T
14
13

" MODELOS DE DECISION CON INFORMACION GENERAL "

Memoria que para optar al grado de Doctor, presenta el licenciado en Ciencias Matemáticas, María Teresa Lamata Jimenez.

Maite Lamata

Directores

Profesor Dr. D. Miguel Delgado Calvo-Flores

V° B°

M. Delgado

Profesor Dr. D. Serafín Moral Callejón

V° B°

Serafín Moral

UNIVERSIDAD DE GRANADA. FACULTAD DE CIENCIAS. Diciembre 1985

Es dificil expresar el agradecimiento que siento y la deuda que tengo con los directores de ésta memoria: D. Miguel Delgado y D.Serafín Moral. Ellos me han orientado, apoyado y animado siempre. Además, sin su trabajo no habria podido terminar ésta tesis. Por todo ello: ¡Gracias¡.

Tambien, quiero expresar mi gratitud al grupo de trabajo en I.O con datos difusos de la Universidad de Granada, en uno de cuyos seminarios surgió el problema que ha motivado ésta memoria, y que siempre ha sido el grupo de amigos que ha estado dispuesto a oirme y ayudarme.

INDICE

- INTRODUCCION GENERAL-

A nadie escapa la importancia que tiene la información sobre los factores no controlables por el decisor asi como su correcto manejo e interpretación en cualquier proceso de toma de decisiones.

Todo fenomeno real puede clasificarse en una de las siguientes categorias:

- Aquellos que tienen resultados que se pueden predecir con certeza y son bien conocidos
- Aquellos sobre cuyos resultados existe algun tipo de incertidumbre, imprecisión o desconocimiento.

Si los factores no controlables por el decisor pertenecen al primer tipo, el problema de decisión se simplifica transformandose en una mera ordenación de recompensas. Ahora bien, en el momento en que el decisor tiene algún desconocimiento o incertidumbre sobre el estado de estos factores, aparece de modo inmediato el riesgo de la elección que todo decisor tratará de valorar y si puede disminuir, para lo cual tendrá que hacer el uso mas correcto posible de su información y, en su caso, intentar aumentar esta.

Durante mucho tiempo la probabilidad ha sido la única herramienta disponible para modelizar situaciones que comportasen algún tipo de incertidumbre. Desde un punto de vista - bayesiano se define la probabilidad de un suceso como una medida del grado de incertidumbre que posee el observador o experimentador sobre la realización de dicho suceso. Debido al caracter netamente subjetivista de este enfoque, los partidarios de la filosofia bayesiana afirman que esta es suficiente para describir cualquier tipo de incertidumbre, llegando a asegurar los mas acerrimos que la probabilidad es la única representación válida de la incertidumbre.

Estas afirmaciones han dado lugar a dos grandes escuelas (podriamos llegar a llamarlas movimientos filosóficos) - caracterizadas por su forma de considerar la incertidumbre o lo que es equivalente, la información (posiblemente incompleta) que se posea sobre un fenómeno o suceso:

- Bayesiana.- Se admite que cualquier información puede traducirse en términos de probabilidad.
- No bayesiana.- No se acepta la afirmación de que el modelo probabilístico es el único válido para describir los distintos casos de - experimento con incertidumbre.

Hay que destacar que, mientras la probabilidad ha sido la única herramienta disponible para describir incertidumbre, la filosofia bayesiana ha constituido el único enfoque para emplear o valorar toda la información disponible sobre un fenómeno cualquiera que sea el caracter de esta. De hecho, - la determinación de la ley de probabilidad asociada a una información no netamente probabilística se hacia, fundamentalmente, mediante el Principio de Máxima Entropia (ver Jaynes (1957), Tribus (1969) o Thomas (1979)).

Por otra parte, los no bayesianos solo podian emplear la información cuando esta era claramente de tipo probabilístico, debiendo despreciarla en caso contrario.

En las dos últimas décadas, y como fruto de la búsqueda de modelos cada vez mas adaptados a la realidad, se han - desarrollado (y se continuan investigando) formas alternativas de representación de la incertidumbre o, lo que es equivavalente, de la información disponible sobre un fenómeno.

La Teoria de la Evidencia (Shafer (1976)) y la Teoria de la Posibilidad (Zadeh (1978)) son claros exponentes de lo que acabamos de decir. Ambas pueden encuadrarse, a su vez, - de tro de la Teoria de Medidas difusas,introducidas por Su-- geno en 1974 como generalización de las medidas aditivas clásicas.

La idea de base de esta memoria es estudiar y desarrollar herramientas para abordar problemas de decisión bajo esquemas de información generales, tratando de conseguir que - la información disponible pueda procesarse de acuerdo con su particular naturaleza.

El modelo mas general que puede plantearse es aquel en que el conocimineto sobre los estados de la naturaleza se -- describe mediante una medida difusa. Ahora bien, la excesiva generalidad del mismo dificulta el desarrollo de las herra-- mientas necesarias para su amplia utilización. Por este motivo nos hemos restringido al campo de la Teoria de la Evidencia, donde se equilibran la generalidad y la estructura formal, de modo que pueden obtenerse potentes resultados de amplio espectro de aplicación.

El modelo de decisión que consideramos puede representarse mediante el cuadruple (Ω,D,r,I), donde:

- Ω es el conjunto de estados de la Naturaleza,
- D es el conjunto de acciones disponibles para el decisor,
- $r:\Omega\times D \longrightarrow R$ es una función valor de las consecuencias de las acciones del decisor frente a los posibles estados de la Naturaleza,
- I representa la información disponible sobre el verdadero estado de la Naturaleza.

Supondremos que Ω y D son finitos para evitar problemas de medibilidad, integrabilidad y convergencia completamente ajenos al tema de la memoria. Este criterio será empleado a lo largo de toda ella. Admitiremos que la información I se representa mediante una evidencia sobre , sobre cuyo origen no hacemos ninguna hipótesis, pudiendo proceder solo de un conocimiento "a priori" o de este mas la aportación de un experimento.

Hay que señalar que admitir que las consecuencias se encuentran valoradas numericamente conlleva, en general, la aceptación de un conjunto de hipótesis de "racionalidad" que permitan la obtención de tal función (ver Fishburn (1970) y Chankong et al. (1983)), aspecto en el que no entraremos.

Planteado de esta manera el modelo, se trata de encontrar formas de ordenar las acciones empleando r e I. Teniendo en cuenta las propiedades de las evidencias, probaremos que es posible definir $G:D \longrightarrow R^2$ de tal manera que G recoja la valoración de las acciones y la información sobre los estados. Dada $G(d) = (G_1(d), G_2(d))$ comprobaremos que $G_1(d)$ es una "estimación pesimista de la recompensa posible", mientras que $G_2(d)$ es una estimación optimista de la recompensa posible".

En estas condiciones, la ordenación de alternativas puede llevarse a cabo por dos procedimientos:

1) empleando relaciones de orden definidas en R^2,
2) construyendo funciones valor sobre R^2,

en cualquiera de los cuales deberá tenerse en cuenta el optimismo o pesimismo del decisor, dadas las caracteristicas de G.

La memoria se estructura en dos grandes bloques. En el primero, constituido por los capítulos 1, 2 y 3, se estudian los fundamentos de la Teoria de Medidas Difusas y de la Evidencia y se desarrollan las herramientas necesarias para abordar la obtención de criterios de decisión en el modelo considerado, lo que constituye el objeto del segundo bloque (capítulo 4).

En el primer capítulo se hace un estudio general de las medidas difusas y sus propiedades. El concepto fundamental es el de dualidad, que establece una aplicación biyectiva del conjunto de medidas duales en si mismo. Partimos de la hipótesis de que una medida y su dual contienen la misma información sobre un mismo suceso, si bien codificada según un criterio diferente. El caracter de este criterio se pone de manifiesto cuando se trata de medidas ordenadas, en cuyo caso la inferior responde a una visión pesimista del tema, mientras que la superior se genera con un criterio optimista. En este sentido podemos afirmar que una medida difusa no asigna a un suceso un único valor, sino un intervalo de incertidumbre. Este es el caso, por ejemplo, de las parejas necesidad-posibilidad y creencia-plausibilidad. El intervalo se reduce a un punto cuando se trata de una medida de probabilidad.

No nos hemos limitado a realizar un estudio de las medidas difusas asociadas a la Teoria de la Evidencia por dos motivos. El primero es que pensamos que un planteamiento global del tema de las medidas difusas es necesario para obtener una buena visión y comprensión de las propiedades de las mismas. El segundo es que es necesario hacer un estudio comparativo de las distintas clases de medidas difusas para justificar la elección de las evidencias para representar la información en un modelo de decisión.

En el segundo capítulo se desarrollan las herramientas necesarias para el tratamiento de los problemas de decisión con experimentación. En primer lugar se estudia la transformación de la información mediante aplicaciones uni o multiva luadas. A continuación se estudian las medidas difusas bidimensionales y los problemas de consicionamiento e independencia.

Es usual que en Teoria de la Posibilidad se considere que la no-interacción es el concepto análogo al de independencia de medidas de probabilidad. Sin embargo, nosotros consideramos que son dos conceptos con distinto significado intuitivo, que han de definirse de modo distinto dando caracterizaciones de los mismos que sean aplicables por separado a cualquier evidencia que se considere, en particular a las medidas de probabilidad y a las posibilidades.

La definición de evidencia condicionada que aqui presentamos está basada en los resultados de Dempster (Dempster (1967)), y se engloba dentro de un concepto mas general como es el de combinación de evidencias. Condicionar una información a un conjunto es equivalente a combinar dos informaciones: la originalmente disponible y la que consiste en afirmar que el resultado del experimento pertenece al subconjunto mencionado.

Para combinar evidencias solo existe la denominada Regla de Dempster, aplicable cuando las informaciones que se combinan pueden considerarse independientes. El desarrollo de reglas de combinación mas generales nos permitirá el abordar modelos con experimentación menos restringidos.

Definiremos la posibilidad condicionada como una particularización de la evidencia condicionada. En la literatura especializada pueden encontrarse diferentes caracterizaciones del condicionamiento de posibilidades: Zadeh (1978), Hisdal (1979) y Nguyen (1979). Nuestra definición es comparada con cada una de ellas.

En el capítulo 3 abordamos el problema de reducción de los valores de una función utilizando la información proporcionada por una medida difusa, es decir, la definición de una "esperanza". De partida se consideran dos opciones: la in tegral de Sugeno y una generalización de la esperanza definida por Choquet para capacidades (Choquet (1953)) considerada por nosostros bajo el nombre de esperanza monotona.

La integral de Sugeno tiene un inconveniente fundamental: no generaliza la esperanza matemática clásica y no es lineal respecto a la suma y multiplicación de constantes. Por este motivo empleamos la esperanza monotona.

Comenzamos haciendo un estudio general de este operador para pasar despues a analizar la forma que toma sobre distintas clases de medidas difusas. En el caso particular de una evidencia, resulta que reproduce los operadores esperanza superior e inferior (que notamos $I_*(./.)$ e $I^*(./.)$ respectivāmente) introducidos por Dempster (1967). Se termina obteniendo expresiones de la esperanza monotona para medidas de posibilidad-necesidad y λ-medidas de Sugeno.

El capítulo 4 está dedicado al analisis del modelo de decisión que consideramos y a la construcción de reglas de decisión razonables de acuerdo con las hipótesis estructurales aceptadas y el optimismo o pesimismo del decisor.

Si admitimos que la información I vienen representada por una evidencia de Asignación Basica de Probabilidad M, entonces, haciendo uso de la esperanza monotona podemos esta blecer $g:R^n \longrightarrow R^2$ de tal manera que para cada r R, $g(r) = [I_*(r/M), I^*(r/M)]$, que permite asociar a cada decisión d la imagen de su vector de recompensas (aplicación G antes mencionada). Se estudian las propiedades de g y la justificación de su empleo para la obtención de reglas de decisión razonables.

A continuación pasamos a analizar los criterios de decisión que se obtienen al considerar ordenes en R^2 o emplear funciones valor $h:R^2 \longrightarrow R$. En el primer caso centramos nuestra atención en la dominancia según el cuadrante positivo y reglas de tipo lexicográfico, mientras que en el segundo contemplamos tres funciones valor; max(.,.), min(.,.) y la combinación lineal de coeficientes positivos.

Teniendo en cuenta la interpretación de g(r) y las téc nicas de ordenación que empleamos, se puede asegurar que es- tos criterios nos permiten atender al optimismo o pesimismo del decisor.

El modelo propuesto generaliza los ambientes de riesgo e incertidumbre y algunos de los criterios de decisión mas u suales en la Teoria de la Decisión clásica (el de la esperan za el de Wald o el Hurwickz) pueden obtenerse como casos par ticulares de los aqui obtenidos.

Terminamos el capítulo analizando algunos problemas a biertos y futuras vias de investigación. Entre ellas destaca el estudio de los problemas con experimentación, suponiendo que tanto la información "a priori" como la proporcionada -- por el experimento son de tipo general y no necesariamente - probabilisticas. Aqui aparecen problemas de combinación de - información que introducimos mediante algunos ejemplos.

- CAPITULO I -

Medidas difusas.
Teoria de la Evidencia.

0.-INTRODUCCION.

En éste capítulo estudiaremos los conceptos y propiedades de las medidas difusas que nos serán precisos para el desarrollo de nuestro problema de decisión bajo la hipótesis de que la incertidumbre sobre los estados de la Naturaleza, no está representada, necesariamente, mediante una distribución de probabilidad.

Comenzaremos, en el apartado I, estableciendo las bases de la teoria de Sugeno (1974) sobre medidas difusas, pero restringida al caso de conjuntos finitos, con objeto de evitar problemas de medibilidad y continuidad. Es bien sabido, por ejemplo que en el caso infinito, una medida de posibilidad no es necesariamente una medida difusa,(ver Puri y Ralescu (1982)).

La teoria de la evidencia de Shafer (1976) está basada en los trabajos de Dempster (1967) sobre probabilidades superiores e inferiores, proporcionando una representación de la incertidumbre mas general que la teoria de la probabilidad, la posibilidad o las λ-medidas de Sugeno. Es por ello que, el problema de decisión objeto de ésta memoria, lo plantearemos en términos de evidencia.

Esta teoria que sintetizamos en el apartado II, está teniendo un gran desarrollo en los últimos años dentro del campo de la teoria de los Conjuntos Difusos, como muestran los trabajos de Zadeh (1984), Dubois y Prade (1982-1985), Yager (1985).

Banon (1981) ha hecho una clasificación de las medidas difusas en la que la medida mas general que se considera son las evidencias. Nosotros, en el apartado III hemos completado éste cuadro con familias de valoraciones que son tambien medidas difusas aunque han sido desarrolladas fundamentalmente fuera de la teoria de Sugeno, como son por ejemplo las capacidades de orden dos de Choquet (1953).

Para ello, nos hemos basado en el concepto de dualidad, de forma que mas que medidas, hemos clasificado parejas de medidas difusas duales. Mantenemos el criterio de que una medida difusa y su dual contienen la misma información, salvo que codificada de forma distinta.

Dentro de las parejas duales haremos hincapie en las parejas ordenadas, donde el valor de una de las medidas es siempre menor o igual que la de su dual. Esto da cierto caracter de coherencia en el sentido de que una de ellas representa la incertidumbre con caracter pesimista y la otra con caracter optimísta. Este es por ejemplo, el caso de las medidas de Necesidad y Posibilidad, Creencia y Plausibilidad,etc.

I.-DEFINICIONES GENERALES SOBRE MEDIDAS DIFUSAS.

En éste apartado vamos a reunir los conceptos fundamentales sobre medidas difusas, centrando nuestra atención en aquellos aspectos que sean necesarios para el desarrollo de los capítulos posteriores. Un estudio mas amplio y detallado podrá encontrarse en Dubois y Prade (1980), Banon (1981).

Sea X un conjunto, que como ya hemos indicado supondremos finito; y además, consideramos un algebra $\mathcal{A} \subset \mathcal{P}(X)$.

Para comenzar daremos la definición de medida difusa Sugeno (1974), restringida a conjuntos finitos.

Definición 1.1.

Sea $h: \mathcal{A} \to [0,1]$; diremos que h es una medida difusa en $(X, \mathcal{A})$ si y solo si cumple:

$$\begin{aligned} &1)\ h(\phi)=0 \qquad h(X)=1 \\ &2)\ A \subset B \Rightarrow h(A) \leqslant h(B) \end{aligned} \qquad (1)$$

Vemos como una medida difusa es una función de conjunto que se diferencia de las medidas de probabilidad en que se ha generalizado la condición de aditividad considerando simplemente la monotonía.

Nota 1. La propiedad de continuidad de las medidas difusas · Si $\{A_i\}$ es una sucesión monótona $\Rightarrow \lim\limits_{i\to\infty} g(A_i) = g(\lim\limits_{i\to\infty} A_i)$ se omite, por considerar X finito.

Nota 2. Para evitar problemas de medibilidad y dado que X es finito, consideramos siempre que $\mathcal{A} = \mathcal{P}(X)$; y, por tanto, las medidas difusas estaran definidas en $(X, \mathcal{P}(X))$. Al conjunto de todas ellas le llamaremos MD(X).

Definición 1.2.

Sean $h, h^* \in MD(X)$. Diremos que son duales si y solo si

$$h^*(\bar{A}) = 1 - h(A) \qquad A \in \mathcal{P}(X) \qquad (2)$$

Este concepto de dualidad, es de gran importancia pues permite obtener representaciones alternativas de la incertidumbre, Así, cuando existe incertidumbre sobre un fenómeno, valorada a través de una medida difusa, la expresión (2) nos permitirá representarla tambien mediante su medida dual.

De ésta forma, una medida y su dual contienen la misma información sobre el fenómeno, salvo que la codificación de una y otra se hace desde distinto punto de vista; lo que nos lleva a la idea de considerar las medidas difusas por parejas.

Ejemplo 1.1.

Las medidas de probabilidad son casos particulares de medidas difusas, ya que:

$$P: \mathcal{P}(X) \rightarrow [0,1]$$

verificando que:

1) $P(\phi)=0 \qquad P(X)=1$

2) Si $A \subset B \qquad P(A) \leqslant P(B)$

y además, dada una P(A), su medida dual vendrá dada por

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

es decir $P^* \equiv P$.

t-norma y t-conorma.

Estos conceptos nos serán de gran utilidad en el desarrollo de las medidas difusas, por ello, recogemos en éste apartado las definiciones mas significativas relativas a los mismos. Para un estudio mas detallado puede verse Schweitzer y Sklar (1963) o Dubois y Prade (1982).

Es obvio que, de acuerdo con la definición 1.1, la clase de las medidas difusas ha de ser muy amplia y, como consecuencia, ha de ser muy dificil establecer criterios de caracterización y métodos de manejo válidos para todos sus elementos. Por éste motivo, resulta interesante la consideración de subclases con propiedades notables.

Sabemos que las medidas de probabilidad cumplen que:

$A, B \in \mathcal{A}$ con $A \cap B = \phi$ entonces:

$$P(A \cup B) = P(A)+P(B)$$

Será interesante considerar familias de medidas difusas que verifiquen una relación de la forma:

$$A,B\in\mathcal{A} \qquad A\cap B=\phi \quad \Rightarrow g(A\cup B)=g(A)*g(B) \qquad (3)$$

siendo $*$ un operador cerrado en $[0,1]$.

Estas medidas, análogamente a lo que ocurre con la probabilidad, seran tales que el grado de incertidumbre de $A\cup B$, con A y B disjuntos, unicamente dependerá del grado de incertidumbre de A y del grado de incertidumbre de B a través de$*$.

