1,..,l.

A esta asignación, definida por la expresión (21), la notaremos $m_1 \oplus m_2$.

Shafer afirma que, asociada a esta regla, la medida del conflicto, de la contradicción, entre dos asignaciones básicas de probabilidad se puede realizar mediante cualquier función creciente de

$$\kappa = \sum_{A_i \cap B_j = \phi} m_1(A_i).m_2(B_j)$$

destacando a $-\log(1-\kappa)$ a la que llama medida del peso del conflicto entre m_1 y m_2, notándola $Con(m_1, m_2)$.

Nota.- Las definiciones anteriores se han dado a nivel de asignaciones básicas de probabilidad, pero se extienden inmediatamente a las distintas representaciones de la evidencia: si Bel_1 y Bel_2 son las funciones de creencia asociadas a m_1 y m_2,--

respectivamente, la función de creencia asociada a $m_1 \oplus m_2$ se nota como $Bel_1 \oplus Bel_2$; y es el resultado de la combinación, según la regla de Dempster de Bel_1 y Bel_2. Lo mismo cabría decir para las funciones de plausibilidad, $Pl_1 \oplus Pl_2$, y de comunalidad, --- $Q_1 \oplus Q_2$, asociadas a $m_1 \oplus m_2$.

Ejemplo 1.7.- (Probabilidad condicionada)

La regla de Demspter permite una interesante justificación de la definición de probabilidad condicionada. Supongamos m_1 y m_2 - en $U = \{a_1, .., a_n\}$ dadas por

$$m_1(A) = \begin{cases} p_i & \text{si } A = \{a_i\} \text{ , } a_i \in U \\ 0 & \text{en otro caso} \end{cases}$$

siendo $0 \leqslant p_i \leqslant 1$, $\sum_{i=1}^{n} p_i = 1$

$$m_2(A) = \begin{cases} 1 & \text{si } A = B \\ 0 & \text{en otro caso} \end{cases}$$

donde $B \subset U$ es un conjunto no vacío arbitrario, pero fijo.

m_1 representa la evidencia que tenemos sobre un elemento - cuando, sobre el, disponemos de la distribución de probabilidad $\{p_i\}_{i \in \{1,..,n\}}$ (ver ejemplo 1.4). m_2 representa la evidencia - que tenemos cuando sabemos que dicho elemento pertenece a $B \subset U$.

La regla de Dempster nos dice que no existe contradicción total entre ambas evidencias si $1 - \sum_{x_i \notin B} p_i < 1$; es decir si ----- $P(B) > 0$, donde P es la medida de probabilidad asociada a la distribución $\{p_i\}$. En este caso la asignación básica resultante es

$$m(A) = \begin{cases} p_i/P(B) & \text{si } A = \{a_i\} \text{ , } a_i \in B \\ 0 & \text{en otro caso} \end{cases}$$

que es una evidencia de tipo probabilístico, cuya medida de probabilidad asociada es precisamente la medida de probabilidad de m_1, P, condicionada a B: $P(./B)$.

Recogemos a continuación algunas de las propiedades más importantes de la regla de Dempster que hemos encontrado en el libro de Shafer (1.976), junto con otras (3 y 4) que no estaban enunciadas anteriormente y nos ha parecido importante señalar.

1.-Asociativa: $(m_1 \oplus m_2) \oplus m_3 \equiv m_1 \oplus (m_2 \oplus m_3)$

2.-Conmutativa: $m_1 \oplus m_2 \equiv m_2 \oplus m_1$

3.-Elemento neutro; si $m_\circ$ representa la ignorancia total (ver ejemplo 1.6), es decir

$$m_\circ(A) = \begin{cases} 1 & \text{si } A = U \\ 0 & \text{en otro caso} \end{cases}$$

entonces $m_\circ \oplus m \equiv m$, para toda asignación m.

4.-Si m_a representa la certeza de que el elemento desconocido coincide con $a \in U$; es decir

$$m_a(A) = \begin{cases} 1 & \text{si } A = \{a\} \\ 0 & \text{en otro caso} \end{cases}$$

y m es una asignación básica cualquiera, con medida de plausibilidad asociada Pl; entonces m_a y m no son totalmente contradictorias cuando $Pl(a) > 0$, en cuyo caso $m_a \oplus m \equiv m_a$.

5.-$Q_1 \oplus Q_2\ (A) = k\ Q_1(A)\ Q_2(A)$, donde $k = 1-/(\sum_{A \cap B \neq \phi} m_1(A).m_2(B))$

La regla de Dempster ha demostrado funcionar bien en numerosas situaciones, como es el caso del ejemplo 1.7. Sin embargo, varias razones nos inclinan a no dar una validez universal a dicha regla. Ya indica Shafer en el Capítulo 8 de su libro que no se pueden combinar evidencias sin tener en cuenta la forma efectiva en que estas interaccionan. Supongamos, por ejemplo, que disponemos de dos asignaciones básicas de probabilidad idénticas $m_1 \equiv m_2$, que han sido obtenidas de la misma fuente. Sería ilógico combinarlas, según la regla de Dempster ya que, en general, de acuerdo con ella $m \oplus m \neq m$. Shafer ha estudiado algunos requisitos para la validez de esta regla. La combinación de eviden cias cualesquiera se presenta, pues, como un problema abierto. - No es objetivo de esta memoria el estudio exhaustivo del mismo y nos limitaremos a desarrollar los elementos necesarios que justifiquen el método de combinar distribuciones de posibilidad que - introduciremos en el siguiente capítulo.

En primer lugar, vamos a tratar un caso de fácil resolución.

<u>Definición 1.15.-</u> Sea m una asignación básica de probabilidad diremos que m representa una evidencia de tipo preciso si y solo si existe $B \subset U$ tal que

$$m(A) = \begin{cases} 1 & \text{si } A = B \\ 0 & \text{en otro caso} \end{cases} \tag{22}$$

Este tipo de evidencias las notaremos m_B, y representan la información que tenemos cuando se sabe que un $x \in U$, desconocido, pertenece a $B \subset U$ con seguridad.

Vamos a poner de relieve como estas evidencias se "combinan - bien ". La combinación de una evidencia de tipo preciso representada por una asignación, m_B, con una evidencia arbitraria dada - por m, se define como aquella que viene representada por

$$m * m_B(A) = \frac{\sum_{B \cap C = A} m(C)}{\sum_{B \cap C \neq \phi} m(C)}$$

con $m * m_B(A) = 0$ si A no está incluído en B.

Siendo la condición de compatibilidad

$$1 - \sum_{B \cap C = \phi} m(C) > 0 \quad ,$$

en el caso en que $\sum_{B \cap C = \phi} m(C) = 1$, m y m_B serán totalmente incompatibles y no tiene sentido combinarlas.

Esta regla de combinación coincide con la regla de Dempster, pero este caso particular no presenta los problemas achacables a ésta . En primer lugar, esta definición parece intuitivamente --

bien fundamentada lo que no ocurre en general. Además en este caso es evidente que el resultado de la combinación no depende de la relación existente entre ambas evidencias; por ejemplo ------ $m_B * m_B \equiv m_B$. Recordemos, también, la justificación de la probabilidad condicionada por la regla de Dempster donde una de las - evidencias era de tipo preciso y estábamos en este caso particular.

El problema en el caso general es mucho más complejo. Nosotros introducimos la siguiente definición que puede ayudar a resolverlo.

<u>Definición 1.16.-</u> Sean m_1 y m_2 dos asignaciones básicas de -- probabilidad. Diremos que la evidencia representada por m_1 está incluída en la representada por m_2 ($m_1 \subset m_2$) si y solo si:

para cada $A \subset U$, existe $m_A: \mathcal{T}(A) \longrightarrow [0,1]$, tal que

$$\sum_{B \subset A} m_A(B) = m_1(A) \qquad y \qquad \sum_{B \subset A \subset U} m_A(B) = m_2(B)$$

Esta definición está basada en la idea intuitiva de que una - información adicional sobre un $x \in U$, desconocido, que sea compatible con la ya existente, debe de producir una atomización de la evidencia. Cuando no sabemos nada sobre x, la evidencia está concentrada en U; a medida que obtenemos más información sobre x la evidencia se debe de ir concentrando sobre subconjuntos de menor tamaño, incluídos en los anteriores.

<u>Propiedades de la relación de inclusión</u>

De las siguientes propiedades no demostraremos aquellas que - consideremos inmediatas.

1) Si m_1 y m_2 son asignaciones básicas de probabilidad con ----- $m_1 \subset m_2$ y Pl_1 y Pl_2 son sus medidas de plausibilidad asociadas,- entonces, para cada $C \subset U$

$$Pl_1(C) \geqslant Pl_2(C)$$

<u>Demostración</u>

En efecto, siguiendo la notación introducida en la definición de inclusión (definición 1.16),

para cada $C \subset U$, $Pl_2(C) = \sum_{C \cap B \neq \phi} m_2(B) = \sum_{C \cap B \neq \phi} \left(\sum_{B \subset A \subset U} m_A(B) \right)$.

Ahora bien, si $A \subset U$ es tal que $B \subset A$ y $C \cap B \neq \phi$, entonces también $C \cap A \neq \phi$, de modo que:

$$\sum_{C \cap B \neq \phi} \left(\sum_{B \quad A \subset U} m_A(B) \right) \leq \sum_{C \cap A \neq \phi} \left(\sum_{B \subset A \subset U} m_A(B) \right) =$$

$\sum_{C \cap A \neq \phi} m_1(A) = Pl_1(C)$, lo que concluye la demostración. #

2) Si Bel_1 y Bel_2 son las medidas de creencia asociadas a m_1 y m_2, respectivamente, con $m_1 \subset m_2$; entonces , para cada $C \subset U$

$$Bel_1(C) \leq Bel_2(C)$$

- La demostración es inmediata teniendo en cuenta que

$$Bel_i(C) = 1 - Pl_i(C) \quad , i=1,2$$

3) Si Q_1 y Q_2 son las funciones de comunalidad asociadas a m_1 y m_2, con $m_1 \subset m_2$ entonces, para cada $C \subset U$, se tiene

$$Q_1(C) \geq Q_2(C)$$

<u>Demostración</u>

En efecto, sea $C \subset U$, entonces

$$Q_2(C) = \sum_{C \subset B \subset U} m_2(B) = \sum_{C \subset B \subset U} \left(\sum_{B \subset A \subset U} m_A(B) \right) =$$

$$\sum_{C \subset A \subset U} \left(\sum_{C \subset B \subset A} m_A(B) \right) \leq \sum_{C \subset A \subset U} \left(\sum_{B \subset A} m_A(B) \right) =$$

$\sum_{C \subset A \subset U} m_1(A) = Q_1(C)$, como queríamos probar #

4) Reflexiva.- Para toda asignación básica m, se tiene $m \subset m$.

5) Antisimétrica.- Si m_1 y m_2 son dos asignaciones básicas tales que $m_1 \subset m_2$ y $m_2 \subset m_1$, entonces $m_1 \equiv m_2$.

Demostración

Por la propiedad 1 si $m_1 \subset m_2$ y $m_2 \subset m_1$, entonces $Pl_1(C) = Pl_2(C)$, para todo $C \subset U$; y como las medidas de plausibilidad, según hemos visto, determinan univocamente su asignación básica se tiene $m_1 \equiv m_2$ #

6) Transitiva.- Si m_1, m_2 y m_3 son asignaciones básicas de probabilidad, entonces

$$m_1 \subset m_2 \quad y \quad m_2 \subset m_3 \implies m_1 \subset m_3$$

7) Existencia de mínimo.- La asignación que representa la ignorancia total

$$m_\circ(A) = \begin{cases} 1 & \text{si } A = U \\ 0 & \text{si } A \neq U \end{cases}$$

está contenida en cualquier otra asignación m : $m_\circ \subset m$, para toda m.

8) Elementos maximales.- Las asignaciones básicas de tipo probabilístico, m_p, es decir aquellas que cumplen (ver ejemplo 1.4)

$$m_p(A) > 0 \implies \text{existe } a \in U \text{ tal que } A = \{a\}$$

son elementos maximales ; si m es tal que $m_p \subset m$, entonces $m_p \equiv m$.

Así pues, la inclusión que hemos definido es una relación de orden parcial en el conjunto de las asignaciones básicas de probabilidad.

Si el conjunto de las asignaciones básicas con esta relación de orden parcial tuviese estructura de semirretículo superior; esto es para cada par de asignaciones básicas m_1 y m_2 existe otra m, que es el supremo de m_1 y m_2, es decir

- $m_1 \subset m$ y $m_2 \subset m$
- Si m' es tal que $m_1 \subset m'$ y $m_2' \subset m'$, entonces $m \subset m'$

entonces el problema de la conjunción estaría resuelto: a cada -

par de m_1 y m_2 le asociaríamos la mínima asignación básica que - las contiene $m = \text{Sup}(m_1, m_2)$. Desgraciadamente esto no ocurre -- así, y el conjunto de los mayorantes comunes de dos asignaciones básicas m_1 y m_2 puede ser vacío y en caso de no serlo, puede no tener mínimo.

Ejemplo 1.9.- Sea $U = \{x_1, x_2\}$ y consideremos en él las distribuciones de posibilidad π_1 y π_2 dadas por:

	x_1	x_2
$\pi_1(x_i)$	1	0.2
$\pi_2(x_i)$	0.3	1

cuyas asignaciones básicas de probabilidad son

$m_1(\phi) = 0$, $m_1(\{x_1\}) = 0.8$, $m_1(\{x_2\}) = 0$, $m_1(\{x_1,x_2\}) = 0.2$

$m_2(\phi) = 0$, $m_2(\{x_1\}) = 0$, $m_2(\{x_2\}) = 0.7$, $m_2(\{x_1,x_2\}) = 0.3$.

Es inmediato comprobar que m_1 y m_2 no tienen mayorantes comunes. A la vista de los valores de π_1 y π_2, podemos observar que esta ausencia de mayorantes comunes se debe a que ambas distribuciones son altamente contradictorias.

Ejemplo 1.10.- Sea $U = \{x_1, x_2, x_3\}$ y las dos distribuciones de posibilidad en U dadas por

	x_1	x_2	x_3
$\pi_1(x_i)$	1	0.7	0.4
$\pi_2(x_i)$	1	0.3	0.8

que tienen como asignaciones básicas de probabilidad m_1 y m_2 tales que

$m_1(\{x_1\}) = 0.3$, $m_1(\{x_1,x_2\}) = 0.3$, $m_1(\{x_1,x_2,x_3\}) = 0.4$,
$m_1(A) = 0$ en otro caso.

$m_2(\{x_1\}) = 0.2$, $m_2(\{x_1,x_3\}) = 0.5$, $m_2(\{x_1,x_2,x_3\}) = 0.3$,
$m_2(A) = 0$ en otro caso.

Estas tienen mayorantes comunes como, por ejemplo

$m_3(\{x_1\}) = 0.6$, $m_3(\{x_1,x_3\}) = 0.1$, $m_3(\{x_1,x_2,x_3\}) = 0.3$,
$m_3(A) = 0$ en otro caso.

$m_4(\{x_1\}) = 0.3$, $m_4(\{x_1,x_2\}) = 0.3$, $m_4(\{x_1,x_3\}) = 0.4$,
$m_4(A) = 0$ en otro caso.

Pero no puede existir un mínimo, m, entre sus mayorantes ya que este debería de cumplir

- $m(\{x_2\}) = m(\{x_3\}) = m(\{x_2,x_3\}) = 0$; pues $Pl(x_1) \geq Pl_3(x_1)=1$
- $m(\{x_1\}) = 0.3$; ya que debe incluir a m_1 y estar incluida en m_4.
- $m(\{x_1,x_2,x_3\}) = 0.3$;ya que debe incluir a m_2 y estar incluida en m_3.
- $m(\{x_1,x_2\}) = 0$; ya que $0.3 = m(\{x_1,x_2,x_3\}) \leq Pl(x_2) \leq Pl_2(x_2) = 0.3$
- $m(\{x_1,x_3\}) \leq 0.1$; ya que $Pl(x_3) \leq Pl_1(x_3) = 0.4$.

y, por tanto, $\sum_{A \subset U} m(A) < 1$, con lo que m no podría ser una asignación básica de probabilidad.

Los ejemplos anteriores proporcionan una base para definir el grado de compatibilidad entre dos asignaciones básicas de probabilidad.

<u>Definición 1.17.-</u> Sean m_1 y m_2 dos A.B.P., llamaremos grado de compatibilidad entre m_1 y m_2 a

$$C(m_1,m_2) = \text{Sup}\{\alpha\in[0,1] \ / \ \text{existe } t \in H(m_1) \cap H(m_2) \text{ con } \sum_{A \subset U} t(A) = \alpha\},$$

donde $H(m\ ) = \{t: \mathcal{P}(U) \longrightarrow [0,1] /$ para cada $A \subset U$, existe m_A,

$$m_A: \mathcal{P}(A) \longrightarrow [0,1], \text{con } \sum_{B \subset A} m_A(B) = m(A) \text{ y}$$

$$t(B) \leq \sum_{B \subset A \subset U} m_A(B)\}$$

como consecuencia de esta definición, dos asignaciones con mayoranres comunes son compatibles con grado uno. Cuando no sea así, el grado de compatibilidad sería, permitiéndonos un abuso del lenguage, "la máxima masa de evidencia procedente de m_1 y m_2 que permite la existencia de un mayorante, no normalizado común.

<u>Proposición 1.4.-</u> $C(m_1,m_2) = 0 \Leftrightarrow$ (para todo $A \subset U$,

$m_1(A) > 0 \Longrightarrow m_2(B) = 0$, cualquiera que sea $B \subset U$ con $B \cap A \neq \phi$)

Demostración.- Es inmediata, a partir de la definición 1.17 #

<u>Nota.-</u> Si m_B, $B \subset U$, representa una evidencia de tipo preciso (ver definición 1.15), entonces $C(m,m_B) = Pl(B)$ para cualquier otra A.B.P. m y, además, "el mayorante no normalizado común t" es único; es decir existe una única $t \in H(m_B) \cap H(m)$ tal que:

$$\sum_{A \subset U} t(A) = Pl(B) \quad y \quad (1/Pl(B)).t = m_B * m.$$

El tema de combinación de evidencias nos parece muy interesante y requiere una mayor profundización, lo que dejamos para futuras investigaciones. En este capítulo nos hemos limitado a desarrollar los elementos que utilizaremos en el Capítulo II de esta memoria, que era nuestro objetivo fundamental.

5.-MEDIDAS DE ENTROPIA ASOCIADAS A UNA EVIDENCIA

Yager (1.983) ha introducido una definición de entropía para una evidencia en un conjunto finito U, que generaliza la noción de entropía de Shannon. La definición que recogemos a continuación es una caracterización de la misma que dió Yager en la referencia anteriormente citada.

<u>Definición 1.18.-</u> Sea U un conjunto finito y supongamos que se dispone de una evidencia en U con asignación básica m y medida de plausibilidad Pl; se define la entropía de dicha evidencia

como

$$E(m) = - \sum_{A \subset U} m(A)L(Pl(A)) \geqslant 0$$

donde L(.) nota la función logaritmo neperiano.

Las propiedades más importantes de E(m) son (ver Yager(1.983))

1) Si m representa una evidencia probabilística, entonces

$$E(m) = - \sum_{a \in U} m(a)L(m(\{a\}))$$

Dicho de otro modo, en este caso E(m) coincide con la entropía de Shannon de la distribución de probabilidad $\{p(a)\}$ que viene dada por

$$p(a) = Pl(\{a\}) = Bel(\{a\}) = m(\{a\}) \text{ , para todo } a \in U.$$

2) $$e^{E(m)} = \prod_{A \subset U} Pl(A)^{-m(A)}$$

3) E(m) es finita para toda asignación m.

4) Si m representa una evidencia posibilística (ver ejemplo 1.5) entonces E(m) = 0.

5) Supongamos que m tiene k elementos focales $A_1,..,A_k$ con ------ $m(A_i) = a_i \in [0,1]$. Entonces E(m) es maximal si los elementos focales A_i son disjuntos; es decir si $i \neq j \implies A_i \cap A_j = \phi$.

6) $E(m) = 0 \iff (m(A) > 0 \text{ y } m(B) > 0 \implies A \cap B \neq \phi)$

7) Supongamos que m tiene k elementos focales disjuntos; entonces, E(m) es maximal si y solo si $m(A_i) = 1/k$, $i = 1,..,k$; en cuyo caso

$$E(m) = - \sum_{i=1}^{k} (1/k).L(1/k) = L(k)$$

Esta medida de entropía, en palabras de Yager, "proporciona una medida de la disonancia de la evidencia". Sin embargo, no po demos afirmar que sea una buena medida del grado de incertidumbre que tendría un individuo sobre el verdadero valor del $x \in U$ desconocido, conociendo la evidencia representada por m. Así, por ejemplo, la asignación básica que representa la ignorancia total ($m_\circ$) y aquellas, m_a, de la forma

$$m_a(A) = \begin{cases} 1 & \text{si } A=\{a\} \\ 0 & \text{en otro caso} \end{cases}$$

donde $a \in U$, arbitrario, pero fijo; que representan la certeza de que $x = a \in U$, tienen la misma entropía. Y, sin embargo, la incertidumbre será grande cuando dispongamos de $m_\circ$ y nula cuando dispongamos de m_a. Para complementar, en este sentido, el papel de la medida de entropía, Yager (1.983) introduce la medida de especificidad asociada a una evidencia.

Definición 1.19.- Sea una evidencia en U representada por la A.B.P. m. Se denomina especificidad de esta evidencia al valor

$$S(m) = \sum_{\substack{A \subset U \\ A \neq \phi}} \frac{m(A)}{|A|}$$

Las siguientes propiedades pueden encontrarse demostradas en Yager (1.983)

1) Sea una evidencia en U, representada por m, entonces

$$\frac{1}{|U|} \leqslant S(M) \leqslant 1$$

2) La especificidad de una evidencia en un conjunto U alcanza su máximo valor ($1/|U|$) para la asignación que representa la ignorancia total: $m_\circ$.

3) Dado un conjunto U, la especificidad S alcanza su máximo valor (uno) para las evidencias de tipo probabilístico (las concentradas en los conjuntos unitarios).

4) Supongamos que una evidencia viene representada por la asignación básica

$$m(B) = \begin{cases} \alpha & \text{si } B = A \\ 1 - \alpha & \text{si } B = U \\ 0 & \text{en otro caso} \end{cases}$$

con $\alpha \in |0,1|$ y $A \subset U$ arbitrarios pero fijos; entonces

$$S(m) = (\alpha/|B|) + (1-\alpha)/|U|$$

5) Si m_1 y m_2 representan evidencias de tipo posibilístico con distribuciones de posibilidad asociadas π_1 y π_2, que cumplen

$$\pi_1(a) \leqq \pi_2(a) \quad , \text{ para todo } a \in U, \text{ entonces}$$

$$S(m_1) \geqq S(m_2)$$

A estas propiedades nosotros añadimos la siguiente:

6) Si m_1 y m_2 representan sendas evidencias en un conjunto U y $m_1 \subset m_2$, entonces $S(m_1) \leqq S(m_2)$

Demostración

Por definición, si $m_1 \subset m_2$ para todo $A \subset U$ existe

$$m_A : \mathcal{P}(A) \longrightarrow [0,1] \text{ tal que}$$

$$m_1(C) = \sum_{D \subset C} m_C(D) \quad \text{y} \quad m_2(C) = \sum_{D \subset C} m_D(C)$$

por tanto

$$S(m_1) = \sum_{A \subset U} \left(m_1(A)/|A| \right) = \sum_{A \subset U} \left[\left(\sum_{B \subset A} m_A(B) \right) / |A| \right]$$

$$\sum_{A \subset U} \left[\sum_{B \subset A} \left(m_A(B)/|B| \right) \right] = \sum_{B \subset U} \left[1/|B| \left(\sum_{B \subset A \subset U} m_A(B) \right) \right] =$$

$$\sum_{B \subset U} \left(m_2(B)/|B| \right) = S(m_2)$$

#

De las propiedades anteriores se deduce que, mientras que la entropía mide la inconsistencia de una evidencia, la especificidad mide la "precisión" de la misma: si la evidencia se concentra en conjuntos de cardinal pequeño la especificidad es alta y si lo hace en conjuntos de cardinal grande esta es baja. Yager afi ma que la combinación de ambas medidas propor ciona una medida de la calidad de la evidencia: cuanto menor sea la entropía, $E(m)$, y mayor la especificidad, $S(m)$, tendremos una evidencia de "mejor calidad".

Al objeto de evitar el manejo de dos índices en la valoración de la calidad de la evidencia, nos cuestionamos la posibilidad de construir medidas que resuman las de entropía y especificidad. El resto de este apartado se dedica a esta cuestión.

Definición 1.20.- Sea una evidencia en un conjunto U, representada por la asignación básica m y medida de plausibilidad Pl; llamaremos entropía inferior de dicha evidencia a

$$H_*(m) = - \sum_{A \subset U} m(A).L(\, Pl(A)/|A| \,) \qquad (23)$$

El nombre de entropía inferior lo justificaremos más adelante

Observemos que, de acuerdo con (23):

$$H_*(m) = - \sum_{A \subset U} m(A).L(\, Pl(A) \,) - \sum_{A \subset U} m(A).L(\, 1/|A| \,) =$$

$$E(m) + \{ - \sum_{A \subset U} m(A).L(\, 1/|A| \,) \}$$

Si $S(m) = \sum_{A \subset U} \frac{m(A)}{|A|}$ es una medida de la precisión de m, entonces se puede considerar que $S'(m) = \sum_{A \subset U} m(A).L(\, 1/|A| \,)$ también mide esta precisión y que $I(m) = -S'(m)$ es una medida de la imprecisión de m.

Así pues, $H_*(m) = E(m) + I(m)$, es decir es la suma de una medida de disonancia y una medida de imprecisión. Si disponemos de una evidencia, la incertidumbre proviene de estos dos conceptos, luego H_* se puede considerar como una medida total de esta incertidumbre.

La siguiente proposición se deduce inmediatamente de la definición 1.20

<u>Proposición 1.5.-</u> Si m_p representa una evidencia de tipo probabilístico, entonces $H_*(m)$ coincide con la entropía de Shannon de la distribución de probabilidad $\{p(a)\}_{a \in U}$ asociada a dicha evidencia $(p(a) = m(\{a\})\ ,\ a \in U)$.

La siguiente proposición justifica el nombre que le hemos dado a H_*.

<u>Proposición 1.6.-</u> Sea una evidencia en U representada por m, entonces $H_*(m)$ es menor o igual que la entropía de Shannon de la distribución de probabilidad que se obtiene eqidistribuyendo la evidencia, $m(A)$, de cada conjunto $A \subset U$ entre sus elementos; esto es la distribución de probabilidad en U que viene dada por

$$p(a) = \sum_{a \in A \subset U} \frac{m(A)}{|A|}$$

<u>Demostración</u>

La entropía de Shannon de $\{p(a)\}_{a \in U}$ vale

$$H(p) = - \sum_{a \in U} p(a).L(p(a)) = - \sum_{a \in U} \left[\sum_{a \in A \subset U} \frac{m(A)}{|A|} \right] L(p(a)) =$$

$$- \sum_{A \subset U} \frac{m(A)}{|A|} \left[\sum_{a \in A} L(p(a)) \right] .$$

Como $-L(x)$ es una función convexa es

$$\frac{- \sum_{a \in A} L(p(a))}{|A|} \qquad -L\left(\sum_{a \in A} \frac{p(a)}{|A|} \right) ,$$

y por tanto

$$H(p) \geqslant -\sum_{A \subset U} m(A).L\left(\frac{\sum_{a \in A} p(a)}{|A|}\right) \geqslant -\sum_{A \subset U} m(A).L\left(\frac{Pl(A)}{|A|}\right) =$$

$= H_*(m)$; ya que

$$\sum_{a \in A} p(a) = \sum_{a \in A}\left[\sum_{a \in B \subset U} \frac{m(B)}{|B|}\right] = \sum_{B \cap A \neq \phi} \frac{m(B)}{|B|} \, |B \cap A| \leqslant$$

$$\sum_{B \cap A \neq \phi} m(B) = Pl(A)$$

#

Vamos a definir a continuación una entropía para una evidencia, cuyo valor maximice el de las entropías de Shannon de todas las evidencias de tipo probabilístico que incluyan a la dada.

Definición.- Sea una evidencia en un conjunto U, representada por m, llamaremos entropía superior de dicha evidencia a

$$H^*(m) = -\sum_{A \subset U} m(A).L\left(\frac{m(A)}{|A|}\right)$$

Esta entropía es en algún sentido dual de la anterior.

Proposición 1.7.- Si m representa una evidencia de tipo probabilístico, $H^*(m)$ coincide con la entropía de Shannon de la distribución de probabilidad asociada: $p(a) = m(\{a\}), \forall a \in U$.

Demostración.- Es inmediata, sin más que considerar que en una evidencia probabilística los únicos conjuntos focales son los unitarios.

Proposición 1.8.- Si m representa una evidencia en un conjunto U; entonces $H^*(m)$ es mayor o igual que la entropía de Shannon de toda evidencia de tipo probabilístico que incluya a m.

Demostración

Si $\{p(a)\}_{a \in U}$ es la distribución de probabilidad en U asociada a una evidencia de tipo probabilístico, m_p, que incluya a m; entonces para cada $A \subset U$ existe $p_A: A \longrightarrow [0,1]$, tal que

$$\sum_{a \in A} p_A(a) = m(A) \qquad y \qquad \sum_{A \subset U} p_A(a) = p(a)$$

Entonces, $\quad H^*(m) = - \sum_{A \subset U} m(A).L\left(\frac{m(A)}{|A|}\right) =$

$$- \sum_{A \subset U} \left[\sum_{a \in A} p_A(a).L\left(\frac{m(A)}{|A|}\right)\right] \geqslant - \sum_{A \subset U} \left(\sum_{a \in A} p_A(a).L(p_A(a))\right)$$

Ya que para $\{p_A(a)\}_{a \in A}$ y m(A) fijos, la solución del problema

$$\text{Min} \qquad - \sum_{a \in A} p_A(a).L(t(a))$$

Sujeto a:

$$0 \leqq t(a) \leqq 1$$

$$\sum_{a \in A} t(a) = m(A)$$

es $t(a) = p_A(a), \quad \forall a \in A$.

Por otra parte y, puesto que $p(a) \geqq p_A(a)$, $\forall a \in A$, se tiene

$$- \sum_{A \subset U} \left[\sum_{a \in A} p_A(a).L(p_A(a))\right] \geqslant - \sum_{A \subset U} \left[\sum_{a \in A} p_A(a).L(p(a))\right] =$$

$$= - \sum_{a \in U} L(p(a)) \left(\sum_{a \in A \subset U} p_A(a)\right) \quad = - \sum_{a \in U} L(p(a)).p(a) = H(p).$$

Así pues, en definitiva $H^*(m) \geqq H(p)$, como queríamos probar #

Algunas de las propiedades más importantes de estas entropías vienen recogidas en las siguientes proposiciones. No incluímos la demostración de aquellas que consideramos inmediatas.

<u>Proposición 1.9.-</u> Si m es una asignación básica en U,

$$0 \leqq E(m) \leqq H_*(m) \leqq H^*(m)$$

Proposición 1.10.- $H_*(m)$ y $H^*(m)$ son finitas para toda asignación básica m.

Proposición 1.11.- $H_*(m) = 0 \iff H^*(m) = 0 \iff$ m representa la certeza de que $x = a$, $a \in U$ fijo; es decir m es de la forma

$$m(A) = m_a(A) = \begin{cases} 1 & \text{si } A = \{a\} \\ 0 & \text{en otro caso} \end{cases} \qquad a \in U, \text{ fijo} \quad (24)$$

Demostración

a) Si $H^*(m) = 0$, como es $0 \leqq H_*(m) \leqq H^*(m)$, entonces, también $H_*(m) = 0$.

b) Si $H_*(m) = 0$, entonces

$$-\sum_{A \subset U} m(A).L\left(\frac{Pl(A)}{|A|}\right) = 0$$ y, puesto que todos los sumandos son positivos, se tiene:

$$\forall A \subset U \text{ , } m(A).L\left(\frac{Pl(A)}{|A|}\right) = 0 \text{ y, por tanto,}$$

$$m(A) > 0 \implies L\left(\frac{Pl(A)}{|A|}\right) = 0 \implies \frac{Pl(A)}{|A|} = 1 \implies$$

$$Pl(A) = |A| = 1.$$

Así pues, la evidencia se concentra en los conjuntos unitarios. Por otra parte, no puede existir $a, b \in U$, $a \neq b$, con $m(\{a\}) > 0$ y $m(\{b\}) > 0$, ya que entonces se tendría $m(\{a\}) > 0$ y $Pl(\{a\}) \leqq$ $\leqq 1 - m(\{b\}) < 1$, en contra de lo demostrado con anterioridad.- Por tanto m(.) ha de ser de la forma que se indica en (24).

c) Si m es tal que $\exists\, a \in U$ con

$$m(A) = m_a(A) = \begin{cases} 1 & \text{si } A = \{a\} \\ 0 & \text{en otro caso} \end{cases}$$

$$H^*(m) = -\sum_{A \subset U} m(A).L(\, m(A)/|A|\,) = -m(\{a\}).L(\, m(\{a\})/1\,) =$$

$$1.0 = 0$$ #

Proposición 1.12.- Si m representa una evidencia en un conjunto U, entonces

a) $H_*(m) \leqq L(|U|)$, alcanzándose esta cota superior para cualquier evidencia que se concentre en una partición de U, $\{A_i\}_{i=1,..,k}$ de la siguiente forma

$$m(A) = \begin{cases} |A_i|/|U| & \text{si } A = A_i,\ i=1,..,k \\ 0 & \text{en otro caso} \end{cases} \tag{25}$$

b) $H^*(m) \leqq L\left(|U|\, 2^{|U|-1}\right)$; obteniéndose esta cota superior para la asignación

$$\tilde{m}(A) = \frac{|A|}{|U|\, 2^{|U|-1}} \quad , \text{ para todo } A \subset U$$

Demostración

a) De acuerdo con la proposición 1.6 $H_*(m) \leqq H(p)$, siendo p una cierta distribución de probabilidad en U. Por otra parte, por las propiedades de la entropía de Shannon $H(p) \leqq L(|U|)$. Así pués $H_*(m) \leqq L(|U|)$, para toda m.

Sea ahora m una asignación básica de probabilidad de la forma dada por (25). Obviamente:

$$H_*(m) = -\sum_{i=1}^{k} m(A_i).L\left(\frac{Pl(A_i)}{|A_i|}\right)$$

Como $A_i \cap A_j = \phi$ si $i \neq j$

$$-\sum_{i=1}^{k} m(A_i).L\left(\frac{Pl(A_i)}{|A_i|}\right) = -\sum_{i=1}^{k} \frac{|A_i|}{|U|} L\left(\frac{|A_i|}{|U|}.\frac{1}{|A_i|}\right) =$$

$$- L\left(\frac{1}{|U|}\right)\left\{\frac{\sum_{i=1}^{k}|A_i|}{|U|}\right\} = L(|U|) \text{ ; y en definitiva}$$

$H_*(m) = L(|U|)$, como queríamos demostrar.

b) Consideremos el siguiente problema de programación matemá tica en $R^{|U|}$:

$$\text{Max} \quad -\sum_{\substack{A \subset U \\ A \neq \phi}} m(A) . L \frac{m(A)}{|A|}$$

s.a.

$$m(A) \geqq 0$$

$$\sum_{\substack{A \subset U \\ A \neq \phi}} m(A) = 1$$

Puesto que el objetivo es una funciín cóncava y las restricciones son lineales, la condición necesaria y suficiente para - que $m: \mathcal{P}(U) - \{\phi\} \longrightarrow [0,1]$ sea solución es que existan $\mu \in R$ y $\{ \lambda_A \geqq 0 , A \subset U\}$, tales que

$$\text{i)} \quad L\left(\frac{m(A)}{|A|}\right) - 1 + \mu + \lambda_A = 0 \quad , \text{para todo } A \subset U, A \neq \phi \ .$$

$$\text{ii)} \quad m(A) \geqq 0 \text{ y } \lambda_A . m(A) = 0 \ , \text{para todo } A \subset U, A \neq \phi, \qquad (25)$$

$$\text{iii)} \quad \sum_{\substack{A \subset U \\ A \neq \phi}} m(A) = 1$$

Despejando M(A) en (i):

$$m(A) = k . |A| . e^{-\lambda_A} \qquad ; \ k = e^{(1-\mu)} > 0$$

de modo que $m(A) > 0$ y por tanto $\lambda_A = 0$, $A \in \mathcal{P}(U) - \{\phi\}$

En estas condiciones y, teniendo en cuenta (iii), se obtiene

$$k = \frac{1}{\sum\limits_{\substack{A \subset U \\ A \neq \phi}} |A|} = \frac{1}{\sum\limits_{A \subset U} |A|} = \frac{1}{\sum\limits_{r=0}^{|U|} r \binom{|U|}{r}} = \frac{1}{|U| \, 2^{|U|-1}}$$

habiendo hecho uso de la conocida igualdad $\sum\limits_{r=0}^{n} r \binom{n}{r} = n . 2^{n-1}$, que puede encontarse en cualquier texto de combinatoria.

Así pues, si fijamos $m(\phi) = 0$, obtenemos en definitiva:

$$m(A) = \frac{|A|}{|U|\, 2^{|U|-1}} \quad , \text{ para todo } A \subset U.$$

y, por tanto

$$H^*(m) = - \sum_{A \subset U} \frac{|A|}{|U|\, 2^{|U|-1}} \, L \left(\frac{|A|}{|U|\, 2^{|U|-1}} \cdot \frac{1}{|A|} \right) =$$

$$L(\, |U|\, 2^{|U|-1}) \geqq H^*(m) \quad , \text{ para toda } m;$$

lo que completa la demostración #

Cuando, como en este caso, se dispone de varias medidas para un mismo concepto intuitivo se impone dar un sistema de axiomas para estas medidas. Nosotros en esta memoria nos limitaremos a recoger, en el Capítulo II, la axiomática de S. Moral (1.984),- para el caso particular de evidencias posibilísticas. La realización de un estudio similar para evidencias generales es una - línea por la que pensamos continuar este trabajo.

- CAPITULO II -

Teoría de la Posibilidad. Informaciones Difusas.

0.-INTRODUCCION

En este capítulo realizamos un estudio general de las medidas de posibilidad. Desde el artículo de Zadeh (1.978): "Fuzzy Sets as a Basis for a Theory of Possibility", las medidas de posibilidad han adquirido gran importancia y numerosos autores han dedicado sus esfuerzos a estudiarlas.

Si tenemos un elemento, x, desconocido, perteneciente a un referencial U y afirmamos que x cumple una determinada propiedad difusa (o lo que es lo mismo, que x pertenece al subconjunto difuso de U de los elementos que cumplen dicha propiedad), estamos dando lo que llamaremos una información difusa.

En el mencionado artículo de Zadeh, éste estableció que, asociadas a dichas informaciones difusas, existe siempre una distribución de posibilidad, que llamaremos información posibilística. Dicha información posibilística es matematicamente más manejable que la información original y, por eso, nos vamos a centrar en ella basicamente; si bien no perderemos de vista la información difusa original sobre la que trabajaremos siempre que sea posible.

Se le ha achacado (ver p.e. Lindley (1.982)) a la Teoría de la Evidencia el ser poco útil en la práctica debido a su marcado carácter teórico (no existen unas reglas definidas y generales que establezcan la forma en que un individuo ha de construir una evidencia particular, a partir de una determinada información o cómo ha de actuar para interpretar y aplicar los resultados obtenidos) . La gran contribución de Zadeh (1.978) es proporcionar una de estas reglas: una evidencia posibilística (o una distribución de posibilidad, equivalentemente) sobre un elemento se obtiene al establecer una propiedad de enunciado difuso sobre dicho elemento, construyéndose dicha distribución de una forma concreta y facilmente realizable en la práctica.

Nuestra forma de asignar distribuciones de posibilidad a informaciones difusas difiere ligeramente de la propuesta por Zadeh. Esto se debe a que nosotros aceptamos el siguiente punto de vista: si un individuo obtiene que un elemento, $x \in U$, desconocido, cumple una determinada propiedad difusa, éste establece un grado con que cada elemento, $a \in U$, es posible valor de x, en base al cumplimiento relativo de dicha propiedad difusa por parte de todos los elementos de U. Este problema es el que consideramos en el primer apartado de este capítulo.

En el apartado segundo estudiamos la combinación de informaciones posibilísticas. En él proponemos un método de realizar la conjunción de distribuciones de posibilidad, basado en la definición de inclusión de evidencias que dimos en el capítulo anterior.

En el tercer apartado introducimos una axiomática que recoge las condiciones que debe de cumplir un funcional para que pueda considerarse una medida de la cantidad de información para distribuciones de posibilidad.

Esta axiomática se obtiene generalizando la dada por Kampé de Feriet (1.970, 1.973) para informaciones " crisp " ($x \in A$, con $A \subset U$). A partir de ella, consideramos varios ejemplos de medidas de información, que utilizaremos en el siguiente capítulo.

El cuarto, y último, apartado está dedicado a un estudio de las informaciones difusas sobre subconjuntos $A \in \mathcal{T}(U)$. Nos centraremos en aquellas que se establecen afirmando una propiedad difusa sobre todos los elementos de A. El motivo de estudiar estas informaciones es que serán la base para nuestra formulación del principio de consistencia posibilidad-probabilidad, objetivo fundamental del tercer capítulo.

2.-INFORMACION DIFUSA E INFORMACION POSIBILISTICA

Sea U un conjunto que, como siempre, consideramos finito y un elemento, $x \in U$, desconocido

Definición 2.1.- Llamaremos información difusa sobre x a una expresión de la forma p: " x es A ", donde A es un subconjunto difuso de U ($A \in \underset{\sim}{\mathcal{F}}(U)$). Es decir lo que se entiende por una restricción difusa sobre los valores de x (ver p.e. Zadeh (1.978)).

Ejemplo 2.1.- Sea $U = \{1,2,..,10\}$ y A el conjunto de los números pequellos de U, caracterizado por una determinada función de pertenencia. Si decimos " x es A ", es decir " x es pequeño" estamos dando una información sobre el elemento, x, desconocido; información que es difusa, en el sentido de que el conjunto al que afirmamos que pertenece x " no está bien definido ".

Nota.- Zadeh (1.978) llamó información difusa a la distribución de posibilidad que asociaba a una restricción difusa. Nosotros hemos preferido llamar información difusa a la restricción difusa, propiamente dicha, y a la distribución de posibilidad asociada la llamaremos información posibilística correspondiente a dicha información difusa.

Nuestro criterio de llamarle información difusa a la propia información difusa viene avalado por el siguiente comentario de Kampé de Feriet (1,974) : ".... La información es dada por un informador a un receptor bajo la forma de una proposición, en el sentido de la lógica formal, es decir de una frase que asocia por medio del verbo ser un nombre (objeto) y un predicado (propiedad, atributo; generalmente expresado por un adjetivo)".

En nuestro caso, al ser dicho predicado difuso, le llamamos información difusa.

En el capítulo anterior vimos que si $x \in U$ es desconocido, pero sabemos que $x \in B$, donde B es un subconjunto clásico de U, entonces se tiene una evidencia representada por una asignación básica de probabilidad dada por

$$m(A) = \begin{cases} 1 & \text{si } A = B \\ 0 & \text{si } A \neq B \end{cases}$$

y denominada evidencia de tipo preciso.

Puesto que la información difusa p: " x es A " proporciona una evidencia (imprecisa por supuesto) sobre $x \in U$, nuestra idea es tratar de asociarle, también, una asignación básica de probabilidad (o cualquier otro tipo de representación).

Antes de entrar de lleno en este tema es necesario destacar una particularidad de las informaciones difusas: si disponemos de una información de tipo preciso, es decir de una restricción de la forma p:" $x \in B$ ", con $B \in \mathcal{P}(U)$; dicha restricción tiene siempre sentido, excepto cuando $B = \phi$, ya que en ese caso no habrá ningún elemento de U que pueda ser x, dado p (ninguno pertenece al vacío). Sin embargo, cuando A es difuso, no basta con que $A \neq \phi$ para que la información p:" x es A " tenga plenamente sentido sobre U.

Si A es un conjunto con $\mu_A(a)$ próximo a cero para todo $a \in U$ entonces, dada p, será difícil aceptar que x pueda ser un elemento cualquiera de U: un individuo que obtenga dicha informacióncíón la rechazará o se planteará si la hipótesis inicial, $x \in U$, es realmente cierta.

<u>Ejemplo 2.2.-</u> Sea $U = \{D_1, D_2, D_3\}$ un conjunto de dados que se suponen normales. Si damos la información " D es redondo ", necesariamente tendremos que poner en duda, bien la realidad de esa información o, bien, que nos estemos refiriendo efectivamente al conjunto U.

A informaciones como la anterior, que postulan para un $x \in U$ una propiedad que es " muy poco cumplida " por todos los elementos de U, las llamaremos informaciones inconsistentes respecto a U. Estas resultarán inaceptables o plantearán serias dudas sobre $x \in U$.

Evidentemente, dado el carácter difuso de las meismas, no tiene sentido afirmar tajantemente que una restricción difusa sea consistente o inconsistente y, por tanto, parece más correcto definir un grado o medida de consistencia.

Dicha medida la estableceremos como una aplicación

$$C: \underset{\sim}{\mathfrak{T}}(U) \longrightarrow [0,1]$$

tal que para todo $A \in \underset{\sim}{\mathfrak{T}}(U)$, $C(A)$ mida la consistencia (respecto a U) de la proposición p_A: " x es A ".

Para determinar cómo debe de ser C impondremos una serie de condiciones que, intuitivamente, resultan plausibles para esta aplicación

P1) $C(\phi) = 0$

Esta propiedad es evidente ya que un elemento pertenece al conjunto vacío no tiene ningún sentido.

P2) $A \subset B \Rightarrow C(A) \leq C(B)$

Si todos los elementos de U pertenecen con más grado a B que a A, la expresión " x es B " debe de ser, por lo menos, tan aceptable como "x es A ".

P3) Si existe $a \in U$ con $\mu_A(a) = 1$, entonces $C(A) = 1$.

Si $a \in U$ es tal que $\mu_A(a) = 1$ no existe duda sobre el cumplimiento de " x es A " en U.

P4) Si existe $b \in U$ tal que $\mu_B(b) \geq \mu_A(a)$ $\forall a \in U$, entonces $C(B) \geq C(A)$.

Esta propiedad, que puede interpretarse como una versión fuerte de P2, impone que C(A) solo depende del máximo grado de cumplimiento de la propiedad p_A sobre U.

Las propiedades P1 y P2 son los axiomas de una energía de -- las estudiadas por De Lucca, Termini (1.979). La propiedad P4 - implica la P2 (es más fuerte); y permite debilitar la P3 a

P'3) $C(U) = 1$.

Las propiedades P1,P'3 y P4 determinaa a C salvo una trans-- formación monótona como se demuestra en la siguiente proposi--- ción.

<u>Proposición 2.1.-</u> Sea $C: \underset{\sim}{\mathcal{F}}(U) \longrightarrow [0,1]$. C cumple P1, P'3 y P4 si y solo si existe $f:[0,1] \longrightarrow [0,1]$ no decreciente con --- $f(0) = 0$, $f(1) = 1$ y tal que

$$C(D) = f\left(\underset{b\in U}{\mathrm{Max}} \; \mu_D(b)\right) \qquad , \forall D \in \underset{\sim}{\mathcal{F}}(U)$$

<u>Demostración</u>

a) Vamos a demostrar, en primer lugar, que si $C: \underset{\sim}{\mathcal{F}}(U) \to [0,1]$ verifica P1, P'3 y P4 entonces existe f en las condiciones de - la proposición. Para ello la vamos a construir de la siguiente forma

Para cada $t \in [0,1]$ tomamos

$$f(t) = C(A_t) \in [0,1] \tag{1}$$

siendo A_t el subconjunto difuso de U dado por:

$$\mu_{A_t}(b) = \begin{cases} 0 & \text{si } b \neq a \\ t & \text{si } b = a \end{cases} \tag{2}$$

con $a \in U$ arbitrario, pero fijo.

El valor de f(t) no depende del $a \in U$ elegido. Sea $D_t \in \underset{\sim}{\mathcal{F}}(U)$ con función de pertenencia

$$\mu_{D_t}(b) = \begin{cases} 0 & \text{si } b \neq d \\ t & \text{si } b = d \end{cases}$$

con $d \in U$, $d \neq a$.

De acuerdo con P4, $C(A_t) \leqq C(D_t)$ y $C(D_t) \leqq C(A_t)$ con lo que $f(t) = C(A_t) = C(D_t)$.

-Sean $t, t' \in [0,1]$ tales que $t' \geqq t$. Por definición $f(t) = C(A_t)$ y $f(t') = C(A_{t'})$. De acuerdo con (2),

$$\mu_{A_{t'}}(a) \geqq \mu_{A_t}(b) \text{ , para todo } b \in U.$$

Entonces por la propiedad P4, $C(A_{t'}) \geqq C(A_t)$ y, por tanto, $f(t') \geqq f(t)$ lo que prueba que f es monótona no decreciente.

-Sea $D \in \tilde{\mathcal{J}}(U)$ y $d \in U$ tal que $\mu_D(d) = \underset{b \in U}{\text{Max}} \mu_D(b)$; y consideremos $D' \in \tilde{\mathcal{J}}(U)$ con función de pertenencia

$$\mu_{D'}(b) = \begin{cases} \mu_D(b) & \text{si } b = d \\ 0 & \text{si } b \neq d \end{cases}$$

De acuerdo con P4 $C(D) = C(D')$; y por la definición de f,

$C(D') = f(\mu_D(d)) = f(\underset{b \in U}{\text{Max}} \mu_D(b))$. En definitiva

para todo $D \in \tilde{\mathcal{J}}(U)$, $C(D) = f(\underset{b \in U}{\text{Max}} \mu_D(b))$

Se obtiene, además, puesto que f es no decreciente

$$C(D) = f(\underset{b \in U}{\text{Max}} \mu_D(b)) = \underset{b \in U}{\text{Max}} f(\mu_D(b))$$

-Por P'3 $C(U) = 1$ y, puesto que $\underset{b \in U}{\text{Max}} \mu_U(b) = 1$, se tiene $f(1) = 1$.

-Por construcción $f(0) = C(\phi) = 0$, con lo que concluye esta parte de la demostración.

b) sea $C: \tilde{\mathcal{J}}(U) \longrightarrow [0,1]$ tal que $C(D) = f(\underset{b \in U}{\text{Max}} \mu_D(b))$ con $f:[0,1] \longrightarrow [0,1]$ verificando las condiciones de la proposición.

Vamos a probar que C cumple P1, P'3 y P4.

-$C(\phi) = f(\underset{b \in U}{\text{Max}}\ \mu(b)) = f(0) = 0$, de modo que se satisface P1.

-$C(U) = f(\underset{b \in U}{\text{Max}}\ \mu_U(b)) = f(1) = 1$, así que también se cumple P'3.

-Sean D, D' $\in \tilde{\mathcal{P}}(U)$ y supongamos que existe $d \in U$ tal que $\mu_D(d) \geqq \mu_{D'}(b)$, $b \in U$. Por ser f no decreciente

$$C(D) = f(\underset{b \in U}{\text{Max}}\ \mu_D(b)) \geqslant f(\mu_D(d)) \geqslant f(\underset{b \in U}{\text{Max}}\ \mu_{D'}(b)) =$$

$= C(D')$, lo que demuestra que se cumple P4. #

<u>Nota 1.-</u> A prtir de ahora solo consideraremos informaciones difusas p_A: " x es A " con $A \in \tilde{\mathcal{P}}(U) - \{\phi\}$; ya que el caso $A = \phi$, resulta trivial.

<u>Nota 2.-</u> Si suponemos que C es continua respecto a la topología de la convergencia uniforme en $\tilde{\mathcal{P}}(U) \equiv [0,1]^U$, entonces f ha de ser continua y reciprocamente. La continuidad de C es intuitivamente exigible ya que cabe esperar que pequeñas variaciones en la función de pertenencia del conjunto que determina una información difusa, produzcan pequeñas variaciones en la consistencia de la misma.

<u>Nota 3.-</u> De acuerdo con la nota anterior y por comodidad, en todos los desarrollos posteriores consideraremos que la medida de consistencia se calcula con $f = I_{[0,1]}$, la función identidad en el intervalo [0,1].De este modo, $C(D) = \underset{b \in U}{\text{Max}}\ \mu_D(b)$ medirá la consistencia de la información p_D; " x es D ".

Una vez estudiado el concepto de consistencia de una información difusa, vamos a tratar de asociar a esta alguna de las representaciones de la evidencia.

Zadeh (1.978) ha dado una solución a este problema, asignando a una restricción difusa p_A: " x es A " una distribución de posibilidad, π , en U, definida por $\pi(a) = \mu_A(a) \quad \forall a \in U$.

$\pi(a)$ se interpreta como el grado con que $a \in U$ es posible valor de x, dado p_A. Por esta razón, a veces, nos referiremos a la distribución de posibilidad asociada a una restrcción difusa denominándola " determinación o discriminación que p_A establece sobre el x desconocido ".

Esta distribución de posibilidad induce, en $\mathcal{P}(U)$, la medida de posibilidad

$$\Pi(B) = \sup_{a \in B} \pi(a) = \sup_{a \in B} \mu_A(a) \text{ , siendo } B \in \mathcal{P}(U),$$

que no está, en general, normalizada ya que

$$\Pi(U) = \sup_{a \in U} \pi(a) = \sup_{a \in U} \mu_A(a) = C(A) \leqq 1$$

Como $\Pi(U) = C(A)$ se interpreta como la posibilidad de $x \in U$, dado p_A, si $C(A) < 1$ la información p_A pone en duda nuestra hipótesis inicial de que $x \in U$ con grado uno. Esto está en contradicción con la siguiente interpretación de Π debida a Dubois, - Prade (1.980) y que reproducimos textualmente:

《El significado de Π asegura: "es imposible que x pertenezca al complementario de Sop(A) en U". Esto es, " es necesario - que x pertenezca a Sop(A) " porque suponemos que estamos seguros que x toma sus valores en U y solo en U. Sin embargo, x puede ser cualquier elemento de A con una posibilidad dada. Π no representa " es posible que x pertenezca a Sop(A) ", sino " cada - elemento de Sop(A) y solo de Sop(A) es una valor posible de --- x ".》

Un primer método de resolver este problema de normalización puede ser considerar la evidencia que tiene como asignación básica a m, dada por

$$\begin{aligned} m(B_i) &= \mu_A(a_i) - \mu_A(a_{i+1}), \quad i:1,..k-1 \\ m(U) &= \mu_A(a_k) + (1 - C(A)) \end{aligned} \tag{3}$$

donde $U = \{a_1,..,a_k\}$ con $\mu_A(a_1) \geqq \mu_A(a_2) \geqq \dots \geqq \mu_A(a_k)$

y $B_i = \{a_1,..,a_i\}$.

Esta asignación básica coincide con la que se asociaría a , salvo en su valor para U que se ve aumentado en 1 - C(A), para que $\sum_{B \subset U} m(B) = 1$. Permitiéndonos un abuso del lenguage, (3) - se puede interpretar diciendo que la " masa de creencia " cuya localización no viene determinada por p_A, la confinamos en U.

La medida de plausibilidad asociada a la asignación básica - de (3) es la medida de posibilidad

$$\Pi'(A) = \Pi(A) + 1 - C(A)$$

Sin embargo, nosotros consideramos que existe un método más natural (es decir, más acorde con la forma de pensar de un individuo normal) de representar la evidencia contenida en una in-- formación difusa y cuya validez se basa en el cumplimiento de - la siguiente hipótesis.

<u>Hipótesis.-</u> Un individuo, ante una información difusa, p_A: " x es A ", sobre un $x \in U$ desconocido, establece para cada --- $a \in U$ el grado con que dicho elemento a es posible valor de x, en base al cumplimiento relativo de la propiedad que define a A por parte de todos los elementos de U.