Es inmediato que existen una infinidad de operaciones binarias en $[0,1]$, si bien solo las que cumplen ciertas propiedades o condiciones de regularidad, serán útiles para la caracterización de medidas difusas.

De modo análogo, puede ser interesante considerar medidas tales que

$$A,B\in\mathcal{A} \qquad A\cup B=X \qquad g(A\cap B)=g(A)\otimes g(B)$$

siendo $\otimes$ una operación binaria en $[0,1]$ que verifique ciertas propiedades. En éste caso es la incertidumbre de $A\cap B$ la que depende unicamente de la incertidumbre de A y de la incertidumbre de B a traves de $\otimes$.

Definición 1.3.

Una aplicación

$$*:[0,1]\times[0,1] \quad \rightarrow \quad [0,1]$$

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

1) Propiedades de frontera

$$1*1=1 \quad , \quad 0*a=a*0=a \quad , \quad \forall a\in[0,1]$$

2) Monotonia

$$a\leq b \quad , \quad c\leq d \qquad a*c \;\; b*d \quad , \forall a,b,c,d\in[0,1]$$

3) Conmutatividad

$$a*b=b*a \quad , \quad \forall a,b \in [0,1]$$

4) Asociatividad

$$a*(b*c)=(a*b)*c \quad , \quad \forall a,b,c \in [0,1]$$

las principales conormas triangulares son:

– Operador Máximo: $(a,b) \rightarrow \max (a,b)$

–Suma Probabilística: $(a,b) \rightarrow a+b-a.b=S_p(a,b)$

–Suma Acotada: $(a,b) \rightarrow \min (1,a+b)=S_A(a,b)$

–Operador T_w: definido por las condiciones frontera y por

$$T_w(a,b)=1 \qquad \forall a,b \in (0,1)$$

Es inmediato comprobar que:

$$\max (a,b) \leqslant a+b-a.b \leqslant \min (1,a+b) \leqslant T_w^*(a,b) \quad \forall (a,b) \in [0,1]^2$$

<u>Definición</u> 1.4.

Una aplicación

$$\otimes : [0,1] \times [0,1] \rightarrow [o,1]$$

se dice que es una norma triangular (t-norma), si y solo si verifica las condiciones 2,3 y 4 anteriores y la propiedad - de frontera

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

las principales normas triangulares son:

–Operador Mínimo: $(a,b) \rightarrow \min (a,b)$

–Operador Producto: $(a,b) \rightarrow a.b$

–Operador S_A': $(a,b) \rightarrow \max (a.b-1,0)$

–Operador T_w': $T_w(a,b)=0 \quad \forall (a,b) \in [0,1)^2$ con la condición de frontera 1').

Es fácil comprobar que:

$$T_w(a,b) \leqslant \max (0,a+b-1) \leqslant a.b \leqslant \min (a,b)$$

Definición 1.5.

Sea g una medida difusa en X, diremos que g está basada en la conorma * si y solo si cumple:

$$A,B\subset X \quad , \quad A\cap B=\phi \qquad g(A\cup B)= g(A)*g(B)$$

Al conjunto de todas las medidas difusas asi caracterizadas le denominaremos $MD_{*}(X)$.

Definición 1.6.

Diremos que la medida difusa g' está basada en la norma $\otimes$ si cumple:

$$A,B\subset X \quad , \quad A\cup B=X \qquad g'(A\cap B)= g'(A) \otimes g'(B)$$

Llamaremos $MD_{\otimes}(X)$ al conjunto de las medidas difusas que están basadas en $\otimes$.

Propiedad.

Si g es una medida difusa basada en la conorma *, entonces su dual g', es una medida difusa basada en la norma *', definida por:

$$a*'b=1-(1-a) * (1-b)$$

A ésta norma *' se le llama norma dual de la conorma *.

—Si f es una medida difusa basada en la norma $\otimes$, entonces su dual f', es una medida difusa basada en la conorma *', definida por:

$$a \otimes b=1-(1-a) \otimes (1-b)$$

Aésta conorma $\otimes'$, se le llama conorma dual de la norma $\otimes$.

La siguiente tabla (Dubois (1982)) recoge las relaciones de dualidad entre las normas y conormas antes reseñadas:

conorma	norma
máximo	mínimo
suma probabilística	producto
suma acotada	S'_A
T_W	T'_W

Medida de probabilidad

Si como conorma * elegimos la suma acotada, obtenemos:

$$A,B\subset X \quad , \quad A\cap B=\phi \quad \Rightarrow \quad g(A\cup B)= \min\ (1,g(A)+G(B))$$

Por tanto:

$$g(A\cup\overline{A})=g(X)=1= \min\ (1,g(A)=g(\overline{A})) \Rightarrow$$

$$g(A)+g(\overline{A}) \geqslant 1$$

Si como norma $\otimes$ elegimos la dual de la anterior, el operador T_W= Max (0,a+b-1), entonces:

$$A,B\subset X \quad , \quad A\cup B=X \qquad g'(A\cap B)= \max\ (0,g'(A)+g'(B)-1)$$

$$g'(A\cap\overline{A})=g'(\phi)=0= \max\ (0,g'(A)+g'(\overline{A})-1)$$

$$g'(A)+g'(\overline{A}) \leqslant 1$$

Caso particular de éstas medidas son las probabilidades para las cuales

$$P(A)+P(\overline{A})=1 \qquad \forall\, A\in\mathcal{P}(X)$$

Medidas de Posibilidad

Las medidas difusas basadas en la t-conorma max(.,.) verifican:

$$g(A\cup B)= \max(g(A),g(B)) \qquad \forall\, A,B\in\mathcal{P}(X)$$

como facilmente puede comprobarse, no limitandose ésta relación al caso $A \cap B = \phi$.

Estas medidas han sido denominadas por Zadeh medidas de posibilidad, cuya clase notaremos PO(X).

Si Π es una medida de posibilidad sobre X, para todo -- $A \in \mathcal{P}(X)$ el valor $\Pi(A)$ se entiende como la "posibilidad de ocurrencia del suceso A" de acuerdo con una información difusa (ver Zadeh (1978)).

Es obvio que:

$$\forall A \in \mathcal{P}(X), \ \Pi(A \cup \bar{A}) = \Pi(X) = 1 = \max (\Pi(A), \Pi(\bar{A})) \qquad \forall \Pi \in PO(X)$$

de modo que $\Pi(A)$ o $\Pi(\bar{A})$ es igual a 1. Esto se interpreta diciendoque " de entre dos sucesos complementarios uno tiene - que ser siempre totalmente posible ", cualquiera que sea la información de que se disponga.

Dada $\Pi \in PO(X)$ podemos definir la denominada distribución de posibilidad asociada, π, como una aplicación $\pi : X \to [0,1]$ tal que $\pi(a) = \Pi(\{a\})$.

Es inmediato que:

$$\Pi(A) = \sup_{a \in A} \pi(a) \qquad \forall A \in \mathcal{P}(X)$$

<u>Medidas de Necesidad</u>

Son las duales de las anteriores, es decir, estan basadas en la t-norma dual del máximo que es el operador mínimo. Asi:

$$\forall A,B \in \mathcal{A} \qquad g'(A \cap B) = \min (g'(A), g'(B))$$

Notaremos NE(X) el conjunto de las medidas de necesidad sobre X. Siendo inmediato que:

$$N(A \cap \bar{A}) = \min (N(A), N(\bar{A})) = 0 \qquad \forall A \in \mathcal{P}(X), \ \forall N \in NE(X)$$

lo que significa que $N(A)$ y $N(\bar{A})$ no pueden ser, simultanea-- mente distintos de cero. Esto equivale a decir que " a lo su mo uno entre dos sucesos complementarios es algo necesario " cualquiera que sea la información disponible.

Por estar las medidas de posibilidad y de necesidad definidas a partir de t-conormas y t-normas duales, entre ella existirá la siguiente relación: para cada $\Pi \varepsilon PO(X)$ existe --- $N \varepsilon NE(X)$ tal que:

$$\forall A \qquad N(A) = 1- (\bar{A}) = \inf_{x \notin A} \{1-\pi(x)\}$$

y reciprocamente.

De quí, que la necesidad de un suceso A se considere como el grado de imposibilidad del suceso contrario.

Propiedad

UN suceso algo necesario debe de ser completamente posible.

En efecto:

Si un suceso es algo necesario $N(A)>0$, y por tanto, como $\min (N(A),N(\bar{A})) = 0$ deducimos que $N(\bar{A})=0$ de donde por dualidad resulta $\Pi(A)=1$.

Nota

A partir de ahora y como es habitual $\max\{a,b\}$ y $\min\{a,b\}$ lo notaremos $a \vee b$ y $a \wedge b$, respectivamente.

λ-medidas de Sugeno

Sugeno (1972) definió una clase de medidas difusas denominadas λ-medidas y caracterizadas por el siguiente axioma.

$A,B \varepsilon \mathcal{A}$ si $A \cap B=\phi$ entonces:

$$g_\lambda(A \cup B) = g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A) \circ g_\lambda(B) \qquad \text{con } \lambda > -1$$

correspondiendo a la conorma parametrizada.

$$a*b = \min(1, a+b+ab)$$

(ver Dubois y Prade (1980)).

—La medida dual de g_λ es g_μ, donde $\mu = -\lambda/(1+\lambda)$; en otras palabras

$$\forall A \subset \mathcal{A} \qquad g_{-\lambda/(1+\lambda)}(A) = 1 - g_\lambda(\overline{A})$$

—Si $\lambda \varepsilon (-1,0)$ la medida dual de g_λ, g_μ, es tal que $\mu \varepsilon (0,+\infty)$ y reciprocamente.

—Para $\lambda=0$ obtenemos las medidas de probabilidad. A la familia de λ-medidas la notaremos por $G\lambda(X)$.

II.- TEORIA DE LA EVIDENCIA DE SHAFER

En éste apartado, estudiamos algünas de las propiedades y definiciones mas importantes de la Teoria de la Evidencia. Un desarrollo mas detallado de los fundamentos de la misma puede encontrarse en Dempster (1967) y Shafer (1976).

El concepto fundamental en Teoria de la Evidencia es el de Asignación Básica de Probabilidad (A.B.P).

Definición 2.1

Una aplicación $m: \mathcal{P}(X) \to [0,1]$, se dice que es una asignación básica de probabilidad, si y solo si cumple que:

1) $m(\phi)=0$

2) $\sum_{A \subset X} m(A) = 1$ (4)

A la cantidad m(A) se la llama número básico de probabilidad de A.

Definición 2.2

Dada una A.B.P sobre X, diremos que $A\subset X$ es un elemento focal si y solo si $m(A)>0$.

—Las A.B.P no son en general monótonas, por tanto, no serán medidas difusas. Ahora bien, estas asignaciones básicas de probabilidad tienen siempre asociadas una pareja de medidas duales que si lo son.

Definición 2.3

Llamamos medida de creencia asociada a una asignación básica de probabilidad m , a la aplicación

$$\mathrm{Bel}: \mathcal{P}(X) \to [0,1]$$

dada por:

$$\mathrm{Bel}\,(A) = \sum_{B\subset A} m(B) \qquad (5)$$

Definición 2.4

Llamamos medida de plausibilidad asociada a una asignación básica de probabilidad m, a la aplicación:

$$\mathrm{Pl}: \mathcal{P}(X) \to [0,1]$$

dada por:

$$\mathrm{Pl}(A) = \sum_{B\cap A\neq \phi} m(B) \qquad (6)$$

Es inmediato comprobar que éstas aplicaciones son medidas difusas duales, es decir

$$\mathrm{Pl}(A) = 1-\mathrm{Bel}(\overline{A}) \qquad (7)$$

En el trabajo de Dempster (1967), éstas medidas eran llamadas P_* y P^* y se introducian por medio de una aplicación multivaluada de un espacio probabilístico en un conjunto.

Si consideramos un espacio $(\Omega,\mathcal{A},P)$ y una aplicación multivaluada $\Gamma:\Omega \to X$, es decir, una aplicación de Ω en $\mathcal{P}(X)$: podremos definir en X, asociada a Γ, las valoraciones P_* y P^* de la siguiente forma:

$$P_*(A) = \frac{P(A_*)}{P(X^*)} \qquad \text{con} \quad P(X^*)\neq \phi$$

$$P^*(A) = \frac{P(A^*)}{P(X^*)} \qquad \text{con} \quad P(X^*)\neq \phi$$

y siendo:

$$A_* = \{w\in \Omega \ / \ \Gamma w\neq\phi, \Gamma w \subset A\}$$

$$A^* = \{w\in \Omega \ / \ \Gamma w \cap A\neq\phi\}$$

Posteriormente, Shafer (1976), definió las medidas de creencia asociadas a las asignaciones básicas de probabilidad, comprobando que coincidian con las P_*. Las P^*, por ser duales de las anteriores, coincidiran con las medidas de plausibilidad.

Las propiedades mas importantes de éstas medidas son:

A) Las medidas de creencia son medidas difusas, verificando que:

$$\forall A,B \subset X \qquad Bel(A\cap B) \geqslant Bel(A)+Bel(B)-Bel(A\cup B) \qquad (8)$$

$$Bel(A\cap B) \geqslant \max\ (0,Bel(A)+Bel(B)-1) \qquad (9)$$

$$Bel(A\cap B) \leqslant \min\ (Bel(A),Bel(B)) \qquad (10)$$

La suma de la creencia de un suceso y la de su complementario será, en general, menor o igual que uno, no teniendo que darse la igualdad. Es decir:

$$Bel(A) + Bel(\overline{A}) \leqslant 1 \qquad (11)$$

la demostración de ésta desigualdad es inmediata ya que

$$Bel(A \cup \overline{A}) \geqslant Bel(A) + Bel(\overline{A}) - Bel(A \cap \overline{A})$$

y por tanto

$$1 \geqslant Bel(A) + Bel(\overline{A})$$

B) Las medidas de plausibilidad, son medidas difusas y verifican:

$$\forall A,B \subset X \qquad Pl(A \cup B) \leqslant Pl(A)+Pl(B)-Pl(A \cap B) \qquad (12)$$

$$Pl(A \cup B) \geqslant \max\ (Pl(A),Pl(B)) \qquad (13)$$

$$Pl(A \cup B) \geqslant \min\ (1,Pl(A)+Pl(B)) \qquad (14)$$

La plausibilidad de un suceso, mas la plausibilidad de su complementario es igual o mayor que uno.

$$\forall A \subset X \qquad Pl(A) + Pl(\overline{A}) \geqslant 1 \qquad (15)$$

la demostración es inmediata a partir de (12), y de los axiomas de las medidas difusas.

C) La relación que liga creencias y plausibilidades es la siguiente:

Si Bel y Pl son las medidas de creencia y plausibilidad asociadas a una cierta A.B.P m, entonces:

$$Bel(A) \leqslant Pl(A) \qquad \forall A \subset X$$

la demostración es inmediata con solo tener en cuenta la definición de medidas duales y (15).

Por otra parte, las medidas de creencia y de plausibilidad, pueden ser caracterizadas, como capacidades de Choquet de orden infinito, como muestran los siguientes teoremas.

<u>Teorema</u> 2.1

Una aplicación Bel : $\mathcal{P}(X) \to [0,1]$ es una medida de creencia si y solamente si cumple:

1) $Bel(\phi)=0$

2) $Bel(X)=1$

3) $\forall m \in N$, $\forall A_i \in X$, $i=1,\ldots,m$ (16)

se verifica que:

$$Bel(\bigcup_{i=1}^{m} A_i) \geq \sum_{i=1}^{m} Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots +$$

$$(-1)^{m+1} Bel(\bigcap_{i=1}^{m} A_i)$$

—La asignación básica de probabilidad asociada a ésta medida se puede obtener como:

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} Bel(B) \qquad \forall A \subset X \qquad (17)$$

siendo $|A|$ el cardinal del conjunto A.

<u>Teorema</u> 2.2

Una aplicación Pl : $\mathcal{P}(X) \to [0,1]$ es una medida de plausibilidad si y solo si cumple:

1) $Pl(\phi)=0$

2) $Pl(X)=1$

3) $\forall m \in N$, $\forall A_i \in X$, $i=1,\ldots,m$ (18)

se verifica que:

$$Pl(\bigcap_{i=1}^{m} A_i) \leq \sum_{i=1}^{m} Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \ldots +$$

$$(-1)^{m+1} Pl(\bigcup_{i=1}^{m} A_i)$$

Como consecuencia de (5), (7), y (15) se puede comprobar que existe una correspondencia biunívoca entre las asignaciones básicas de probabilidad, las medidas de creencia y las de Plausibilidad, cuando están definidas sobre el mismo conjunto X. Por tanto, podemos considerar que cada una de ellas contiene la misma información pero basandose en diferentes criterios para su representación.

Bel(A) nos mide la mínima creencia que podemos tener sobre la pertenencia a A de un cierto elemento desconocido, mientras que m(A) mide la creencia que, exactamente, se asigna a A, sin contar la que se asigna a sus partes propias; y Pl(A) el máximo grado de creencia que se puede tener sobre la pertenencia a A de un elemento desconocido.

Teniendo en cuenta éstas ideas, cuando a lo largo de la memoria nos refiramos a una evidencia sobre X, estaremos haciendo alusión a cualquiera de sus representaciones, que podrán ser empleadas indistintamente.

Evidencia probabilística

Una medida de probabilidad P en X, es a la vez una medida de Plausibilidad y de creencia que tiene por elementos focales en la A.B.P a los singletones o conjuntos unitarios no imposibles, es decir, todos aquellos $\{x\}$, $x \in X$ tales que:

$$m(\{x\}) = P(\{x\}) = p(x) \neq 0$$

donde p es la distribución de probabilidad asociada a la medida P.

Reciprocamente, toda evidencia cuyos únicos elementos focales sean los conjuntos unitarios es tal que las medidas de creencia y plausibilidad asociadas, coinciden con una medida de probabilidad. Por ello, a éstas evidencias las llamaremos probabilísticas.

Evidencia Posibilística

Las medidas de necesidad y de posibilidad son medidas de creencia y de plausibilidad respectivamente. Esta evidencia se caracteriza por tener sus elementos focales anidados, es decir, existen $A_i \in \mathcal{P}(X)$, i=1,...,n tales que:

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

Si Π es una medida de posibilidad, π su distribución asociada, y si $X = \{x_1,\ldots,x_n\}$ está ordenado de forma que:

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

entonces la asignación básica de probabilidad asociada a Π, m_π , es:

$$m_\pi(A) = \begin{cases} \pi(x_i)-\pi(x_{i+1}) & \text{si } A=\{x_1,\ldots,x_i\},\ i=1,..,n-1 \\ \pi(x_n) & \text{si } A=X \\ 0 & \text{en otro caso} \end{cases} \qquad (19)$$

De forma inmediata podriamos comprobar que sus elementos focales están anidados.

Reciprocamente, toda asignación básica tal que sus elementos focales estén anidados, tienen medidas de creencia y de plausibilidad asociadas que son medidas de necesidad y de posibilidad, respectivamente.

Para mas detalles sobre éste tipo de medidas puede consultarse Dubois y Prade (1981) y Moral (1985).

Ignorancia

La asignación básica de probabilidad dada por:

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

representa la información nula sobre las características de un elemento desconocido $x\varepsilon X$. Shafer resalta que ésta representación de la ignorancia es diferente de la que se realiza en la Teoria Bayesiana; donde la ignorancia se supone caracterizada por la distribución uniforme sobre X.

La medida de creencia asociada a ésta asignación básica de probabilidad es:

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

y la de plausibilidad

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

Puede comprobarse facilmente que $Pl_o(\cdot)$, es una medida de posibilidad con distribución asociada

$$\pi(a)=1 \quad , \quad \forall\, a\, \varepsilon X$$

Inclusión de evidencias

Esta definición ha sido dada independientemente en Moral (1985) y Yager (1985). Un estudio detallado puede encontrarse en M.Delgado y S.Moral (1985).

Definición 2.5

Sean m_1 y m_2 dos asignaciones básicas de probabilidad. Diremos que la evidencia representada por m_1 está incluida - en la representada por m_2 y notaremos $m_1 \subset m_2$, si y solo si:

$\forall A \subset X$, $m_A : \mathcal{P}(A) \to [0,1]$ tal que

$$m_1(A) = \sum_{B \subset A} m_A(B) \qquad (20)$$

$$m_2(B) = \sum_{A \supset B} m_A(B) \qquad (21)$$

La idea que inspira ésta caracterización es que cual -- quier información adicional sobre un $x \varepsilon X$ dseconocido, con la condición de ser compatible con la ya existente, debe de producir una atomización de la evidencia.

Así, cuando nuestra información es nula (ignorancia), - toda la masa de evidencia está concentrada en el total X, y a medida que ganamos información, la evidencia se nos va repartiendo entre las distintas partes de X.

Propiedad

Si m_1 y m_2 son A.B.P tales que $m_1 \subset m_2$ y Pl_1 y Pl_2 son - sus medidas de plausibilidad asociadas, se verifica que:

$$\forall C \subset X \qquad Pl_1(C) \geqslant Pl_2(C)$$

En efecto:

De acuerdo con (21):

$$\forall C \subset X \qquad Pl_2(C) = \sum_{C \cap B \neq \phi} m_2(B) = \sum_{C \cap B \neq \phi} \left(\sum_{A \supset B} m_A(B) \right)$$

Ahora bien, si $A \subset X$ es tal que $B \subset A$ y $C \cap B \neq \phi$, entonces tambien $C \cap A \neq \phi$, con lo que:

$$\sum_{C\cap B\neq\phi}\Big(\sum_{A\supset B} m_A(B)\Big) \leqslant \sum_{C\cap B\neq\phi}\Big(\sum_{A\supset B} m_A(B)\Big) =$$

$$\sum_{C\cap A\neq\phi} m_1(A) = Pl_1(C)$$

Propiedad

Si m_1 y m_2 son dos A.B.P. tales que $m_1 \subset m_2$ y tienen como medidas de creencia asociadas Bel_1 y Bel_2 respectivamente entonces se verifica que:

$$\forall C \subset X \qquad Bel_1(C) \leqslant Bel_2(c)$$

la demostración es inmediata con solo tener en cuenta que la medida de creencia es dual de la de plausibilidad.

Desgraciadamente el recíproco delas anteriores propiedades no se cumple, en general, siendo cierto solo en algunos casos particulares.

Propiedad

Sean π_1 y π_2 dos distribuciones de posibilidad sobre X. La evidencia representada por π_1 está contenida en la representada por π_2 si y solo si:

$$\forall a \in X \qquad \pi_1(a) \geqslant \pi_2(a)$$

Propiedad

Si Pl y P son una medida de plausibilidad y probabilidad respectivamente, con asignaciones básicas de probabilidad asociadas m y m_p, entonces:

$$Pl(A) \geqslant P(A) \ , \quad \forall A \subset X \Rightarrow m \subset m_p$$

Para su demostración consultar M.Delgado y S.Moral (1985).

III.- ESTUDIO DE FAMILIAS DE MEDIDAS DIFUSAS

El objeto de éste apartado es completar el esquema de Banon (1981), con nuevas familias de medidas difusas que consideramos de interés. Nuestra clasificación, aunque basada en la de Banon, se diferencia de ésta en que introduce y se basa en el concepto de dualidad entre medidas.

3.1. Parejas de medidas duales

Como ya hemos dicho, para toda $g \in MD(X)$ existe su dual $\overline{g} \in MD(X)$. De éste modo podemos considerar la familia PD(X) formada por las parejas $(g,\overline{g})$ donde g y $\overline{g}$ son medidas difusas duales, es decir:

$$g(A) + \overline{g}(\overline{A}) = 1 \qquad \forall A \subset X$$

En general, el conocimiento de una de ellas nos es suficiente para el conocimiento del par; por tanto algunas veces para referirnos a los elementos de PD(X) daremos simplemente una de las dos posibles medidas. Por otra parte, ésto nos permite asegurar que PD(X) y MD(X) pueden ponerse siempre en correspondencia biunívoca.

Es inmediato comprobar que PD(X) es una familia convexa, ya que:

$$(\overline{\lambda g_1 + (1-\lambda)g_2}) = \lambda\overline{g}_1 + (1-\lambda)\overline{g}_2 \qquad \forall \lambda \in [0,1]$$

3.2. Parejas de medidas ordenadas

Definimos la familia PO(X) como el conjunto de las parejas (g_*,g^*) donde g_* y g^* son duales y además

$$g_*(A) \leqslant g^*(A) \qquad \forall A \subset X$$

Por dualidad, toda medida difusa g, asocia siempre a cada conjunto A dos valores de incertidumbre g(A) y -------

$\overline{g}(A)=1-g(\overline{A})$. En la familia PO(X) hemos considerado aquellas parejas en las que el inferior del par de valores correspon de siempre a la misma medida.

Asi, si $(g_*,g^*)\varepsilon PO(X)$ podemos decir que g_* codifica la información con un caracter que llamaremos pesimísta, mientras que, g^* lo hará de forma "optimísta".

Podemos definir una proyección de PD(X) en PO(X), que asocia a un par de medidas difusas cualesquira un elemento de PO(X) del siguiente modo:

$$\psi: PD(X) \mapsto PO(X)$$
$$(g,\overline{g}) \mapsto (g_*,g^*)$$

donde:

$$g_*(A) = \text{Min}\ \{g(A),\overline{g}(A)\} \qquad A\varepsilon\mathcal{P}(X)$$

$$g^*(A) = \text{Max}\ \{g(A),\overline{g}(A)\} \qquad A\varepsilon\mathcal{P}(X)$$

Propiedad

Es inmediato comprobar que g_* y g^* son medidas difusas. Obviamente $g_*(A)\leqslant g^*(A)\quad \forall A$.
Por otra parte, son duales ya que:

$$1-g_*(\overline{A}) = 1-\text{Min}\ \{g(\overline{A}),\overline{g}(\overline{A})\} = \max\ \{1-g(\overline{A}),1-\overline{g}(\overline{A})\}=$$

$$\text{Max}\ \{\overline{g}(A),g(A)\} = g^*(A)$$

Propiedad

La familia PO(X) es convexa ya que si $g_*(A)\leqslant g^*(A)$ y $h_*(A)\leqslant h^*(A)\quad \forall A\subset X$, entonces:

$$\forall\lambda\varepsilon[0,1]\quad \lambda g_*(A) + (1-\lambda)h_*(A) \leqslant \lambda g^*(A) + (1-\lambda)h^*(A)$$

Como hemos indicado si tenemos un par de medidas difusas ordenadas, (g_*,g^*), a cada conjunto A se le asocian dos valores de incertidumbre: $g_*(A)$ y $g^*(A)$ con $g_*(A) \leqslant g^*(A)$. - Tendremos una medida mas precisa cuanto mas pequeñas sean - las diferencias $g^*(A)-g_*(A)$. Por ésta motivo, cabe hacer la siguiente definición.

<u>Definición</u> 3.1

Dada una pareja de medidas difusas ordenadas (g_*,g^*), definimos su grado de imprecisión como el valor:

$$Im(g_*,g^*) = \frac{\sum_{A\subset X} [g^*(A) - g_*(A)]}{2^{|X|} - 2}$$

<u>Teorema</u> 3.1

Si P es una medida de probabilidad, $Im(P,P)=0$

La demostración es inmediata sin mas que tener en cuenta -- que las probabilidades son medidas difusas autoduales.

<u>Teorema</u> 3.2

La $Im(g_*,g^*)=1$, si y solo si g_* y g^* tienen la forma:

$$g_*(A) = \begin{cases} 1 & \text{si } A=X \\ 0 & \text{si } A\neq X \end{cases} \qquad g^*(A) = \begin{cases} 0 & \text{si } A=\phi \\ 1 & \text{si } A\neq \phi \end{cases}$$

es decir, coinciden con las medidas de creencia y de plausi bilidad, (Bel_0,Pl_0), asociadas a la A.B.P que representa la ignorancia.

Demostración

Es inmediata sin mas que tener en cuenta la definición de Im(.,.) y de la ignorancia.

Como consecuencia de éstos resultados y en cierto sentido, podremos afirmar que el grado de imprecisión mide "lo lejana" que está una cierta pareja de medidas duales de una medida autodual. De ésta forma, diremos que cuanto mas proximo esté a cero el grado de imprecisión, mejor será nuestra representación de la incertidumbre.

3.3. Parejas de medidas duales aditivo-coherentes

Definimos la familia AC(X) como el conjunto de todas las parejas ordenadas (g_*,g^*), que verifican:

$$g^*(A\cup B) \leqslant g^*(A) + g^*(B)$$

$$g_*(A\cup B) \geqslant g_*(A) + g_*(B)$$

$\forall\ A,B\subset X$ con $A\cap B=\varphi$

Estas medidas estan consideradas en Huber (1981).

En AC(X) no pueden ocurrir situaciones como las que muestra el siguiente ejemplo:

Ejemplo 3.1

Sea $X=\{x_1,x_2,x_3\}$ y el par de medidas ordenadas en X, (g_*,g^*), dado por:

$g^*\{x_1\}=g^*\{x_2\}=g^*\{x_3\}=0.3$

$g^*\{x_1,x_2\}=g^*\{x_1,x_3\}=g^*\{x_2,x_3\}=0.9$

$g^*\{x_1,x_2,x_3\}=1$ y;

$g_*\{x_1\}=g_*\{x_2\}=g_*\{x_3\}=0.1$

$g_*\{x_1,x_2\}=g_*\{x_1,x_3\}=g_*\{x_2,x_3\}=0.7$

$g_*\{x_1,x_2,x_3\}=1$

Observese que, la medida superior de $\{x_1\}$ y de $\{x_2\}$, vale 0.3, mientras que la inferior de $\{x_1,x_2\}$ vale 0.7, lo cual parece una contradicción con la idea de incertidumbre aditiva en algún sentido.

Propiedad

La familia AC(X) es convexa.

Demostración:

$$\begin{aligned}[\lambda g_1^* + (1-\lambda)g_2^*](A\cup B) &= \lambda g_1^*(A\cup B) + (1-\lambda)g_2^*(A\cup B)\\ &\leqslant \lambda(g_1^*(A) + g_1^*(B)) + (1-\lambda)(g_2^*(A) + g_2^*(B))\\ &= (\lambda g_1^* + (1-\lambda)g_2^*)(A) + (\lambda g_1^* + (1-\lambda)g_2^*)(B)\end{aligned}$$

Propiedad

Dada la pareja (g_*,g^*), si pertenece a AC(X) y es tal que $g_*\equiv g^*$, entonces g_* y g^* coinciden con una única medida de probabilidad P.

En efecto, si

$$g_*(A\cup B) = g^*(A\cup B) \qquad \forall A,B$$

entonces $g_*\equiv g^*$ es una medida aditiva y normalizada, es decir una probabilidad.

3.4. Familia de medidas que acotan una probabilidad

Definimos la familia PA(X) como el conjunto de parejas duales (g_*,g^*) tales que existe una medida de probabilidad verificando:

$$g_*(A) \leqslant P(A) \leqslant g^*(A) \qquad \forall A\subset X$$

siendo ésta una condición de consistencia alternativa a la -

de la familia AC(X).

Esta familia y la anterior no coinciden, Pero, $AC(X) \cap PA(X)$, contendrá a las familias que consideraremos -- posteriormente.

La caracterización de éstas medidas fué dada por Huber (1981), mediante el siguiente teorema:

<u>Teorema</u> 3.3

Si (g_*, g^*) son medidas difusas duales, entonces, $(g_*, g^*) \in PA(X)$ si y solo si cumple:

$$\sum a_i I_{A_i} \leqslant 1 \ , \quad a_i \geqslant 0 \ , \quad A_i \subset X \quad \Rightarrow \sum a_i g_*(A_i) \leqslant 1$$

donde I_{A_i} es la función indicador de A_i.

<u>Propiedad</u>

La familia PA(X) es convexa.

En efecto:

Si $\lambda \in [0,1]$ y

$$g_{*1}(A) \leqslant P_1(A) \leqslant g_1^*(A)$$

$$g_{*2}(A) \leqslant P_2(A) \leqslant g_2^*(A)$$

entonces

$$\lambda g_{*1}(A) + (1-\lambda) g_{*2}(A) \leqslant \lambda P_1(A) + (1-\lambda) P_2(A)$$

$$\leqslant \lambda g_1^*(A) + (1-\lambda) g_2^*(A)$$

Si P_1 y P_2 son medidas de probabilidad, su combinación convexa tambien lo es, es decir:

$$\lambda P_1(A) + (1-\lambda) P_2(A) = P_3(A)$$

y $P_3(A)$ estará acotada por la combinación convexa de (g_{*1}, g_1^*) y (g_{*2}, g_2^*).

lo que prueba que $\lambda(g_{*1},g_1^*) + (1-\lambda)(g_{*2},g_2^*)$ pertenece a -- PA(X).

Propiedad

Si una pareja de PA(X) tiene sus componentes autoduales entonces, ambas medidas coinciden con la medida de probabilidad a la que acotan.

3.5 Medidas representables

Definimos la familia MR(X) como el conjunto de todas - las parejas (g_*,g^*) que sean representables, es decir, para las que existe $\mathcal{P}$ (familia de medidas de probabilidad), tal que:

$$g^*(A) = \sup_{P\in\mathcal{P}} P(A)$$

$$g_*(A) = \inf_{P\in\mathcal{P}} P(A)$$

Estas medidas fueron introducidas por Dempster (1967) y posteriormente caracterizadas por Wolf (1977). Giles ha llegado a la misma caracterización, estudiandolas desde un punto de vista diferente.

Teorema3.4

Una pareja de medidas difusas (g_*,g^*) es representable por alguna familia $\mathcal{P}$ si y solo si cumple:

$$I_A \leqslant \sum a_i I_{A_i} - a \qquad \text{con } a_i \geqslant 0 \text{ y } a\in R \qquad g^*(A) \leqslant \sum a_i g^*(A_i) - a$$

Propiedad

La familia de medidas representables es convexa.

Es obvio que:

si (g_{*1}, g_1^*) es representable por $\mathcal{P}_1$ y

(g_{*2}, g_2^*) es representable por $\mathcal{P}_2$

entonces, la combinación convexa de las dos parejas es representable por:

$$\mathcal{P} = \{\lambda P_1 + (1-\lambda) P_2 \ / \ P_1 \varepsilon \mathcal{P}_1 \ , \ P_2 \varepsilon \mathcal{P}_2 \}$$

Nota

La familia MR(X)está incluida en la intersección de --- AC(X) y de PA(X), pero no coincide con ella: como muestra el siguiente ejemplo extraido de Huber.

Sea un conjunto con cuatro elementos y supongamos que - $g_*(A)$ y $g^*(A)$ dependen solo del cardinal de A según la si - guiente tabla;

A	0	1	2	3	4
g_*	0	0	1/2	1/2	1
g^*	0	1/2	1/2	1	1

Vemos que $(g_*, g^*) \in AC(X)$, pero existe una única probabilidad acotada entre g_* y g^* que es $P(A) = |A| \ / \ 4$.

Por tanto (g_*, g^*) no es representable.

Nota

Un elemento (g_*, g^*) de MR(X), puede admitir distintas representaciones.

3.6. Capacidades de orden 2

Definimos la familia C2(X), como el conjunto de pares - (g_*, g^*) de medidas difusas, tales que:

$$g^*(A\cup B) \leqslant g^*(A) + g^*(B) - g^*(A\cap B)$$

$$g_*(A\cup B) \geqslant g_*(A) + g_*(B) - g_*(A\cap B)$$

es decir:

g^* es una capacidad alternante de orden 2, y

g_* es una capacidad monótona de orden 2.

(ver Choquet (1953-1954)

Propiedad

Se puede comprobar que $C2(X) \subset MR(X)$.

Para su demostración ver Huber (1981).

Propiedad

El conjunto de las capacidades de orden dos es una familia convexa.

Omitiremos la demostración por ser análoga a la de la familia AC(X).

3.7. Otras familias de medidas duales

En éste apartado consideraremos las familias de medidas difusas que han sido definidas con anterioridad.

- Evidencias

Notaremos EV(X) la familia de medidas duales (Bel,Pl), tales que: Bel y Pl son las medidas difusas de creencia y plausibilidad, asociadas a una misma asignación básica de probabilidad.

Propiedad

La familia EV(X) es convexa.

Demostración

Si Bel_1 , Pl_1 están asociadas con la A.B.P m_1

Bel_2 , Pl_2 están asociadas con la A.B.P m_2

entonces el par

$$(\lambda Bel_1 + (1-\lambda)Bel_2 \ , \ \lambda Pl_1 + (1-\lambda)Pl_2)$$

está asociado a $\lambda m_1 + (1-\lambda)m_2$.

Por otra parte:

1) $[\lambda m_1 + (1-\lambda)m_2](\phi) = \lambda m_1(\phi) + (1-\lambda)m_2(\phi) = 0$

$[\lambda m_1 + (1-\lambda)m_2](X) = \lambda m_1(X) + (1-\lambda)m_2(X) = 1$

2) $\sum_{A\subset X}[\lambda m_1 + (1-\lambda)m_2](A) = \lambda\sum_{A\subset X} m_1(A) + (1-\lambda)\sum_{A\subset X} m_2(A)$

$= \lambda + (1-\lambda) = 1$

lo que prueba que $\lambda m_1 + (1-\lambda)m_2$ es una A.B.P. y por tanto la propiedad.

Evidencias consonantes

Notaremos EC(X) la familia de parejas de medidas difusas ordenadas (N,Π), siendo Π una medida de posibilidad y N una necesidad. Esta familia no es convexa, si bien su clausura convexa es EV(X).

Evidencias de tipo "crisp"

Notaremos CR(X) a la familia de parejas (N,Π) donde N y Π son medidas de necesidad y de posibilidad que unicamente pueden tomar los valores $\{0,1\}$. Están asociadas a A.B.P - con un único elemento focal $A\subset X$. Esta familia tampoco es -- convexa, pero su clausura convexa coincide con EV(X). (ver Dubois y Prade (1985)).

λ- medidas de Sugeno

Notaremos $G\lambda(X)$, al conjunto de las parejas (g_λ, g_μ) con $\lambda \varepsilon [0,\infty)$ y $\mu = -\lambda / 1+\lambda \varepsilon [-1,0]$.

Al igual que pasa con las familias anteriores se puede - demostrar que $G\lambda(X)$ no es convexa.

Banon (1978) probó que $G\lambda(X) \subset EV(X)$.

medidas de probabilidad

Notaremos $PR(X)$ al conjunto de medidas $(P.P)$, donde P es una medida de probabilidad; por simplicidad designaremos al par $(P,P) \equiv P$.

Evidentemente, $PR(X) \subset G\lambda(X)$ y además es una familia -- convexa.