Como consecuencia se debe de considerar una distribución de posibilidad normalizada. Una vez que hemos aceptado una información difusa (lo que depende de su consistencia), si actuamos según nuestra hipótesis considerando que la posibilidad de cada - elemento se establece según el grado de cumplimiento de una propiedad difusa, en relación con los demás elementos del conjunto; el que más cumpla dicha propiedad debe de ser posible con grado máximo. Por ejemplo, si en el conjunto de los alumnos de una -- clase afirmamos que uno de ellos es alto, entonces el más alto de esa clase será posible con grado uno, aunque no sea alto con grado uno.

En lo que sigue actuaremos de acuerdo con esta idea y, por - tanto, asignaremos una distribución de posibilidad $\hat{\pi}$ tal que -- $\hat{\Pi}(U) = 1$.

Dicha distribución puede obtenerse, en general, de la forma:

$$\hat{\pi}(a) = \Psi(\mu_A(a) , C(A))$$

donde $A \in \tilde{\mathcal{P}}(U)$ determina la información que se considera y Ψ es una función de normalización, es decir

$$\Psi:\{(x,y)/x,y \in [0,1], y \neq 0 , x \leq y\} \longrightarrow [0,1] \quad \text{,cumpliendo}$$

1) $\Psi(0 , y) = 0 \quad , \forall y \in (0,1]$

2) $\Psi(y , y) = 1 \quad , \forall y \in (0,1]$

3) Si $x < z \leq y$, entonces $\Psi(x , y) < \Psi(z , y)$

4) Si $x \leq y < z$, entonces $\Psi(x , y) > \Psi(x , z)$

En lo que sigue consideraremos siempre la distribución de posibilidad que se obtiene para $\Psi(x , y) = x/y$; ya que creemos que se adapta mejor a la idea expuesta en nuestra hipótesis y es bastante cómodo trabajar con ella. Es decir, consideraremos

$$\hat{\pi}(a) = \frac{\mu_A(a)}{\sup\limits_{b \in U} \mu_A(b)} = \frac{\mu_A(a)}{C(A)} \tag{4}$$

con medida de posibilidad asociada

$$\hat{\Pi}(B) = \frac{\sup\limits_{b \in B} \mu_A(b)}{\sup\limits_{b \in U} \mu_A(b)}$$

que tiene asociada la siguiente asignación básica de probabilidad

$$m(B_i) = \frac{\mu_A(a_i) - \mu_A(a_{i+1})}{C(A)} = \hat{\pi}(a_i) - \hat{\pi}(a_{i+1})$$

$i:1,..,k-1$

$$m(U) = \frac{\mu_A(a_k)}{C(A)} = \hat{\pi}(a_k)$$

donde $U = \{a_1,..,a_k\}$ con $\mu_A(a_1) \geq \ldots \geq \mu_A(a_k)$ y

$B_i = \{a_1,..,a_i\}$

A la distribución de posibilidad $\tilde{\pi}_A$ asociada a la información difusa p_A: " x es A " le llamaremos información posibilística asociada a p_A. Dicha distribución de posibilidad, $\tilde{\pi}_A$, es más útil que la información difusa original, en el sentido de que es más manejable matematicamente.

Este método de asignar posibilidades nos lleva a considerar la siguiente relación de equivalencia en $\tilde{\mathcal{P}}(U) - \{\phi\}$:

<u>Definición 2.2.-</u> Sean A, B $\in \tilde{\mathcal{P}}(U) - \{\phi\}$, diremos que A es equivalente a B ($A \sim B$) si y solo si $\tilde{\pi}_A$ coincide con $\tilde{\pi}_B$. Es decir si existe $k > 0$, tal que $\mu_A(a) = k.\mu_B(a)$, $\forall a \in U$.

Si A y B están en la misma clase de equivalencia, entonces p_A y p_B proporcionan la misma determinación sobre el x desconocido, es decir la misma información posibilística sobre dicho x. Lo que distingue a dos elementos de la misma clase es el grado de consistencia de las restricciones difusas asociadas que, de acuerdo con lo dicho, puede considerarse como una medida de la adaptación de dichas informaciones al conjunto U.

En resumen, si A y B están en la misma clase, p_A y p_B determinan lo mismo a x; pero serán informaciones de distinta calidad, dependiendo esta del grado de consistencia de cada una de ellas.

Esta relación de equivalencia ha sido considerada por R. H. Warren (1.981), pero en otro contexto distinto del nuestro; y por tanto, con otra interpretación.

<u>2.-COMBINACION DE INFORMACIONES DIFUSAS Y POSIBILISTICAS</u>

Consideremos, como siempre, un conjunto finito U y $x \in U$ un elemento desconocido de U. Supongamos que se dispone de dos informaciones difusas, relativas ambas a x:

$$p_A: \text{" x es A "} \qquad\qquad A, B \in \tilde{\mathcal{P}}(U) - \{\phi\}$$
$$p_B: \text{" x es B "}$$

En estas condiciones nos planteamos dos dos problemas:

1) Estudiar si existe compatibilidad, coherencia, entre ambas informaciones.

2) En el caso de que no exista contradicción entre ellas, cómo obtener una sola que resuma a las dos (problema que llamaremos de conjunción o combinación de informaciones)

La solución al primero de ellos consistirá, en general, en establecer una aplicación CO: $\tilde{\mathcal{P}}(U)-\{\phi\} \times \tilde{\mathcal{P}}(U)-\{\phi\} \longrightarrow [0,1]$, tal que para todos $A,B \in \tilde{\mathcal{P}}(U)-\{\phi\}$, $CO(\tilde{A},B)$ mida la consistencia de las informaciones p_A y p_B.

Si consideramos como conjunción de las informaciones p_A y p_B la información $p_{A\cap B}$, tal como propone Zadeh (1.975, 1.978), parece lógico pensar que el grado de compatibilidad de p_A y p_B, $CO(A,B)$, es igual a la consistencia de $p_{A\cap B}$; es decir

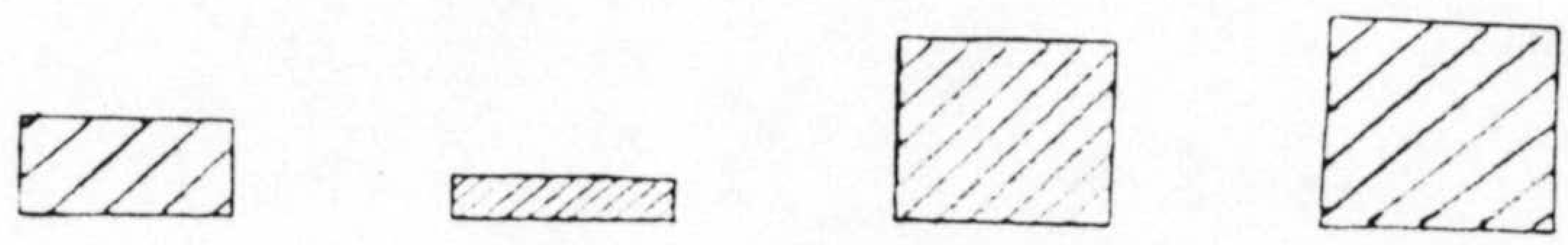
$$CO(A,B) = C(A \cap B) = \sup_{a \in U} (\mu_A(a) \wedge \mu_B(a))$$

Evidentemente CO(.,.) cumple las siguientes propiedades

- $0 \leqq CO(A,B) \leqq 1$
- $CO(A,B) \leqq \text{Min}\, \{C(A), C(B)\}$
- $CO(A,B) = CO(B,A)$
- $CO(A \cap B,C) = CO(A,B \cap C)$
- Si $A \cap B = \phi$, entonces $CO(A,B) = 0$.

Sin embargo, este procedimiento no parece consistente con la hipótesis expuesta en el punto anterior y que nos llevaba a asociar a una información difusa una información posibilística, es decir, una distribución de posibilidad normalizada. Los siguientes ejemplos nos servirán para ilustrar esta idea.

Ejemplo 2.3.- Sea $U = \{A, B, C, D\}$ donde A , B, C y D son las siguientes figuras geométricas



Supongamos que se ha seleccionado un elemento X de U que no conocemos, si bien disponemos de las siguientes informaciones - difusas sobre él:

p_1: " X es rayado "

p_2: " X es alto "

Si los subconjuntos difusos de U, de los elementos rayados y altos vienen dados, respectivamente por

$\mu_1(A) = 0.5 \quad \mu_1(B) = 1 \quad \mu_1(C) = 0.8 \quad \mu_1(D) = 0.5$

$\mu_2(A) = 0.1 \quad \mu_2(B) = 0 \quad \mu_2(C) = 0.4 \quad \mu_2(D) = 0.5$

obtenemos, por el método antes expuesto, la siguiente distribución de posibilidad asociada a la conjunción de p_1 y p_2:

$\pi(A) = 1/5 \quad \pi(B) = 0 \quad \pi(C) = 4/5 \quad \pi(D) = 1$

que coincide con la asociada a p_2, que como puede comprobarse, - es la información menos consistente sobre U (de peor calidad, a la vista de los elementos considerados). Este fenómeno ocurre - siempre. Si se combinan informaciones de distinta consistencia se pondera más la de menos consistencia.

Ejemplo 2.4.- Este ejemplo constituye un caso extremo del anterior.

Sean $U = \{x_1, x_2, x_3\}$ y $A, B \in \mathfrak{F}(U)$ con funciones de pertenencia

$\mu_A(x_1) = 0.5 \quad \mu_A(x_2) = 0.5 \quad \mu_A(x_3) = 0.5$

$\mu_B(x_1) = 1 \quad \mu_B(x_2) = 0.7 \quad \mu_B(x_3) = 0.5$

p_A: " x es A " no contiene ninguna información sobre x: nos dice que cumple una propiedad que es verificada con el mismo -- grado por todos los elementos de U y la distribución de posibilidad π_A es tal que $\pi_A(x_1) = \pi_A(x_2) = \pi_A(x_3) = 1$. La información p_B, sin embargo, nos induciría a pensar que x es x_1.

Si procedemos a la combinación de p_A y p_B obtenemos $p_{A\cap B}$, que coincide con p_A, la más irrelevante de las dos.

La contradicción que, a nuestro criterio, ponen de manifiesto estos ejemplos procede de la utilización de $p_{A\cap B}$ para representar la combinación de p_A y p_B. De acuerdo con nuestra hipótesis, las informaciones difusas p_A: " x es A " y p_B: " x es B " deben de interpretarse como " x es A relativamente " y " x es B relativamente " y a nivel intuitivo, se ve que la combinación de estas expresiones no tiene por qué coincidir con " x es $A \cap B$ relativamente ".

Por otra parte, si disponemos de dos informaciones difusas p_A y p_B con distribuciones de posibilidad π_A y π_B, y consideramos las informaciones p_C y p_D con A equivalente (ver definición 2.2) a C y B equivalente a D, entonces $\pi_A \equiv \pi_C$ y $\pi_B \equiv \pi_D$; sin embargo la distribución de posibilidad asociada a $p_{A\cap B}$, $\pi_{A\cap B}$, no coincide en general con la asociada a $p_{C\cap D}$, $\pi_{C\cap D}$.

Si queremos que el resultado de la combinación sea invariante dentro de una misma clase de equivalencia (si $A \sim C$ y $B \sim D$, entonces el resultado de combinar p_A y p_B sea el mismo que el de combinar p_C y p_D), entonces debemos de trabajar con las distribuciones de posibilidad normalizadas directamente, en vez de con las informaciones originales (observemos que existe una correspondencia biunívoca entre el conjunto cociente $\underset{\sim}{\mathcal{P}}(U)-\{\phi\}/\sim$ y el conjunto de las distribuciones de posibilidad normalizadas en U).

Como consecuencia de esto sustituiremos el problema de la combinación de informaciones difusas por el problema de la combinación de informaciones posibilísticas.

Si queremos combinar dos distribuciones de posibilidad, π_1 y π_2, en U, disponemos, en principio, de dos caminos:

1) Considerar que las medidas de posibilidad, Π_1 y Π_2, asocia

das a las distribuciones π_1 y π_2 son dos medidas de plausibilidad que representan sendas evidencias sobre un mismo x desconocido; y que el problema de combinar distribuciones de posibilidad es un caso particular del problema de combinación de la evidencia.

2) Considerar que la combinación de dos distribuciones de posibilidad es otra distribución de posibilidad que recoge la información contenida en ambas.

La diferencia esencial entre ambas opciones estriba en admitir, bien que la combinación de dos evidencias de tipo posibilístico es una evidencia de tipo general, o bien, que ha de ser, de nuevo, una evidencia de tipo posibilístico.

El primer planteamiento es, por tanto, más general que el segundo. Sin embargo este último proporciona, como veremos, soluciones intuitivamente aceptables y fáciles de obtener, por lo cual, nos centraremos en él. El estudio del primer camino y su comparación con el segundo será el objeto de investigaciones futuras.

En primer lugar, vamos a estudiar la particularización al caso posibilístico de la relación de inclusión de evidencias que introdujimos en el capítulo anterior (definición 1.16).

Proposición 2.2.- Sean π_1 y π_2 dos distribuciones de posibilidad sobre U. La evidencia representada por π_1 está contenida en la representada por π_2 si y solo si $\pi_1(a) \geq \pi_2(a)$,$\forall a \in U$.

Demostración

a) Si la evidencia representada por π_1 está contenida en la representada por π_2, entonces $\pi_1(a) \geq \pi_2(a)$, $\forall a \in U$, es una consecuencia inmediata de la primera propiedad de evidencias incluídas que demostramos en el capítulo anterior.

b) Supongamos $\pi_1(a) \geq \pi_2(a)$, $\forall a \in U$, y sean m_1 y m_2 las asignaciones básicas asociadas a π_1 y π_2, respectivamente. Demostre mos que $m_1 \subset m_2$, es decir que para todo $A \subset U$ existe $m_A: \mathcal{P}(A) \longrightarrow [0,1]$ tal que

$$\sum_{B \subset A} m_A(B) = m_1(A)$$

$$\sum_{B \subset A \subset U} m_A(B) = m_2(B) \quad , \quad B \subset U.$$

Probaremos dicha existencia por un procedimiento constructivo:

Evidentemente para todo $A \subset U$ con $m_1(A) = 0$ tomaremos $m_A(B) = 0$, $B \subset A$.

Si notamos $U = \{a_1,..,a_k\}$ con $\pi_1(a_1) \geqq \pi_1(a_2) \geqq \ldots \geqq \pi_1(a_k)$ y $A_i = \{a_1,..,a_i\}$ entonces $m_1(A) > 0$ si y solo si existe $i \in \{1,2,..,k\}$ tal que $A = A_i$. Basta, por tanto, definir m_{A_i} con $i=1,..,k$.

Notemos ahora $U = \{b_1,..,b_k\}$ con

$$\pi_2(b_1) \geqq \pi_2(b_2) \geqq \ldots \geqq \pi_2(b_k) \text{ y } B_i = \{b_1,..,b_i\}.$$

-Comencemos definiendo m_{A_k}. Sea i_k el menor elemento de $\{1,2,..,k\}$ que cumple

$$\sum_{i=i_k}^{k} m_2(B_i) \leqq m_1(A_k).$$

Tomamos entonces

$$m_{A_k}(B_i) = m_2(B_i) \quad i = i_k,..,k$$

$$m_{A_k}(B_{i_{k-1}}) = m_1(A_k) - \sum_{i=i_k}^{k} m_2(B_i)$$

$$m_{A_k}(B) = 0 \qquad \text{si } B \neq B_i \quad i=i_{k-1},..,k$$

Evidentemente $\sum_{B \subset A_k} m_{A_k}(B) = m_1(A_k)$.

Además, puesto que $m_1(A_k) = \pi_1(a_k) \geqq \pi_2(a_k) = \sum_{a_k \in B_i} m_2(B_i)$,

si $a_k \in B_i$ entonces es $i_k \leqq i$ y $m_{A_k}(B_i) = m_2(B_i)$.

- A continuación definimos $m_{A_{k-1}}$. Sea i_{k-1} el menor elemento de $\{1,2,..,i_k\}$ que cunple

$$\sum_{i=i_{k-1}}^{k} m_2(B_i) \leqq \pi_1(a_{k-1}) = m_1(A_k) + m_1(A_{k-1})$$

Entonces hacemos

1) Si $i_{k-1} = i_k$

$$m_{A_{k-1}}(B_{i_k-1}) = m_1(A_{k-1})$$

$$m_{A_{k-1}}(B) = 0 \quad , \text{ si } B \neq B_{i_{k-1}}$$

2) Si $i_{k-1} < i_k$

$$m_{A_{k-1}}(B_{i_k-1}) = m_2(B_{i_k-1}) - m_{A_k}(B_{i_k-1}) \geqq 0$$

$$m_{A_{k-1}}(B_i) = m_2(B_i) \ , \ i_{k-1} \leqq i < i_k - 1$$

$$m_{A_{k-1}}(B_{i_{k-1}-1}) = m_1(A_{k-1}) - \sum_{i=i_{k-1}}^{i_k-1} m_{A_{k-1}}(B_i) \geqq$$

$$\geqq 0.$$

$$m_{A_{k-1}}(B) = 0 \quad , \text{ en otro caso.}$$

$m_{A_{k-1}}$ está bien definido ya que si $i \leqq i_k - 1$, $a_k \notin B_i$ (recordemos que hemos demostrado $a_k \in B_i \Rightarrow i_k \leqq i$); y, como, $A_{k-1} = U - \{a_k\}$ se tiene $B_i \subset A_{k-1}$.

Consecuencia inmediata es que:

$$\sum_{B \subset A_{k-1}} m_{A_{k-1}}(B) = m_1(A_{k-1}).$$

Notemos que al ser

$$\pi_1(a_{k-1}) \geqq \pi_2(a_{k-1}) = \sum_{a_{k-1} \in B_i} m_2(B_i)$$

entonces si $a_{k-1} \in B_i$ es $i_{k-1} \leqq i$.

- De forma recurrente para $j \in \{1,2,..,k-2\}$ denominaremos -- i_{k-j} al menor elemento de $\{1,2,..,i_{k-j+1}\}$ tal que

$$\sum_{i=i_{k-j}}^{k} m_2(B_i) \leqq \pi_1(a_{k-j}) = \sum_{i=k-j}^{k} m_1(A_i)$$

Y entonces se define

1) Si $i_{k-j}=i_{k-j+1}$

$$m_{A_{k-j}}(B_{i_{k-j}-1}) = m_1(A_{k-j})$$

$$m_{A_{k-j}}(B) = 0 \quad , \text{ si } B \neq B_{i_{k-j}-1}$$

2) Si $i_{k-j} < i_k$

$$m_{A_{k-j}}(B_{i_{k-j+1}-1}) = m_2(B_{i_{k-j+1}-1}) -$$

$$- \sum_{i=k-j+1}^{k} m_{A_i}(B_{i_{k-j+1}-1})$$

$$m_{A_{k-j}}(B_i) = m_2(B_i), \; i_{k-j} \leqq i < i_{k-j+1}-1$$

$$m_{A_{k-j}}(B_{i_{k-j}-1}) = m_1(A_{k-j}) - \sum_{i=i_{k-j}}^{i_{k-j+1}-1} m_{A_{k-j}}(B_i) \geqq 0$$

$$m_{A_{k-j}}(B) = 0 \quad , \text{ en otro caso.}$$

Se demuestra, entonces, que $\sum_{B \subset A_{k-j}} m_{A_{k-j}}(B) = m_1(A_{k-j})$

y que si $a_{k-j} \in B_i$ entonces es $i \geqq i_{k-j}$.

- Para definir m_{A_1} se considera $i_1 = 1$ y

$$m_{A_1}(B_1) = m_1(A_1) \leqq m_2(B_1)$$

$$m_{A_1}(B) = 0 \text{ , si } B \neq B_1 \text{ .}$$

- Para concluir la demostración solo queda probar que

$$\sum_{B \subset A} m_A(B) = m_2(B) \quad , \quad \forall B \subset U$$

o, lo que es equivalente,

$$\sum_{B_i \subset A_j} m_{A_j}(B_i) = m_2(B_i) \quad \text{, para todo } i \in \{1,2,..,k\}$$

Esta propiedad es inmediata, por construcción, para $i \geqq 2$.

Para $i=1$ observemos que

$$\sum_{i=1}^{k} \sum_{B_i \subset A_j} m_{A_j}(B_i) = \sum_{j=1}^{k} \sum_{B_i \subset A_j} m_{A_j}(B_i) =$$

$$\sum_{j=1}^{k} m_1(A_j) = 1, \text{ y por otra parte}$$

$$\sum_{i=1}^{k} \sum_{B_i \subset A_j} m_{A_j}(B_i) = \sum_{B_1 \subset A_j} m_{A_j}(B_1) + \sum_{i=2}^{k} m_2(B_i) =$$

$$\sum_{B_1 \subset A_j} m_{A_j}(B_1) + 1 - m_2(B_1).$$

de donde

$$\sum_{B_1 \subset A_j} m_{A_j}(B_1) = m_2(B_1) \text{ como queríamos demostrar} \qquad \#$$

Como consecuencia de esta proposición la relación de orden - que definimos para las evidencias nos determina una relación de orden en el conjunto de las distribuciones de posibilidad normalizadas sobre U y que se puede expresar : la distribución de posibilidad π_1 está contenida en la distribución π_2 ($\pi_1 \subset \pi_2$) si y solo si $\pi_1(a) \geqq \pi_2(a)$, $\forall a \in U$.

Este resultado está de acuerdo con las consideraciones de -- Zadeh (1.978) quien afirma que la información contenida en p_A: " x es A " está incluida en la información de p_B: " x es B "-- ($I(p_A) \subset I(p_B)$) si y solo si $\pi_A(a) \geqq \pi_B(a)$, $\forall a \in U$; (recordemos que para Zadeh la información contenida en una restricción difu coincide con la distribución de posibilidad asociada).

Sin embargo, hay que hacer notar que esta relación solo tiene pleno sentido cuando trabajamos con distribuciones de posibilidad normalizadas. Para distribuciones no normalizadas, pueden encontrarse casos extraños, como el siguiente:

<u>Ejemplo 2.5.-</u> Sea $U = \{x_1, x_2, x_3\}$ y dos distribuciones π_1 y - π_2 en U, con

$$\pi_1(x_1) = 1 \quad , \pi_1(x_2) = \pi_1(x_3) = 0.1$$

$$\pi_2(x_1) = \pi_2(x_2) = \pi_2(x_3) = 0.1$$

En este caso sería $\pi_1 \subset \pi_2$ pero es obvio que π_1 determina mejor que π_2 el elemento desconocido: da más información.

<u>Notación.-</u> El conjunto de distribuciones de posibilidad normalizadas sobre un conjunto finito U, lo notaremos $\prod(U)$.

<u>Propiedades de la inclusión de distribuciones de posibilidad</u>

De las siguientes propiedades no consideraremos las demostraciones que sean inmediatas.

1) Reflexiva.- $\forall \pi \in \prod(U)$, $\pi \subset \pi$

2) Antisimétrica.- Si $\pi_1 \subset \pi_2$ y $\pi_2 \subset \pi_1$, entonces $\pi_1 \equiv \pi_2$

3) Transitiva.-$\pi_1 \subset \pi_2$ y $\pi_2 \subset \pi_3 \Rightarrow \pi_1 \subset \pi_3$

4) Existencia de mínimo.- La distribución de posibilidad π_U:

$$\pi_U(a) = 1 \ , \ \forall a \in U$$

es tal que $\pi_U \subset \pi, \quad \forall \pi \in \prod(U)$

5) Elementos maximales.- Las distribuciones de posibilidad - de la forma

$$\pi_a(b) = \begin{cases} 1 & \text{si } b = a \\ 0 & \text{si } b \neq a \end{cases} \qquad \text{donde } a \in U \text{ arbitrario,pero fijo.}$$

son tales que $\pi_a \subset \pi \quad \Rightarrow \quad \pi_a \equiv \pi$

6) Si π_1 y π_2 son dos distribuciones de posibilidad tal que existe $\pi \in \prod(U)$ con $\pi_1 \subset \pi$ y $\pi_2 \subset \pi$ (tienen, al menos, un mayorante común) entonces existe el supremo de π_1 y π_2.

<u>Demostración</u>

Si existe $\pi \in \prod(U)$ tal que $\begin{array}{l} \pi(a) \leqq \pi_1(a) \\ \pi(a) \leqq \pi_2(a) \end{array} \quad \forall a \in U$

entonces, por la hipótesis de normalización, tiene que existir $a_0 \in U$ tal que $\pi(a_0) = \pi_1(a_0) = \pi_2(a_0) = 1$ y, por tanto,

$\pi_1 \wedge \pi_2 : U \longrightarrow [0,1]$ dada por

$$(\pi_1 \wedge \pi_2)(a) = \text{Min} \{ \pi_1(a), \pi_2(a) \} \ , \ \forall a \in U$$

es una distribución de posibilidad normalizada.

Es inmediato probar que $\pi_1 \wedge \pi_2 = \text{Sup}(\pi_1, \pi_2)$ #

En vista de la propiedad 6, parece que que una forma natural de combinar, en condiciones generales, dos informaciones posibilísticas π_1 y π_2 sea considerar $\pi_1 \wedge \pi_2$. El problema es que no siempre π_1 y π_2 tienen un mayorante común que garantice que --- $\pi_1 \wedge \pi_2$ sea una distribución de posibilidad normalizada.

La no existencia de mayorante común se puede atribuir a la - falta de coherencia, a la contradicción, que exista entre π_1 y π_2. Para solucionar este problema, definimos de forma análoga a como lo hicimos para las evidencias de tipo general, una medida de compatibilidad de dos distribuciones de posibilidad.

Definición 2.3.- Sean π_1 y π_2 dos distribuciones de posibilidad. Se define la compatibilidad de π_1 y π_2 como

$$C(\pi_1,\pi_2) = \mathrm{Sup}\{\alpha \in [0,1] / \exists \pi \in \prod(U), \text{ con } \pi_1(a) \geq \alpha\pi(a) \leq \pi_2(a) \ , \ \forall a \in U\}$$

Como consecuencia inmediata de la definición, obtenemos las siguientes propiedades

1) $0 \leq C(\pi_1,\pi_2) \leq 1$, $\forall \pi_1,\pi_2 \in \prod(U)$

2) $C(\pi_1,\pi_2) = \sup\limits_{a \in U} \{\pi_1(a) \wedge \pi_2(a)\}$

3) π_1 y π_2 tienen un mayorante común si y solo si $C(\pi_1,\pi_2)=1$.

4) $C(\pi_1,\pi_2) = 0 \Leftrightarrow (\pi_1(a) > 0 \Rightarrow \pi_2(a) = 0 \ , \ \forall a \in U)$

5) Si $A,B \in \tilde{\mathcal{P}}(U)$, entonces $CO(A,B) \leq C(\pi_A,\pi_B)$, donde π_A y π_B son las distribuciones de posibilidad normalizadas asociadas a p_A: " x es A " y p_B: " x es B ", respectivamente.

6) $C(\pi_U,\pi) = 1$, $\forall \pi \in \prod(U)$

Para definir la conjunción de dos distribuciones de posibilidad cuando la compatibilidad no es uno, haremos uso de la propiedad que demostraremos a continuación.

Proposición 2.3.- Si π_1 y π_2 son dos distribuciones de posibilidad con $C(\pi_1,\pi_2) > 0$, entonces existe Min $\mathcal{H}$, donde $\mathcal{H} \subset \prod(U)$

$$\mathcal{H} = \{\pi \in \prod(U) / \pi_1(a) \geq C(\pi_1,\pi_2)\pi(a) \leq \pi_2(a) \ , \forall a \in U\}$$

y ese mínimo es

$$\pi(a) = \frac{\pi_1(a) \wedge \pi_2(a)}{C(\pi_1,\pi_2)} \tag{5}$$

Demostración

a) $\pi \in \mathcal{H}$, ya que $C(\pi_1,\pi_2)\pi(a) = \pi_1(a) \wedge \pi_2(a)$ y es inmediato que se cumple $\pi_1(a) \geq \pi_1(a) \wedge \pi_2(a) \leq \pi_2(a)$, $\forall a \in U$.

b) Si $\pi' \in \mathcal{H}$, entonces de la definición de $\mathcal{H}$, se obtiene

$$\pi'(a)C(\pi_1,\pi_2) \leqq \pi_1(a) \quad y \quad \pi'(a)C(\pi_1,\pi_2) \leqq \pi_2(a)$$

por tanto $\pi'(a)C(\pi_1,\pi_2) \leqq \pi_1(a) \wedge \pi_2(a)$

de donde se deduce

$$\pi'(a) \leqq \frac{\pi_1(a) \wedge \pi_2(a)}{C(\pi_1,\pi_2)} = \pi(a)$$

es decir $\pi \subset \pi'$, con lo que π ha de ser el mínimo de como queríamos demostrar. #

Nota.- Dadas π_1 y π_2, $C(\pi_1,\pi_2)$ puede considerase como el máximo nivel al cual existen mayorantes comunes de ambas distribuciones. El mínimo de dichos mayorantes viene dado por (5) y puede considerarse como una especie de supremo de π_1 y π_2. Esto -- nos lleva a introducir la siguiente definición.