Medidas de Dirac

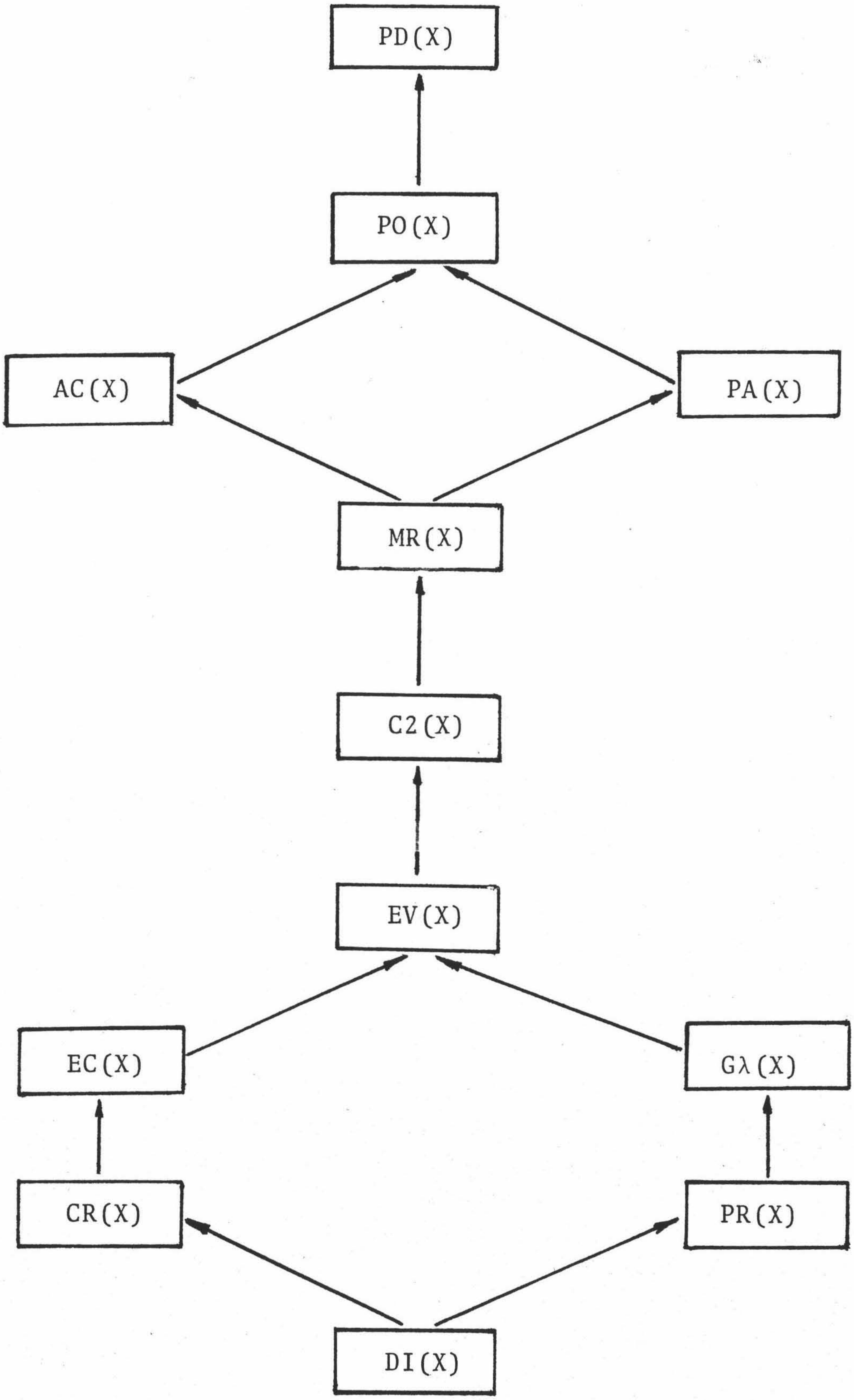
Notaremos $DI(X)$ al conjunto formado por las medidas autoduales δ_a , $a \varepsilon X$, dadas por:

$$\delta_a(A) = \begin{cases} 1 & \text{si} \quad a \varepsilon A \\ 0 & \text{si} \quad a \not\varepsilon A \end{cases}$$

es decir, δ_a es una medida de probabilidad degenerada en a.

Es inmediato comprobar que $DI(X) = PR(X) \cap CR(X)$, no - es una familia convexa, pero en ésta caso su clausura convexa es $PR(X)$.

- Para terminar y a modo de resumen, damos el siguiente diagrama donde se relacionan éstas medidas. El sentido de las - flechas indica inclusión.


PD(X)
PO(X)
AC(X)
PA(X)
MR(X)
C2(X)
EV(X)
EC(X)
Gλ(X)
CR(X)
PR(X)
DI(X)

Un problema de decisión será tanto mas general cuanto -- mas amplia sea la familia de medidas que consideremos para la representación de la información disponible sobre un fenómeno Sin embargo, dispondremos de estructuras mas ricas, con mas - propiedades, cuando consideremos familias mas restringidas.

Nosotros, hemos planteado el problema de decisión en -- EV(X), porque creemos que en ella se compensan bien éstos dos aspectos: es bastante general, permitiendo al mismo tiempo, - el desarrollo de todas las herramientas necesarias para el -- problema de decisión.

- CAPITULO II -

Medida difusa bidimensional.
Medida difusa condicionada.

0.- INTRODUCCION

En el capítulo anterior hemos estudiado las medidas difusas como representaciones o soportes de la información disponible sobre un determinado fenómeno.

Es concebible la existencia de varias fuentes de información diferentes sobre un mismo hecho, surgiendo de modo inmediato, el problema de combinación de las mismas.

La combinación y como caso particular el condicionamiento de informaciones, serán esenciales para la resolución de los problemas de decisión bajo la hipótesis de existencia de experimentación. Estos temas serán abordados de la forma mas general posible en éste capítulo, considerandolos en las distintas familias de medidas difusas definidas en el capítulo anterior.

Si g es una medida difusa, y $B \subset X$, la medida difusa condicionada $g(./B)$ puede interpretarse como la combinación de de nuestra información original, g, con una información posterior que afirma que el resultado del experimento está en B. Por tanto, el condicionamiento es un caso particular del problema de la combinación de la información. Desgraciadamente, éste es un tema que aun no está muy desarrollado. Existen pro cedimientos particulares, como la regla de Dempster (1967) - para evidencias independientes; o como los desarrollados en Zadeh (1975) y Moral (1985) para medidas de posibilidad. El desarrollo de reglas generales de combinación nos permitirá resolver el problema de decisión con experimentación, cuando nuestra observación del resultado del experimento no sea totalmente precisa.

El apartado 1 está dedicado ha hacer una breve reseña - del tema de la combinación de evidencias y presentar la re-- gla de Dempster.

Para la definición de medida difusa condicionada (apartado 2) partimos de la formulación de Dempster (1967) para - evidencias, generalizandolo al caso de medidas difusas ordenadas. Nos encontramos que ésta definición no se puede extender al caso de una medida difusa general ya que previamente necesitamos conocer el carácter pesimísta u optimísta de dicha medida.

Un problema importante e intuitivamente relacionado con los que analizamos en éste capítulo es el paso de la información de un espacio a otro relacionado con el.

En el apartado tercero, comenzamos estudiando la transformación de una medida difusa mediante una aplicación univaluada o multivaluada como base para el estudio de las medidas difusas bidimensionales.

Posteriormente, analizamos detenidamente el condicionamiento y la independencia de evidencias, encontrando dos formas o grados de independencia: la débil y la fuerte. En el - caso particular de una medida de posibilidad, comparamos --- nuestra definición de posibilidad condicionada con las dadas por Zadeh (1978), Hisdall (1978-1979) y Nguyen (1978-79).

Para terminar, destacamos que la no interacción en distribuciones de posibilidad, es en general considerada, como el concepto análogo al de independencia de la Teoria de la -

Probabilidad. Nosotros, tratamos éstos dos conceptos como nociones bien diferenciadas aplicables a todas las medidas difusas, y en particular, a las medidas de probabilidad y de posibilidad. Consideraremos, que las medidas son independientes - cuando el conocimiento de una de ellas, no influye sobre la otra. Las medidas serán no interactivas (descomponibles según nuestra terminologia), cuando la medida global es la mínima - que contiene a las marginales.

1.- COMBINACION DE EVIDENCIAS. LA REGLA DE DEMPSTER

En éste apartado vamos a estudiar sucintamente el problema de combinar dos evidencias (informaciones), sobre un mismo experimento, representadas por sendas A.B.P m_1 y m_2. Para mas detalles sobre el tema puede verse Dempster (1967), Shafer -- (1985) y Delgado y Moral (1985).

La primera y mas conocida forma de abordar la combinación de evidencias está constituida por la regla de Dempster.

Suponiendo que m_1 y m_2 tienen elementos focales $A_1,..,A_k$ y $B_1,...,B_l$, 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 Demps-- ter, proporciona

$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i).m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i).m_2(B_j)} \qquad (22)$$

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

A ésta asignación, definida por la expresión (22), la no taremos $m_1 \oplus m_2$.

La regla se ha enunciado en términos de asignaciones básicas de probabilidad, pero se extiende 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 cabria decir para -- las funciones de plausibilidad, $Pl_1 \oplus Pl_2$, asociadas a $m_1 \oplus m_2$.

Ejemplo 2.1

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

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

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

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

donde $B \subset X$ es un conjunto no vacio arbitrario, pero fijo.

m_1 representa la evidencia que tenemos sobre un elemento cuando, sobre el, disponemos de la distribución de probabilidad $\{p_i\},\{i \in 1,\ldots,n\}$. m_2 representa la evidencia que tenemos cuando sabemos que dicho elemento pertenece a $B \subset X$.

La regla de Dempster nos dice que no existe contradic -- ción 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 éste caso, la asignación bá sica resultante es:

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

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 mas importantes de la regla de Dempster.

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_o representa la ignorancia, entonces

$m_o \oplus m \equiv m$, para toda asignación m.

4.- Si m_a representa la A.B.P asociada a la medida de Dirac 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 $m_a \oplus m \equiv m_a$.

La regla de Dempster ha demostrado funcionar bien en numerosas situaciones, como es el caso del ejemplo 2.1. Sin embargo, varias razones nos inclinan a no dar 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 éstas 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, de acuerdo con ella, $m \oplus m \neq m$. Shafer ha estudiado algunos requisitos para la validez de ésta regla La combinación de evidencias cualesquiera se presenta pues, como un problema abierto.

A continuación y aunque no es el objeto de ésta memoria vamos a tratar un caso de fácil resolución.

Vamos a poner de relieve como las evidencias de tipo crisp se "combinan bien". La combinación de una evidencia de tipo crisp representada por una asignación, m_B, con una evidencia arbitraria dada por m, se define como aquella que viene reprensentada 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 incluido en B.

Siendo la condición de compatibilidad

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

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 éste caso particular no presenta los problemas achacables a ésta. En primer lugar, ésta definición parece intuitivamente bien fundamentada, lo que no ocurre en general. Además, en éste 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, tambien, la justificación de la probabilidad condicionada por -

la regla de Dempster, donde una de las evidencias era de tipo crisp y estábamos en éste caso particular.

2.- MEDIDA DIFUSA CONDICIONADA

Las definiciones de probabilidad superior e inferior condicionadas, dadas por Dempster (1967)

$$P^*(A/B) = \frac{P^*(A\cap B)}{P^*(B)}$$

$$P_*(A/B) = 1 - P^*(\overline{A}/B) = 1 - \frac{P^*(\overline{A}\cap B)}{P^*(B)} \qquad (23)$$

siendo $P^*(B) \neq 0$.

son formalmente trasladables a la clase de medidas difusas ordenadas PO(X), en la forma siguiente:

$$g^*(A/B) = \frac{g^*(A\cap B)}{g^*(B)}$$

$$g_*(A/B) = 1 - g^*(\overline{A}\cap B) \;/\; g^*(B) \qquad (24)$$

siendo $(g_*, g^*) \varepsilon$ PO(X) y $g^*(B) \neq 0$.

Ahora bien, el par $(g_*(./B), g^*(./B))$, no necesariamente pertenece a la clase PO(X), como muestra el siguiente contraejemplo.

Sea $X = \{x_1, x_2, x_3\}$

$g_*\{x_1\} = g_*\{x_2\} = g_*\{x_3\} = 0.1$

$g_*\{x_1,x_2\} = g_*\{x_1,x_3\} = g_*\{x_2,x_3\} = 0.6$

$g_*\{x_1,x_2,x_3\} = 1$

$g^*\{x_1\} = g^*\{x_2\} = g^*\{x_3\} = 0.2$

$g^*\{x_1,x_2\} = g^*\{x_1,x_3\} = g^*\{x_2,x_3\} = 0.8$

$g^*\{x_1,x_2,x_3\} = 1$

si condicionamos a $\{x_1,x_2\}$, obtenemos:

$$g^*\{x_1/\{x_1x_2\}\} = \frac{g^*\{x_1\}}{g^*\{x_1,x_2\}} = \frac{0.2}{0.8} = 1/4$$

$$g_*\{x_1/\{x_1x_2\}\} = 1 - g^*\{\{x_2,x_3\}/\{x_1,x_2\}\} =$$

$$1 - \frac{g^*\{x_1\}}{g^*\{x_1,x_2\}} = 1 - \frac{0.2}{0.8} = 3/4$$

Por ello, cabe pensar en transformar la anterior definición de forma que se cumpla la condición de ordenación. Para ello, recordando que de cualquier par de medidas se puede siempre obtener un elemento de PO(X), proponemos:

$$g^{**}(A/B) = \text{máx} \left\{ \frac{g^*(A\cap B)}{g^*(B)}, 1 - \frac{g^*(\bar{A}\cap B)}{g^*(B)} \right\}$$

$$g_{**}(A/B) = \min \left\{ \frac{g^*(A\cap B)}{g^*(B)}, 1 - \frac{g^*(\bar{A}\cap B)}{g^*(B)} \right\}$$

Vemos como, $(g_{**}(./B)\ ,\ g^{**}(./B))$ es la imagen por la aplicación ψ (definida en el apartado 3.2) del par

$(g_*(./B)\ ,\ g^*(./B))$ y por tanto un elemento de PO(X).

Ahora bien, a pesar de haber logrado una definición de condicionamiento interna en PO(X), encontramos que los valores $(g_{**}(./B)\ ,\ g^{**}(./B))$ no tienen una interpretación intuitiva sencilla.

La definición (24), adquiere pleno sentido cuando nos --

restringimos a la clase de las capacidades de orden 2, C2(X); ya que, en éste caso puede enunciarse:

Teorema 2.1

Si $(g_*, g^*) \in C2(X)$, entonces:

1).- $g_{**}(./B) = g_*(./B)$ y $g^{**}(./B) = g^*(./B)$

2).- $(g_*(./B)\ ,\ g^*(./B)) \in C2(X)$

Demostración

1. Basta comprobar que:

$$\frac{g^*(A \cap B)}{g^*(B)} \geq 1 - \frac{g^*(\bar{A} \cap B)}{g^*(B)}$$

o lo que es igual

$$g^*(B) \leq g^*(\bar{A} \cap B) + g^*(A \cap B)$$

lo que se deduce de la definición de capacidad alternante de orden dos y de que:

$$(\bar{A} \cap B) \cup (A \cap B) = B$$

$$g^*((\bar{A} \cap B) \cap (A \cap B)) = g^*(\phi) = 0$$

2. Hemos de verificar en primer lugar que:

$$g^*(A \cup B/C) \leq g^*(A/C) + g^*(B/C) - g^*(A \cap B/C)$$

lo cual es equivalente a

$$\frac{g^*((A \cup B) \cap C)}{g^*(C)} \leq \frac{g^*(A \cap C)}{g^*(C)} + \frac{g^*(B \cap C)}{g^*(C)} + \frac{g^*(A \cap B \cap C)}{g^*(C)}$$

es decir:

$$g^*((A \cup B) \cap C) = g^*((A \cap C) \cup (B \cap C))$$

$$\leqslant g^*(A\cap C) + g^*(B\cap C) + g^*[(A\cap C)\cap(B\cap C)]$$

lo que se deduce de la definición de capacidad alternante de orden dos.

Por dualidad se puede deducir que:

$$g_*(A\cup B/C) \geqslant g_*(A/C) + g_*(B/C) - g_*(A\cap B/C)$$

"lo mas cercano" al teorema de la probabilidad total, son -- las cotas que se obtienen en la siguiente proposición.

Proposición 2.1

Si $(g_*,g^*) \in C2(X)$ y $g^*(B)\neq 0$, $g^*(\overline{B})\neq 0$

entonces:

$$g^*(A) \leqslant g^*(A/B)\ g^*(B) + g^*(A/\overline{B})\ g^*(\overline{B})$$

$$g_*(A) \geqslant [1-g^*(\overline{A}/B)\ g^*(B)] + [1-g^*(\overline{A}/\overline{B})\ g^*(\overline{B})]$$

la demostración es inmediata sin mas que considerar la definición de $g^*(./B)$ y tener en cuenta que el par pertenece a - C2(X).

Como complemento al teorema 2.1 puede comprobarse que - la definición (2.3) es interna en todas las familias incluidas en C2(X), en el sentido de que al condicionar una pareja de una clase obtenemos otra de la misma.
En concreto:

En el caso particular de una evidencia, es decir, un -- par (Bel,Pl)ε EV(X), se verifica que; las medidas condicionadas (Bel(./B),Pl(./B)) coinciden con la medidas de creencia y plausibilidad asociadas a la A.B.P que resulta de combinar por la regla de Dempster la asignación básica de probabilidad asociada a (Bel,Pl) y la A.B.P de tipo crips, como -- puede verse por ejemplo, en Wierzchon(1982).

Por tanto, la definición de evidencia condicionada equivale a combinar la información original con la información adicional m_B que expresa que el resultado del experimento está en B.

Dada una λ-medida, la condicionada $g_\lambda(./C)$ es una μ-medida -- con $\mu = \lambda g_\lambda(C)$ si $\lambda < 0$

Para una medida de posibilidad se obtiene:

$$\Pi(A/B) = \frac{\Pi(A \cap B)}{\Pi(B)}$$

y para la medida de necesidad asociada

$$N(A/B) = 1 - \frac{\Pi(\bar{A} \cap B)}{\Pi(B)} \qquad \text{con } \Pi(B) \neq 0$$

que tambien son una posibilidad y una necesidad respectiva-- mente.

3.- TRANSFORMACION DE MEDIDAS DIFUSAS MEDIANTE APLICACIONES: MEDIDAS DIFUSAS BIDIMENSIONALES

3.1. Transformación de medidas difusas mediante aplicaciones

Sea g una medida difusa en X y $f: X \rightarrow Y$ una aplicación. Definimos la transformada de la medida g mediante f como una relación sobre Y, dada por:

$$g_f(A) = g[f^{-1}(A)] \qquad \forall A \subset Y$$

como propiedades inmediatas podemos destacar:

A) g_f es una medida difusa sobre Y, como puede comprobarse -

facilmente a partir de su definición.

B) La transformación de medidas respeta la dualidad en efecto:

si g y $\overline{g}$ son duales, entonces:

$$g_f(A) + \overline{g}_f(\overline{A}) = g[f^{-1}(A)] + \overline{g}[f^{-1}(\overline{A})]$$

$$= g[f^{-1}(A)] + \overline{g}[\overline{f^{-1}(A)}] = 1$$

luego g y $\overline{g}$ son duales.

C) Si (g_*, g^*) pertenece a cualquier subclase de medidas difusas, de las ya estudiadas, entonces la transformada del par - mediante la aplicación f, (g_{*f}, g^*_f), tambien pertenece a la - misma subclase.

3.2. ~~T~~ransformación de medidas difusas mediante aplicaciones multivaluadas

Consideremos una aplicación multivaluada $\Gamma: X \to Y$, es decir, que asocia a cada $x \varepsilon X$ un subconjunto de Y, Definimos la transformada del par de medidas difusas duales ordenadas ---- (g_*, g^*), como el par:

$$\begin{aligned} g_{*\Gamma}(A) &= g_*\{x \varepsilon X \ / \ \Gamma(x) \subset A\} \\ g^*_\Gamma(A) &= g^*\{x \varepsilon X \ / \ \Gamma(x) \cap A \neq \phi\} \end{aligned} \qquad (25)$$

ésta definición está inspirada, en la transformación de una - probabilidad mediante una aplicación multivaluada, considerada en Dempster (1967).