Definición 2.4.- Da dadas dos distribuciones de posibilidad, π_1 y π_2, llamamos conjunción de ambas a la distribución de posibilidad $\pi_{1,2}$ dada por

$$\pi_{1,2}(a) = \frac{\pi_1(a) \wedge \pi_2(a)}{C(\pi_1,\pi_2)} \qquad , \forall a \in U;$$

siempre que $C(\pi_1,\pi_2) > 0$.

Notación.- En lo que sigue, la conjunción de las distribuciones π_i y π_j será notada $\pi_{i,j}$

Ejemplo 2.6.- Vamos a ver como se comporta esta definición - para los casos considerados en los ejemplos 2.3 y 2.4.

En el primero de ellos, las distribuciones de posibilidad -- asociadas a " X es alto " y " X es rayado " tienen una compatibilidad de 0.5 y su combinación da lugar a $\pi_{1,2}$ con

$$\pi_{1,2}(A) = 1/4 \quad \pi_{1,2}(B) = 0 \quad \pi_{1,2}(C) = 1 \quad \pi_{1,2}(D) = 5/8$$

donde se observa que desaparece el efecto de ponderar la información menos consistente.

Para el caso del ejemplo 2.4, la combinación de π_A y π_B proporciona

$\pi_{A,B}(x_1) = 1$ $\quad\quad$ $\pi_{A,B}(x_2) = 0.7$ $\quad\quad$ $\pi_{A,B}(x_3) = 0.5$

que coincide con π_A, como intuitivamente cabía esperar (la información p_B era irrelevante).

Propiedades

Las siguientes propiedades de la combinación de distribuciones de posibilidad son de inmediata demostración.

1)Una distribución de posibilidad, π_1, está incluída en otra, π_2, si y solo si el resultado de la combinación de ambas, $\pi_{1,2}$, coincide con π_2.En otras palabras

$$\pi_1 \subset \pi_2 \quad \Leftrightarrow \quad \pi_{1,2} \equiv \pi_2$$

2) Si la distribución de posibilidad π_1 está incluída en la distribución π_2, y ambas presentan la misma compatibilidad con una tercera π_3, entonces el resultado de combinar π_1 con π_3, $\pi_{1,3}$, está incluído en la combinación de π_2 con π_3, $\pi_{2,3}$. Es decir

$$\pi_1 \subset \pi_2 \ \text{ y } C(\pi_1,\pi_2)=C(\pi_2,\pi_3) \quad \Rightarrow \quad \pi_{1,3} \subset \pi_{2,3}.$$

3) Elementos absorventes.- Si π_a, $a \in U$, viene dada por

$$\pi_a(b) = \begin{cases} 1 & \text{si } b = a \\ 0 & \text{si } b \neq a \end{cases}$$

(representa la certeza de que x=a), entonces , si π es una distribución de posibilidad cualquiera,

$C(\pi,\pi_a) > 0 \quad \Leftrightarrow \quad \pi(a) > 0$; y en ese caso, la conjunción de π y π_a coincide con π_a.

4) La combinación de π_U ($\pi_U(a) = 1, \forall a \in U$) y π da lugar a π para toda $\pi \in \prod(U)$.

5) Conmutativa.- $\pi_{1,2} \equiv \pi_{2,1}$, para todas π_1, $\pi_2 \in \prod(U)$.

-La conjunción de distribuciones de posibilidad no es asocia tiva, en general: si disponemos de tres distribuciones de posibilidad π_1, π_2 y π_3 y notamos por $\pi_{(1,2),3}$ el resultado de la conjunción de $\pi_{1,2}$ y π_3 y por $\pi_{1,(2,3)}$ el resultado de la conjunción de π_1 y $\pi_{2,3}$, entonces puede ocurrir $\pi_{(1,2),3} \neq \pi_{1,(2,3)}$

Se puede afirmar que $\pi_{(1,2),3} \equiv \pi_{1,(2,3)}$ cuando $C(\pi_1,\pi_2) = C(\pi_2,\pi_3) = C(\pi_{1,2},\pi_3) = C(\pi_1,\pi_{2,3})$. Es decir la existencia de conflicto entre las informaciones es lo que provoca la no asociatividad.

Como consecuencia la combinación de más de dos distribuciones de posibilidad no puede hacerse de forma iterativa y hay que definirla directamente. Esto lo haremos por un procedimiento similar al empleado para el caso de dos distribuciones.

<u>Definición 2.5.-</u> Si $\pi_1,..,\pi_n$ son n distribuciones de posibilidad en un mismo conjunto U, denominamos compatibilidad de estas distribuciones a

$$C(\pi_i/i=1,..,n) = \sup_{a \in U} \{\min_{1 \leq i \leq n} \pi_i(a)\}$$

- En el caso de que $C(\pi_i/i=1,..,n) > 0$ denominamos combinación de π_i, i=1,..,n a la distribución de posibilidad

$$\pi_{(i/i=1,..,n)}(a) = \frac{\min_{1 \leq i \leq n} \pi_i(a)}{C(\pi_i/i=1,..,n)}$$

Para terminar, vamos a señalar los pasos lógicos que, de acuerdo con los desarrollos anteriores, han de llevarse a cabo para combinar dos informaciones de la forma p_A: " x es A " ; p_B: " x es B ".

1.- Se calculan las consistencias C(A) y C(B). Si son suficientemente grandes, se aceptan p_A y p_B y se construyen las informaciones posibilísticas asociadas π_A y π_B.

2.- A continuación, se calcula $C(\pi_A,\pi_B)$ para ver si existe contradicción entre π_A y π_B.

3.- Si no existe contradicción, se calcula la distribución - de posibilidad

$$\pi_{A,B}(a) = (\pi_A(a) \wedge \pi_B(a))/C(\pi_A,\pi_B)$$

que es la información posibilística que consideraremos para su posteior utilización.

3.-CANTIDAD DE INFORMACION CONTENIDA EN UNA RESTRICCION DIFUSA Y EN UNA DISTRIBUCION DE POSIBILIDAD

Hasta el momento, hemos analizado las caracteríctícas y propiedades de las restricciones (informaciones) difusas. En este apartado vamos a tratar de establecer una medida de la cantidad de información contenida en una tal restricción difusa.

Vamos a comenzar presentando el concepto de medida de información de Kampé de Feriet que será la base de nuestras construcciones.

Kampé de Feriet (1.970, 1.973) considera un conjunto Ω arbitrario y que una información sobre $x \in \Omega$ es de la forma p_A: " x se encuentra en A ", con $A \in \mathcal{S} \subset \mathcal{P}(\Omega)$, siendo $\mathcal{S}$ una σ-álgebra - de sucesos observables o propiedades enuciables de los elementos de Ω.

Los axiomas que debe de cumplir una medida J para que J(A) - pueda considerarse como la cantidad de información que proporciona p_A son:

AI) $J:\mathcal{S} \longrightarrow \bar{R}_0^+ = R_0^+ \cup \{+\infty\}$

AII) $A,B \in \mathcal{S}$, $B \subset A$ $\Rightarrow$ $J(B) \geq J(A)$

y un tercer axioma opcional de normalización

AIII) $J(\Omega) = 0$, $J(\phi) = +\infty$

Si J cumple estos tres axiomas, se dice que es una medida de información en $(\Omega,\mathcal{S})$ y a la terna $(\Omega,\mathcal{S},J)$ se le llama espacio - de información de sucesos.

Nosotros vamos a considerar, como siempre, un conjunto, U, finito. Admitiremos que sobre $x \in U$ son enunciables todas las propiedades de la forma p_A: " x es A " donde $A \in \tilde{\mathcal{P}}(U)-\{\phi\}$ y vamos a tratar de medir la cantidad de información de p_A. Según hemos discutido con anterioridad, nuestra filosofía es que la información realmente relevante sobre x contenida en una restricción difusa es la distribución de posibilidad asociada; por tanto, vamos a definir las medidas de información en el conjunto $\Pi(U)$ de las distribuciones de posibilidad normalizadas sobre U. La cantidad de información de p_A será la de distribución asociada π_A.

El axioma de Kampé de Feriet más importante es el segundo: si $B \subset A$, entonces el afirmar " $x \in B$ " debe de informar más que " $x \in A$ ". Para las distribuciones de posibilidad podemos establecer una propiedad análoga: si la información posibilística π, está contenida en π' ($\pi'(a) \leq \pi(a)$, $\forall a \in U$) entonces π' debe de contener más información que π, por lo que hacemos la siguiente definición.

<u>Definición 2.6.-</u> Sea U un conjunto y $\Pi(U)$ el conjunto de las distribuciones de posibilidad normalizadas sobre U. Se dice que I es una medida de información en $(U, \Pi(U))$ si y solo si cumple las propiedades

i) $I: \Pi(U) \longrightarrow R_0^+$

ii) $I(U) = 0$

iii) $\pi \subset \pi' \quad \Rightarrow \quad I(\pi) \leq I(\pi')$

<u>Nota.-</u> Es conveniente que I sea continua respecto a la topología de la convergencia uniforme en $\Pi(U)$, pero es una propiedad que no vamos a exigir en la definición.

Analogamente definimos una medida de incertidumbre o indeterminación en $(U, \Pi(U))$.

<u>Definición 2.7.-</u> H es una medida de incertidumbre o indeterminación en $(U, \Pi(U))$ si y solo si cumple las propiedades

i') $H: \prod(U) \longrightarrow R_0^+$

ii') $H(\pi_a) = 0$, donde

$$\pi_a(b) = \begin{cases} 1 & \text{si } b = a \\ 0 & \text{en otro caso} \end{cases} \qquad a \in U, \text{ arbitrario pero fijo.}$$

iii') $\pi \subset \pi' \quad \Rightarrow \quad H(\pi) \geqq H(\pi')$

Se deduce inmediatamente de ambas definiciones que si H es - una medida de incertidumbre $I_H(\pi) = H(U) - H(\pi)$ es nna medida de información que cumple

$$I_H(\pi_a) = I_H(\pi_c), \forall a,c \in U.$$

Reciprocamente, si I es una información que cumple la ante-- rior propiedad, $H_I(\pi) = I(\pi_a) - I(\pi)$ es una medida de información, cualquiera que sea $a \in U$.

Nota.- Las propiedades de una medida de incertidumbre son parecidas a las que caracterizan una medida de energía de un conjunto difuso. Czogala et al. (1.982) consideraron dichas medidas de energía como medidas de incertidumbre o, como ellos lo - llaman, " índice de calidad " con aplicación en campos tales como la toma de decisiones con controlador difuso y la predicción en sistemas difusos.

No obstante, existen dos diferencias entre ambos tipos de medidas:

a) Las medidas de energía están definidas en el conjunto de las aplicaciones de U en [0,1] , mientras que las de incertidumbre lo están solamente para aquellas que tienen su máximo valor igual a uno.

b) El mínimo valor (cero) de una medida de energía se obtiene para la aplicación identicamente nula (asociada al vacío), - mientras que para una medida de incertidumbre se obtiene para - las distribuciones π_a, $a \in U$.

Proposición 2.4.- Cualquier función de la forma

$$E_{f,g}(\pi) = g\left(\sum_{a \in U} f(\pi(a))\right)$$

con

- $g:[1,+\ ) \longrightarrow R_0^+$, no decreciente con $g(1) = 0$.
- $f:[0,1] \longrightarrow [0,1]$, no decreciente con $f(0) = 0$ y $f(1) = 1$

es una medida de indeterminación.

Demostración.- Es inmediato comprobar que las aplicaciones - $E_{f,g}$ cumplen los axiomas que definen una medida de entropía.

Nota.- Si, además, f y g son continuas, $E_{f,g}$ será continua - respecto a la topología de la convergencia uniforme en $\prod(U)$.

Nota.- Obsérvese que una medida de incertidumbre de este tipo puede escribirse como $E(\pi) = g(\ E(\pi)\ )$ siendo E una energía Sum-Prod con ponderación constante (Trillas, Riera (1.978); --- De Lucca, Termini (1.978)). De hecho las medidas de incertidumbre están más cercanas a las medidas de energía de un conjunto difuso que a las de entropía, ya que, en esencia, estas últimas miden cuán difuso es dicho conjunto.

Si A es un conjunto "crisp", entonces

$$E_{f,g}(\pi_A) = g(|A|)$$

De acuerdo con esto, si queremos una medida de incertidumbre que, para cada conjunto "crisp", A, $E(\pi_A)$ coincida con la entropía de Shannon de un experimento aleatorio en el que se obtienen los elementos de A de forma equiprobable, hemos de considerar $g \equiv \log_2$. Obtenemos, entonces, la familia

$$E_f(\pi) = \log_2 \left(\sum_{a \in U} f(\pi(a)) \right)$$

Si consideramos las funciones

$$f_n(t) = t^n \quad , n \in N^*$$

$$f_n^*(t) = t^{1/n}, \ n \geq 2$$

se generan las medidas

$$E_n(\pi) = \log_2 \left(\sum_{a \in U} \pi^n(a) \right) ; \ E_n^*(\pi) = \log_2 \left(\sum_{a \in U} \pi^{1/n}(a) \right)$$

La diferencia más notable entre E_n y E_n^* ($n \geq 2$) es que, para distribuciones de posibilidad con idéntico valor de E_1, E_n se hace mayor para las distribuciones " menos homogéneas " y E_n^* para las "más homogéneas ". Esta propiedad de f_n ha sido estudiada por Czogala et al. (1.982), dando varios ejemplos ilustrativos.

Proposición 2.5.- Si m_π es la asignación básica asociada a la distribución π, entonces $E_{H_*}(\pi) = H_*(m_\pi)$ es una medida de incertidumbre en $\prod(U)$ (donde H_* es la medida de entropía inferior que definimos en el primer capítulo)

Demostración

Probar que $E_{H_*}(.)$ verifica i') y ii') (ver definición 2.) es inmediato.

Para probar iii') observemos que, de acuerdo con la definición de H_*:

$$E_{H_*}(\pi) = H_*(m_\pi) = - \sum_{A \subset U} m(A) L\left(\frac{\Pi(A)}{|A|}\right)$$

$$E_{H_*}(\pi') = H_*(m_{\pi'}) = - \sum_{A \subset U} m(A) L\left(\frac{\Pi'(A)}{|A|}\right)$$

Según la proposición 2.2 se tiene que $\pi \subset \pi'$ implica que

$$\pi(a) \geq \pi'(a) \quad , \forall a \in U$$

de modo que notando

$$U = \{a_1,..,a_k\} \quad \text{con } \pi(a_1) \geq \ldots \geq \pi(a_k),\ A_i = \{a_1,..,a_i\}$$

$$U = \{b_1,..,b_k\} \quad \text{con } \pi'(b_1) \geq \ldots \geq \pi'(b_k),\ B_j = \{b_1,..,b_j\}$$

Tiene que cumplirse $\pi(a_i) \geq \pi'(b_i), \forall i \in \{1,..,k\}$

Tenemos con esta notación

$$E_{H_*}(\pi) = -\sum_{i=1}^{k-1} (\pi(a_i)-\pi(a_{i+1}))L(1/i)-\pi(a_k)L(1/k) =$$

$$= \sum_{i=1}^{k-1} (\pi(a_i) - \pi(a_{i+1}))L(i) + \pi(a_k)L(k)$$

$$E_{H_*}(\pi') = \sum_{i=1}^{k-1} (\pi'(b_i) - \pi'(b_{i+1}))L(i) + \pi'(b_k)L(k)$$

Consideremos las funciones $t, t' : [0,1] \longrightarrow R_0^+$, dadas por

$$t(r) = \begin{cases} L(i) & \pi(a_i) \leqq r < \pi(a_{i+1}) \\ L(k) & 0 \leqq r \leqq \pi(a_k) \end{cases}$$

$$t'(r) = \begin{cases} L(i) & \pi'(b_i) \leqq r < \pi'(b_{i+1}) \\ L(k) & 0 \leqq r \leqq \pi'(b_k) \end{cases}$$

Entonces se cumple

$$E_{H_*}(\pi) = \int_0^1 t(x)dx \qquad\qquad E_{H_*}(\pi') = \int_0^1 t'(x)dx$$

Para $r \in [0,1]$ sea

$$i(r) = \begin{cases} i & \text{si } \pi(a_i) \leqq r < \pi(a_{i+1}) \\ k & \text{si } 0 \leqq r \leqq \pi(a_k) \end{cases}$$

$$j(r) = \begin{cases} j & \text{si } \pi'(b_j) \leqq r < \pi'(b_{j+1}) \\ k & \text{si } 0 \leqq r \leqq \pi'(b_k) \end{cases}$$

Como $\pi(a_i) \geqq \pi'(b_i)$ $\forall i \in \{1,..,k\}$, es entonces $i(r) \geqq j(r)$ cualquiera que sea $r \in [0,1]$. Por tanto $t(r) \geqq t'(r)$, para todo $r \in [0,1]$; y entonces

$$E_{H_*}(\pi) = \int_0^1 t(x)dx \geqq \int_0^1 t'(x)dx = E_{H_*}(\pi')$$, como queríamos probar.

#

<u>Nota.-</u> Desgraciadamente no se puede definir de forma análoga

una medida de incertidumbre por medio de H* (entropía superior) ya que no se puede garantizar el cumplimiento de la propiedad - iii'). Probaremos esta afirmación mediante un contraejemplo.

<u>Ejemplo 2.7.-</u> Sea $U = \{x_1, x_2, x_3, x_4, x_5\}$ y dos distribuciones, π y π' con

U	x_1	x_2	x_3	x_4	x_5
$\pi(x_i)$	1	0.8	0.6	0.4	0.2
$\pi'(x_i)$	1	0.8	0.6	0.4	0.4

Es inmediato que $\pi' \subset \pi$. De acuerdo con su definición

$$H^*(m_\pi) = -0.2 \sum_{i=1}^{5} \log(0.2/i)$$

$$H^*(m_{\pi'}) = -0.2 \sum_{i=1}^{3} \log(0.2/i) - 0.4\log(0.4/5)$$

de modo que $\quad H^*(m_\pi) > H^*(m_{\pi'})$

El ejemplo que acabamos de ver no quiere decir que H* sea -- una medida de entropía inaceptable, ya que en la mayoría de los casos, si $\pi' \subset \pi$ entonces $H^*(m_\pi) \leqq H^*(m_{\pi'})$. Ahora bien, como medida de ntropía por si sola es "poco fina"; recordemos que ha - sido definida en el capítulo anteior, fundamentalmente, para -- que sirva de complemento a H_*. Al ser una cota superior, en evidencias particulares está muy sobrevalorada la incertidumbre, y eso da lugar a la existencia de ejemplos como el que acabamos - de examinar.

Tal y como hemos hemos indicado antes, las siguientes aplicaciones en $\prod(U)$

$$I_n(\pi) = E_n(\pi_U) - E_n(\pi) = \log_2 \left[\frac{k}{\sum_{a \in U} \pi^n(a)} \right] \ , \ n \geqq 1$$

$$I_n^*(\pi) = E_n^*(\pi_U) - E_n^*(\pi) = \log_2 \left[\frac{k}{\sum_{a \in U} \pi^{1/n}(a)} \right], \ n \geqq 2$$

$$I_{H_*}(\pi) = E_{H_*}(\pi_U) - E_{H_*}(\pi) = - \sum_{i=1}^{k-1} (\pi(a_i) - \pi(a_{i+1}))L(i/k)$$

donde $U = \{a_1,..,a_k\}$ con $\pi(a_1) \geqq ... \geqq \pi(a_k)$
son medidas de información en $(U, \prod(U))$.

-Sería interesante definir medidas de información condicionadas en $(U, \prod(U))$ (que midan la información que proporciona π, conocida π'). Desgraciadamente para esto nos haría falta una estructura de retículo en $\prod(U)$ de la que no disponemos.

4.-INFORMACIONES DIFUSAS Y POSIBILISTICAS SOBRE LOS ELEMENTOS DE UN CONJUNTO

Sea U un conjunto finito y supongamos que, en lugar de desconocer un elemento de U, lo que desconocemos es un subconjunto A de U ($A \in \mathcal{P}(U)$).

De acuerdo con los desarrollos anteriores, una información difusa sobre A será una expresión de la forma p_H: " A es H " con $H \in \tilde{\mathcal{P}}(\mathcal{P}(U))$, que tendrá asociada una distribución de posibilidad en $\mathcal{P}(U)$. Ahora bien, existen otro tipo de informaciones sobre A, que consisten en afirmar que todos y cada uno de sus elementos cumplen una determinada propiedad. Esta última versión nos será de gran utilidad en el siguiente capítulo, por lo que introducimos la siguiente definición.

<u>Definición 2.8.-</u> Si $A \subset U$, desconocido, llamaremos información difusa sobre los elementos de A, a una expresión de la forma

$$p: \text{" } \forall x \in A, \text{ x es B " donde } B \in \tilde{\mathcal{P}}(U).$$

Vamos a estudiar este tipo de informaciones transformándolas en informaciones difusas sobre A, propiamente dichas, es decir en expresiones de la forma

$$\bar{p}: \text{" A es } H_B \text{ ", con } H_B \in \tilde{\mathcal{P}}(\mathcal{P}(U))$$

Esto puede hacerse observando que la restricción p indica --

que todos los elementos de A están en B, lo que, de alguna manera puede expresarse como " A está incluído en B ", si bien la definición de inclusión que hemos de considerar no es "crisp". De este modo, a p le podemos asociar $\overline{p}$ con H_B el subconjunto difuso de $\mathcal{P}(U)$ de aquellos elementos que están incluídos en B. A todo $F \subset U$ se le asigna un grado de inclusión en B, que será el grado de pertenencia de F a H_B.

Para esto puede ser útil la definición de inclusión débil debida a Dubois, Prade.

<u>Definición 2.9 (Dubois, Prade 1.980).-</u> Sean $C,D \in \underset{\sim}{\mathcal{P}}(U)$. Se dice que C está debilmente incluído en D con grado α ($C \vdash_\alpha D$) si y solo si $\forall a \in U$, $\text{Max}\ \{1 - \mu_C(a), \mu_D(a)\} \geqq \alpha$.

-Si C y D $\in \underset{\sim}{\mathcal{P}}(U)$, el grado con que C está incluído en D es

$$\text{Sup}\ \{\alpha \in [0,1] / C \vdash_\alpha D\}$$

En nuestro caso tenemos $F \in$ (U) y $B \in$ (U), entonces

$$F \vdash_\alpha B \text{ si y solo si } \text{Max}\{1 - \mu_F(a), \mu_B(a)\} \geqq \alpha, \forall a \in U$$

o, equivalentemente, $\mu_B(a) \geqq \alpha$, $\forall a \in U$.

De acuerdo con esto, si $B \in \underset{\sim}{\mathcal{P}}(U)$ y H_B es el subconjunto difuso de $\mathcal{P}(U)$ tal que el grado de pertenencia de $F \in \mathcal{P}(U)$ a H_B es igual al grado con que F está incluído en B; entonces según la definición 2.9, se tiene que

$$\forall F \in \mathcal{P}(U) \ , \quad \mu_{H_B}(F) = \text{Sup}\ \{\alpha \in [0,1] / a \in F, \mu_B(a) \geqq \alpha\} =$$

$$= \operatorname*{Inf}_{a \in F} \mu_B(a)$$

con $\operatorname*{Inf}_{a \in \phi} \mu_B(a) = 1$.

Considerando esta función de pertenencia para H_B, la distribución de posibilidad asociada a p, π, viene dada por

$$\overline{\pi}(F) = \operatorname*{Min}_{a \in F} \mu_b(a)$$

Ahora bien, si por hipótesis, se supone que el conjunto A es distinto del vacío se obtiene:

$$\cdot\bar{\pi}(F) = \frac{\mu_{H_B}(F)}{\displaystyle\sup_{C\in\mathcal{P}(U)-\{\phi\}} (\mu_{H_B}(C))} = \frac{\displaystyle\min_{a\in F} \mu_B(a)}{\displaystyle\sup_{\substack{C \subset U \\ C \neq \phi}} \left[\min_{a\in C} \mu_B(a)\right]} =$$

$$= \frac{\displaystyle\min_{a\in F} \mu_B(a)}{\displaystyle\max_{b\in U} \mu_B(b)} = \min_{a \in F} \left[\frac{\mu_B(a)}{\displaystyle\max_{b\in U} \mu_B(b)}\right] = \min_{a \in F} \left[\frac{\mu_B(a)}{C(B)}\right] \text{, si } F\neq\phi$$

$$\cdot\bar{\pi}(\phi) = 0.$$

Nota.- En las aplicaciones que vamos a realizar de este tipo de informaciones, tiene sentido suponer que $A \neq \phi$. Por tanto, en lo que sigue, consideraremos la distribución de posibilidad que se obtiene en esta hipótesis.

Sea $\bar{\Pi}$ la medida de posibilidad inducida por $\bar{\pi}$ en $\mathcal{P}(\mathcal{P}(U))$ y - consideremos la aplicación $\Pi: \mathcal{P}(U) \longrightarrow [0,1]$ dada por

$$\Pi(F) = \bar{\Pi}(T(F)), \text{con } T(F) = \{C \subset U/C \cap F \neq \phi\} .$$

Proposición 2.6.- Π es una medida de posibilidad en $\mathcal{P}(U)$ tal que

$$\Pi(F) = \max_{a \in F} \bar{\pi}(\{a\}) = \max_{a \in F}\left[\frac{\mu_B(a)}{C(B)}\right] \quad ,\forall F \subset U$$

Demostración.- Es inmediata sin más que considerar la definición de Π.

Sea π la distribución de posibilidad sobre U asociada a Π. A partir de ella podemos obtener $\bar{\pi}$ en $\mathcal{P}(U)$, por medio de la expresión

$$\bar{\pi}(F) = \min_{a \in F} \pi(a)$$

La interpretación que se le puede dar a π y Π es la siguiente. $\Pi(F) = \Pi(T(F))$ que, según indicamos en el primer punto, se considera que es la posibilidad de que A sea igual a algún elemento de T(F) (que está formado por los subconjuntos de U que - tienen intersección no vacía con F); es decir $\Pi(F)$ se puede con siderar como la posibilidad de que algún elemento de F esté en A. Y analogamente si $a \in U$, $\pi(a)$ se puede interpretar como la posibilidad de que a pertenezca al suconjunto $A \subset U$, desconocido.

A modo de conclusión, vamos a resumir en una serie de puntos lo que hemos expuesto anteriormente.

- Tenemos $A \subset U$, $A \neq \phi$, y una información sobre A de la forma p: " $x \in A$, x es B ", $B \in \underset{\sim}{\mathfrak{P}}(U)$.

- Esta información la transformamos en una de la forma p: "A pertenece a H_B " donde $H_B \in \underset{\sim}{\mathfrak{P}}(\mathfrak{P}(U))$ con

$$\mu_{H_B}(F) = \underset{a \in F}{\mathrm{Min}}\ \mu_B(a) \quad , F \subset U.$$

- A esta información difusa, sabiendo $A \neq \phi$, le asociamos una distribución de posibilidad, $\overline{\pi}$, y su correspondiente medida $\overline{\Pi}$, con

$$\overline{\pi}(F) = \underset{a \in F}{\mathrm{Min}} \left[\frac{\mu_B(a)}{C(B)} \right]$$

- A partir de $\overline{\pi}$ y $\overline{\Pi}$ se pueden definir una distribución de posibilidad, π, en U y su correspondiente medida Π. Si $\overline{\pi}(F)$ se interpreta como la posibilidad de que el A desconocido coincida con F, ahora $\Pi(F)$ se interpreta como la posibilidad de que algún elemento de F esté en A y $\pi(a)$, $a \in U$, como la posibilidad de que $a \in A$. Esta distribución π, o su correspondiente medida, Π, permiten obtener $\overline{\pi}$ y $\overline{\Pi}$. Las expresiones que las relacionan son

$$\Pi(F) = \underset{a \in F}{\mathrm{Max}}\ \overline{\pi}(\{a\}) \qquad ; \quad \pi(a) = \overline{\pi}(\{a\})$$

$$\overline{\pi}(F) = \underset{a \in F}{\mathrm{Min}}\ \pi(a) = \underset{a \in F}{\mathrm{Min}}\ \Pi(\{a\})$$

A $\overline{\pi}$ le llamaremos información posibilística global asociada a p, sobre A.