La pareja $(g_{*\Gamma}, g^*_\Gamma)$, es un elemento de PO(X), siempre que $\Gamma(x) \neq \phi$, $\forall x \varepsilon X$. En lo que sigue, supondremos que nos encontramos en éste caso, aunque no se señale explicitamente.

Proposición 2.2

Si (g_*, g^*) son duales en X, entonces, cualquiera que sea , - $(g_{*\Gamma}, g^*_\Gamma)$ son duales en Y.

Demostración:

Por definición:

$$g_{*\Gamma}(A) + g^*_\Gamma(\bar{A}) = g_*\{x \in X \ / \ \Gamma(x) \subset A\} + g^*\{x \in X \ / \ \Gamma(x) \cap \bar{A} \neq \phi\}$$

$$= g_*\{x \in X \ / \ \Gamma(x) \subset A\} + g^*\{x \in X \ / \ \Gamma(x) \not\subset A = 1$$

Proposición 2.3

Si (g_{*2}, g^*_2) acota a la pareja (g_{*1}, g^*_1) en el sentido de:

$$g_{*2}(A) \leqslant g_{*1}(A) \leqslant g^*_1(A) \leqslant g^*_2(A)$$

entonces $(g_{*2\,\Gamma}, g^*_{2\,\Gamma})$ acota a la pareja $(g_{*1\,\Gamma}, g^*_{1\,\Gamma})$, es decir:

$$g_{*2\,\Gamma}(B) \leqslant g_{*1\,\Gamma}(B) \leqslant g^*_{1\,\Gamma}(B) \leqslant g^*_{2\,\Gamma}(B)$$

Demostración

Es inmediata a partir de (25).

A continuación vamos a estudiar como actua ésta transforma-- ción para las diferentes familias de medidas.

1) Ya hemos señalado anteriormente que:

$$(g_*, g^*) \in PO(X) \Rightarrow (g_* \ , g^*) \in PO(Y)$$

2) Si $(g_*, g^*) \in AC(X)$, entonces, $(g_{*\Gamma}, g^*_\Gamma) \in AC(Y)$

en efecto, para $A, B \subset Y$ tales que $A \cap B = \phi$, se tiene:

$$g_{*\Gamma}(A \cup B) = g_*(\{x \in X \ / \ \Gamma(x) \subset (A \cup B)\})$$

$$= g_*(\{x \in X \ / \ \Gamma(x) \subset A\} \cup \{x \in X \ / \ \Gamma(x) \subset B\})$$

$$\geqslant g_*(\{x \in X \ / \ \Gamma(x) \subset A\}) + g_*(\{x \in X \ / \ \Gamma(x) \subset B\})$$

$$= g_{*\Gamma}(A) + g_{*\Gamma}(B)$$

$$
\begin{aligned}
g^*_\Gamma(A\cup B) &= g^*(\{x\varepsilon X\ /\Gamma(x)\cap(A\cup B)\neq\phi\}) \\
&= g^*(\{x\varepsilon X\ /\Gamma(x)\cap A\neq\phi\}\{x\varepsilon X\ /\Gamma(x)\cap B\neq\phi\ ) \\
&\leqslant g^*(\{x\varepsilon X\ /\Gamma(x)\cap A\neq\phi\}) + g^*(\{x\varepsilon X\ /\Gamma(x)\cap B\neq\phi\}) \\
&= g^*_\Gamma(A) + g^*_\Gamma(B)
\end{aligned}
$$

3) Si $(g_*,g^*)\varepsilon$ PA(X), entonces, $(g_{*\Gamma},g^*_\Gamma)\varepsilon$ PA(Y)

si se verifica que: $g_*(A) \leqslant P(A) \leqslant g^*(A)$, entonces, por la proposición 2.3, se deduce que

$$g_{*\Gamma}(B) \leqslant P_{*\Gamma}(B) \leqslant P^*_\Gamma(B) \leqslant g^*_\Gamma(B) \qquad (26)$$

Teniendo en cuenta Dempster (1967), que la transformada de una probabilidad por una aplicación multivaluada, es un e lemento $(P_{*\Gamma},P^*_\Gamma)\varepsilon EV(Y)\subset PA(Y)$, entonces, ha de existir ---- $P'\varepsilon P(Y)$ de forma que:

$$P_{*\Gamma}(B) \leqslant P'(B) \leqslant P^*_\Gamma(B)$$

teniendo en cuenta (25), deducimos que:

$$g_{*\Gamma}(B) \leqslant P'(B) \leqslant g^*_\Gamma(B)$$

de donde concluimos $(g_{*\Gamma},g^*_\Gamma)\varepsilon$ PA(Y).

4) Si $(g_*,g^*)\varepsilon$ MR(X), entonces, $(g_{*\Gamma},g^*_\Gamma)\varepsilon$ MR(Y)

si $(g_*,g^*)\varepsilon$ MR(X), existe una familia $\mathcal{P}$ de medidas de probabilidad tal que:

$$g^*(A) = \sup_{P\varepsilon\mathcal{P}} P(A) \qquad \forall P\varepsilon\mathcal{P}$$

$(P_{*\Gamma}.P^*_\Gamma)$ es una evidencia, y por tanto, una medida representable, debiendo existir una familia $\mathcal{G}_P$ en Y, tal que:

$$P^*(B) = \sup_{P\varepsilon\mathcal{G}_P} P'(B)$$

ahora bien,

$$g^*_\Gamma(A) = g^*\{x\varepsilon X \;/\Gamma(x)\cap A\neq\phi\} = \sup_{P\varepsilon\mathcal{P}} P\{x\varepsilon X\;/\Gamma(x)\cap A\neq\phi\}$$

$$= \sup_{P\varepsilon\mathcal{P}} P^*_\Gamma(A) = \sup_{P\varepsilon\mathcal{P}}\Big(\sup_{P'\varepsilon\mathcal{P}_P} P'(A)\Big) = \sup_{P'\varepsilon\cup\mathcal{P}_P} P'(A)$$

lo que prueba que $(g_{*\Gamma}, g^*_\Gamma)$ es representable mediante $(\underset{P}{\cup}\mathcal{P}_P)$.

5) Si $(g_*,g^*)\,\varepsilon\,C2(X)$, entonces, $(g_{*\Gamma},g^*_\Gamma)\,\varepsilon\,C2(Y)$

$$g_{*\Gamma}(A\cup B) = g_*\{x\varepsilon X\;/\Gamma(x)\subset A\cup B\} =$$

$$= g_*\{(x\varepsilon X\;/\Gamma(x)\subset A)\;\;(x\varepsilon X\;/\Gamma(x)\subset B)\}$$

$$\geqslant g_*\{x\varepsilon X\;/\Gamma(x)\subset A\} + g_*\{x\varepsilon X\;/\Gamma(x)\subset B\} -$$

$$- g_*\{x\varepsilon X\;/\Gamma(x)\subset(A\cap B)\}$$

$$= g_{*\Gamma}(A) + g_{*\Gamma}(B) - g_{*\Gamma}(A\cap B)$$

$$g^*_\Gamma(A\cup B) = g^*\{x\varepsilon X\;/\Gamma(x)\cap(A\cup B)\neq\phi\}$$

$$= g^*(\{x\varepsilon X\;/\Gamma(x)\cap A\neq\phi)\cup(x\varepsilon X\;/\Gamma(x)\cap B\neq\phi)\}$$

$$\leqslant g^*\{x\varepsilon X\;/\Gamma(x)\cap A\neq\phi\} + g^*\{x\varepsilon X\;/\Gamma(x)\cap B\neq\phi\} +$$

$$+ g^*\{x\varepsilon X\;/\Gamma(x)\cap(A\cap B)\neq\phi\}$$

$$= g^*_\Gamma(A) + g^*_\Gamma(B) + g^*_\Gamma(A\cap B)$$

6) Si $(Bel,Pl)\,\varepsilon\,EV(X)$, con A.B.P m, entonces;
$(Bel_\Gamma,\; Pl_\Gamma)\,\varepsilon\,EV(Y)$ con A.B.P $m_\Gamma(A) = \sum_{\Gamma(B)=A} m(B)$, donde $\Gamma(B) = \underset{x\varepsilon B}{\cup}\Gamma(x)$

$$\sum_{A\cap C\neq\phi} m_\Gamma(A) = \sum_{A\cap C\neq\phi}\Big(\sum_{\Gamma(B)=A} m(B)\Big) = \underset{\Gamma^*_C}{} m(B) = \sum_{B\cap\{x\varepsilon X/\Gamma(x)\cap C\neq\phi\}\neq\phi} m(B) =$$

donde: $\Gamma^*_C = \{B\subset X/\; x\varepsilon B \text{ con } \Gamma(x)\cap C\neq\phi\}$

$$= Pl\,\{x\varepsilon X\;/\Gamma(x)\cap C\neq\phi\} = Pl_\Gamma(C)$$

7) Si $(g_*,g^*)\,\varepsilon\,C2(X)$, entonces, $(g_{*\Gamma},g^*)\,\varepsilon\,C2(Y)$

Es una consecuencia inmediata de la anterior, ya que, - $G\lambda(X) \subset EV(X)$.

No se puede asegurar que la transformación sea interna en $G\lambda(X)$, porque la familia $EV(Y)$ se puede obtener a partir - de ella mediante aplicaciones multivaluadas, como se deduce de la inclusión de $PR(X)$ en $G\lambda(X)$.

8) Si $(N,\Pi) \in EC(X)$, entonces $(N_\Gamma,\Pi_\Gamma) \in EC(Y)$.

Esta propiedad se deduce inmediatamente del hecho de -- que si m es la A.B.P asociada a una evidencia consonante ,- con elementos focales anidados, $A_1 \subset A_2 \subset \ldots \subset A_n$, entonces, su transformada es una evidencia sobre Y con A.B.P m_Γ , que -- tiene como elementos focales $\Gamma(A_1) \subset \Gamma(A_2) \subset \ldots \subset \Gamma(A_n)$, que -- tambien estarán anidados ya que $\Gamma(A_i) = \bigcup_{x \in A_i} \Gamma(x)$

9) Si $(g_*,g^*) \in PR(X)$, entonces, $(g_{*\Gamma},g^*_\Gamma) \in EV(Y)$.

Recordemos que Dempster (1967) introdujo las evidencias como transformaciones de las probabilidades mediante una a-- plicación multivaluada,(empleando las A.B.P).

10) Si $(g_*,g^*) \in CR(X)$, entonces,$(g_{*\Gamma},g^*_\Gamma) \in CR(Y)$.

Si $(g_*,g^*) \in CR(X)$, entonces, su A.B.P ,m, tendrá un único elemento focal A, y la transformación de m mediante Γ será una A.B.P sobre Y,m_Γ, que tendrá como único elemento fo-- cal a $\Gamma(A) = \bigcup_{x \in A} \Gamma(x)$

11) Si $(g_*,g^*) \in DI(X)$, entonces, $(g_{*\Gamma},g^*_\Gamma) \in CR(Y)$.

Si $(g_*,g^*) \in DI(X)$, su A.B.P estará focalizada en $\{a\}$ - donde $a \in X$, y m_Γ tendrá a $\Gamma(a) \subset Y$ como único elemento focal, es decir, será un elemento de CR(Y).

3.3. Medidas difusas bidimensionales

Llamaremos medida difusa bidimensional g a una medida - difusa sobre el producto cartesiano de dos conjuntos $(X_1 \times X_2)$.

A la hora de estudiar las medidas bidimensionales, se-- rán de particular importancia las proyecciones sobre X_1 y so bre X_2.

$$P_1: X_1 \times X_2 \to X_1 \quad ; \quad P_2: X_1 \times X_2 \to X_2$$

$$(x_1,x_2) \to x_1 \qquad\qquad (x_1,x_2) \to x_2$$

y las extensiones de X_1 y X_2 en $(X_1 \times X_2)$ que vienen definidas como las aplicaciones multivaluadas:

$$\Gamma_1: X_1 \to X_1 \times X_2 \quad ; \quad \Gamma_2: X_2 \to X_1 \times X_2$$

$$x_1 \to \{(x_1,x_2)/x_2 \varepsilon X_2\} \qquad\qquad x_2 \to \{(x_1,x_2)/x_1 \varepsilon X_1\}$$

Definición 2.1

Si g es una medida bidimensional en $(X_1 \times X_2)$, llamamos marginales de g, a las medidas difusas g_1, g_2 en X_1 y X_2, res pectivamente, que se obtienen transformando g por P_1 y P_2. - Es decir:

$$g_1(A_1) = g(P_1^{-1}(A_1)) = g\{(x_1,x_2)/x_1 \varepsilon A_1\} = g(A_1 \times X_2)$$

$$g_2(A_2) = g(P_2^{-1}(A_2)) = g\{(x_1,x_2)/x_2 \varepsilon A_2\} = g(X_1 \times A_2)$$

Nota

En el caso particular de una probabilidad, la defini - ción2.1 conduce al concepto clásico de distribución marginal En el caso de una distribución de posibilidad, se obtienen, las distribuciones de posibilidad marginales definidas por - Zadeh (1978). Cuando se trata de una evidencia ésta defini-- ción coincide con la de proyección de una evidencia dada por

Shafer (1976) y Dubois y Prade (1985).

Definición 2.2

Si (h_1,h^1) son un par de medidas difusas en X_1, se llama extensión de éste par a $X_1 \times X_2$ a la pareja de medidas difusas bidimiensionales que se obtienen, transformando (h_1,h^1), mediante la aplicación Γ_1. Es decir:

$$H_1(A) = h_1\{x_1 \varepsilon X_1 / \Gamma_1(x_1) \subset A\}$$
$$= h_1\{x_1 \varepsilon X_1 / (x_1,x_2) \varepsilon A \quad \forall x_2 \varepsilon X_2\}$$

$$H^1(A) = h^1\{x_1 \varepsilon X_1 / \Gamma_1(x_1) \cap A \neq \phi\}$$
$$= h^1\{x_1 \varepsilon X_1 / \exists x_2 \varepsilon X_2 \text{ con } (x_1,x_2) \varepsilon A\}$$

Nota

Cuando se trata de una posibilidad, ésta definición conduce al concepto de extensión cilíndrica, dado por Zadeh --- (1978). En el caso de una evidencia, tambien se obtienen las extensiones cilíndricas definidas por Shafer (1976) y Dubois y Prade (1985).

Dado un par de medidas difusas ordenadas (g_*,g^*), el conocimiento de sus marginales no nos permitirá en general, reconstruir la medida bidimensional original. Sin embargo, si podemos acotar éstas medidas como muestra la siguiente proposición.

Proposición 2.4

Dado el par ordenado bidimensional (g_*,g^*) en $X_1 \times X_2$, entonces se verifica:

$$\forall A \subset X_1 \times X_2 \quad ; \qquad G_1(A) \leqslant g_*(A) \leqslant g^*(A) \leqslant G^1(A)$$
$$G_2(A) \leqslant g_*(A) \leqslant g^*(A) \leqslant G^2(A)$$

donde (G_1,G^1) , (G_2,G^2) son las extensiones de las marginales

Demostración

$$G_1(A) = g_1\{x_1 \varepsilon X_1/\Gamma_1(x_1)\subset A\} = g_1\{x_1 \varepsilon X_1/(x_1,x_2)\varepsilon A \ \forall x_2 \varepsilon X_2\}$$
$$= g_*(\{x_1 \varepsilon X_1/(x_1,x_2)\varepsilon A, \ \forall x_2 \varepsilon X_2\} \times X_2) \leq g_*(A)$$

ya que:

$$(\{x_1 \varepsilon X_1/(x_1,x_2)\varepsilon A \quad \forall x_2 \varepsilon X_2\} \times X_2)\subset A$$

$$G^1(A) = g^1\{x_1 \varepsilon X_1/ \ \exists x_2 \varepsilon X_2 \text{ con } (x_1,x_2)\varepsilon A\}$$
$$= g^*(\{x_1 \varepsilon X_1/ \ \exists x_2 \varepsilon X_2 \text{ con } (x_1,x_2)\varepsilon A\} \times X_2) \geq g^*(A)$$

ya que:

$$A\subset (\{x_1 \varepsilon X_1/ \ \exists x_2 \varepsilon X_2 \text{ con } (x_1,x_2)\varepsilon A\} \times X_2)$$

Corolario

Dado el par ordenado bidimensional (g_*,g^*) en $(X_1{\times}X_2)$,- se verifica que:

$$\max (G_1(A),G_2(A)) \leq g_*(A) \leq g^*(A) \leq \min (G^1(A),G^2(A))$$
$$A \subset X_1{\times}X_2$$

En el caso de que se alcancen las cotas del corolario - anterior para cada subconjunto A de $X_1{\times}X_2$, podemos reconstru ir el par de medidas (g_*,g^*), a partir de sus marginales. Es te caso viene recogido en la siguiente proposición.

Definición 2.3

Dado un par ordenado (g_*,g^*) en $X_1{\times}X_2$, se dice que es - fuertemente descomponible si y solamente si:

$$\max (G_1(A),G_2(A)) = g_*(A) \leq g^*(A) = \min (G^1(A),G^2(A))$$
$$A \subset X_1{\times}X_2$$

En el caso de una medida de posibilidad Π, la descomposición fuerte, implica la no-interacción definida por Zadeh.

$$\Pi(A_1 \times A_2) = \min\ (\Pi_1(A_1), \Pi_2(A_2))$$

ya que, ésta última definición solo exige la igualdad de la cota en las medidas superiores para los rectángulos $(A_1 \times A_2)$, $A_1 \subset X_1$, $A_2 \subset X_2$, y no para cualquier conjunto $A \subset X_1 \times X_2$, como es nuestro caso.

La definición anterior es bastante fuerte, ya que, por ejemplo, implica que la medida superior de un conjunto A, sea igual a la medida del mínimo rectángulo que lo contiene, ---- $P_1(A) \times P_2(A)$, verificandose que:

$$G^i(A) = G^i(P_1(A) \times P_2(A)) \qquad i=1,2$$

Como consecuencia de ello la debilitaremos, introduciendo el concepto de descomposición débil.

Definición 2.4

Dado el par (g_*, g^*) en $X_1 \times X_2$ diremos que es debilmente descomponible, cuando se verifique:

$$\begin{aligned} g^*(A_1 \times A_2) &= \min\ (G^1(A_1 \times A_2), G^2(A_1 \times A_2)) \\ &= \min\ \{g^1(A_1), g^2(A_2)\} \end{aligned}$$

Proposición 2.5

(g_*, g^*) es debilmente descomponible, si y solamente si:

$$\begin{aligned} g_*(\overline{A_1 \times A_2}) &= \max\ \{G_1(\overline{A_1 \times A_2}), G_2(\overline{A_1 \times A_2})\} \\ &= \max\ \{g_1(\overline{A}_1), g_2(\overline{A}_2)\} \end{aligned}$$

Demostración

Es inmediata por dualidad.

Nota

En el caso particular de una medida de posibilidad, la - definición anterior es equivalente a la no-interacción.

Nota

Una medida de probabilidad bidimensional no será nunca debilmente descomponible, excepto cuando sea una medida de Dirac.

Nota

Una medida debilmente descomponible no es, en general, reconstruible a partir de sus marginales, excepto en el caso en que el conocimiento de la medida superior sobre los rectángulos, permita calcular la medida de cualquier conjunto, como ocurre por ejemplo, con las medidas de posibilidad.