A π le llamaremos información posibilística particular sobre los elementos de A, asociada a p.

Nota 1.- Existe otro método de establecer π, dada p, que nos permitirá comprender mejor su interpretación.

Consideremos $A \in \mathcal{P}(U)$, $A \neq \phi$, desconocido y una información difusa sobre A, de la forma p: "$\forall x \in A$, x es B " con $B \in \underset{\sim}{\mathcal{P}}(U)$. Si a es un elemento cualquiera perteneciente a A, sobre este a disponemos de la información p_a: " a es B ". La distribución de posibilidad asociada a p_a coincide con π. Luego π es la información posibilística que se desprende de p, para cualquier elemento de A. $\pi(b)$, $b \in U$, representa la posibilidad de que b coincida con a, siendo este un elemento cualquiera de A; es decir, la posibilidad de que b pertenezca a A.

Nota 2.- Si tenemos un conjunto U y $A \subset U$ desconocido, no toda información de la forma p: " A es H " con $H \in \underset{\sim}{\mathcal{P}}(\mathcal{P}(U))$ da lugar a una distribución de posibilidad en $\mathcal{P}(U)$ que se pueda obtener equivalentemente, por medio de otra en U.

Por ejemplo, esto no ocurriría así con informaciones de la forma " A tiene muchos elementos ".

Existen otras para las que si parece lógico que esto sea así, por ejemplo, si afirmamos q: " $\exists x \in A$, tal que x es B ", esta información se puede transformar en $\overline{q}$: " A pertenece a T_B " con

$$\mu_{T_B}(F) = \operatorname{Sup}_{a \in F} \mu_B(a)$$

Si $\overline{\pi}$ es la distribución de posibilidad asociada a $\overline{q}$, sería

$$\overline{\pi}(F) = \operatorname{Sup}_{a \in F} \pi(a) \quad \text{con} \quad \pi(a) = \frac{\mu_B(a)}{C(B)}$$

es decir $\overline{\pi} \equiv \Pi$.

Nota 3.- El estudio que acabamos de hacer se puede extender al caso en que el elemento desconocido sea un subconjunto difuso de U; es decir $A \in \underset{\sim}{\mathcal{P}}(U)$. En ese caso, una información difusa p: " $\forall x \in U$, si x es A, entonces x es B ", tendrá asociada una información en $\underset{\sim}{\mathcal{P}}(U)$

$$p: \text{" A pertenece a } H_B \text{ "}$$

con $H_B \in \underset{\sim}{\mathcal{P}}(\underset{\sim}{\mathcal{P}}(U))$ dado por

$$\mu_{H_B}(F) = \mathrm{Sup}\,\{\alpha \in [0,1] \,/\, F \vdash_{\alpha} B\}$$

La transformación de p en $\overline{p}$ (realizada de modo análogo al caso anterior) permite obtener una distribución de posibilidad en (U).

Nota 4.- Sea X una aplicación entre los conjuntos U y V, que supondremos desconocida. Si damos una información de la forma

$$p: \text{" } \forall b \in U,\ X(b) \text{ es } B \text{ " con } B \in \tilde{\mathcal{P}}(U)$$

estamos dando una información difusa sobre los elementos de Im(X) $\subset$ V (que es desconocida) del tipo que acabamos de estudiar. Si sabemos que $U \neq \phi$, entonces Im(X) $\neq \phi$, por lo que nos situaremos en el caso en que el subconjunto desconocido no puede ser vacío.

Esta información, según hemos visto, lleva asociada una distribución de posibilidad, $\overline{\pi}$, en $\mathcal{P}(V)$:

$$\overline{\pi}(F) = \underset{a \in F}{\mathrm{Min}} \left[\frac{\mu_B(a)}{C(B)}\right] \quad , \forall F \subset V \ ,\ F \neq \phi\ ;\ \overline{\pi}(\phi) = 0.$$

El valor de $\overline{\pi}(F)$ se interpreta como la posibilidad de que Im(X) = F, dada p.

Dicha distribución, $\overline{\pi}$, puede obtenerse a partir de la distribución de posibilidad π en V, dada por

$$\pi(a) = \frac{\mu_B(a)}{C(B)} \qquad , \quad \forall a \in V.$$

Si $a \in V$, $\pi(a)$ se interpreta como la posibilidad de que a pertenezca a Im(X), dada p; es decir el grado con que a es posible valor de X. A consecuencia de esta interpretación, notaremos también $\pi(a)$ como $\pi(X = a)$.

Disponemos también de Π que es la medida de posibilidad asociada a π en V, y que viene dada por

$$\Pi(F) = \underset{a \in F}{\mathrm{Max}}\ \pi(a) = \underset{a \in F}{\mathrm{Max}} \left[\frac{\mu_B(a)}{C(B)}\right] \quad , \forall F \subset V.$$

Dicho valor se interpreta como la posibilidad de que algún elemento de F pertenezca a Im(X), dada p; es decir la posibilidad de que algún valor de X caiga en F. Y, por tanto, $\Pi(F)$ lo notaremos también como $\Pi(X \in F)$.

- CAPITULO III -

Consistencia posibilidad-probabilidad. Relaciones entre posibilidad y probabilidad.

0.-INTRODUCCION

En este capítulo desarrollamos el tema fundamental de esta memoria: el estudio del principio de consistencia posibilidad-probabilidad y, en general, las relaciones entre Posibilidad y Probabilidad.

Comenzamos, en el primer apartado, haciendo una revisión bibliográfica sobre el tema. Analizamos los trabajos de Zadeh (1.978) y Dubois, Prade (1.982 y 1.983) que constituyen las aportaciones más importantes sobre el principio de consistencia. También estudiamos los artículos de Hohle (1.982), Lindley (1.982) y Natvig (1.983) que consideran las relaciones entre los conceptos de posibilidad y probabilidad. Los dos últimos afirman, desde un punto de vista bayesiano, que la Teoría de la Probabilidad proporciona un modelo suficiente para describir todos los casos de incertidumbre.

En el segundo apartado, presentamos nuestra formulación de la consistencia entre posibilidad y probabilidad. Para ello, comenzamos exponiendo en qué contexto creemos nosotros que tiene sentido hablar de consistencia. Este aspecto no ha sido concretado hasta ahora, lo que ha impedido su formulación sin ambigüedades. Hay dos aspectos que destacamos de nuestra formulación:

-La consistencia tiene sentido cuando cuando la distribu ción de posibilidad que se considere, proviene de una información difusa sobre todos los resultados de un determinado experimento aleatorio (ver apartado cuarto del segundo capítulo.

-La consistencia es una propiedad difusa y, por tanto, - una definición de ella ha de hacerse mediante una medida que determine un grado de cumplimiento.

Nosotros introducimos una axiomática que debe de cumplir un funcional para que pueda considerarse una medida de consistencia. Las medidas de consistencia de Zadeh y Dubois, Prade cumplen esta axiomática y son compa adas con otras medidas que hemos obtenido.

En el apartado tercero, consideramos el problema de asignar una distribución de probabilidad a una de posibilidad. En primer lugar, analizamos el procedimiento de Y.Leng (1.980), basado en el principio de máxima entropía: si disponemos de una --- cierta información sobre un experimento y necesitamos obtener - una distribución de probabilidad, entonces, la forma más objetiva de establecerla es considerando la distribución de máxima entropía de entre las que son compatibles con la dada.

Nosotros proponemos una variación del procedimiento de Leng que esencialmente consiste en sustituir una condición de tipo - preciso (la consistencia de la distribución de probabilidad con la de posibilidad dada ha de ser α, siendo α un número suficientemente próximo a uno) por una de tipo difuso (ambas distribuciones son consistentes), ya que la consistencia es una propiedad difusa. El problema de programación original se transforma, entonces, en un problema de programación difuso. Proponemos resolver dicho problema por el procedimiento de J.L. Verdegay -- (1.980, 1.982); obteniendo la solución explícita utilizando varios tipos de medidas de consistencia.

Por último, en el cuarto apartado, consideramos un problema que, en cierto sentido, es dual del anterior: asignar una distribución de posibilidad a una distribución de probabilidad. -- Este problema se encuentra planteado originalmente en Dubois, -

Prade (1.982). Nosotros proponemos una versión de este problema que difiere notablmente de la de Dubois, Prade y que puede formularse como: determinar la distribución de posibilidad que proporciona máxima información dentro del conjunto de aquellas que son consistentes con la distribución de probabilidad dada.

Esta versión la justifacamos en base a un ejemplo sobre test de hipótesis, que muestra como en algunas ocasiones, en Estadística Matemática, se utilizan conceptos que alcanzan su pleno sentido dentro de la Teoría de la Posibilidad

Una vez elegida la medida de consistencia y de información, el problema se nos reduce a un problema de programación difusa que resolvemos en varios casos concretos.

1.-ANTECEDENTES DEL PRINCIPIO DE CONSISTENCIA POSIBILIDAD-PROBABILIDAD

A) El principio de consistencia según Zadeh.- El principio de consistencia posibilidad-probabilidad fue introducido por Zadeh en 1.978 (Zadeh (1.978)). Como el mismo indica, el principio de consistencia posibilidad-probabilidad no es una "ley precisa o una relación que sea intrínseca a los conceptos de posibilidad y probabilidad "; sino que es la expresión de una débil conexión de tipo heurístico entre ambos conceptos.

Antes de pasar a una discusión más detallada del mencionado principio, presentaremos la definición de distribución de posibilidad dada por Zadeh.

Definición 3.1.- Sea A un conjunto difuso (en un universo de discurso U) caracterizado por su función de pertenencia μ_A, interpretando el grado de pertenencia, $\mu_A(a)$, como la compatibilidad de a con el concepto representado por A. Sea X una variable que toma valores en U, y consideremos que A actúa como una restricción difusa, R(X), asociada a X. Entonces la proposición " X es A "; que se expresa

$$R(X) = A$$

asocia una distribución de posibilidad, π_X, a X que se postula igual a R(X); esto es

$$\pi_X = R(X)$$

Correspondientemente, la función distribución de posibilidad asociada a X (o la función de distribución de posibilidad de -- π_X) se nota por π_X y se define como numericamente igual a la -- función de pertenencia de A; esto es

$$\pi_X \equiv \mu_A$$

Por tanto $\pi_X(a)$, la posibilidad de que X = a, se postula --- igual a $\mu_A(a)$.

Nota.- Como ya hemos comentado anteriormente, nosotros traba jamos con la hipótesis de que la distribución de posibilidad a- sociada a una restricción difusa ha de normalizarse.

Zadeh afirma que, según la definición de $\pi_X(a)$, el grado de posibilidad puede ser cualquier número en el intervalo [0,1], - más bien que solo 0 o 1. Estos grados intermedios de posibili-- dad caracterizan situaciones entre lo imposible y lo totalmente posible y se encuentran frecuentemente en expresiones del tipo: " es casi imposible que un hombre mida 2m, 20cm", " es bastante posible que llueva mañana ".

Frecuentemente, cuando se dice "es muy posible...", " es po- co posible..." se confunde con " es muy probable...", " es poco probable... "; sin embargo estas expresiones tienen significa-- dos distintos: unas se refieren a la posibilidad de unos suce-- sos y las otras a su probabilidad de ocurrencia.

Para aclararnos las diferencias y conexiones entre ambas ex- presiones (diferencias y conexiones entre posibilidad y probabi lidad) Zadeh(1.978) propone el siguiente ejemplo.

Ejemplo 3.1.- Consideremos la proposición " Hans toma X hue- vos para desayunar ", con X tomando valores en U = {1,2,3,...}. Podemos asociar una distribución de posibilidad a X, interpre-- tando $\pi_X(a)$ ($a \in U$) como el grado de facilidad con que Hans pue de comer a huevos. Supongamos que disponemos, también, de una distribución de probabilidad asociada a X: $p_X(a)$ = probabilidad con que Hans toma a huevos.

Sean los valores numéricos de π_X y p_X como se muestran en la siguiente tabla.

$a \in U$	1	2	3	4	5	6	7	8	9	...
$\pi_X(a)$	1	1	1	1	0.8	0.6	0.4	0.2	0	...
$p_X(a)$	0.1	0.8	0.1	0	0	0	0	0	0	...

La probabilidad está relacionada con la frecuencia con que Hans come un número determinado de huevos y la posibilidad con la capacidad de Hans para comérselos; por tanto, ambas distribuciones representan distintas realidades.

Sin embargo, existe una relación entre ambos conceptos que surge de las siguientes observaciones:

- Un alto grado de posibilidad no implica un alto grado de probabilidad, como prueba el hecho de que la posibilidad con que Hans se come tres huevos es uno y la probabilidad es, sin embargo, 0.1. Esto se explica diciendo que si una persona puede comerse tres huevos, sin problemas, esto no quiere decir que lo haga frecuentemente.

- Si un suceso es imposible, debe de ser improbable. Esto se cumple en nuestro ejemplo; es lógico, ya que si Hans tiene dificultad para comerse un número determinado de huevos, no lo hará frecuentemente.

En base a estas consideraciones, Zadeh formula su principio de consistencia posibilidad-probabilidad, que se expresa de la siguiente forma:

- Si una variable X toma los valores $\{a_1,..,a_k\}$ con posibilidades respectivas $\pi = (\pi_1,\pi_2,..,\pi_k)$ y probabilidades $p = (p_1,p_2,..,p_k)$, entonces el grado de consistencia de la distribución de probabilidad p con la distribución de posibilidad π --

viene expresado por

$$\gamma = \pi_1 p_1 + \dots + \pi_n p_n$$

Este valor , γ, no se deduce a partir de ninguna propiedad o relación que deba existir entre π y p sino que, en palabras del propio Zadeh, " es una formulación aproximada de la observación heurística de que una disminución de la posibilidad de un suceso conlleva una disminución de su probabilidad si bien, no se cumple la propiedad recíproca".

B) El principio de consistencia de Dubois, Prade.- El principio de consistencia de posibilidad-probabilidad ha sido ampliamente estudiado por Dubois, Prade (Dubois, Prade (1.982)). Estos autores comienzan haciendo una distinción entre dos usos de la palabra posible:

a)Cuando se entiende por posible aquello " que puede ser hecho " (modalidad " de re ").

b)Cuando posible se refiere a aquello "que puede ocurrir " (modalidad " de dicto ").

Estos dos tipos de posibilidad están relacionados: lo que no puede ser hecho no puede ocurrir, pero no reciprocamente. Dubois, Prade utilizan el ejemplo de Zadeh para ilustrar esta diferencia. La posibilidad " de re " está relacionada con la capacidad de Hans para comer un número determinado de huevos y es la que considera Zadeh. La posibilidad " de dicto " se relaciona con la ocurrencia de un determinado suceso y esta no coincide con la estudiada por Zadeh: por ejemplo, la posibilidad que se asig a a que Hans se coma cuatro huevos es 1; sin embargo, su probabilidad es cero y, por tanto, es un suceso que nunca puede ocurrir y su posibilidad " de dicto" ha de ser cero.

Este tipo de posibilidad (" de dicto ") y su relación con la probabilidad es la que consideran Dubois, Prade. Su estudio no trata solo el problema de la consistencia, sino también, el desarrollo de un procedimiento para obtener una distribución de posibilidad a partir de una distribución de probabilidad (pro--

blema al que llaman " interpretación posibilística de histogramas "). A continuación recogemos las ideas de estos autores.

Sea U un conjunto finito y Π y P dos medidas, de posibilidad y probabilidad respectivamente, en $\mathcal{P}(U)$. Entonces Π y P se consideran consistentes cuando

$$\Pi(A) \geqq P(A) \quad , \forall A \in \mathcal{P}(U) \tag{1}$$

Esta desigualdad se justifica intuitivamente haciendo referencia a la idea de que cuanto más probable es un suceso, más posible hemos de considerar su ocurrencia (considerando, claro está, posibilidades " de dicto ").

<u>Proposición 3.1 (Dubois (1.980).-</u> Sean π y p dos distribuciones de posibilidad en U (finito) y definamos

$$C(\alpha) = \{a \in U / \ p(a) \geqq \alpha \} \ , \ \alpha \in [0, \sup_{a \in U} p(a)].$$

Si se cumple

$$\pi(a) \geqq P(\ C(p(a))\ ) = \sum_{b \in C(p(a))} p(b) \tag{2}$$

entonces la medida asociada a π , Π, verifica (1), es decir es consistente con P.

Esta condición no es, en general, necesaria y su expresión es demasiado complicada. Para obviar estos inconvenientes, hemos desarrollado la siguiente proposición.

<u>Proposición 3.2.-</u> Sean π y p dos distribuciones de posibilidad y probabilidad, respectivamente, en U y notemos

$$U = \{a_1,\ldots,a_n\} \text{ con } \pi(a_1) \geqq \ldots \geqq \pi(a_n)$$

Condicion necesaria y suficiente para que se cumpla (1) es que

$$\pi(a_i) \geqq p(a_1)+\ldots+p(a_i) \qquad i=1,\ldots,n \tag{3}$$

<u>Demostración</u>

a)condición necesaria.- Por la forma en que hemos notado U, basta considerar que se cumple (1) para los conjuntos de la for ma $A_i = \{a_1,..,a_i\}$; ya que

$$\Pi(A_i) = \pi(a_i) \quad y \quad P(A_i) = p(a_1) + \ldots + p(a_i)$$

de manera que $\Pi(A_i) \geqq P(A_i)$, $\forall i \in \{1,..,n\}$,implica el cumplimiento de (2).

b) Condición suficiente.- Consideremos $A \subset U$ arbitrario y sea $i(A) = \text{Max} \{ i \in \{1,..,n\} / a_i \in A \}$. Entonces,

$$\Pi(A) = \sup_{a \in A} \pi(a) = \pi(a_{i(A)})$$

$$P(A) = \sum_{a \in A} p(a) \leqq p(a_1) + \ldots + p(a_{i(A)})$$

ya que si $i > i(A)$, entonces a_i no pertenece a A.

Como, por (3), $\pi(a_{i(A)}) \geqq p(a_1) + \ldots + p(a_{i(A)})$, se deduce $\Pi(A) \geqq P(A)$, como queríamos demostrar. #

Existe una justificación de (1) como condición de consistencia , en términos de la teoría de la evidencia, y que se basa en la siguiente proposición.

<u>Proposición 3.3.-</u> Sea U un conjunto finito y m_π y m_p dos asignaciones básicas en U que representan sendas evidencias, posibilística y probabilística respectivamente, con distribuciones asociadas π y p. Entonces π y p son consistentes según Dubois, Prade si y solo si la evidencia m_π está incluída en m_p.

<u>Demostración</u>

a) Demostremos que si π y p son consistentes, entonces m_π está incluída en m_p.

Sea $U = \{a_1,...,a_n\}$ con $\pi(a_1) \geqq \ldots \geqq \pi(a_n)$ y

$A_i = \{a_1,...,a_i\}$.

Entonces, se tiene

$$m_\pi(A) = \begin{cases} 0 & \text{si } A \neq A_i, \forall i \in \{1,..,n\} \\ \pi(a_1) & \text{si } A = A_1 \\ \pi(a_i)-\pi(a_{i-1}) & \text{si } A = A_i,\ i = 2,..,n \end{cases}$$

y, por otra parte,

$$m_p(A) = \begin{cases} p(a_i) & \text{si } A = \{a_i\} \\ 0 & \text{en otro caso} \end{cases}$$

Consideremos ahora la familia de aplicaciones

$$\{(m_\pi)_{A_i} : U \longrightarrow [0,1]\}_{i \in \{1,...,n\}}$$

construídas recurrentemente, de acuerdo con

- $i=1$. Sea $j(A_1) \in \{1,..,n\}$ el mayor valor para el cual

$$\sum_{i=1}^{j(A_1)} p(a_i) \leqq \pi(a_1).$$

Notemos que la condición (3) garantiza $j(A_1) \geqq 1$.

Hacemos entonces

$$(m_\pi)_{A_1}(\{a_i\}) = p(a_i) \qquad , 1 \leqq i \leqq j(A_1)$$

$$(m_\pi)_{A_1}(\{a_{j(A_1)+1}\}) = \pi(a_{j(A_1)+1}) - \sum_{i=1}^{j(A_1)} p(a_i)$$

$$(m_\pi)_{A_1}(A) = 0 \quad , \text{en otro caso.}$$

- $i = k \geqq 2$. Supuesto construído $(m_\pi)_{A_{k-1}}$, definimos $j(A_k)$ como el mayor valor para el cual se cumple

$$\sum_{i=1}^{j(A_k)} p(a_i) \leqq \pi(a_k)$$

Notemos que $j(A_{k-1}) \leqq j(A_k)$ y, además, la condición (3) nos garantiza que $j(A_k) \geqq k$.

Hacemos entonces

- Si $j(A_{k-1}) = j(A_k)$

$$(m_\pi)_{A_k}(\{a_{j(A_k)}\}) = \pi(a_k) - \pi(a_{k-1})$$

$$(m_\pi)_{A_k}(A) = 0 \quad , \text{ si } A \neq \{a_{j(A_k)}\}$$

- Si $j(A_k) > j(A_{k-1})$

$$(m_\pi)_{A_k}(\{a_{j(A_{k-1})+1}\}) = p(a_{j(A_{k-1})+1}) - $$
$$- (m_\pi)_{A_{k-1}}(\{a_{j(A_{k-1})+1}\}) \geq 0$$

$$(m_\pi)_{A_k}(\{a_i\}) = p(a_i) \quad , \text{ si } j(A_{k-1})+1 < i \leq j(A_k)$$

$$(m_\pi)_{A_k}(\{a_{j(A_k)+1}\}) = \pi(a_k) - \sum_{i=j(A_{k-1})}^{j(A_k)} (m_\pi)_{A_k}(\{a_i\}) \geq 0$$

$$(m_\pi)_{A_k}(A) = 0 \quad , \text{ en otro caso.}$$

Con esta construcción se garantiza

$$\sum_{A \subset A_k} (m_\pi)_{A_k}(A) = m_\pi(A_k) \quad (\text{ya que } j(A_k) \geq k \ , \ k=1,..,n)$$

y $$\sum_{i=1}^{k} (m_\pi)_{A_k}(A) = m_p(A), \ \forall A \subset U \ ;$$

es decir $m_\pi \subset m_p$, como queríamos demostrar.

b) Si las evidencias m_π y m_p cumplen $m_\pi \subset m_p$, entonces la -- condición de consistencia de Dubois, Prade es una consecuencia inmediata de la propiedad 1) de la inclusión de evidencias, teniendo en cuenta que Π y P son las medida de plausibilidad asociadas a m_π y m_p, respectivamente. #

Permitiéndonos un abuso del lenguage y, en base a la propiedad demostrada, podremos decir que la distribución de posibilidad π es consistente con la de probabilidad p, cuando " la asignación básica que representa a p se puede obtener repartiendo, de alguna forma, la evidencia que m_π asigna a cada conjunto --- $A \subset U$ entre los conjuntos unitarios que están incluídos en A; - cuando la evidencia m_p se puede obtener mediante un proceso de atomización de m . En vista de esto, sería preferible disponer de p, en lugar de disponer de π; ya que, si π y p son consistentes, la evidencia representada por m está contenida en la re-- presentada por m_p.

El problema de la interpretación posibilística de los histogramas está planteado en los siguientes términos:

Sea $\{h(a) \in R^+ \; ; \; a \in U\}$ un histograma (h(a) representa el número de veces que ha aparecido $a \in U$) y supongamos que la --- mu stra empleada ha sido lo suficientemente grande, como para - igualar frecuencias y probabilidades; es decir

$$p_h(a) = \frac{h(a)}{\sum\limits_{b \in U} h(b)} \quad , \; a \in U$$

El problema consiste en construir, a partir del histograma,- una distribución de posibilidad en U, de forma que sea consis-- tente con p_h, y esté relacionada con ella en el siguiente sentido: si interpretamos la posibilidad de un suceso como la ausencia de sorpresa cuando dicho suceso ocurre, entonces un suceso de poca probabilidad (ocurre raramente) deberá tener poca posibilidad, ya que su aparición nos sorprenderá; por el contrario, un suceso de alta probabilidad ocurrirá a menudo y, por tanto, su aparición nos sorprenderá poco, de manera que deberá tener - posibilidad alta.

A partir de esta idea, una primera caracterización de la distribución de posibilidad asociada a p_h puede ser:

$$\pi_h(a) = \frac{h(a)}{\operatorname*{Sup}\limits_{b \in U} h(b)} = \frac{p_h(a)}{\operatorname*{Sup}\limits_{b \in U} p_h(b)} \quad , \; \forall a \in U.$$

Es evidente que se cumple

$$\pi_h(a) \geqq p_h(a) \quad , \forall a \in U.$$

Sin embargo, si consideramos las medidas Π_h y P_h asociadas a las distribuciones anteriores no ocurre, en general, que

$$\Pi_h(A) \geqq P_h(A) \quad , \ \forall\, A \subset U$$

de modo que no se cumple el principio de consistencia (1).

Hay, por tanto, que investigar otras formas de construir una distribución de posibilidad, a partir de una de probabilidad - dada.

Dubois, Prade proponen un método que justifican en términos de la medida de necesidad, N, asociada a una distribución de posibilidad ($N(A) = 1 - \Pi(\overline{A})$). En Dubois, Prade (1.983) puede encontrarse una exposición de este procedimiento con varios --- ejemplos ilustrativos.Se considera un conjunto U, finito, y una distribución de probabilidad, p, en U. Supongamos $U=\{b_1,..,b_n\}$ con $p(b_1) \geqq \ldots \geqq p(b_n)$ y sea $B_i = \{b_1,\ldots,b_i\}$. Entonces, si la necesidad de B_i se interpreta como el exceso de probabilidad de los elementos de B_i con respecto al elemento que tiene mayor probabilidad fuera de B_i; obtendremos

$$N(B_i) = \sum_{j=1}^{i} (p(b_j)-p(b_{i+1})), \quad j \leqq n-1$$

$$N(B_n) = N(U) = 1$$

$$N(B) = \max_{B_i \subset B} N(B_i) \quad , \forall B \subset U.$$

Se considera, entonces, asociada a p la medida de posibili-- dad dual de N ($\Pi_p(A) = 1 - N(\overline{A})$), cuya distribución de posibilidad tiene la forma

$$\begin{aligned} \pi_p(b_1) &= 1 \\ \pi_p(b_{i+1}) &= 1 - \sum_{j=1}^{i}(p(b_j)-p(b_{i+1})) \end{aligned} \qquad (4)$$

Esta construcción establece una aplicación biunívoca entre - las distribuciones de posibilidad y probabilidad en un conjunto U, ya que p viene determinada a partir de π_p por

$$p(b_i) = (1/i)\pi_p(b_i) - \sum_{j=i+1}^{n} \frac{1}{j(j-1)} \pi_p(b_j) \quad ,i=1,..,n \qquad (5)$$

Esta medida de posibilidad Π_h asociada a p es consistente, - según Dubois, Prade, con la medida P; es decir $P(A) \leqq \Pi_h(A)$, para todo $A \subset U$.

En Dubois, Prade (1.982) se encuentra, también, la siguiente justificación de (4): si nosotros queremos aproximar una evidencia de tipo general con asignación básica m, mediante una medida de probabilidad p_m, el método más natural de hacerlo consiste en equidistribuir los valores de la asignación básica,m(A), para cada conjunto $A \subset U$, entre los elementos de dicho conjunto; dando lugar a

$$p_m(a) = \sum_{a\in A\neq\phi} m(A) \frac{1}{|A|} \quad , a \in U.$$

Si m_π representa una evidencia posibilística con distribu-- ción π y notamos $U = \{a_1,..,a_n\}$ con $\pi(a_1) \geqq \dots \geqq \pi(a_n)$, se obtiene

$$p_\pi(a_i) = (1/i)\pi(a_i) - \sum_{j=i+1}^{n} \pi(a_j) \frac{1}{j(j-1)} \qquad (6)$$

expresión que coincide con (5).

Por tanto, este método,restringido a las evidencias posibi-- lísticas,coincide con la aplicación inversa de

$$\begin{array}{ccc} P(U) & \longrightarrow & \Pi(U) \\ p & \longmapsto & \pi_p \end{array}$$

donde $P(U)$ es el conjunto de las distribuciones de probabilidad sobre U.

Como consecuencia de todo lo anterior: $p_{\pi_p} \equiv p$ y $\pi_{p_\pi} \equiv \pi$.

Para terminar, Dubois, Prade (1.982) completan el ejemplo de

Zadeh con las posibilidades " de dicto " asociadas a la distribución de probabilidad con que la variable X: " huevos que Hans se come en el desayuno " , toma sus valores. Se obtiene la siguiente tabla.

U	1	2	3	4	5	6	7	8	9 ..
probabilidades	0.1	0.8	0.1	0	0	0	0	0	0 ..
posibilidades " de re "	1	1	1	1	0.8	0.6	0.4	0.2	0 ..
posibilidades " de dicto "	0.3	1	0.3	0	0	0	0	0	0 ..

c) El principio de consistencia en U. Hohle (1.982).- Hohle (1.982) estudia las medidas difusas en general, considerando las relaciones existentes entre ellas. Como caso particular, trata las medidas de posibilidad y probabilidad, estableciendo algunas conclusiones sobre el principio de consistencia posibilidad-probabilidad.

En primer lugar, demuestra un torema de representación de medidas difusas generales que recogemos, adaptando su notación a la nuestra y considerando el conjunto referencial U finito.

Teorema 3.1.- a) Sea h una medida difusa en $(U,\mathcal{P}(U))$, entonces existe una única medida de probabilidad D en $\mathcal{P}(\mathcal{P}(U))$ tal que $h(A) = D\{\omega \in \mathcal{P}(\mathcal{P}(U))/A \in \omega\}$, $\forall A \subset U$
con $D\{\omega/(\ \forall A,B \subset U, (A \subset B \text{ y } A \in \omega) \Rightarrow B \in \omega), \phi \notin \omega, U \in \omega\} = 1$ (7)

b) Pl es una medida difusa de plausibilidad en $(U , \mathcal{P}(U))$ si y solo si existe una única medida de probabilidad D en $\mathcal{P}(\mathcal{P}(U))$ tal que $Pl(A) = D\{\omega \in \mathcal{P}(\mathcal{P}(U))/A \in \omega\}$, $\forall A \subset U$
con $D\{\omega/(\forall A,B \in \mathcal{P}(U),\ A \cup B \in \omega \Leftrightarrow A \in \omega \quad B \in \omega), \phi \notin \omega, U \in \omega\} =$
$= 1$ (8)

c) Bel es una medida de creencia en $(U,\mathcal{P}(U))$ si y solo si existe una única medida de probabilidad D en $\mathcal{P}(\mathcal{P}(U))$ tal que

$$Bel(A) = D\{\omega \in \mathcal{P}(\mathcal{P}(U))/A \in \omega\} , \forall A \subset U$$

y $D\{\omega/(\ \forall A,B \subset U,\ A \cap B \in \omega \Leftrightarrow A \in \omega \text{ y } A \in \omega\ ), \phi \notin \omega, U \in \omega\}=1$ (9)

d) P es una medida de probabilidad en $(U,\mathcal{P}(U))$ si y solo si existe una medida de probabilidad D en $\mathcal{P}(\mathcal{P}(U))$ tal que

$$P(A) = D\{\omega \in \mathcal{P}(\mathcal{P}(U)) / A \in \omega\} \text{ y D cumple (8) y (9).}$$

Nota.- Un elemento $\omega \in \mathcal{P}(\mathcal{P}(U))$ se considera una realización de un experimento, respecto a un conjunto de sucesos $\mathcal{P}(U)$: para una realización $\omega \in \mathcal{P}(\mathcal{P}(U))$, el suceso $A \subset U$ ocurre si y solo si $A \in \omega$.

Como consecuencia de este teorema, U. Hohle obtiene una serie de resultados, en los cuales basa posteriormente sus conclusiones y que vamos a resumir.

Definición 3.2.- Sea $\otimes$ una conorma, se dice que $\otimes$ es distributiva si y solo si $\forall a,b,c \in [0,1]$,

$$a \otimes (1 - (1 - b) \otimes (1 - c)) = a \otimes b + a \otimes c - a \otimes (b \otimes c)$$

Definición 3.3.- Sea g una medida difusa en U que está basada en la conorma distributiva $\otimes$, diremos entonces que g es una medida de $\otimes$-posibilidad.

Si $\otimes$ es la conorma del máximo,obtenemos las medidas de posibilidad consideradas a lo largo de este trabajo.

Teorema 3.2.- Toda medida de $\otimes$-posibilidad es una medida de plausibilidad.

Proposición 3.4.- Supongamos que $\mathcal{P}(U)$ contiene más de dos elementos ($|U| > 1$), entonces no existe ninguna medida difusa en (U, (U)) tal que sea, simultaneamente, una medida de $\otimes$-posibilidad y una medida de probabilidad.

Proposición 3.5.- Sean Pl_1 y Pl_2 dos medidas de plausibilidad en $(U,\mathcal{P}(U))$ y g una combinación convexa de Pl_1 y Pl_2; entonces

a) Si g es una medida de posibilidad entonces Pl_1 y Pl_2 también lo son.

b) g es una medida de probabilidad $\Leftrightarrow$ Pl_1 y Pl_2 son, ambas, medidas de probabilidad.

Las conclusiones establecidas por U. Hohle, apartir de estos resultados y que están directamente relacionadas con nuestro tema son:

1) De la proposición 3.4 se infiere que la posibilidad y la probabilidad son nociones inconsistentes.

2) Como el conjunto de todas las medidas de plausibilidad en (U) es convexo, desde un punto de vista geométrico el conjunto de las medidas de plausibilidad puede considerarse como un cono convexo. Como consecuencia de la proposición 3.5, las medidas de posibilidad y probabilidad serán puntos extremos de dicho cono convexo..

3) Debido a lo anterior, el principio de consistencia posibilidad-probabilidad no tiene sentido; y se sugiere reemplazar el estudio de este principio por los de consistencia probabilidad-plausibilidad y probabilidad-creencia.

A nuestro criterio, las conclusiones de U. Hohle han sido obtenidas muy a la ligera y atendiendo más a justificaciones formales que a consideraciones intuitivas.

d) Otros estudios sobre las relaciones entre posibilidad y probabilidad.-

-B. Natvig (1.983) en su artículo " Possibility versus Probability " no menciona explicitamente el principio de consistencia posibilidad-probabilidad, pero estudia las relaciones entre ambos tipos de medidas. Natvig analiza una serie de ejemplos (entre los cuales se encuentra el que utilizó Zadeh (1.978) para introducir el principio de consistencia) y en cada uno de ellos, reduce la distribución de posibilidad considerada a una familia de distribuciones de probabilidad.

Como consecuencia, plantea la cuestión de si las reglas que rigen el comportamiento humano en presencia de incertidumbre, pueden extraerse siempre de las leyes de la probabilidad (probabilidad bayesiana, claró está. No considerada como límite de las frecuencias relativas, sino como grado subjetivo de incertidumbre).

Como respuesta, nosotros indicaremos que el término " posibilidad " es utilizado habitualmente en el lenguage humano para expresar una idea distinta a la de probabilidad. Es muy común, por ejemplo, escuchar la frase: " es posible, pero poco probable que ocurra ... ". En segundo lugar, la Teoría de la Posibilidad proporciona métodos de actuación frente a la incertidumbre de distinta naturaleza que los de la Teoría de la Probabilidad y, por tanto, la primera no puede considerarse como un caso particular de la segunda.

Es obvio que, si sobre el resultado de un experimento no determinístico disponemos de una distribución de probabilidad, tendremos mejor información (mejor descrito el experimento) que si disponemos de una distribución de posibilidad sobre dicho resultado; lo cual no quiere decir que si estamos en el segundo caso, sea conveniente considerar las posibilidades como probabilidades y trabajar con ellas como tales.

Por último, hay que señalar que el enfoque empleado por Natvig es muy limitado, ya que solo considera el problema de calcular la posibilidad de un conjunto difuso, dada una distribución de posibilidad.

- Lindley (1.982) demuestra que si un individuo expresa su incertidumbre sobre cada suceso A por un número x y actúa racionalmente (lo que equivale a seguir un conjunto de axiomas), existe una transformación de los valores x en probabilidades. Como consecuencia afirma que " solo la probabilidad es una representación sensible de la incertidumbre ". Esta afirmación descarta el modelo posibilístico. En ese trabajo aparecen una serie de comentarios de Shafer y Zadeh, entre otros, que muestran su disconformidad con estas ideas.

Como comentario a dicho trabajo, nosotros señalaremos, en primer lugar, que cuando un individuo establece una distribución de posibilidad, ante una determinada información difusa, no esta siguiendo los axiomas de Lindley.

El mismo autor da un ejemplo que nos permite ilustrar esta - situacion. Se supone un geologo al que se le pide que exprese - su incertidumbre sobre la existencia de petroleo en un determinado punto. Entonces, segun la filosofia bayesiana, dicho geologo, tiene que dar un numero de incertidumbre que, teniendo en - cuenta la informacion de que disponga, no sea tan alto como para dar lugar a una perforacion, ni tan bajo como para que se deje el petroleo sin extraer. Pero, por ejemplo, si él no sabe nada sobre el pozo, no hay nada que le impida decir que la exis-- tencia de petroleo es posible con grado maximo, y necesaria con grado cero; lo que no lleva consigo aceptar ningun tipo de axioma. Simplemente esta representando su estado de ignorancia. Tal representacion nos parece mas exacta que el afirmar que existe- probabilidad 0.5 de que haya petroleo (esto mediria su incertidumbre cuando, despues de realizar diversos estudios y, en vista de las informaciones obtenidas, tenga tantas razones para -- pensar que hay petroleo, como para asegurar lo contrario).

2.- FORMULACION DEL PRINCIPIO POSIBILIDAD-PROBABILIDAD.

En lo que sigue, presentaremos nuestro punto de vista sobre- el Principio de Consistencia Posibilidad-Probabilidad. Los objetivos que vamos a tratar de cubrir, son los siguientes;

1) Precisar sin ambiguedades el contexto en el cual se plantea- el Principio de Consistencia.

En cualquiera de los trabajos anteriormente mencionados, se- considera una distribucion de posibilidad y una de probabilidad en un conjunto pero, no se sabe que representan dichas distribuciones. Por ejemplo, no se sabe si la distribucion de posibilidad se refiere a un resultado particular de un experimento aleatorio o, a todos los resultados del mismo. Sin embargo, creemos innecesaria la distincion de Dubois, Prade (1982) entre posibilidades "de re" y posibilidades "de dicto", ya que la diferen - cia entre ambas modalidades no esta en la naturaleza de las medidas propiamente dichas sino, en la fuente de donde se ha obtenido la informacion. Por ejemplo, si tenemos "La altura de Juan es X metros", la distribución "de re" sobre X recoge una información que está contenida en la propia proposición enunciada -- (no solo sabemos que X es un número real, sino que al ser X la altura en metros de una persona, es imposible que X = 0.02 ó -- que X = 3). La modalidad "de dicto", en cambio, es una información posibilística obtenida a partir del conocimiento de una -- distribución de probabilidad sobre los valores de X. Nosotros - vamos a desarrollar la consistencia para una distribución π independientemente de donde provenga la información contenida en ella.

2) Determinar qué se entiende por consistencia y justificar el interés de esta relación entre una distribución de posibilidad y una de probabilidad.

De acuerdo con Zadeh, la consistencia es la expresión matemática de que lo probable ha de ser posible (para Dubois, Prade, además, lo improbable ha de ser imposible) sin que se dé una definición más concreta de este concepto.

Por otra parte, en Zadeh (1.978) no se justifica que $\gamma = \sum_{a \in U} p(a)\pi(a)$ sea una medida de consistencia; ni en Dubois, Prade (1.982) que la condición (1) caracterice la consistencia.

3) Proponer una familia de medidas de consistencia.

Obviamente cabe cuestionarse si, una vez precisado el concepto de consistencia, se puede medir esta de otra forma, además - de las propuestas por Zadeh y Dubois, Prade. En este sentido, - vamos a proponer una axiomática que debe de cumplir una determinada aplicación para que ésta pueda considerarse una medida de consistencia.

2.1) Contexto.- Sea $(\Omega,\mathcal{A},P)$ un espacio probabilístico, U un conjunto finito y X una aplicación medible (variable aleatoria si U es numérico)

$$X:(\Omega,\mathcal{A}) \longrightarrow (U,\mathcal{P}(U))$$

Esta aplicación induce una medida de probabilidad en $(U,\mathcal{P}(U))$

que notaremos también P : $P(A) = P(X^{-1}(A))$, si $A \subset U$.

Como es habitual, supondrenos que $P:\mathcal{P}(U) \longrightarrow [0,1]$ es conocida, sin que ello implique la necesidad de conocer exactamente la aplicación X. Esto es lo que ocurre cuando decimos "sea X -- una variable aleatoria que sigue una distribución binomial".

Si disponemos de una distribución de posibilidad en U sobre X, esta puede referirse a dos cosas:

a) A un valor observable particular del experimento aleatorio. Por ejemplo la que proviene de una información difusa de - la forma " $X(\omega)$ es A ", con $A \in \mathcal{P}(U)$, $\omega \in \Omega$ arbitrario, pero fijo.

b) A todos los valores observables del experimento aleatorio. Por ejemplo, la que proviene de una información difusa sobre -- los valores de X de la forma: "$\forall\omega \in \Omega$, $X(\omega)$ es A " , con A subconjunto difuso de U.

Vamos a estudiar estos dos casos por separado.

a) Distribución de posibilidad sobre un resultado particular. En este caso no tiene sentido ni la noción de consistencia considerada por Zadeh ni la considerada por Dubois, Prade. Por --- ejemplo, si tenemos el experimento de lanzar un dado con U = -- $\{1,2,3,4,5,6\}$ y $p(a) = 1/6, \forall a \in U$, sobre un lanzamiento particular nos pueden decir que ha salido el 2; y nos habrán proporcionado una información que induce sobre U la distribución de posibilidad

$$\pi(a) = \begin{cases} 1 & \text{si } a = 2 \\ 0 & \text{si } a \neq 2 \end{cases}$$

No es contradictorio que $p(3) = 1/6 > 0$ y, sin embargo, según la distribución π el 3 sea "imposible"; ya que π no se refiere a un lanzamiento arbitrario, sino a un resultado particular.

Esta situación está estudiada en los trabajos de E. Takeda - et al. (1975, 1.976). A continuación, resumimos las conclusio--

nes más importantes.

Se define un mensaje difuso sobre un experimento aleatorio - como una expresión de la forma p_A: " A ha ocurrido " ($A \in \underset{\sim}{\mathcal{F}}(U)$) o, equivalentemente, " $X(\omega)$ pertenece a A ". Si disponemos de - la distribución de probabilidad con que X toma los valores en U $p(a) = P[X(\omega) = a]$, $a \in U$; entonces, de acuerdo con la defini-- ción de Zadeh (Zadeh (1.968)), la distribución de probabilidad de X, condicionada al mensaje difuso p_A es:

$$p(a/A) = \frac{\mu_A(a)p(a)}{\sum_{b \in U} \mu_A(b)p(b)} \quad .$$

Siguiendo a Theil (1.967), que considera que la cantidad de información de un mensaje que transforma las probabilidades "a priori" $\{p(a)\}_{a \in U}$ en las "a posteriori" $\{q(a)\}_{a \in U}$, vale

$$\sum_{a \in U} q(a)L(q(a)/p(a)),$$

se define la cantidad de información de un mensaje difuso de la forma p_A: " $X(\omega)$ pertenece a A ", como

$$I(A) = \sum_{a \in U}\left[\frac{p(a)\mu_A(a)}{\sum_{b \in U} p(b)\mu_A(b)} L\left(\frac{\mu_A(a)}{\sum_{b \in U} p(b)\mu_A(b)}\right)\right]$$

$$= (1/P(A)) \sum_{a \in U} p(a)\mu_A(a)L(\mu_A(a)) + L(1/P(A))$$

$$\text{con } P(A) = \sum_{a \in U} p(a)\mu_A(a) > 0$$

Analogamente, se definen las probabilidades condicionadas a la distribución de posibilidad π:

$$p(a/\pi) = \frac{\pi(a)p(a)}{\sum_{b \in U} p(b)\pi(b)} \tag{10}$$

y la cantidad de información aportada por la distribución de po̲sibilidad π:

$$I(\pi) = \sum_{a \in U}\left[\frac{\pi(a)p(a)}{\sum_{b\in U}\pi(b)p(b)} L\left(\frac{\pi(a)}{\sum_{b\in U} p(b)\pi(b)}\right)\right]$$

Nota 1.- La expresión (10) es coherente con la regla de ---- Dempster de combinar la evidencia: si m_π y m_p son las asignaciones básicas asociadas a π y p, entonces la combinación de ambas según la regla de Dempster, da lugar a una evidencia de tipo -- probabilístico $m_{p(./\pi)}$ con distribución de probabilidad $p(./\pi)$ que cumple (10).

Nota 2.- Okuda et al. (1.974) han estudiado otro concepto de información difusa y de cantidad de información difusa, distinto del anteriormente presentado y que se formula en un contexto distinto, aunque, en ciertos aspectos, más general del de - Takeda et al. (1.976). No lo recogeremos aquí ya que se sale de la línea general de este trabajo.

b) Distribución de probabilidad sobre todos los resultados - del experimento aleatorio.- El conjunto de todos los valores observables del experimento aleatorio es Im(X) y, por tanto, en - este caso, tendremos distribuciones de posibilidad sobre Im(X), que no son otra cosa que un caso particular de las que hemos estudiado en el punto cinco del capítulo anterior, para la imagen de una aplicación genérica.

Así pues, dichas distribuciones estarán asociadas a informaciones difusas de la forma "$\forall\omega \in \Omega$, $X(\omega)$ es A ", con $A \in \tilde{\mathcal{P}}(U)$, y se podrán considerar a dos niveles:

- A nivel global se tiene una distribución de posibilidad $\bar{\pi}$ en $\mathcal{P}(U)$. Si $F \subset U$, $\bar{\pi}(F)$ se interpreta como la posibilidad de que Im(X) = F.

- A nivel local se tiene una distribución de posibilidad, π, en U. Si $a \in U$, $\pi(a)$ se interpreta como la posibilidad de -- que $a \in$ Im(X) y se nota $\pi(X = a)$.π es una distribución de posibilidad válida para cada uno de los resultados del experimento aleatorio.

Entre ambas distribuciones ha de existir la siguiente relación

$$\pi(F) = \min_{a \in F} \pi(a) \quad , \forall F \subset U.$$

Se puede observar como, en todos los ejemplos que aparecen en las distintas referencias, las distribuciones de posibilidad que se consideran en U son de este tipo (a nivel local), aunque esto no se mencione explicitamente. En el ejemplo de Zadeh, cuando se dice que la posibilidad de que Hans coma nueve huevos es cero, se está diciendo que nueve no puede ser el resultado de X: " número de huevos que Hans se come para desayunar ", es decir, que no puede ser elemento de Im(X). Cuando se dice que las posibilidades $\pi(1) = \pi(2) = \pi(3) = \pi(4) = 1$, se afirma que 1, 2, 3 y 4 son posibles resultados del experimento, pueden estar en Im(X).

En este contexto es en el que vamos a definir y justificar la consistencia posteriormente.

2.2) <u>Concepto de consistencia.-</u> Sea $X:\Omega \longrightarrow U$ una aplicación medible, entre el espacio probabilístico $(\Omega,\mathcal{A},P)$ y el medible $(U,\mathcal{P}(U))$ y supongamos que conocemos la distribución de probabilidad p:

$$p(a) = P[X = a] \quad , a \in U.$$

Tenemos, entonces, lo que llamaremos una información probabilística sobre los resultados del experimento aleatorio.

Supongamos que, también, disponemos de una información posibilística sobre todos los resultados del experimento aleatorio, que se traduce en una distribución π en U, donde $\pi(a) = \pi(X = a)$ es la posibilidad de que $a \in Im(X)$.

El siguiente ejemplo muestra que puede existir contradicción entre dos informaciones de este tipo. Observemos que es un caso extremo, pero nos obliga a cuestionarnos la coherencia, compatibilidad o contradicción entre posibilidad y probabilidad.

<u>Ejemplo 3.2.-</u> Sean $U = \{1,2,3,4\}$ y π y p dos distribuciones de posibilidad y probabilidad, respectivamente, dadas por

$\pi(1) = \pi(2) = 1$, $\pi(3) = \pi(4) = 0$

$p(1) = p(2) = 0$, $p(3) = p(4) = 1/2$.

Si $\bar{\pi}$ es la distribución de posibilidad en $\mathcal{P}(U)$ asociada a π, entonces los conjuntos $F \subset U$ con posibilidad mayor que cero son $\{1\}$, $\{2\}$ y $\{1,2\}$ y, según la interpretación de $\bar{\pi}$, $Im(X)$ solo -- puede ser uno de ellos. Pero esto es contradictorio con la distribución de posibilidad que tenemos ya que ha de ser $P(Im(X)) = P(\Omega) = 1$, y sin embargo, $P(\{1\}) = P(\{2\}) = P(\{1,2\}) = 0$.

Como consecuencia de lo anterior, definimos la consistencia posibilidad-probabilidad como la coherencia, la no contradicción, que exista entre dos informaciones sobre los valores observables de un experimento aleatorio: una de tipo probabilístico (p) y otra de tipo posibilístico (π).

Esta coherencia o falta de contradicción es una propiedad difusa y, como tal, vendrá determinada por una medida que a cada par de distribuciones π y p les asocie un grado de consistencia. En lo que sigue, vamos a determinar distintas medidas de este - tipo, cada una de las cuales medirá la consistencia desde distintos puntos de vista.

2.3) Medidas de consistencia.- En primer lugar, vamos a obtener valoraciones concretas, justificando por qué pueden considerarse como medidas de consistencia. Basándonos en las propiedades comunes de estas medidas, obtendremos un conjunto de axiomas -- que nos servirán para definir lo que entenderemos por medidas - de consistencia.

- Sean π y p dos distribuciones en U y sea $\bar{\pi}$ la distribución de posibilidad en $\mathcal{P}(U)$ asociada a π. $\bar{\pi}(F) = \operatorname{Min}_{a \in F} \pi(a)$ se considera como la posibilidad de que $F = Im(X)$ y sabemos que $P(Im(X)) = 1$. Por lo tanto, para que no exista ningún tipo de - contradicción entre ambas informaciones, debería de existir un $F \subset U$ tal que $\bar{\pi}(F) = P(F) = 1$ (es decir, $p(a) > 0 \Rightarrow \pi(a)=1$).

Podemos considerar que este caso extremo tiene consistencia uno e inspirándonos en él, medir la consistencia por el grado con que se cumpla la proposición: "existe un $F \subset U$ con posibilidad, $\pi(F)$, y probabilidad, $P(F)$, próximas a uno.

Esta propiedad recoge la idea intuitiva de Zadeh de que si resultado es probable ha de ser posible. En efecto, si $b \in U$ es tal que $p(b)$ es alta, entonces b debe pertenecer a todo $F \subset U$ con $P(F)$ próximo a uno. Si, a la vez, $\pi(b)$ es próximo a cero, entonces para todo F con probabilidad alta $\pi(F) = \operatorname{Min}_{a \in F} \pi(a) \leqq$

$\leqq \pi(b)$ próximo a cero, y no habría consistencia.

El grado con que $\pi(F)$ $(P(F))$ es próximo a uno será una función creciente de $\pi(F)$ $(P(F))$ y, si consideramos que el conectivo "y" de la propiedad de consistencia enunciada viene definido por una norma triangular $\perp$, obtenemos una familia de medidas de consistencia que son las aplicaciones

$$C_{\perp} : \prod(U) \times \mathcal{P}(U) \longrightarrow [0,1]$$

(donde $\prod(U)$ es el conjunto de las distribuciones de posibilidad sobre U y $\mathcal{P}(U)$ el de las de probabilidad) con

$$C_{\perp}(\pi,p) = \operatorname*{Max}_{F \subset U} \left(f(P(A)) \perp g(\operatorname*{Min}_{a \in F} \pi(a)) \right) \tag{11}$$

y $f, g: [0,1] \longrightarrow [0,1]$ son monótonas crecientes con $f(0) = g(0) = 0$ y $f(1) = g(1) = 1$.

Recordando la expresión (2) del primer capítulo para la integral seminormada de F. Suárez, obtenemos (teniendo en cuenta que toda norma es una seminorma):

$$C_{\perp}(\pi,p) = N\text{-}\int_{\perp} g(\pi) \circ f(P),$$

habiendo notado $g(\pi)$ $(f(P))$ la composición de g y π (de f y P).

En el caso particular en que la norma $\perp$ es la norma del mínimo, $\wedge$, se llega a la integral de Sugeno:

$$C_{\wedge}(\pi,p) = ⨍ g(\pi) \circ f(P)$$

-Como medidas de consistencia, encierta forma, duales de las-anteriores, podemos considerar las basadas en la integral semiconormada de F. Suárez.

$$C_{\otimes}(\pi,p) = C-\int_{\otimes} g(\pi)\circ f(P) \tag{12}$$

donde $\otimes$ es una conorma triangular y f y g como en (11).

-Si π es una distribución de posibilidad para todos los valores de Im(X), entonces $\pi\circ X$ es una variable aleatoria que toma - valores en [0,1]. Para $\omega \in \Omega$, $\pi(X(\omega))$ representa la posibilidad de que $X(\omega) \in Im(X)$; y como sabemos que $X(\omega) \in Im(X), \forall\omega \in \Omega$, entonces $\pi(X(\omega))$ ha de ser alta, para todo ω,si la información dada por π es coherente con Im(X).Una predicción del valor de - $\pi\circ X$, utilizando la distribución p, podrá considerarse como una medida de la consistencia entre π y p.

Conocida p, la distribución de probabilidad $p\pi$ con que $\pi\circ X$ - toma sus valores en [0,1] será:

$$p\pi(\alpha) = \sum_{\substack{a \in U \\ \pi(a)=\alpha}} p(a).$$

Con ella y empleando el método bayesiano, puede obtenerse -- una predicción del tipo anteriormente citado.

Si $L(\alpha,\alpha')$ es una función de pérdida en $[0,1] \times [0,1]$convexa y creciente en $|\alpha - \alpha'|$ (ver De Groot (1.970)), entonces una medida de consistencia (predicción óptima) vendrá dada por

$$C_L(\pi,p) = \alpha^* \quad \Leftrightarrow \quad E[L(\alpha^*,\pi\circ X)] = \inf_{\alpha\in[0,1]} (E[L(\alpha,\pi\circ X)]).$$

Si L es una pérdida cuadrática, obtenemos como predicción óptima

$$C_Z(\pi,p) = E[\pi\circ X] = \sum_{a \in U} p(a)\pi(a), \tag{13}$$

que es la medida de consistencia de Zadeh (1.978). En efecto, según indicamos en el primer capítulo, la contradicción entre - dos asignaciones básicas m_1 y m_2 se puede medir mediante una --

función creciente de

$$\kappa = \sum_{A_i \cap B_j = \phi} m_1(A_i).m_2(B_j)$$

donde A_i son los elementos focales de m_1 y B_j los de m_2.

Si aplicamos esto a las asignaciones básicas m_π y m_p, asociadas a π y a p, resulta ser

$$\kappa = \sum_{A_i \cap B_j = \phi} m_p(A_i).m_\pi(B_j) = 1 - \sum_{a \in U} p(a)\pi(a)$$

y, por tanto, si la contradicción se mide mediante una función creciente de $1 - \sum_{a \in U} p(a)\pi(a)$, la coherencia se puede medir - mediante una función creciente de

$$C_Z(\pi,p) = 1 - \kappa = \sum_{a \in U} p(a)\pi(a)$$

En el caso particular de que $A \in \tilde{\mathcal{P}}(U)$, con $C(A) = 1$ y π_A sea la distribución de posibilidad en U, asociada a la información difusa p_A: "$\forall\omega \in \Omega$, $X(\omega)$ es A ", entonces

$$C_Z(\pi_A,p) = P(A) = \sum_{a \in U} p(a)\pi_A(a)$$

donde P(A) es la probabilidad del subconjunto difuso A, según - la definición de Zadeh (1.968).