En todo caso las marginales, permiten conocer la medida - de los rectángulos que en bastantes ocasiones no será suficiente.

3.4. INDEPENDENCIA DE EVIDENCIAS Y MEDIDAS DE POSIBILIDAD

Hemos visto que la definición de medida condicionada alcanza pleno sentido en la familia C2(X). Nosotros, nos vamos a limitar dentro de ésta clase a las evidencias, ya que son - las que posteriormente consideraremos.

Definición 2.5

Sea (Bel,Pl) una evidencia bidimensional en $X_1 \times X_2$ y --- $A_2 \subset X_2$ con $Pl_2(A_2) > 0$.

Definimos la marginal en X_1 condicionada a A_2, como:

$$Pl_1(A_1/A_2) = \frac{Pl\ (A_1 \times A_2)}{Pl_2(A_2)} \qquad A_1 \subset X_1$$

$$Bel_1(A_1/A_2) = 1 - \frac{Pl\ (\bar{A}_1 \times A_2)}{Pl_2(A_2)}$$

que es una evidencia en X_1.

<u>Nota</u>

La evidencia condicionada se puede obtener paso a paso de la siguiente forma:

1) Construimos en X_2 la evidencia de tipo crisp concentrada en A_2 de A.B.P $\{m_{A_2}\}$.

2) Extendemos ésta evidencia a $X_1 \times X_2$ mediante Γ_2, obteniendo la A.B.P

$$M_{A_2}(A) = \begin{cases} 1 & \text{si } A = X_1 \times A_2 \\ 0 & \text{en otro caso} \end{cases}$$

3) Combinamos la evidencia original m con M_{A_2}, por medio de la regla de Dempster.

4) Consideramos la marginal en X_1 de la evidencia obtenida - anteriormente, cuya A.B.P es:

$$m_1(A_1/A_2) = \sum_{B \cap (X_1 \times A_2) = A_1 \times A_2} m(B) \Big/ \sum_{B \cap (X_1 \times A_2) \neq \phi} m(B)$$

$$= \sum_{B \cap (X_1 \times A_2) = A_1 \times A_2} m(B) \Big/ Pl_2(A_2)$$

Las medidas de Plausibilidad y Creencia asociadas a ésta A.B.P coinciden con las de nuestra definición, es decir constituyen la evidencia marginal condicionada a A_2.

En el caso de que Π sea una medida de posibilidad, la - definición anterior se traduce en:

$$\Pi_1(A_1/A_2) = \frac{\Pi(A_1 \times A_2)}{\Pi_2(A_2)} \qquad \Pi_2(A_2) > 0$$

Si $A_2=\{x_2\}$ obtenemos $\Pi_1(A_1/x_2) = \dfrac{\Pi(A_1\times\{x_2\})}{\pi_2(x_2)}$,

siendo la distribución de posibilidad condicionada asociada:

$$\pi_1(x_1/x_2) = \frac{\pi(x_1,x_2)}{\pi_2(x_2)}$$

en la literatura pueden encontrarse diversas definiciones de posibilidad condicionada, entre las que citamos:

- Zadeh (1978)

$$\pi_1^Z(x_1/x_2) = \pi(x_1,x_2)$$

- Hisdal (1978,79)

$$\pi_1^H(x_1/x_2) = \begin{cases} \pi(x_1,x_2) & \text{si} \quad \pi_2(x_2) > \pi(x_1,x_2) \\ [\pi(x_1,x_2),1] & \pi_2(x_2) = \pi(x_1,x_2) \end{cases}$$

donde $[\pi(x_1,x_2),1]$ denota cualquier valor del intervalo cerrado de extremos $\pi(x_1,x_2)$ y 1. Por tanto, la posibilidad -- condicionada π_1^H solo está determinada a partir de π_1 en el - dominio en que $\pi_2(x_2) > \pi(x_1,x_2)$, donde coincide con la defi nición de Zadeh.

Esta definición viene justificada por el cumplimiento - de:

$$\pi(x_1,x_2) = \pi_2(x_2) \wedge \pi_1^H(x_1/x_2)$$

que trata de reproducir la relación:

$$P[X_1=x_1, X_2=x_2] = P[X_2=x_2].P[X_1=x_1/X_2=x_2]$$

al reemplazar el operador producto por el mínimo.

- Nguyen (1978-79)

$$\pi_1^N(x_1/x_2)=\begin{cases} \pi(x_1,x_2) & \pi_1(x_1) \leqslant \pi_2(x_2) \\ \pi(x_1,x_2)\dfrac{\pi_1(x_1)}{\pi_2(x_2)} & \pi_1(x_1) > \pi_2(x_2) \end{cases}$$

ésta característica se obtiene, al suponer que:

$$\pi_1^N(x_1/x_2) = \pi(x_1,x_2).\ \alpha(\pi_1(x_1),\pi_2(x_2)),$$

donde:

$$\alpha : [0,1 \times 0,1] \rightarrow R^+ \quad \text{con:}$$

i) $\pi(x_1,x_2).\alpha(\pi_1(x_1),\pi_2(x_2)) \varepsilon [0,1] \qquad \forall (x_1,x_2) \varepsilon X_1 \times X_2$

ii) $(\pi_1(x_1) \wedge \pi_2(x_2)).\alpha(\pi_1(x_1),\pi_2(x_2))=\pi_1(x_1) \qquad \forall x_1 \varepsilon X_1$

la propiedad i) se impone ya que, las posibilidades condicionadas han de pertenecer al intervalo [0,1].

la propiedad ii) se exige para que ésta definición sea consistente con la de no-interacción de Zadeh en el siguiente sentido: si π tiene marginales no interactivas, $(\pi(x_1,x_2) = \pi_1(x_1) \wedge \pi_2(x_2))$, entonces, $\pi_1(x_1/x_2)=\pi_1(x_1) \ \forall x_2 \varepsilon X_2$

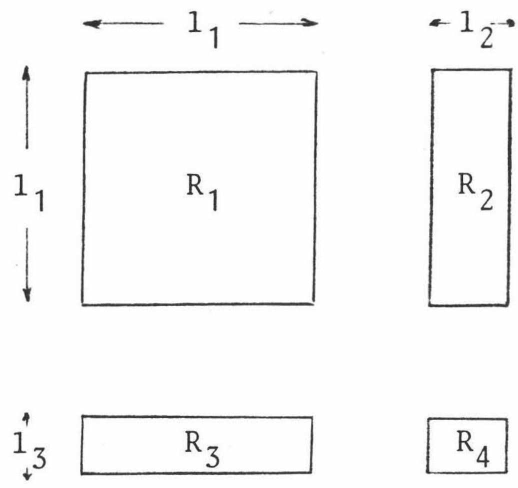
Para ilustrar el comportamiento de las definiciones anteriores, consideraremos el siguiente ejemplo:

Ejemplo 2.2

Supongamos el conjunto $X=\{R_1,R_2,R_3,R_4\}$ formado por los siguientes rectángulos.

Si identificamos cada rectángulo con las dimensiones de su altura l_i y su base l_j; puede considerarse X como el producto cartesiano de $X_1=\{l_1,l_2\}$ y $X_2=\{l_1,l_3\}$:

$R_1=\{l_1,l_1\}$; $R_2=\{l_1,l_2\}$; $R_3=\{l_3,l_1\}$; $R_4=\{l_3,l_2\}$.



Supongamos un rectángulo $R \epsilon X$, que no conocemos, y la si guiente información difusa sobre el: "R es casi cuadrado". Esta información difusa tendrá una distribución de posibilidad asociada (ver Zadeh (1978)), que podemos admitir que toma los valores:

$$\pi(l_1,l_1)=1 \quad ; \quad \pi(l_1,l_2)=0.5 \quad ; \quad \pi(l_3.l_1)=0.2 \quad ; \quad \pi(l_3,l_2)=0.6$$

Supongamos ahora que sabemos que la base del rectángulo es l_2, entonces, sobre la altura dispondremos de una distribución de posibilidad que es la condicionada $\pi(./l_2)$. Según, las distintas definiciónes consideradas, tomará la forma:

1) $\pi_1^Z(l_1/l_2)=0.5 \quad ; \quad \pi_1^Z(l_3/l_2)=o.6$

que no es una distribución de posibilidad normalizada en X_1. No da lugar a una medida de posibilidad.

2) $\pi_1^H(l_1/l_2)=0.5 \quad ; \quad \pi_1^H(l_3/l_2) \in [0.6,1]$

que no está normalizada, excepto cuando $\pi_1^H(l_3/l_2)=1$, en cuyo caso, cabe observar que la diferencia en el cumplimiento de la propiedad por parte de (l_1,l_2) y (l_3,l_2) es mucho me-- nor que la diferencia entre las posibilidades correspondientes.

3) $\pi_1^N(l_1/l_2)=5/6 \quad ; \quad \pi_1^N(l_3/l_2)=0.6$

que tampoco es una distribución normalizada. Además, aparece la siguiente paradoja: El hecho de que el rectángulo de altura l_1 con base l_1 sea "mas cuadrado" que el de altura l_3 con base l_1, hace que la posibilidad de la altura l_1, sabiendo que la base es l_2, sea mayor que la de la altura l_3, condicionado tambien a l_2; aun siendo "menos cuadrado" el rectángulo (l_1,l_2) que el (l_3,l_2).

4) Según nuestra definición:

$$\pi_1(l_1/l_2)=5/6 \qquad y \qquad \pi_1(l_3/l_2)=1$$

que es una distribución de posibilidad normalizada en X_1, bien determinada y sin la paradoja de la definición de Nguyen.

A continuación presentamos el concepto de independencia para evidencias. En numerosas ocasiones se ha interpretado el concepto de no-interacción (descomponibilidad débil), entre distribuciones de posibilidad como el análogo a la independencia de distribuciones de probabilidad. En nuestro trabajo, consideraremos ambos conceptos bien diferenciados y aplicables a cualquier evidencia, y en particular a posibilidades y probabilidades.

Definición 2.6

Sea (Bel.Pl) una evidencia en $X_1 \times X_2$.Diremos que tiene marginales independientes, si y solo si:

$$\forall A_2 \subset X_2 \text{ con } Pl_2(A_2) > 0 \qquad Pl_1(./A_2) \equiv Pl_1(.)$$

o equivalentemente

$$Bel_1(./A_2) \equiv Bel_1(.)$$

Como consecuencia inmediata, se puede obtener la siguiente caracterización de la independencia.

Proposición 2.6

(Bel,Pl) tiene marginales independientes, si y solo si:

$$Pl(A_1 \times A_2)=Pl_1(A_1)\times Pl_2(A_2) \qquad \forall A_1 \subset X_1 \ , \forall A_2 \subset X_2 \qquad (27)$$

Desgraciadamente, no existe en general, una caracterización similar a la anterior en función de las creencias ni de las A.B.P , como muestra el siguiente ejemplo:

Ejemplo 2.3

Sea $X_1=X_2=\{x_1,x_2\}$ y una evidencia cuya A.B.P en $X_1 \times X_2$ es:

$m(X_1 \times X_2)=0.72$

$m(\{(x_1,x_1)\})=0.1$

$m(\{(x_1,x_1),(x_2,x_1)\})=0.1$

$m(\{(x_1,x_1),(x_2,x_1),(x_2,x_2)\})=0.08$

las evidencias marginales serán:

$m_1(\{x_1\})=0.1 \qquad m_2(\{x_1\})=0.2$

$m_1(\{x_1,x_2\})=0.9 \qquad m_2(\{x_1,x_2\})=0.8$

y es inmediato comprobar que existe independencia.

Sin embargo, se tiene que:

$$m(\{x_1,x_2\}\times\{x_1\})=0.1 \neq 0.9\times 0.2 = m_1(\{x_1,x_2\})\times m_2(\{x_1\})$$

y

$$Bel(\{(x_2,x_2)\})=0.8 \neq 0.8\times 0.9 = Bel_1(\{x_2\})\times Bel_2(\{x_2\})$$

una caracterización de independencia de marginales análoga a (27), pero en términos de creencias y A.B.P se obtiene, -- cuando la evidencia se concentra en los rectángulos.

Definición 2.7

Sea una evidencia bidimensional en $X_1 \times X_2$ con A.B.P m.

Diremos que dicha evidencia se concentra en los rectángulos, si y solo si:

$$A \subset X_1 \times X_2,\ m(A) > 0 \Longrightarrow \exists\, A_1 \subset X_1,\ A_2 \subset X_2,\ \text{tal que, } A = A_1 \times A_2$$

o equivalentemente:

$$\sum_{\substack{A_1 \subset X_1 \\ A_2 \subset X_2}} m(A_1 \times A_2) = 1 \qquad (28)$$

<u>Teorema</u> 2.2

Sea una evidencia bidimensional (Bel,Pl) en $X_1 \times X_2$, que se concentra en los rectángulos, y sea m su A.B.P . Entonces se verifican las siguientes relaciones.

1) $$\mathrm{Bel}(A_1 \times A_2) = \sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} m(B_1 \times B_2) \ ; \ m(A_1 \times A_2) = \sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} (-1)^{|A_1 - B_1| + |A_2 - B_2|} \mathrm{Bel}(B_1 \times B_2)$$

2) $$\mathrm{Bel}(A_1 \times X_2 \cup X_1 \times A_2) = \mathrm{Bel}(A_1 \times X_2) + \mathrm{Bel}(X_1 \times A_2) - \mathrm{Bel}(A_1 \times A_2)$$

3) $$\mathrm{Pl}(A_1 \times X_2 \cup X_1 \times A_2) = \mathrm{Pl}(A_1 \times X_2) + \mathrm{Pl}(X_1 \times A_2) - \mathrm{Pl}(A_1 \times A_2)$$

Demostración

1) La primera parte es consecuencia inmediata de la definición anterior. Para la segunda tenemos que:

$$\sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} (-1)^{|A_1 - B_1| + |A_2 - B_2|} \mathrm{Bel}(B_1 \times B_2) = \sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} \left[(-1)^{|A_1 - B_1| + |A_2 - B_2|} \Big(\sum_{\substack{C_1 \subset B_1 \\ C_2 \subset B_2}} m(C_1 \times C_2)\Big)\right]$$

$$= \sum_{\substack{C_1 \subset A_1 \\ C_2 \subset A_2}} m(C_1 \times C_2) \left[\sum_{\substack{D_1 \subset A_1 - C_1 \\ D_2 \subset A_2 - C_2}} (-1)^{|D_1| + |D_2|}\right] = m(A_1 \times A_2)$$

ya que:

$$\sum_{\substack{D_1\subset A_1-C_1\\ D_2\subset A_2-C_2}} (-1)^{|D_1|+|D_2|} = \sum_{D_1\subset A_1-C_1} (-1)^{|D_1|} \sum_{D_2\subset A_2-C_2} (-1)^{|D_2|} =$$

$$= \begin{cases} 1 & \text{si } \ A_1-C_1=\phi \ ; \ A_2-C_2=\phi \\ 0 & \text{en otro caso} \end{cases}$$

con solo tener en cuenta que:

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

2) $Bel(A_1\times X_2)+Bel(X_1\times A_2)-Bel(A_1\times A_2)=$

$$= \sum_{\substack{C_1\subset A_1\\ C_2\subset X_2}} m(C_1\times C_2) + \sum_{\substack{C_1\subset X_1\\ C_2\subset A_2}} m(C_1\times C_2) - \sum_{\substack{C_1\subset A_1\\ C_2\subset A_2}} m(C_1\times C_2) =$$

$$= \sum_{(C_1\subset A_1)\cup(C_2\subset A_2)} m(C_1\times C_2) = \sum_{(C_1\times C_2)\subset(A_1\times X_2\cup X_1\times A_2)} m(C_1\times C_2) =$$

$$= Bel\ (A_1\times X_2\cup X_1\times A_2)$$

3) Se obtiene de la anterior por dualidad.

Definición 2.8

Dada una evidencia bidimensional en $X_1\times X_2$, diremos que las evidencias marginales son fuertemente independientes, si y solo si, son independientes y la evidencia se concentra en los rectángulos.

La definición anterior nos permite obtener las siguientes caracterizaciones de la independencia fuerte.

Teorema 2.3

Sea (Bel,Pl) una evidencia en $(X_1 \times X_2)$, con A.B.P m; entonces, las siguientes condiciones son equivalentes:

1.- Las evidencias marginales son fuertemente independientes

2.- $\forall A_i \subset X_i$, $i=1,2$, $m(A_1 \times A_2) = m_1(A_1) \times m_2(A_2)$

3.- $\forall A_i \subset X_i$ $i=1,2$, $Pl(A_1 \times A_2) = Pl_1(A_1) \times Pl_2(A_2)$

y la evidencia se concentra en los rectángulos.

4.- $\forall A_i \subset X_i$ $i=1,2$, $Bel(A_1 \times A_2) = Bel_1(A_1) \times Bel_2(A_2)$

y la evidencia se concentra en los rectángulos.

Demostración

Probaremos las siguientes implicaciones $1 \Leftrightarrow 3 \Rightarrow 4 \Rightarrow 2 \Rightarrow 3$.

$1 \Leftrightarrow 3$.- Es inmediato por la definición de independencia fuerte y de la caracterización de la independencia dada por la relación (27).

$3 \Rightarrow 4$.- En efecto:

$$Bel(A_1 \times A_2) = 1-Pl(\overline{A_1 \times A_2}) = 1-Pl(\overline{A}_1 \times X_2 \cup X_1 \times \overline{A}_2) =$$

$$= 1-[Pl(\overline{A}_1 \times X_2) + Pl(X_1 \times \overline{A}_2) - Pl(\overline{A}_1 \times \overline{A}_2)] =$$

$$= 1-[Pl_1(A_1) \times Pl_2(X_2) + Pl_1(X_1) \times Pl_2(A_2) - Pl_1(A_1) \times Pl_2(A_2))$$

$$= 1-Pl_1(\overline{A}_1) - Pl_2(\overline{A}_2) + Pl_1(\overline{A}_1) \times Pl_2(\overline{A}_2) =$$

$$= (1-Pl_1(\overline{A}_1)) \times (1-Pl_2(\overline{A}_2)) = Bel_1(\overline{A}\ ) \times Bel_2(\overline{A}_2)$$

$4 \Rightarrow 2$.- Teniendo en cuenta la propiedad 1 del capítulo anterior:

$$m(A_1 \times A_2) = \sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} (-1)^{|A_1 - B_1| + |A_2 - B_2|} Bel(B_1 \times B_2) =$$

$$= \sum_{\substack{B_1 \subset A_1 \\ B_2 \subset A_2}} (-1)^{|A_1-B_1| + |A_2-B_2|} \; Bel_1(B_1) \times Bel_2(B_2) =$$

$$= \left(\sum_{B_1 \subset A_1} (-1)^{|A_1-B_1|} Bel_1(B_1) \right) \left(\sum_{B_2 \subset A_2} (-1)^{|A_2-B_2|} Bel_2(B_2) \right) =$$

$$= m_1(A_1) \times m_2(A_2)$$

$2 \Rightarrow 3$.-Si $m(A_1 \times A_2) = m_1(A_1) \times m_2(A_2)$, entonces,

$$\sum_{\substack{A_1 \subset X_1 \\ A_2 \subset X_2}} m(A_1 \times A_2) = \sum_{\substack{A_1 \subset X_1 \\ A_2 \subset X_2}} m_1(A_1) \times m_2(A_2) = \sum_{A_1 \subset X_1} m_1(A_1) \sum_{A_2 \subset X_2} m_2(A_2) = 1$$

y por tanto, de acuerdo con (28), la evidencia se concentra en los rectángulos.