Si $L(\alpha,\alpha') = |\alpha - \alpha'|$ obtenemos como medida de consistencia la mediana de la variable aleatoria $\pi \circ X$ (ver De Groot (1.970))

- La condición de consistencia de Dubois, Prade, que justificamos anteriormente utilizando la teoría de la evidencia de Shafer, puede expresarse como una aplicación de $\prod(U) \times P(U)$ en -- [0,1] de la siguiente forma

$$C_{DP}(\pi,p) = \begin{cases} 1 & \text{si } \Pi(A) \geq P(A),\ \forall A \subset U \\ 0 & \text{en otro caso} \end{cases} \qquad (14)$$

donde $\Pi(A) = \max_{a \in A} \pi(a)$ es la medida de probabilidad asociada a π.

Así pues, la condición de consistencia de Dubois, Prade puede identificarse con una medida de consistencia de tipo "crisp".

- Considerando las propiedades comunes de todas las valoraciones anteriormente presentadas y, teniendo en cuenta que una medida de consistencia ha de medir el grado de coherencia de dos informaciones, hemos seleccionado tres propiedades como axiomas que ha de cumplir una tal medida.

Definición 3.4.- Una aplicación

$$C:\prod(U) \times P(U) \longrightarrow [0,1]$$

se dice que es una medida de consistencia si y solo si cumple los tres axiomas siguientes

A1) $\pi \subset \pi' \Rightarrow C(\pi,p) \geq C(\pi',p), \forall \pi,\pi' \in \prod(U), \forall p \in P(U)$.

A2) Si $\pi \in \prod(U)$, $p,p' \in P(U)$, de tal forma que existen $a,b \in U$ tales que $p(c) = p'(c)$ si $c \in U$, $c \neq a$ y $c \neq b$; entonces

$$\left.\begin{array}{l}\pi(a) \geq \pi(b) \\ p(a) \geq p'(b)\end{array}\right\} \Rightarrow \quad C(\pi,p) \geq C(\pi,p').$$

A3) Si $p \in P(U)$ y $\pi \in \prod(U)$, entonces

$$(p(a) > 0 \Rightarrow \pi(a) = 0) \quad \Rightarrow \quad C(\pi,p)= 0$$
$$(p(a) > 0 \Rightarrow \pi(a) = 1) \quad \Rightarrow \quad C(\pi,p)= 1$$

Justificación de los axiomas

A1) Si $\pi \subset \pi'$, entonces π es menos restrictiva que $\pi'(\pi(a) \geq \pi'(a))$ y, por tanto, al informar "menos", debe ser menos contradictoria con cualquier otra información.

Es inmediato comprobar que todas las valoraciones anteriormente consideradas verifican A1).

A2) Si consideramos una única información posibilística, π, en U y dos probabilísticas, p y p', que coinciden en todos los elementos de U excepto en a y b, ha de ser $p(a) + p(b) = p'(a) + p'(b)$; es decir las dos distribuciones "reparten" la misma pro babilidad en $\{a,b\}$. Si p asigna más probabilidad que p' al elemento de mayor posibilidad, entonces debe de haber menos contradicción entre p y π que entre p' y π.

Las medidas $C_\perp$ verifican A2). En efecto:

- Si $F \subset U$ y $a,b \in F$ o $a,b \notin F$, entonces $P(F) = P'(F)$, de -- donde, $\overline{\pi}(F) \perp P(F) = \overline{\pi}(F) \perp P'(F)$.

-Si $F \subset U$ y ($a \in F$ y $b \notin F$), entonces se cumple $P(F) \geqq P'(F)$ y como consecuencia, $\overline{\pi}(F) \perp P(F) \geqq \overline{\pi}(F) \perp P'(F)$.

-Si $F \subset U$ y ($b \in F$ y $a \notin F$), sea $F^*=F \cup \{a\} - \{b\}$. Entonces $\overline{\pi}(F^*) \geqq \overline{\pi}(F)$ y $P(F^*) \geqq P'(F^*) \geqq P'(F)$ (ya que $p'(a) \geqq p'(b)$) y, por tanto, $\overline{\pi}(F^*) \perp P(F^*) \geqq \overline{\pi}(F) \perp P'(F)$.

Así pues, para todo $F \subset U$ existe $F^* \subset U$ ($F^* = F$ en los dos - primeros casos) con $\overline{\pi}(F) \perp P'(F) \leqq \overline{\pi}(F^*) \perp P(F^*)$, luego

$$\operatorname*{Sup}_{F \subset U} (\overline{\pi}(F) \perp P(F)) \geqq \operatorname*{Sup}_{F \subset U} (\overline{\pi}(F) \perp P'(F)) ,$$

es decir, $C_\perp(\pi,p) \geqq C_\perp(\pi,p')$, como queríamos probar.

Las consistencia del tipo C_L verifican A2). En efecto:

Si $L(\alpha,\alpha') = g(|\alpha - \alpha'|)$ con g creciente y convexa, entonces

$E_p[L(\alpha,\pi\circ X)] = E_{p'}[L(\alpha,\pi\circ X)] + (p(a) + p'(a))(g(|\alpha-\pi(a)|) -$

$- g(|\alpha-\pi(b)|))$.

Por otra parte, bajo las mismas hipótesis, se puede demos--- trar que $h(\alpha) = g(|\alpha-\pi(a)|) - g(|\alpha-\pi(b)|)$ es decreciente en --- $[0,1]$. Por tanto, para todo $\alpha \leqq \alpha^*_{p'} = C_L(\pi,p')$, se cumple

$$E_p[L(\alpha,\pi\circ X)] \geqq E_p[L(\alpha^*_{p'},\pi\circ X)] .$$

Ahora bien, si $\alpha^*_p = C_L(\pi,p)$, es decir $E_p[L(\alpha^*_p,\pi\circ X)] =$ $= \operatorname*{Inf}_{\alpha\in[0,1]} \{E_p[L(\alpha,\pi\circ X)]\}$, tiene que ser $\alpha^*_p \geqq \alpha^*_{p'}$, como queríamos demostrar.

A3) El primer caso de este axioma ($p(a) > 0 \Rightarrow \pi(a) = 0$)-

es, precisamente, el del ejemplo 3. y no vamos a insistir más sobre él. El segundo caso también lo hemos considerado anteriormente: si todos los elementos de probabilidad positiva son posibles congrado uno, no existe ninguna contradcción entre las dos informaciones.

Para las medidas C y C_L, puede encontarse $F \subset U$ con $\pi(F) = P(F) = 1$ y $\pi \circ X$ tiene una distribución de probabilidad degenerada en el uno.

<u>Nota.-</u> Es inmediato comprobar que las medidas $C_\otimes$ y C_{DP} cumplen A1), A2) y A3).

<u>Comparación entre C_Z y $C_\wedge$</u>.- Vamos a estudiar si existe alguna relación entre las medidas de consistencia C_Z y $C_\wedge$, que destacamos como casos particulares más importantes de C_L y $C_\wedge$.

Tenemos que $C_Z(\pi,p) = \sum_{a \in U} p(a)\pi(a) = \int_U \pi dP$

$$C_\wedge(\pi,p) = \int_U \pi \circ P$$

Por tanto (por un resultado debido a Sugeno (1.974)) se tiene

$$\forall \pi \in \prod(U), \forall p \in \mathcal{P}(U) \quad , \quad |C_Z(\pi,p) - C_\wedge(\pi,p)| \leqq 1/4$$

cota que proviene de la desigualdad:

$$|C_Z(\pi,p) - C_\wedge(\pi,p)| \leqq C_\wedge(\pi,p)(1-C_\wedge(\pi,p))$$

$C_\wedge(1-C_\wedge)$ alcanza su máximo cuando $C_\wedge(\pi,p) = 1/2$, y este máximo es igual a $1/4$, que es la cota anterior.

Esta cota nos indica que si $C_\wedge \longrightarrow 0$ (o a 1) entonces --- $C_\wedge \simeq C_Z$. Analogamente si $C_Z \longrightarrow 0$ (o a 1) entonces $C_Z \simeq C_\wedge$.

$C_\wedge$ es menos sensible que C_Z a las fluctuaciones de p en las colas (recordemos que la integral de Sugeno es una especie de - mediana). En Wierzchon (1.982) se estudia detenidamente este aspecto de la integral difusa. Esto queda de manifiesto en el siguiente ejemplo.

Ejemplo 3.4.- Sea $U = \{a_1, a_2, a_3\}$. Vamos a considerar en U la distribución de posibilidad π, dada por

$$\pi(a_1) = 0 \qquad \pi(a_2) = 0.5 \qquad \pi(a_3) = 0$$

y las dos distribuciones de probabilidad

$$p(a_1) = 0.5 \qquad p(a_2) = 0.5 \qquad p(a_3) = 0$$

$$p'(a_1) = 0 \qquad p'(a_2) = 0.5 \qquad p'(a_3) = 0.5$$

Intuitivamente p' debe de ser más compatible con π que p, y esto queda de manifiesto utilizando C_Z ($C_Z(\pi,p) = 0.25$ y ----- $C_Z(\pi,p') = 0.75$). Sin embargo, $C_\wedge(\pi,p) = C_\wedge(\pi,p') = 0.5$.

Además de las estudiadas, sería interesante investigar otras medidas de consistencia. Los trabajos de De Lucca, Termini ---- (1.979) y Trillas, Riera (1.978) sobre energías y entropías en conjuntos difusos, pueden ser de gran utilidad. Las medidas de energía del conjunto difuso de función de pertenencia π, ponderadas por la distribución p, bajo ciertas condiciones, son medidas de consistencia. Entre las que se de esta forma se encuentra la medida de Zadeh, C_Z.

Si $\mathcal{C}(U)$ es el conjunto de las medidas de consistencia en --- $\prod(U) \times P(U)$, las siguientes propiedades nos permiten generar -- nuevas medidas, a partir de las conocidas.

1.- Convexidad.- $\lambda \in [0,1]$, C_1 y $C_2 \in \mathcal{C}(U) \Rightarrow$
$\Rightarrow \lambda C_1 + (1-\lambda)C_2 \in \mathcal{C}(U)$.

2.- Multiplicatividad.- $\forall C_1, C_2 \in \mathcal{C}(U)$, $C_1 . C_2 \in \mathcal{C}(U)$.

3.- Si $u:[0,1] \longrightarrow [0,1]$ es una función monótona no decreciente con $u(0) = 0$ y $u(1) = 1$, entonces

$$C \in \mathcal{C}(U) \quad \Rightarrow \quad u_\circ C \in \mathcal{C}(U)$$

4.- $\forall C_1, C_2 \in \mathcal{C}(U)$, $\mathrm{Min}(C_1,C_2)$ y $\mathrm{Max}(C_1,C_2) \in \mathcal{C}(U)$.

La demostración de estas propiedades es inmediata.

3.-ASIGNACION DE UNA DISTRIBUCION DE PROBABILIDAD A UNA DISTRIBUCION DE POSIBILIDAD

Este problema ha sido planteado por Y. Leung (1.980). Su trabajo está basado en el principio de máxima entropía (Jaynes --- (1.957) y Thomas (1.979)): si sobre un experimento aleatorio X, que toma valores en un conjunto U, no se dispone de una distribución de probabilidad "a priori" , entonces la forma "más objetiva" de obtener una es considerar aquella que maximiza la entropía, es decir la solución del problema

$$\text{Max} \quad H: -k \sum_{a \in U} p(a)L(p(a))$$

$$\text{s.a.} \quad \sum_{a \in U} p(a) = 1$$

Si disponemos de alguna otra información adicional; por ejemplo, (siendo $U \subset R$) conocemos n momentos

$$E[f_r(X)] = \sum_{a \in U} p(a)f_r(a) = E_r \quad , r=1,..,n$$

la distribución a considerar sería aquella que resolviese:

$$\text{Max} \quad H: -k \sum_{a \in U} p(a)L(p(a))$$

$$\text{s.a.} \quad \sum_{a \in U} p(a) = 1$$

$$\sum_{a \in U} p(a)f_r(a) = E_r, \; r=1,..,n.$$

En general, si sabemos que p cumple una determinada propie-- dad ($p \in \mathcal{F} \subset \mathcal{P}(U)$) , la distribución "a priori" se ha de obtener como solución del problema

$$\text{Max} \quad H: -k \sum_{a \in U} p(a)L(p(a))$$

$$\text{s.a.} \quad p \in \mathcal{F} \; .$$

Si tenemos una información difusa sobre X, es decir, disponemos de una distribución de posibilidad π,sobre los resultados - del experimento aleatorio X, Y. Leung propone (siguiendo la filosofía del principio de máxima entropía) considerar aquella -- distribución "a priori" que maximice la entropía:

$$\text{Max} \quad H: -k \sum_{a \in U} p(a)L(p(a))$$

$$\text{s.a.} \quad \sum_{a \in U} p(a) = 1 \qquad (15)$$

$$\sum_{a \in U} p(a)\pi(a) = \alpha$$

con $\alpha \in [0,1]$, suficientemente grande.

Leung justifica esta última restricción ($\sum_{a \in U} p(a)\pi(a) = \alpha$) haciendo referencia al principio de consistencia posibilidad--- probabilidad de Zadeh.

En el citado artículo, el problema (15) se resuelve utilizando la función lagrangiana

$$L = -k \sum_{a \in U} p(a)L(p(a)) + \lambda_0(1- \sum_{a \in U} p(a)) +$$

$$+ \lambda_1(\alpha - \sum_{a \in U} p(a)\pi(a)).$$

Como solución del sistema:

$$\frac{\partial L}{\partial p(a)} = 0 \quad , \sum_{a \in U} p(a) = 1 \ , \ \sum_{a \in U} p(a)\pi(a) = \alpha,$$

Leung obtiene

$$p(a) =[\exp(-(\lambda_1/k)\lambda(a))].(z(\lambda_1))^{-1}$$

donde $z(\lambda) = \sum_{a \in U} \exp(-(\lambda/k)\pi(a))$

y λ_0 y λ_1 se obtienen como solución de

$$\alpha = - \frac{\partial}{\partial \lambda_1} L(z(\lambda_1))$$

$$\lambda_0 = k(L(z(\lambda_1)) - 1)$$

Vamos a tratar de expresar la solución de este problema de -

este problema de forma más explícita.

Tenemos que $\frac{\partial}{\partial p(a)} = -k(1 + L(p(a))) - \lambda_0 - \lambda_1 \pi(a) = 0,$

y, despejando p(a), obtenemos

$p(a) = \exp(\lambda_0/k + (\lambda_1 \pi(a))/k - 1).$

Si notamos

$$v = \exp((\lambda_0/k) - 1) > 0$$

$$u = \exp(\lambda_1/k) > 0 \ ,$$

p(a) se expresa como

$$p(a) = v.u^{\pi(a)} \ , \text{ con } u > 0,\ v > 0.$$

Puesto que, $\sum_{a \in U} p(a) = 1$, obtenemos

$$v = \frac{1}{\sum_{a\in U} u^{\pi(a)}} \qquad \text{y} \quad p(a) = \frac{u^{\pi(a)}}{\sum_{b\in U} u^{\pi(b)}}$$

donde u se calcula como solución de la ecuación

$$\alpha = \sum_{a \in U} p(a)\pi(a) = \frac{\sum_{a\in U} \pi(a)u^{\pi(a)}}{\sum_{a\in U} u^{\pi(a)}} \tag{16}$$

con $\{\pi(a)\}_{a\in U}$ conocidos.

Si consideramos $f: R^+ \longrightarrow [0,1]$, dada por

$$f(u) = \frac{\sum_{a\in U} \pi(a)u^{\pi(a)}}{\sum_{a\in U} u^{\pi(a)}}$$

entonces, f es una función continua y derivable, con derivada

$$f'(u) = \frac{\left[\sum_{a\in U} \pi(a)^2 u^{\pi(a)-1}\right]\left[\sum_{a\in U} u^{\pi(a)}\right]}{\left[\sum_{a \in U} u^{\pi(a)}\right]^2} \quad -$$

$$- \frac{\left[\sum_{a\in U} \pi(a)u^{\pi(a)}\right]\left[\sum_{a\in U} \pi(a)u^{\pi(a)-1}\right]}{\left[\sum_{a\in U} u^{\pi(a)}\right]^2} =$$

$$= \frac{\sum_{\substack{a,b\in U\\ \pi(a)\neq\pi(b)}} \pi(a)(\pi(a)-\pi(b))u^{\pi(a)+\pi(b)-1}}{\left[\sum_{a\in U} u^{\pi(a)}\right]^2} =$$

$$= \frac{\sum_{\substack{a,b\in U\\ \pi(a)<\pi(b)}} (\pi(a) - \pi(b))^2\, u^{\pi(a)+\pi(b)-1}}{\left[\sum_{a\in U} u^{\ (a)}\right]^2} \geq 0$$

Luego f es una función monótona no decreciente y será estrictamente creciente en R^+ si existen $a,b \in U$ con $\pi(a) \neq \pi(b)$.

Es fácil probar que:

- $\lim_{u\to 0} f(u) = \lim_{u\to 0} \frac{\sum_{a\in U} \pi(a)u^{\pi(a)}}{\sum_{a\in U} u^{\pi(a)}} = \min_{a\in U} \{\pi(a)\}$

- $\lim_{u\to +\infty} f(u) = \lim_{u\to +\infty} \frac{\sum_{a\in U} \pi(a)u^{\pi(a)}}{\sum_{a\in U} u^{\pi(a)}} = \max_{a\in U} \{\pi(a)\} = 1$

- $f(1) = \left[\sum_{a\in U} \pi(a)\right] / |U|$

en estas condiciones, la ecuación (16) se expresa como:

$$f(u) = \alpha$$

La existencia de solución para la misma se garantiza, mediante el siguiente resultado.

Proposición 3.6.- Si $\alpha \in [0,1]$ y π es una distribución de posibilidad en U con $1 \neq \min_{a\in U} \{\pi(a)\} \leq \alpha \leq 1$; entonces existe un único $u \in R^+$ tal que $f(u) = \alpha$, es decir $u = f^{-1}(\alpha)$.

Demostración.- Es inmediata, a partir de las propiedades de la función f.

Por otra parte, si no existen $a,b \in U$ tales que $\pi(a) \neq \pi(b)$, tiene que ser $\pi(a) = 1, \forall a \in U$, y, en este caso el problema solo tiene solución para $\alpha = 1$. Dicha solución (de acuerdo con el principio de máxima entropía) es la distribución uniforme en U.

Por tanto, la solución del problema (15) viene dada por:

- $p(a) = \dfrac{u^{\pi(a)}}{\sum_{a \in U} u^{\pi(b)}}$, donde $u = f^{-1}(\alpha)$, si $\underset{a \in U}{\operatorname{Min}} \pi(a) \leq \alpha \leq 1$ (18)

- $p(a) = 1/|U|$, si $\forall a \in U, \pi(a) = 1 = \alpha$.

- No existe solución en otro caso.

La solución (18) presenta un problema cuando $\alpha < \dfrac{\sum_{a \in U} \pi(a)}{|U|}$:

a) Si $\pi(a) = 1, \forall a \in U$, entonces (15) no tiene solución.

b) Si existen $a,b \in U$ con $\pi(a) \neq \pi(b)$, f será estrictamente creciente; y como consecuencia, f^{-1} también lo será. Como se cumple $f^{-1}(\sum_{a \in U} \pi(a)/|U|) = 1$, entonces, si $\alpha < \sum_{a \in U} \pi(a)/|U|$, es $u = f^{-1}(\alpha) < 1$. Para este $u < 1$, la solución (18) cumple

$$\pi(a) < \pi(b) \quad \Rightarrow \quad p(a) > p(b) \tag{19}$$

lo cual es poco coherente, ya que le damos más probabilidad a los elementos menos posibles. Esto sucede porque, en este caso, la distribución de máxima entropía sin restricciones ($p_m(a) = 1/|U|, \forall a \in U$) es tal que $\sum_{a \in U} p_m(a)\pi(a) \geq \alpha$, es decir su sonsistencia con π es mayor que α.

Este problema se evita si sustituímos la restricción ------ $\sum_{a \in U} p(a)\pi(a) = \alpha$, por $\sum_{a \in U} p(a)\pi(a) \geq \alpha$.

La solución al problema modificado:

$$\begin{aligned} &\text{Max} \quad H: -k \sum_{a\in U} p(a)L(p(a)) \\ &\text{s.a.} \quad \sum_{a\in U} p(a) = 1 \\ &\qquad\quad \sum_{a\in U} p(a)\pi(a) \geqq \alpha \end{aligned} \tag{20}$$

coincide con la de (15) si $\alpha \geqq \sum_{a\in U} \pi(a)/|U|$.

Ahora bien, para $\alpha < \sum_{a\in U} \pi(a) /|U|$ la solución a (20) es

$$p(a) = 1/|U| \quad , \forall a \in U,$$

solución válida incluso cuando $\alpha < \dfrac{\text{Min } \pi(a)}{|U|}$,
lo cual parece más coherente.

Este problema admite una nueva formulación en base a nuestro estudio del principio de consistencia posibilidad-probabilidad.

Si conocemos una distribución de posibilidad, π, sobre los resultados de un experimento aleatorio, de acuerdo con la filosofía del principio de máxima entropía la forma más objetiva de asignar probabilidades "a priori" es considerar la solución de

$$\begin{aligned} &\text{Max} \quad H: -k \sum_{a\in U} p(a)L(p(a)) \\ &\text{s.a.} \quad \sum_{a\in U} p(a) =1 \\ &\qquad\quad \text{"p es consistente con } \pi\text{"} \end{aligned} \tag{21}$$

La última restricción es, obviamente, difusa y es equivalente a $p \in A(\pi)$, donde $A(\pi)$ es el subconjuto difuso de $P(U)$ formado por las distribuciones de probabilidad consistentes con π y que estará determinado por su función de pertenencia. Fijada -- una medida de consistencia C, consideraremos que dicha función de pertenencia tiene la forma

$$\mu_{A(\pi)}(p) = C(\pi,p),$$

obteniendo una restricción difusa distinta, según sea la medida C considerada. Al conjunto difuso asociado lo notaremos $A_C(\pi)$.

Una vez determinada $\mu_{A(\pi)}(.)$, resolveremos el problema (21) por el método propuesto por J.L. Verdegay (1980, 1982). Según - dicho método, la solución será un subconjunto difuso $S \in \tilde{\mathcal{F}}(P(U))$ que se obtiene de la siguiente forma:

Para cada $\alpha \in [0,1]$ se considera el α-corte del conjunto de las restricciones

$$(A_C(\pi))_\alpha = \{p/C(\pi,p) \geq \alpha\}$$

y se procede a clcular la distribución de máxima entropía en di cho α-corte, es decir, se resuelve

$$\begin{aligned} &\text{Max} \quad H: -k \sum_{a\in U} p(a)L(p(a)) \\ &\text{s.a.} \quad \sum_{a\in U} p(a) = 1 \\ &\qquad C(\pi,p) \geq \alpha \end{aligned} \tag{22}$$

Notando p_α, la solución de este problema, obtenemos que la función de pertenencia de S es

$$\mu_S(p) = \begin{cases} 0 & \text{si } \{\alpha\in[0,1]/p \equiv p_\alpha\} = \phi \\ \text{Max}\{\alpha\in[0,1]/p\equiv p_\alpha\} & \text{en otro caso.} \end{cases}$$

A continuación analizamos las caracteríCticas de este método para distintas medidas de consistencia.

<u>Caso 1.-</u> Si consideramos la medida de consistencia de Zadeh, C_Z, entonces cada uno de los problemas (22) coincide con nues-- tra modificación del problema propuesto por Y.Leung para un - fijo, que ya hemos resuelto.

<u>Ejemplo 3.4.-</u> Sea $U = \{x_1,x_2,x_3,x_4\}$ y supongamos que se tie-

ne una distribución de posibilidad sobre los resultados de un experimento aleatorio en U dada por

$$\pi(x_1) = 1 \quad \pi(x_2) = 0.9 \quad \pi(x_3) = 0.7 \quad \pi(x_4) = 0.2$$

Entonces, la solución al problema (21) con la medida de consistencia de Zadeh, C_Z, es el subconjunto difuso S tal que

- Si $u \in [1,+\infty)$ la distribución

$$p_u(x_1) = \frac{u}{u + u^{0.9} + u^{0.7} + u^{0.2}}$$

$$p_u(x_2) = \frac{u^{0.9}}{u + u^{0.9} + u^{0.7} + u^{0.2}}$$

$$p_u(x_3) = \frac{u^{0.7}}{u + u^{0.9} + u^{0.7} + u^{0.2}}$$

$$p_u(x_4) = \frac{u^{0.2}}{u + u^{0.9} + u^{0.7} + u^{0.2}}$$

pertenece a S con grado

$$\alpha = \frac{u + o.9u^{0.9} + 0.7u^{0.7} + 0.2u^{0.2}}{u + u^{0.9} + u^{0.7} + u^{0.2}} \in [0.7,1)$$

- La distribución $p_\infty(x_1) = 1$, $p_\infty(x_2) = p_\infty(x_3) = p_\infty(x_4) = 0$, pertenece a S con grado 1.

Observemos que al ir aumentando α desde 0.7 (en que la distribución es la uniforme en U), hasta 1 (en que la distribución es la degenerada en x_1), o equivalentemente al ir aumentando u de 1 a ∞, la distribución p_u se va concentrando en los elementos que tienen probabilidad más alta.

Esto puede comprobarse en la sigiente tabla:

	$p_u(x_1)$	$p_u(x_2)$	$p_u(x_3)$	$p_u(x_4)$
$u = 1$, $\alpha = 0.7$	0.25	0.25	0.25	0.25
$u = 2$, $\alpha = 0.76007$	0.30123	0.28106	0.24467	0.17301
$u = 10$, $\alpha = 0.85469$	0.40750	0.32369	0.20423	0.06458
$u \to \infty$, $\alpha = 1$	1	0	0	0

<u>Caso 2.-</u> Si como medida de consistencia consideramos la integral de Sugeno, el problema (22) puede expresarse como:

$$\begin{aligned} &\text{Max} \quad H = -k \sum_{a \in U} p(a) L(p(a)) \\ &\text{s.a.} \quad \sum_{a \in U} p(a) = 1 \\ &\qquad \oint_U g(\pi) \circ f(P) \geq \alpha \end{aligned} \tag{23}$$

La restricción $\oint_U g(\pi) \circ f(P) \geq \alpha$, es equivalente a

$$\sup_{\beta \in [0,1]} \{ f(P(B_\beta)) \wedge \beta \} \geq \alpha$$

donde $B_\beta = \{a \in U / g(\pi(a)) \geq \beta\}$

Lo cual, a su vez, se cumple si y solo si

$$f(P(B_\beta)) \geq \alpha$$

Si notamos $U = \{a_1, a_2, .., a_n\}$ con $\pi(a_1) \geq \pi(a_2) \geq \dots \geq \pi(a_n)$, la condición anterior se expresará como

$$f(\ p(a_1) + \dots + p(a_{i(\alpha)})\ ) \geq \alpha \tag{24}$$

donde $a_{i(\alpha)}$ viene determinado por

$$g(\pi(a_{i(\alpha)})) \geq \alpha \quad \text{y } g(\pi(a_{i(\alpha)+1})) < \alpha \ .$$

La desigualdad (24) se puede transformar siempre en una restricción lineal, con ayuda de la función $h: [0,1] \longrightarrow [0,1]$ definida por

$$h(\alpha) = \text{Inf } \{t \in [0,1] / f(t) \geqq \alpha\} ,$$

que coincide con f^{-1}, si f tiene inversa. Como f es monótona, - dicha condición (24) se puede expresar

$$p(a_1) + \ldots + p(a_{i(\alpha)}) \geqq h(\alpha) \tag{25}$$

restrcción, que para cada α, es lineal.