Por otra parte,

$$Pl(A_1 \times A_2) = \sum_{A \cap (A_1 \times A_2) \neq \phi} m(A) = \sum_{(B_1 \times B_2) \cap (A_1 \times A_2) \neq \phi} m(B_1 \times B_2) =$$

$$= \sum_{\substack{B_1 \cap A_1 \neq \phi \\ B_2 \cap A_2 \neq \phi}} m_1(B_1) \times m_2(B_2) = \sum_{B_1 \cap A_1 \neq \phi} m_1(B_1) \sum_{B_2 \cap A_2 \neq \phi} m_2(B_2) =$$

$$= Pl_1(A_1) \times Pl_2(A_2)$$

Observemos que, cuando existe independencia fuerte, la evi-- dencia bidimensional se puede obtener a partir de sus marginales. Sin embargo, cuando solo existe independencia, ésto - en general, no sucederá asi. En todo caso, conoceremos Pl en los rectángulos mediante la expresión:

$$Pl\ (A_1 \times A_2) = Pl_1(A_1) \times Pl_2(A_2)$$

y por tanto, la creencia podrá determinarse en los conjuntos que sean complementarios de rectángulos.

En el caso particular de una distribución de posibilidad, la condición de independencia entre marginales queda:

$$\Pi(A_1 \times A_2) = \Pi_1(A_1) \times \Pi_2(A_2)$$

que se puede expresar equivalentemente, en función de distribuciones de posibilidad como:

$$\pi(x_1,x_2)=\pi_1(x_1).\pi_2(x_2) \qquad (29)$$

<u>Proposición</u> 2.7

Para distribuciones de posibilidad bidimensional, solo existe independencia fuerte entre sus marginales, cuando una de ellas sea de tipo crisp. En caso contrario, no puede existir.

Demostración

Sea π una distribución de posibilidad en $X_1 \times X_2$ y π_1 y π_2 sus marginales, con A.B.P asociadas m_π, m_{π_1} y m_{π_2} respectivamente.
Si ninguna de las marginales es de tipo crisp, entonces:

$$A_1 \not\subseteq B_1 \subseteq X_1, \quad A_2 \not\subseteq B_2 \subseteq X_2, \text{ con } m_{\pi_i}(A_i) > 0, \quad m_{\pi_i}(B_i) > 0 \qquad i=1,2$$

Si suponemos, que las evidencias marginales son fuertemente independientes, tendremos:

$$m_\pi(A_1 \times B_2) = m_{\pi_1}(A_1).m_{\pi_2}(B_2) > 0$$

y
$$m_\pi(B_1 \times A_2) = m_{\pi_1}(B_1).m_{\pi_2}(A_2) > 0$$

Sin embargo, ni $A_1 \times B_2 \subset B_1 \times A_2$ ni $B_1 \times A_2 \subset A_1 \times B_2$, en contra de la hipótesis de que m_π es de tipo posibilístico. (elementos focales anidados).

Si una de las marginales es de tipo preciso es inmediato, que la independencia implica la independencia fuerte.

Nota

Dada una distribución de posibilidad bidimensional, el conocimiento de sus marginales, siendo éstas independientes, nos permitirá reconstruir la distribución primitiva, por medio de (29).

- CAPITULO III-

La Esperanza Monótona.
Integral superior e inferior

0.- INTRODUCCION

Como es conocido y tendremos ocasión de discutir mas - ampliamente en el apartado IV, un problema de decisión puede reformularse en términos de ordenar un conjunto de vectores de consecuencias. Una forma de llevar a cabo ésta ordenación es "medir" o "valorar" éstos vectores, de modo coherente con un conjunto de axiomas de comportamiento, asociando a cada uno de ellos un número real mediante el que, con el orden natural de R, se determina la acción optima. Un ejemplo típico de ésta forma de actuar es el critério de la esperanza matemática para el ambiente de riesgo. Bajo la hipótesis de que el conocimiento disponible para el decisor - sobre los Estados de la Naturaleza se describe mediante una distribución de probabilidad, y aceptando un conjunto de axiomas de comportamiento, cada vector se valora por medio - de su esperanza matemática respecto de dicha distribución.

El modelo de decisión que presentamos y analizamos en el capítulo siguiente, tiene como característica fundamental el suponer que la información disponible sobre los estados, viene dada por una evidencia, (con lo que el ambiente de riesgo resulta ser un caso particular), es decir, por un par (Bel,Pl) de medidas de creencia y plausibilidad duales. Bajo éstas hipótesis, disponemos de dos herramientas para valorar un vector de consecuencias (supuestas numéricas):

1) La integral difusa, definida por Sugeno en (1974).

2) La esperanza monótona, definida por Huber en 1981 y Nguyen en 1982 (aunque pueden encontrarse antecedentes de la misma en Choquet (1953)), para medidas difusas y funciones positivas.

El empleo de la integral de Sugeno plantea dos problemas fundamentales; el primero es que no generaliza la esperanza matemática clásica. El segundo radica en que solo se pueden integrar funciones con valores en $[0,1]$, no siendo posible extenderla por linealidad a funciones generales, como probaron Klement y Ralescu (1983). Por ésta razón hemos escogido la esperanza monótona como la herramienta mas adecuada para la valoración de vectores de consecuencias en nuestro modelo de decisión.

Las propiedades de la esperanza monótona de funciones positivas han sido estudiadas por Bolaños, Lamata y Moral (1985). En éste capítulo definimos y analizamos la esperanza monótona para funciones cualesquiera.

Dempster (1967), introdujo las integrales superior e inferior asociadas a una evidencia. En el segundo apartado de éste capítulo, probamos que la integral superior coincide con la esperanza monótona respecto a la medida de plausibilidad asociada a la evidencia, y la integral inferior con la realizada respecto a la medida de creencia. Consideramos las propiedades que se deducen de ésta identificación y estudiamos aquellas que son particulares de éstas integrales. Entre ellas cabe destacar la relación de las mismas con la inclusión de evidencias y la independencia fuerte.

Terminamos el capítulo analizando la forma particular que adopta la esperanza monótona para algunos tipos especia les de medidas.

Como resumen queremos destacar que; la esperanza monótona es un operador tal que, dado un par de medidas difusas ordenadas (g_*, g^*), permite asociar a cada **función** dos valores de esperanza: uno respecto a g_*, (el valor esperado inferior), y otro respecto a g^*, (valor esperado superior), con la particularidad de que el primero siempre es menor o igual que el segundo. Podemos considerar pues que, a cada función asociamos un "intervalo esperado", intervalo que se reduce a un punto (la esperanza matemática clásica) en el caso de una medida de probabilidad.

1.- ESPERANZA MONOTONA RESPECTO A UNA MEDIDA DIFUSA

Sea X un conjunto finito y h una aplicación

$$h: X \to R_0^+$$

Si P es una medida de probabilidad en X, sabemos que

$$E_P(h) = \sum_{a \in X} p(a).h(a) =$$

$$= \int_0^\infty P[h \geqslant x]\,dx \qquad (30)$$

Basándonos en ésta expresión de la esperanza matemática introducimos la siguiente definición de esperanza respecto a una medida difusa.

Definición 3.1

Sea $h: X \rightrightarrows R_0^+$ y g una medida difusa en X. Llamaremos esperanza monótona de h con respecto a g, al valor

$$E_g(h) = \int_0^\infty g(H_\alpha)\,d\alpha \qquad (31)$$

donde

$$H_\alpha = \{a \in X / \ h(a) \geqslant \alpha\}$$

Proposición 3.1

La esperanza monótona $E_g(h)$ existe y es finita, cualesquiera que sean h y g.

Demostración

Es inmediata con solo tener en cuenta que $g(H_\alpha)$ es una función no creciente en α, y con $g(H_\alpha)=0$, $\alpha \geqslant \max_{a \in X} h(a)$; --- siendo éste máximo finito, por ser h una aplicación sobre un conjunto finito.

La esperanza monótona no es aditiva, en general, pero - si es lineal respecto a la suma y producto por un escalar positivo, como demuestra la siguiente proposición.

Proposición 3.2

Si $h:X \to R_o^+$, $c,b \in R$, $b \geqslant 0$ y g es una medida difusa en X, entonces:

$$E_g(c+bh) = c+bE_g(h)$$

Demostración

Por definición

$$E_g(c+bh) = \int_o^\infty g(\{a \in X/ \ (c+bh)(a) \geqslant \alpha\})d\alpha =$$

$$= \int_o^c g(\{a \in X/(c+bh)(a) \geqslant \alpha\})d\alpha +$$

$$+ \int_c^\infty g(\{a \in X/(c+bh)(a) \geqslant \alpha\})d\alpha =$$

$$= \int_o^c g(\{a \in X/h(a) \geqslant \frac{\alpha-c}{b}\})d\alpha + \int_c^\infty g(\{a \in X/h(a) \geqslant \frac{\alpha-c}{b}\})d\alpha =$$

$$= \int_o^c 1d\alpha + \int_c^\infty g(\{a \in X/h(a) \geqslant \frac{\alpha-c}{b}\})d\alpha.$$

Si hacemos el cambio de variable $\frac{\alpha-c}{b}=\beta$, nos queda:

$$E_g(c+bh) = c + \int_0^\infty g(\{a \in X / h(a) \geq \beta\}).b.d\beta =$$
$$= c + bE_g(h)$$

La proposición anterior, es de particular importancia a la hora de definir la esperanza monótona, para una función - cualquiera no necesariamente positiva.

Si $h: X \to R$, como X es finito, existirá $\min_{a \in X} h(x) = s$ $s \in R$. Si consideramos $h' = h - s$, entonces $h': X \to R_0^+$ y podemos calcular su esperanza monótona.

Definimos, entonces, la esperanza de h, como el valor

$$E_g(h) = E_g(h') + s = E_g(h-s) + s$$

Es inmediato comprobar que se sigue cumpliendo la propo sición 3.2.

Nota

La integral difusa de Sugeno (1974) de una función h $h: X \to [0,1]$, respecto a una medida difusa g, se define como:

$$\int_X h \circ g = \sup_{A \subset X} \{\min_{x \in A} h(x) \wedge g(A)\}$$

Kandel (1981-1982), propuso extender éste concepto a -- funciones cualesquiera, de forma lineal analogamente a como hemos hecho con la esperanza monótona.

Esta extensión ha sido duramente criticada por Klement y Ralescu (1983), debido a que la integral de Sugeno no es - lineal, con lo que la extensión no resulta consistente.

Si P es una medida de probabilidad y h una aplicación, $h: X \to [0,1]$, Sugeno (1974), probó que:

$$| E_P(h) - \int h \circ P | \leq 1/4$$

En Bolaños, Lamata y Moral (1985), ésta cota ha sido - generalizada, probando que si g es una medida difusa en X y h una aplicación $h: X \to [0,1]$, entonces:

$$| E_g(h) - \int h\circ g | \leqslant (\int h\circ g)(1 - \int h\circ g) \leqslant 1/4$$

Si (g_*, g^*) es un par de medidas difusas ordenadas, entonces, para cada función $h:X \to R$, tendremos dos valores, -- $E_{g_*}(h)$ y $E_{g^*}(h)$; que, como consecuencia de la proposición anterior verifican:

$$E_{g_*}(h) \leqslant E_{g^*}(h)$$

Podemos considerar, por tanto, que asociado a un par de medidas difusas ordenadas, tendremos un intervalo de esperanza:

$$[E_{g_*}(h), E_{g^*}(h)]$$

Este intervalo, se reducirá a un punto en el caso de una medida de probabilidad, ya que para ellas $g_* \equiv g^*$.

Proposición 3.3

Si g es una medida de probabilidad en X, $E_g(h)$ coincide con la esperanza matemática clásica de h, cualquiera que sea $h:X \to R$.

Demostración

Es inmediata como consecuencia de (30) y de la proposición 3.2.

A continuación recogemos algunas propiedades elementales de E_g.

Proposición 3.4

Si $h(a) = x_o \in R$, $\forall a \in X$; entonces, $E_g(h) = x_o$, cualquiera que sea la medida g.

Demostración

- Si $x_o \geqslant 0$, entonces

$$H_\alpha = \begin{cases} X & \text{si} \quad \alpha \in [0, x_o] \\ \phi & \text{si} \quad \alpha > x_o \end{cases}$$

y, por tanto

$$E_g(h) = \int_o^\infty g(H_\alpha)d\alpha = \int_o^{x_o} g(u)d\alpha = \int_o^{x_o} d\alpha = x_o$$

- Si x_o 0, entonces

$$E_g(h) = E_g(h-x_o)+x_o = 0+x_o = x_o$$

ya que

$$(h-x_o)(a)=0 \qquad \forall a \in X.$$

<u>Proposición</u> 3.5

La esperanza monótona respeta el orden entre las aplica ciones. Es decir, si $h_1, h_2 : X \to R$ son tales que $h_1(a) \leq h_2(a)$, $\forall\, a \in X$, entonces se verifica que:

$$E_g(h_1) \leq E_g(h_2)$$

Demostración

Si $h_1(a) \geqslant 0$, $\forall a \in X$, entonces

$$E_g(h_i) = \int_0^\infty g(H_{i\alpha})d\alpha$$

y la proposición se deduce, teniendo en cuenta, que si: $h_1(a) \leqslant h_2(a)$, $\forall a \in X$, entonces $H_{1\alpha} \subset H_{2\alpha}$, $\alpha \in R_o^+$, y considerando la monotonia de la medida difusa y la integral de Lebesgue.

-Si $\exists\, a \in X$; con $h_1(a) < 0$, entonces:

$$E_g(h_i) = E_g(h_i - s) + s$$

siendo, $s = \min\limits_{a \varepsilon X} h_1(a)$

Basta, por tanto, aplicar la primera parte de la demostración a $h_1 - s \leqslant h_2 - s$, para concluir la demostración.

Proposición 3.6

Si $g_1(A) \leqslant g_2(A)$, $\forall A \subset X$, entonces

$$E_{g_1}(h) \leqslant E_{g_2}(h)$$

cualquiera que sea $h: X \to R$.

Demostración

Si $h \geqslant 0$, es una consecuencia inmediata de la definición ya que, $g_1(H_\alpha) \leqslant g_2(H_\alpha) \quad \forall \alpha \geqslant 0$.

En el caso general se obtiene por linealidad.

Proposición 3.7

Si $h: X \to R$ y $b \varepsilon R_o^-$, entonces:

$$E_g(b.h) = b.E_{\overline{g}}(h)$$

siendo, g y $\overline{g}$ medidas difusas duales.

Demostración

Probaremos la proposición para $h \quad 0$. El caso general,- se obtiene por linealidad. Sea $t = \max\limits_{a \varepsilon X} h(a) \geqslant 0$, entonces

$$tb = \min_{a \varepsilon X} (h.h(a)) \quad y$$

$$E_g(b.h) = E_g(bh - bt) + bt = \int_0^\infty g(\{x \varepsilon X / bh(x) - bt \geqslant \alpha\}) d\alpha + bt =$$

$$= bt + \int_0^\infty g(\{x \varepsilon X / h(x) \leqslant \frac{\alpha + bt}{b}\}) \; d\alpha =$$

$$\int_{o}^{-bt} g(\{x\varepsilon X/h(x) \leqslant \frac{\alpha+bt}{b}\})\ d\alpha + \int_{-bt}^{\infty} g(\{x\varepsilon X/h(x) \leqslant \frac{\alpha+bt}{b}\})\ d\alpha+bt$$

Ahora bien, si $\alpha > -bt$, entonces, $\alpha+bt \geqslant 0$ y $\frac{\alpha+bt}{b} \leqslant 0 \leqslant h(x)$, obteniendose:

$$E_g(bh) = \int_{o}^{-bt} g(\{x\varepsilon X/h(x) \leqslant \frac{\alpha+bt}{b}\})\ d\alpha + \int_{-bt}^{\infty} 0.d\alpha+bt =$$

$$= \int_{o}^{-bt} [1-\overline{g}(\{x\varepsilon X/h(x) > \frac{\alpha+bt}{b}\})]\ d\alpha+bt =$$

$$= \int_{o}^{-bt} 1.d\alpha - \int_{o}^{-bt} \overline{g}(\{x\varepsilon X/h(x) > \frac{\alpha+bt}{b}\})\ d\alpha+bt$$

si hacemos $\frac{\alpha+bt}{b} = \beta$, queda

$$E_g(bh) = -bt - \int_{t}^{o} \overline{g}(\{x\varepsilon X/h(x) > \beta\})b\ d\beta + bt =$$

$$= b \cdot \int_{o}^{t} \overline{g}(\{x\varepsilon X/h(x) > \beta\})\ d\beta = bE_{\overline{g}}(h)$$

<u>Nota</u>

La esperanza monótona ha sido utilizada en Bolaños, Lamata y Moral (1985), para extender medidas difusas a conjuntos difusos. Si $g\varepsilon MD(X)$, podemos definir la medida de un subconjunto difuso $A\varepsilon \underset{\sim}{\mathcal{G}}(X)$, como:

$$\underset{\sim}{g}(A) = E_g(\mu_A) \qquad (32)$$

En éstas condiciones y considerando el complementario del -- conjunto difuso A, como $\overline{A}$, con función de pertenencia

$$\mu_{\overline{A}}(a) = 1-\mu_A(a)$$

se obtiene, como consecuencia de la proposición 3.7, que la extensión (32) conserva la dualidad. Es decir:

$$g(A) + \bar{g}(\bar{A}) = E_g(\mu_A)+E_{\bar{g}}(\mu_{\bar{A}}) = E_g(\mu_A)+E_{\bar{g}}(1-\mu_A) =$$

$$= E_g(\mu_A)+E_{\bar{g}}(1)-E_g(\mu_A) = 1$$

Como ya hemos dicho, la esperanza monótona no es aditiva en general, lo que la diferencia de la integral de Lebesgue, que, como bien se sabe, es siempre un funcional aditivo

La siguiente proposición demuestra que la esperanza monótona solo cumple ésta propiedad en el caso en el que la medida g sea una medida de probabilidad.

Proposición 3.8

Si g es una medida difusa en X, una condición necesaria y suficiente para que g sea una medida de probabilidad, es que:

$$E_g(h_1+h_2) = E_g(h_1)+E_g(h_2), \ \forall h_1,h_2:X \to R$$

Demostración

La condición necesaria es inmediata a partir de la proposición 3.3.

Para probar la condición suficiente, basta con verificar que, dados $A_1,A_2 \in \mathcal{G}(X)$, tales que $A_1 \cap A_2=\phi$, entonces, $g(A_1 \cup A_2)=g(A_1)+g(A_2)$

Designemos por I_1 e I_2 las funciones indicador de A_1 y de A_2 respectivamente. Obtenemos entonces:

$$E_g(I_1+I_2) = \int_0^1 g(A_1 \cup A_2)\ d\alpha = g(A \cup B)$$

$$E_g(I_i) = g(A_i), \quad i=1,2$$

Como, por hipótesis, se cumple que:

$E_g(I_1+I_2) = E_g(I_1)+E_g(I_2)$, se obtiene que

$$g(A_1 \cup A_2) = g(A_1) + g(A_2)$$

En general, solo puede asegurarse que la esperanza monótona es aditiva , respecto a las partes positiva y negativa de una función.