Con esta transformación, la solución de (23) es la siguiente

- Si $i(\alpha)/|U| \geqq h(\alpha)$

$$p_\alpha(a) = 1/|U| \quad , \forall a \in U.$$

- Si $i(\alpha)/|U| < h(\alpha)$

$$p_\alpha(a_k) = \begin{cases} \dfrac{h(\alpha)}{i(\alpha)} & \text{si } k \leqq i(\alpha) \\ \dfrac{1 - h(\alpha)}{|U| - i(\alpha)} & \text{si } k > i(\alpha) \end{cases}$$

<u>Ejemplo 3.5.-</u> En las mismas condiciones del ejemplo ante--- rior y, considerando f y g la identidad en el intervalo [0,1], se obtiene la siguiente tabla

	$p_u(x_1)$	$p_u(x_2)$	$p_u(x_3)$	$p_u(x_4)$
$\alpha = 0.76007$	0.380035	0.380035	0.119965	0.119965
$\alpha = 0.85469$	0.427340	0.427340	0.072655	0.072655
$\alpha = 0.95$	0.950000	0.016667	0.016667	0.016667

<u>Caso 3.-</u> Si utilizamos como medida de consistencia la de Dubois, Prade:

$$C_{DP}(\pi,p) = \begin{cases} 1 & \text{si } \Pi(A) \geqq P(A) \quad , \forall A \subset U \\ 0 & \text{en otro caso} \end{cases}$$

entonces el problema (21) queda

$$\begin{aligned} \text{Max} \quad & H = -k \sum_{a \in U} p(a) L(p(a)) \\ \text{s.a.} \quad & \sum_{a \in U} p(a) = 1 \\ & P(A) \leqq \Pi(A), \forall A \subset U \end{aligned} \tag{26}$$

Observemos que es un problema de programación clásico como - consecuencia de que la medida de consistencia de Dubois, Prade es "crisp".

Si notamos $U = \{b_1,..,b_n\}$ con $\pi(b_1) \leqq \ldots \leqq \pi(b_n)$, la condición de consistencia se expresa (proposición 3.2) como

$$\pi(b_i) \geqq p(b_1) + \ldots + p(b_i) \ , \ i \in \{1,..,n\} \tag{27}$$

En estas condiciones, la solución del problema (26) se puede obtener de modo recurrente de la siguiente forma

$$\begin{aligned} p(b_1) &= \operatorname{Min} \{\pi(b_i)/i, \ i=1,..,n\} \\ p(b_k) &= \operatorname{Min} \left\{ \frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1})}{i-k+1} \ , \ i=k,..,n \right\} ,k=2,..,n \end{aligned} \tag{28}$$

Esta distribución de probabilidad cumple las siguientes propiedades:

a) $p(b_1) + \ldots + p(b_k) \leqq \pi(b_k) \ , \ k \in \{1,..,n\}$ (29)

Demostración.- Es inmediata por la definición de p.

b) $p(b_1) \leqq p(b_2) \leqq \ldots \leqq p(b_n)$ (29)

Demostración.-

$$p(b_k) = \operatorname{Min} \left\{ \frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1})}{i-k+1} \ ; \ i=k,..,n \right\}$$

$$p(b_{k+1}) = \operatorname{Min} \left\{ \frac{\pi(b_i)-p(b_1)-\ldots-p(b_k)}{i-k} \ ; \ i=k+1,..,n \right\}$$

Ahora bien, si $i \in \{k+1,..,n\}$

$$\frac{\pi(b_i)-p(b_1)-\ldots-p(b_k)}{i-k} \geqq$$

$$\geqq \frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1}) - \left(\frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1})}{i-k+1}\right)}{i-k} =$$

$$= \frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1})}{i-k+1}$$

y, por tanto,

$$p(b_{k+1}) \geq \text{Min}\left\{\frac{\pi(b_i)-p(b_1)-\ldots-p(b_{k-1})}{i-k+1} \; ; \; i=k+1,\ldots,n\right\} \geq p(b_k).$$

Como esto se cumple para cada k, la propiedad queda probada#

c) $k \in \{1,\ldots,n-1\}$ $(p(b_k) < p(b_{k+1}) \Rightarrow p(b_k)=\pi(b_k)-p(b_1)\ldots-$
$-p(b_{k-1}))$ (31)

Demostración

Sea $p(b_k) < p(b_{k+1})$, y supongamos que

$$\frac{\pi(b_{i_1})-(b_1)-\ldots-p(b_k)}{i_1-k+1} = \text{Min}\left\{\frac{\pi(b_i)-p(b_1)\ldots-p(b_{k-1})}{i-k+1} \; ; i=k,\ldots,n\right\}$$

con $i_1 \geq k+1$; entonces

$$\frac{\pi(b_{i_1})-p(b_1)\ldots-p(b_k)}{i_1-k} =$$

$$\frac{\pi(b_{i_1})-p(b_1)\ldots-p(b_{k-1}) - \left(\frac{\pi(b_{i-1})-p(b_1)\ldots-p(b_{k-1})}{i_1-k+1}\right)}{i_1-k} =$$

$$\frac{\pi(b_{i_1})-p(b_1)\ldots-p(b_{k-1})}{i_1-k} = p(b_k);$$

y, por tanto, $p(b_{k+1}) \leq p(b_k)$, en contra de nuestra hipótesis.- Luego, si $p(b_k) < p(b_{k+1})$, el mínimo se obtiene para $i = k$; y, como consecuencia

$$p(b_k) = \frac{\pi(b_k)-p(b_1)\ldots-p(b_{k-1})}{k-k+1} = \pi(b_k)-p(b_1)\ldots-p(b_{k-1}).$$

Como queríamos probar. #

A partir de estas propiedades, es inmediato comprobar que p cumple las condiciones de Kuhn-Tucker para el problema (26), de donde se deduce que es solución del mismo.

Ejemplo 3.6.- Para los datos de los ejemplos anteriores, obtenemos:

$$p(x_1) = 4/15 \ , \quad p(x_2) = 4/15 \ , \quad p(x_3) = 4/15 \ , \quad p(x_4) = 0.2.$$

Obsérvese que esta distribución de probabilidad no coincide con la que se asignaría de acuerdo con (5), y que es

$$p(x_1) = 1/2 \ , \quad p(x_2) = 19/60 \ , \quad p(x_3) = 13/120, \ p(x_4) = 0.05$$

Nota.- El método de Dubois y Prade para asignar una distribución de probabilidad, también puede considerarse basado en el principio de máxima entropía, pero aplicado a otro nivel.

Una asignación básica m, en $\mathcal{P}(U)$, da lugar a una distribución de probabilidad en U, cuando los únicos elementos focales son los conjuntos unitarios. En estas condiciones, aplicamos el principio de máxima entropía en cada $A \subset U$, equidistribuyendo m(A) entre sus elementos; obtenemos, entonces, la distribución de probabilidad

$$p(a) = \sum_{a \in A \subset U} \frac{m(A)}{|A|}$$

Si m es de tipo posibilístico con distribución de posibilidad asociada π, de la expresión anterior se obtiene, de modo, inmediato (5).

Al aplicar el principio de máxima entropía a nivel de los conjuntos $A \subset U$, este segundo método de Dubois y Prade produce distribuciones de probabilidad más informativas (con menos entropía) que el primero.

4.-ASIGNACION DE UNA DISTRIBUCION DE POSIBILIDAD A UNA DISTRIBUCION DE PROBABILIDAD

Sea U un conjunto finito y X un experimento aleatorio que toma valores en U. Si disponemos de la distribución de probabilidad con que X toma valores en U, nos proponemos extraer de ella una información posibilística para todos los resultados del experimento aleatorio, es decir, una distribución de posibilidad válida para todos los valores de X.

Dubois y Prade (1.982) proponen asignar a p la distribución de posibilidad

$$\pi(a_1) = 1$$

$$\pi(a_{i+1}) = 1 - \sum_{j=1}^{i} (p(a_j) - p(a_{i+1}))$$

(supuesto $U = \{a_1,..,a_n\}$ con $p(a_1) \geq \ldots \geq p(a_n)$ (ver (4))

Nosotros vamos a proponer una solución a este problema que - se puede considerar, en cierta forma, dual de la propuesta para el problema estudiado en el apartado anteior. En principio, si p es una distribución de probabilidad en U, parece lógico afirmar que son posibles todos los resultados del experimento con - probabilidad positiva; lo que equivale a asignar la distribu--- ción de posibilidad

$$\pi(a) = \begin{cases} 1 & \text{si } p(a) > 0 \\ 0 & \text{si } p(a) = 0 \end{cases} \tag{32}$$

Esta distribución de posibilidad tiene consistencia uno con p, cualquiera que sea la medida C considerada, y por tanto, puede pensarse que es una buena solución del problema. Sin embargo, tiene el inconveniente de que es poco informativa. Desde luego, cualquier distribución de posibilidad que asignemos a p debe de ser consistente con ella, pero cabe preguntarse si, manteniendo la consistencia próxima a uno, aunque no exactamente uno, podríamos asociar distribuciones a p que informaran más sobre los resultados del experimento aleatorio. Consideremos el siguiente - ejemplo para ilustrar esta idea.

Ejemplo 3.7.- Sea $U \subset R$ y una observación de un experimento aleatorio X que toma valores en U. Supongamos que queremos ha-- cer un test sobre la distribución del experimento en base a dicha observación:

$H_o; p \equiv p_o$ $\qquad\qquad$ $H_1: p \in C - \{p_o\};$

donde C es una cierta clase de distribuciones de probabilidad en U.

Sea cual sea el método utilizado, siempre se selecciona un nivel de significación, α, próximo a uno y una región de aceptación $A \subset U$ (con $P_o(A) > \alpha$), tal que si $X \in A$, se acepta H_o y si $X \notin A$ se rechaza.

Consideremos, para la región de aceptación $A \subset U$, la distribución de posibilidad de tipo "crisp", π_A, definida de la forma habitual

$$\pi_A(a) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{si } a \notin A \end{cases}$$

La condición $P_o(A) > \alpha$ la podemos expresar como $C(p_o, \pi_A) > \alpha$ para cualquier medida de consistencia que cumpla $C(p_o, \pi_A) = P_o(A)$ (entre las medidas de consistencia que verifican esta propiedad, están las de Zadeh y la basada en la integral de Sugeno)

La regla de decisión asociada a A permite la siguiente interpretación a partir de π_A: si el resultado del experimento $X(\omega) = a$ es posible, según esta distribución ($\pi_A(a) = 1$, es decir $a \in A$), se acepta H_o; si el resultado $X(\omega) = a$ es imposible según π_A ($\pi_A(a) = 0$, es decir $a \notin A$), se considera que no puede ser H_o (se acepta H_1).

Si $\alpha = 1$, todos métodos de contraste proporcionan la misma región de aceptación: $A_o = \{a/p_o(a) > 0\}$, cuya distribución de posibilidad asociada, π_{A_o}, es la que correspondería a p_o, de acuerdo con (32). Ahora bien, el test correspondiente a A_o es muy poco potente: aceptará muchas veces la hipótesis nula cuando esta sea falsa. Se prefiere, entonces, elegir niveles de confianza (consistencias) próximos a uno y conseguir regiones de aceptación (distribuciones de posibilidad) más potentes (más informativas).

Analogamente a como se hace en el ejemplo, nosotros proponemos asignar a una distribución de probabilidad p aquella que, siendo consistente con ella, proporcione la máxima información sobre los resultados del experimento. Así pues, asignaremos a p la solución del problema de programación difuso

$$\begin{array}{lll} \text{Max:} & I(\pi) & \\ \text{s.a.} & \pi \in \prod(U) & (33) \\ & \text{"}\pi \text{ es consistente con } p\text{"} & \end{array}$$

donde I es una medida de información de las estudiadas en el segundo capítulo.

"π es consistente con p" es una propiedad difusa que expresa que $\pi \in \mathcal{B}(p)$, donde $\mathcal{B}(p)$ es el subconjunto difuso de las distribuciones de posibilidad en U que son consistentes con la distribución de probabilidad dada. Dicho conjunto vendrá determinado por una medida de consistencia C, considerando

$$\mu_{\mathcal{B}(p)}(\pi) = C(\pi,p)$$

Nota.- Si sobre los resultados de un experimento aleatorio disponemos de una información probabilística, dicho experimento estará mejor descrito que si disponemos de una información posibilística.

Por eso, parece lógico, si tenemos una distribución de posibilidad, considerar entre las distribuciones de probabilidad consistentes con ella, aquella que maximiza la entropía (es la menos informativa) y, "dualmente", asignar a una distribución de probabilidad p, la de posibilidad, π, que proporciona máxima información. (el calificativo "dual" no se refiere a relaciones existente entre los problemas de programación matemática que aparecen en estas asignaciones).

El problema (33) lo vamos a resolver, nuevamente, siguiendo el método propuesto en Verdegay (1.980, 1.982), según el cual, la solución ha de ser difusa y se obtiene resolviendo los problemas (α-cortes):

$$\begin{aligned}
&\text{Max} \quad I(\pi) \\
&\text{s.a.} \quad \pi \in \prod(U) \qquad\qquad\qquad\qquad (34) \\
&\qquad C(\pi,p) \geq \alpha \ , \ \alpha \in [0,1]
\end{aligned}$$

que son problemas de programación, no lineales en general. Si notamos su solución por π_α, entonces la de (33), que notaremos S', tendrá como función de pertenecia

$$\mu_{S'}(\pi) = \begin{cases} 0 & , \text{ si } \{\alpha \in [0,1] / \pi \equiv \pi_\alpha\} = \phi \\ \text{Sup } \{\alpha \in [0,1] / \pi \equiv \pi_\alpha\} & , \text{ en otro caso} \end{cases}$$

A continuación, vamos a obtener S' para varios casos particulares de I y C.

<u>Caso 1.-</u> Si consideramos como medida de consistencia la de Dubois y Prade, el problema (33) no es difuso y tiene la forma

$$\begin{aligned}
&\text{Max} \quad I(\pi) \\
&\text{s.a.} \quad \pi \in \prod(U) \qquad\qquad\qquad\qquad (35) \\
&\qquad \Pi(A) \geq P(A), \ A \subset U
\end{aligned}$$

Notando $U = \{a_1,..,a_n\}$ con $p(a_1) \leq \ldots \leq p(a_n)$, la solución de (35), independientemente de la medida de información I considerada, viene dada por

$$\pi_{DP}(a_i) = \sum_{j=1}^{i} p(a_j)$$

(Puesto que $C(\pi_{DP},p) = 1$ y, para cualquier otra π con $C(\pi,p) = 1$, se cumple $\pi \subset \pi_{DP}$ y, por tanto, $I(\pi) \leq I(\pi_{DP})$, para toda medida de información I).

<u>Caso 2.-</u> Sea $I(\pi) = I_1(\pi) = \text{Log}(|U| / \sum_{a\in U} \pi(a))$ y C la medida de consistencia de Zadeh, C_Z. Si notamos $U = \{a_1,...,a_n\}$ con $p(a_1) \leq \ldots \leq p(a_n)$, entonces, la solución a (34) es

- $\pi(a_i) = 1$, si $i \geq i(\alpha)$.

$-\pi_\alpha(a_i) = \left(\alpha - \sum_{j=i(\alpha)} p(a_j)\right) / p(a_{i(\alpha)-1})$, si $i=i(\alpha)-1$

$-\pi_\alpha(a_i) = 0$, si $i < i(\alpha)-1$.

donde $i(\alpha) \in \{1,2,..,n\}$ viene determinado por

- $i(\alpha) = 1$, si $\alpha = 1$

- $\sum_{j=i(\alpha)}^{n} p(a_j) \leq \alpha$ y $\sum_{j=i(\alpha)-1}^{n} p(a_j) > \alpha$, si $\alpha < 1$.

Ejemplo 3.8.- Si $U = \{n \in N / n \leq 20\}$ y p es una distribución binomial $B(20,1/2)$, para $\alpha = 0.95$, obtenemos la distribución de posibilidad

$$\pi(0) = \pi(2) = \ldots = \pi(5) = \pi(15) = \ldots = \pi(20) = 0$$

$$\pi(6) = 0.23293 \quad , \pi(7) = \pi(8) = \ldots = \pi(14) = 1.$$

Caso 3.- Con la medida de consistencia de Zadeh, C_Z, y la medida de información

$$I_n(\pi) = \log\left(|U| / \sum_{a \in U} \pi^n(a)\right) \qquad (n \geq 2)$$

obtenemos, por el método de los multiplicadores de Lagrange,

$$\pi_\alpha(a) = \frac{h_n(a)}{\underset{b \in U}{\mathrm{Max}}\ h_n(b)} \quad , \text{ donde, } h_n(b) = \frac{\sqrt[n-1]{p(b)}}{\sum_{c \in U} \sqrt[n]{p(c)}}$$

Ejemplo 3.8.- Para $U = \{0,1,2,3,4\}$, p la distribución binomial $B(4,1/2)$ y considerando como medida de información I_3, para $\alpha = 0.95$, obtenemos la distribución de posibilidad

$$\pi(0) = \pi(4) = 0.408246 \ , \ \pi(1) = \pi(3) = 0.816497 \ , \ \pi(2) = 1.$$

Caso 4.- En las mismas condiciones del caso (3), pero con medida de información I_n^* $(n \geq 2)$,se obtiene la misma solución que en el caso (2).

Nota.- Son interesantes las distribuciones de posibilidad --

que se obtienen para medidas de consistencia de la forma

$$C_\times(\pi,p) = C(\pi,p).C_{DP}(\pi,p) \quad (C_\times = C.C_{DP})$$

Con ellas, nos garantizamos que la distribución de posibilidad, π, solución de (34) cumple

$$\pi(a) \geq P\{b \in U/p(b) \leq p(a)\} \qquad \text{(ver (2))}$$

<u>Ejemplo 3.9.-</u> Sea $U = \{n \in N/n \leq 20\}$, p la distribución binomial $B(20,1/2)$. Si consideramos como medida de información I_1 y, como medida de consistencia $C_\times = C_Z.C_{DP}$, para $\alpha = 0.95$ obtenemos

$\pi(0) = \pi(20) = 1.90735\times10^{-6}$, $\pi(1) = \pi(19) = 4.00543\times10^{-5}$

$\pi(2) = \pi(18) = 4.02451\times10^{-4}$, $\pi(3) = \pi(17) = 2.57683\times10^{-3}$

$\pi(4) = \pi(16) = 0.0118179$, $\pi(5) = \pi(15) = 0.041389$

$\pi(6) = \pi(14) = 0.86542$, $\pi(a) = 1$ (si $7 \leq a \leq 13$)

De los anteriores desarrollos en los dos últimos apartados, se deduce que la asignación de una medida de probabilidad (a -- una de posibilidad) o de posibilidad (a una de probabilidad), - depende de la elección de la medida de consistencia y de la medida de información empleadas.

Las caracteríсticas y propiedades de estas medidas han sido estudiadas en los apartados correspondientes. Aquí, hemos dado una panorámica de los resultados que se obtienen empleando las medidas más importantes. El decisor, de acuerdo con sus objetivos y criterio, deben de escoger la más adecuada.

REFERENCIAS

BALDWIN, J.F. (1979)
Fuzzy Logic and its Application to Fuzzy Reasoning. En Advances in Fuzzy Set Theory and Applications. M.M. Gupta, R.K. Ragade y R.R. Yager (Eds). North-Holland, 93-116

BALDWIN, J.F. (1979)
A New Approach to Approximate Reasoning Using Fuzzy Logic. Fuzzy Sets and Systems, 2, 309-325

BELLMAN, R. y ZADEH, L.A. (1970)
Decision Making in a Fuzzy Environment. Management Science, 17 B(4), 141-164.

CZOGALA, E., GOTTWALD, S. y PEDRYCZ, W. (1982)
Contribution to Application of Energy Measure of Fuzzy Sets. Fuzzy Sets and Systems, 8, 205-214.

DE GROOT, M. (1970)
Optimal Statistical Decisions. McGraw Hill.

DE FINETTI, B. (1974)
Theory of Probability. Vol. I. J. Wiley

DEMPSTER, A.P. (1967)
Upper and Lower Probabilities by a Multivalued Mapping. Ann. Math. Statis. Vol 38, 325-339.

DE LUCA, A. y TERMINI, S. (1972)
A Definition of a non Probabilistic Entropy in the Setting of Fuzzy Sets Theory. Information and Control, 20, 301-312.

DE LUCA, A. y TERMINI, S. /1979)
Entropy and Energy Measures of a Fuzzy Set. En Advances in Fuzzy Set Theory and Applications. M.M. Gupta, R.K. Ragade y R.R. Yager (Eds). North-Holland, 321-338.

DUBOIS, D. (1980)
Sur les Liens entre les Notions de Probabilite et de Possibilite (Quelques Remarques). CNRS Round Table "Quelques Applications Concretes Utilisant les Derniers Perfectionements de la Theorie du Flou". Lyon, 23-25.

DUBOIS, D. y PRADE, H. (1980)
Fuzzy Sets and Systems: Theory and Applications. Academic Press.

DUBOIS, D. y PRADE, H. (1982)
On Several Representations of an Uncertain Body of Evidence. En Fuzzy Information and Decision Processes. M.M. Gupta y E. Sanchez (Eds). North-Holland. 167-182.

DUBOIS, D. y PRADE, H. (1983)
Unfair Coins and Necessity Measures: Towards a Possibilistic Interpretation of Histograms. Fuzzy Sets and Systems, 10,1, 15-20.

DUBOIS, D. y PRADE, H. (1984)
A Note on Measures of Specificity for Fuzzy Sets. BUSEFAL, 19, 83-89.

KAMPE DE FERIET, J. (1970)
Measure de L'Information Fournie par un evenement. Coll. Int. CNRS 186. Paris, 191-221.

KAMPE DE FERIET, J. (1974)
La Theorie Generalisee de L'Information et la Measure Subjetive de L'Information. En Theories de L'Information J. Kampe de Feriet y F. Picard (Eds). Springer-Verlag. 1-28.

HIGASHI, M. y KLIR, G.J. (1983)
Measures of Uncertainty and Information Based on Possibility Distributions. Int. J. General Systems, 9,1, 43-58.

HISDAL, E. (1978)
Conditional Possibilities Independence and Noninteraction. Fuzzy Sets and Systems,1, 283-298.

HISDAL, E. (1979)
Possibilistically Dependent Variables and a General Theory of Fuzzy Sets. En Advances in Fuzzy Set Theory and Applications. M.M. Gupta, R.K. Ragade y R.R. Yager (Eds). North-Holland, 215-234.

HOHLE, V. (1982)
A Mathematical Theory of Uncertainty. En Fuzzy Set and Possibility Theory. Recent Developments. R.R. Yager (Ed) Pergamon Press. 344-355.

JAYNES, E.T. (1957)
Information Theory and Statistical Mechanics. Phys. Rev. 106, 620-630.

KANDEL, A. (1982)
Fuzzy Techniques in Pattern Reconigtion. John Wiley.

LEUNG, Y. (1980)
Maximum Entropy Estimation with Inexact Information. En- Fuzzy Set and Possibility Theory. Recent Developments. R R. Yager (Ed). Pergamon Press. 32-37.

LINDLEY, D.V. (1982)
Scoring Rules and the Inevitability of Probability. International Statistical Review, 50, 1-26.

MIZUMOTO, M., FUKAMI, S y TANAKA, K. (1979)
Some Methods of Fuzzy Reasoning. En Advances y Fuzzy Set Theory and Applications. M.M. Gupta, R.K. Ragade y R.R.- Yager (Eds). North-Holland. 117-136.

MORAL, S. (1984)
Un Estudio de la Informacion Proporcionada por una Restriccion Difusa. XIV Congreso Nacional de Estadistica, - Investigacion Operativa e Informatica. Granada. 720-730

NATVIG, B. (1983)
Possibility vs. Probability. Fuzzy Sets and Systems, 10, 1, 31-36.

NEGOITA, C.V. y RALESCU, D. (1977)
On Fuzzy Optimization. Kybernetes, 6, 193-195.

NGUYEN, H.T. (1978)
On Conditional Possibility Distributions. Fuzzy Sets and Systems,1, 299-310.

NGUYEN, H.T. (1979)
Toward a Calculus of the Mathematical Notion of Possibility. En Advances in Fuzzy Set Theory and Applications.- M.M. Gupta, R.K. Ragade y R.R. Yager (Eds). North-Hollan 235-246.

NGUYEN, H.T. (1982)
On the Possibilistic Approach to the Analysis of Evidence. En Fuzzy Set and Possibility Theory. Recent Developments. Ronald R. Yager (Ed). Pergamon Press. 395-401.

OKUDA, T., TANAKA, H. y ASAI, K. (1974)
Decision-Making and Information in Fuzzy Events. Bulletin of University of Osaka Prefecture, A, 23, nº 2.

PRADE, H. (1981)
On the Link Between Dempster's Rule of Combination of Evidence on Fuzzy Set Intersection. BUSEFAL, 8, 60-64.

PRADE, H. (1982)
Modal Semantics and Fuzzy Set Theory. En Fuzzy Set and - Possibility Theory. Recent Developments. Ronald R. Yager (Ed). Pergamon Press. 232-246

SHAFER, G. (1976)
A Mathematical Theory of Evidence. Princeton Univ. Press

SUAREZ GARCIA, F. (1983)
Familias de Integrales Difusas y Medidas de Entropia Relacionadas. Tesis Doctoral. Universidad de Oviedo.

SUAREZ GARCIA, F. y BREZMES BREZMES, T. (1984).
Teoremas de Convergencia en Integrales Difusas. XIV Congreso Nacional de Estadistica, Investigacion Operativa e Informatica. Granada. 781-791.

SUGENO, M. (1974)
Theory of Fuzzy Integrals and its Applications. Ph. D. - Thesis. Tpkyo Inst. of Tech.

TAKEDA, E., NAKAJIMA, N. y TAHARA, A. (1975)
Fuzzy Information in Product Spaces. Technology Reports- of the Osaka University, 25, nº 1230

TAKEDA, E., TAHARA, A. y NAKAJIMA, N. (1976).
Information Content of a Fuzzy Message. Rev. Stat. Appl. Res., JUSE, 23, nº 4

TERANO, T. y SUGENO, M. (1975)
Conditional Fuzzy Measures and Their Applications. En - L.A. Zadeh, K.S. Fu, K. Tanaka y M. Shimura (Eds): Fuzzy Sets and Their Applications to Cognitive and Decision Processes. Academic Press. 151-170.

THEIL, M. (1967)
Economics and Information Theory. Morth-Holland.

THOMAS, M.V. (1979)
A Generalized Maximum Entropy Principle. Oper. Res., 27 1188-1196.

TRILLAS, E. y RIERA, T. (1978)
Entropies in Finite Fuzzy Sets. Information Sciences, - 159-168

VERDEGAY, J.L. (1981)
Problemas de Decision en Ambiente Difuso. Tesis Doctoral Universidad de Granada.

VERDEGAY, J.L. (1982)
Fuzzy Mathematical Programming. En Fuzzy Information and Decision Processes. M.M. Gupta y Elie Sanchez (Eds). -- North-Holland. 231-237.

WIERZCHON, S.T. (1982)
On Fuzzy Measure and Fuzzy Integral.En Fuzzy Information and Decision Processes. M.M. Gupta y Elie Sanchez (Eds). North-Holland. 79-86.

XIE, W.X. y BEDROSIAN, S.D. (1984)
An Information Measure for Fuzzy Sets. IEEE Trans. Syst. Man and Cybern. SMC-14, nº 1.

YAGER, R.R. (1981)
Measurement of Properties on Fuzzy Sets and Possibility-Distribution. Proc. III Int. Seminar on Fuzzy Set Theory E.P. Klement (Ed). J. Kepler Univ., Linz, Austria 211-222

YAGER, R.R. (1983)
An Introduction to Applications of Possibility Theory. - Human Systems Management, 3, 246-253.

YAGER, R.R. (1983)
Entropy and Specifity in a Mathematical Theory of Evidence. Int. J. General Systems, 9, 249-260.

ZADEH, L.A. (1968)
Probability Measures of Fuzzy Events. J. Math. Anal. and Applications, 23, 421-427.

ZADEH, L.A. (1975)
Calculus of Fuzzy Restrictions. En Fuzzy Sets and their-Applications to Cognitive and Decision Processes. L.A.-- Zadeh, K.S. Fu, K. Tanaka y M. Shimura (Eds). Academic - Press. 1-39.

ZADEH, L.A. (1978)
Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems,1, 3-28.