Proposición 3.9

Si g es una medida difusa en X, y $h: X \to R$; entonces

$$E_g(h) = E_g(h^+) + E_g(-h^-) = E_g(h^+) - E_{\overline{g}}(h^-)$$

donde:

$$h^+ = \max\{h, 0\} \qquad y \qquad h^- = \max\{-h, 0\}$$

Demostración

Si $h \geqslant 0$ es inmediata

Si $h \ngeq 0$, sea $t = \min\limits_{a \in X} h(a) = \min\limits_{a \in X}(-h^-(a)) < 0$. Entonces

$$E_g(h) = \int_0^\infty g(\{x \in X / h(x) - t \geqslant \alpha\})\, d\alpha + t =$$

$$= t + \int_0^{-t} g(\{x \in X / h(x) - t \geqslant \alpha\}) d\alpha + \int_{-t}^{\infty} g(\{x \in X / h(x) \geqslant \alpha + t\}) d\alpha.$$

Teniendo en cuenta que, para todo $\alpha \leqslant -t$ es $\alpha + t \leqslant 0$, y por tanto son equivalentes $h(x) \geqslant \alpha + t$ y $-h^-(x) = \min\{h(x), 0\} \geqslant \alpha + t$, se tiene:

$$E_g(h) = t + \int_0^{-t} g(\{x \in X / -h^-(x) \geqslant \alpha + t\}) d\alpha + \int_{-t}^{\infty} g(\{x \in X / h(x) \geq t + \alpha\}) d\alpha$$

Si hacemos, $t + \alpha = \beta$ en la segunda integral y consideramos que si $\alpha > -t$, entonces $\{x \in X / -h^-(x) \geqslant \alpha + t\} = \phi$, obtenemos:

$$E_g(h) = t + \int_0^{\infty} g(\{x \in X / -h^-(x) \geqslant \alpha + t\}) d\alpha + \int_0^{\infty} g(\{x \in X / h(x) \geq \beta\}) d\beta$$

ahora bien, si $\beta > 0$, entonces:

$$\{x \in X / h(x) \geq \beta\} = \{x \in X / h^{+}(x) \geq \beta\}$$

con lo que,

$$E_g(h) = t + \int_0^\infty g(\{x \in X / -h^{-}(x) - t \geq \alpha\})\, d\alpha + \int_0^\infty g(\{x \in X / h^{+}(x) \geq \beta\})\, d\beta$$

$$= E_g(-h^{-}) + E_g(h^{+})$$

Huber (1981), ha caracterizado las capacidades de orden dos de Choquet con ayuda de la esperanza monótona, demostrando la siguiente proposición.

Proposición 3.10

Si g es una medida difusa en X, entonces:

a) g es una capacidad monótona de orden dos, si y solo si:

$$E_g(h_1 + h_2) \geq E_g(h_1) + E_g(h_2), \qquad \forall\, h_i : X \to R_o^{+}$$

b) g es una capacidad alternante de orden dos, si y solo si:

$$E_g(h_1 + h_2) \leq E_g(h_1) + E_g(h_2) \qquad \forall\, h_i : X \to R_o^{+}$$

Nota

Es inmediato comprobar, a partir de la proposición anterior, que:

1) si g es una capacidad monótona de orden dos, se verifica:

$$E_g(h_1 + h_2) \geq E_g(h_1) + E_g(h_2), \qquad \forall\, h_i : X \to R, \quad i=1,2$$

2) si g es una capacidad alternante de orden dos, entonces:

$$E_g(h_1 + h_2) \leq E_g(h_1) + E_g(h_2)$$

cualesquiera que sean $h_i : X \to R$, $i=1,2$

Como corolario de la proposición 3.10, obtenemos, en el caso de una evidencia, la siguiente propiedad.

Corolario

Si m es una A.B.P con medidas de plausibilidad y creencia asociadas, Pl y Bel, entonces:

$$E_{Bel}(h_1+h_2) \geq E_{Bel}(h_1) + E_{Bel}(h_2)$$

$$E_{Pl}(h_1+h_2) \leq E_{Pl}(h_1) + E_{Pl}(h_2)$$

cualesquiera que sean $h_i:X \to R, \quad i=1,2$

Nota

Si consideramos que el par (Bel,Pl) asocia a cada función un intervalo de esperanza, $[E_{Bel}(h),E_{Pl}(h)]$, entonces:

$$[E_{Bel}(h_1+h_2),E_{Pl}(h_1+h_2)] \subset [E_{Bel}(h_1),E_{Pl}(h_2)] + [E_{Bel}(h_1),E_{Pl}(h_2)]$$

donde la suma de los intervalos se entiende en su sentido habitual.

Una consecuencia de la linealidad de la esperanza monótona respecto a la multiplicación por un escalar positivo y de la proposición 3.10, es que, para una capacidad de orden dos, $E_g(.)$ es un operador convexo en el espacio de las funciones sobre X, siendo cóncavo cuando g es una capacidad alternante de orden dos, como queda enunciado en la siguiente proposición.

Proposición 3.11

Si $h_1,h_2:X \to R$ y $\lambda \in [0,1]$; entonces

a) Si g es una capacidad monótona de orden dos

$$E_g(\lambda h_1+(1-\lambda)h_2) \geqslant \lambda E_g(h_1)+(1-\lambda)E_g(h_2)$$

b) Si g es una capacidad alternante de orden dos

$$E_g(\lambda h_1+(1-\lambda)h_2) \leqslant \lambda E_g(h_1)+(1-\lambda)E_g(h_2)$$

Nota

Teniendo en cuenta la proposición 3.11, para todo ----- $(g_*,g^*)\varepsilon C2(X)$,se obtiene la siguiente propiedad de los intervalos de esperanza:

$$[E_{g_*}(\lambda h_1+(1-\lambda)h_2),E_{g^*}(\lambda h_1+(1-\lambda)h_2)]\subset$$

$$\lambda[E_{g_*}(h_1),E_{g^*}(h_1)] + (1-\lambda)[E_{g_*}(h_2),E_{g^*}(h_2)]$$

Proposición 3.12

Si g_1 y g_2 son medidas difusas y $\lambda\varepsilon[0,1]$

$$E_{\lambda g_1+(1-\lambda)g_2}(h) = \lambda E_{g_1}(h)+(1-\lambda)E_{g_2}(h)$$

$\forall$ h:X $\to$ R.

Demostración

Se deduce teniendo en cuenta que

$$E_{\lambda g_1+(1-\lambda)g_2}(h) = E_{\lambda g_1+(1-\lambda)g_2}(h^+)-E_{\lambda\bar{g}_1+(1-\lambda)\bar{g}_2}(h^-)$$

$$= \int_o^\infty(\lambda g_1+(1-\lambda)g_2)(H_\alpha^+)d\alpha-\int_o^\infty(\lambda\bar{g}_1+(1-\lambda)\bar{g}_2)(H_\alpha^-)d\alpha$$

$$= \lambda[\int_o^\infty g_1(H_\alpha^+)d\alpha-\int_o^\infty\bar{g}_1(H_\alpha^-)d\alpha]$$

$$+(1-\lambda)[\int_o^\infty g_2(H_\alpha^+)d\alpha-\int_o^\infty g_2(H_\alpha^-)d\alpha]$$

$$= \lambda E_{g_1}(h) + (1-\lambda)E_{g_2}(h)$$

2.- CASOS PARTICULARES DE LA ESPERANZA MONOTONA

En éste apartado vamos a estudiar las expresiones que toma la esperanza monótona cuando consideramos distintos tipos particulares de medidas difusas. Prestaremos una particular atención al caso de las evidencias, por ser éstas las medidas que consideraremos en el capítulo siguiente.

2.1. La esperanza monótona en EV(X). Integrales superior e inferior.

El primer concepto de esperanza respecto a una evidencia se debe a Dempster (1967), quien introdujo las llamadas integrales superior e inferior.

Definición 3.2

Sea $h:X \to R$ y $(Bel,Pl) \varepsilon EV(X)$. En éstas condiciones:

a) Se denomina integral superior de h con respecto a la evidencia, al valor:

$$E^*(h) = \int_{-\infty}^{\infty} x.dF_*(x)$$

con:

$$F_*(x) \doteq Bel(\{a \varepsilon X/h(a) \leqslant x\})$$

b) Se denomina integral inferior de h con respecto a la evidencia, al valor:

$$E_*(h) = \int_{-\infty}^{\infty} x.dF(x)$$

con:

$$F(x) = Pl(\{a \varepsilon X/h(a) \leqslant x)$$

Notación: Si m es la A.B.P asociada a (Bel,Pl), nosotros no taremos a partir de ahora $E^*(h)$ como $I^*(h/m)$ y $E_*(h)$ como -- $I_*(h/m)$.

Como el conjunto X es finito, $X=\{x_1,\ldots,x_n\}$, la función h puede identificarse con el vector $\tilde{h}=(h(x_1),\ldots,h(x_n))\varepsilon R^n$ y por ello a veces escribiremos $I^*(\tilde{h}/m)$ e $I_*(\tilde{h}/m)$. Recíprocamente, cualquier $\tilde{r}\varepsilon R^n$, puede identificarse con una aplicación $r\varepsilon R^n$, dado lo cual tiene sentido considerar $I_*(r/m)$ e - $I^*(r/m)$.

<u>Proposición</u> 3.13. (Dempster (1967))

Si m es una A.B.P en X, y $h:X \rightarrow R$, entonces:

$$I^*(h/m) = \sup_{P\varepsilon\mathcal{P}} [E_P(h)]$$

$$I_*(h/m) = \inf_{P\varepsilon\mathcal{P}} [E_P(h)]$$

donde, $\mathcal{P}$ es la familia de medidas de probabilidad, que verifican:

$$Bel(A) \leqslant P(A) \leqslant Pl(A) \qquad \forall A\subset X$$

y $E_P(h)$ es la esperanza matemática clásica.

Smets(1981), ha probado que éstas integrales se pueden poner de la forma:

$$I_*(h/m) = \sum_{A\subset X} m(A)(\inf_{x\varepsilon A} h(x))$$

$$I^*(h/m) = \sum_{A\subset X} m(A)(\sup_{x\varepsilon A} h(x))$$

<u>Nota</u>

En vista de las anteriores expresiones para las integrales superior e inferior, podemos interpretar la esperanza inferior de una aplicación h, respecto a una A.B.P m, como la esperanza matemática de la función h respecto a la medida de probabilidad que se obtendria de la siguinete forma: toda la masa de evidencia correspondiente al conjunto focal $A \subset X$, se la asignariamos a aquel elemento $a \varepsilon A$, para el cual la función h, alcance su mínimo. Análogamente, se puede interpretar la integral superior: la evidencia correspondiente a A, se concentra en el elemento $b \varepsilon A$, para el cual h alcanza su máximo, dentro de A.

Nota

Si m_o es la A.B.P que representa la ignorancia

$$I^*(h/m_o) = \sum_{A \subset X} m_o(A) \max_{a \varepsilon A} h(a) = \sup_{a \varepsilon X} h(a)$$

$$I_*(h/m_o) = \sum_{A \subset X} m_o(A) \min_{a \varepsilon A} h(a) = \inf_{a \varepsilon X} h(a)$$

es decir, las integrales superior e inferior, cuando no tenemos ninguna información sobre X, coinciden con el mayor y el menor valor de h en X, respectivamente.

A continuación vamos a probar que las integrales superior e inferior respecto a una A.B.P m, coinciden con la esperanza monótona respecto a las medidas de plausibilidad y creencia, respectivamente, asociadas a m.

Proposición 3.14

Si m es una A.B.P en X, con medidas de creencia y plausibilidad asociadas; Bel y Pl, entonces:

$$I_*(h/m) = E_{Bel}(h)$$

$$I^*(h/m) = E_{Pl}(h)$$

$\forall\ h: X \rightarrow R$

Demostración

$$E_{Bel}(h) = E_{Bel}(h^+) - E_{Pl}(h^-)$$

$$= \int_0^\infty Bel\{x \varepsilon X / h^+(x) \geqslant \alpha\}\, d\alpha - \int_0^\infty Pl\{x \varepsilon X / h^-(x) \geqslant \alpha\}\, d\alpha$$

$$= \int_0^\infty \sum_{A \subset H_\alpha^+} m(A)\, d\alpha - \int_0^\infty \sum_{A \cap H_\alpha^- \neq \phi} m(A)\, d\alpha$$

sea ahora $A \subset X$. Pueden ocurrir dos casos:

a) $\inf_{x \varepsilon A} h(x) \geqslant 0$. Entonces:

$$A \cap H_\alpha^- = \{x \varepsilon A / h^-(x) \geqslant \alpha\} = \phi \qquad \forall \alpha > 0. \quad y$$

$$A \subset H_\alpha^+ \Leftrightarrow \alpha \leqslant \inf_{x \varepsilon A} h(x)$$

b) $\inf_{x \varepsilon A} h(x) < 0$. Entonces:

$\exists x \varepsilon A$ tal que $h(x) < \alpha$, $\forall \alpha > 0$ y por tanto $A \not\subset H_\alpha^+$.

Además $A \cap H_\alpha^- \neq \phi$ equivale a $\alpha < \sup_{x \varepsilon A} h^-(x) = -\inf_{x \varepsilon A} h(x)$

En éstas condiciones, intercambiando la sumatoria y la integral en la expresión de $E_{Bel}(h)$ se obtiene:

$$E_{Bel}(h) = \sum_{A \subset X} \int_0^{\inf_{x \varepsilon A} h(x)} m(A)\, d\alpha - \sum_{A \subset X} \int_0^{-\inf_{x \varepsilon A} h(x)} m(A)\, d\alpha$$

donde, la primera sumatoria se extiende a los $A \subset X$, tales -- que, $\inf_{x \varepsilon A} h(x) \geqslant 0$, y la segunda a aquellos tales que $\inf_{x \varepsilon A} h(x) < 0$. Asi pues, y en definitiva

$$E_{Bel}(h) = \sum_{A \subset X} m(A) \inf_{x \varepsilon A} h(x) = I^*(h/m)$$

Para la medida de plausibilidad se puede demostrar la - igualdad por dualidad.

$$E_{Pl}(h) = -E_{Bel}(-h) = -\sum_{A\subset X} m(A) \inf_{x\varepsilon A}(-h(x))$$

$$= -(-\sum_{A\subset X} m(A) \sup_{x\varepsilon A} h(x)) = I^*(h/m)$$

A partir de la proposición anterior, podemos obtener -- las siguientes propiedades de las integrales superior e inferior, sin mas qüe particularizar las correspondientes de la esperanza monótona:

1.- Si $c,b\varepsilon R$, $b \geqslant 0$ y $h:X \to R$, entonces:

$$I^*(c+bh/m) = c+bI^*(h/m) \quad ; \quad I_*(c+bh/m) = c+bI_*(h/m)$$

2.- Si m se concentra en los conjuntos unitarios:

$$I^*(h/m) = I_*(h/m) = \sum_{a\varepsilon X} h(a),m(a)$$

3.- Si $h_1(a) \leqslant h_2(a)$, $\forall a\varepsilon X$, entonces:

$$I^*(h_1/m) \leqslant I^*(h_2/m) \quad ; \quad I_*(h_1/m) \leqslant I_*(h_2/m)$$

4.- Si $h(a)=x_o\varepsilon R$, $\forall a\varepsilon X$, entonces:

$$I^*(h/m) = I_*(h/m) = x_o$$

5.- Si $c,b\varepsilon R$ y $b \leqslant 0$

$$I^*(c+bh/m) = c+bI_*(h/m) \quad ; \quad I_*(c+bh/m) = c+bI^*(h/m)$$

6.- m se concentra en los unitarios, si y solo si:

$I^*(h_1+h_2/m) = I^*(h_1/m) + I^*(h_2/m)$ o equivalentemente

$$I_*(h_1+h_2/m) = I_*(h_1/m) + I_*(h_2/m)$$

cualesquiera que sean $h_1,h_2:X \to R$.

7.- $I^*(h/m) = I^*(h^+/m) - I_*(h^-/m)$

$$I_*(h/m) = I_*(h^+/M) - I^*(h^-/m)$$

8.- $I^*(h_1+h_2/m) \leqslant I^*(h_1/m) + I^*(h_2/m)$

$I_*(h_1+h_2/m) \geqslant I_*(h_1/m) + I_*(h_2/m)$

9.- Si $\lambda \varepsilon [0,1]$, $I^*(\lambda h_1+(1-\lambda)h_2/m) \leqslant \lambda I^*(h_1/m)+(1-\lambda)I^*(h_2/m)$

$I_*(\lambda h_1+(1-\lambda)h_2/m) \geqslant \lambda I_*(h_1/m)+(1-\lambda)I_*(h_2/m)$

10.- Si $\lambda \varepsilon [0,1]$, $I^*(h/\lambda m_1+(1-\lambda)m_2) = \lambda I^*(h/m_1)+(1-\lambda)I^*(h/m_2)$

$I_*(h/\lambda m_1+(1-\lambda)m_2) = \lambda I_*(h/m_1)+(1-\lambda)I_*(h/m_2)$

A continuación vamos a probar una serie de propiedades específicas de las integrales superior e inferior.

Proposición 3.15

Si m_1 y m_2, son dos A.B.P en X, tales que $m_1 \subset m_2$, entonces:

$$[I_*(h/m_2), I^*(h/m_2)] \subset [I_*(h/m_1), I^*(h/m_1)]$$

Demostración

Se deduce inmediatamente, sin mas que considerar, que si $m_1 \subset m_2$, entonces, $Bel_1 \leqslant Bel_2$ y $Pl_1 \geqslant Pl_2$, con lo que:

$$I_*(h/m_2) = E_{Bel_2}(h) \geqslant E_{Bel_1}(h) = I_*(h/m_1)$$

$$I^*(h/m_2) = E_{Pl_2}(h) \leqslant E_{Pl_1}(h) = I^*(h/m_1)$$

Vemos como, a medida que vamos adquiriendo información, los conjuntos focales se van atomizando, dando como resultado, que para una misma aplicación h, la amplitud de los intervalos determinados a partir de las integrales superior e inferior, va disminuyendo, hasta reducirse a un punto, en el caso, en que la asignación básica de probabilidad sea de tipo probabilístico.

La siguiente proposición, caracteriza las esperanzas monótonas que están asociadas a medidas de plausibilidad y creencia.

Proposición 3.16

Sea g una medida difusa y $E_g(.)$ su esperanza monótona. En éstas condiciones

a) g es una medida de plausibilidad, si y solo si, para cualquier familia $\{h_i\}_{i=1,,,p}$ de funciones se cumple:

$$E_g(\inf_i h_i) \leqslant \sum_{I\subset\{1,..,p\}} (-1)^{|I|+1} E_g(\sup_i h_i)$$

b) g es una medida de creencia si y solo si, para cualquier familia $\{h_i\}_{i=1,..,p}$, se cumple:

$$E_g(\sup_i h_i) \geqslant \sum_{I\subset\{1,..,p\}} (-1)^{|I|+1} E_g(\inf_i h_i)$$

Demostración

La suficiencia de las propiedades a) y b), se obtiene considerando para cada familia de subconjuntos de X, $\{A_i\}$ $i\varepsilon\{1,..,p\}$, la familia de funciones indicador de éstos conjuntos, $\{I_{A_i}\}_{i\varepsilon\{1,..,p\}}$.

Las condiciones necesarias para familias de funciones no negativas, son consecuencia de las igualdades:

$$(\{x\varepsilon X/ \inf_{i\varepsilon I} h_i(x) \geqslant \alpha\}) = \bigcap_{i\varepsilon I} \{x\varepsilon X/ h_i(x)\geqslant\alpha\} \ , \ \forall \alpha\varepsilon R_o^+$$

$$(\{x\varepsilon X/ \sup_{i\varepsilon I} h_i(x) \geqslant \alpha\}) = \bigcup_{i\varepsilon I} \{x\varepsilon X/ h_i(x)\geqslant\alpha\} \ , \ \forall \alpha\varepsilon R_o^+$$

aplicadas a la definición de esperanza monótona.

Para una familia de funciones cualesquiera, $\{h_i\}_{i\varepsilon\{1,.p\}}$ hay que tener en cuenta que si

$$h_i^+ = \sup\{h_i,0\} \qquad y \qquad h_i^- = \sup\{h_i,0\},$$

entonces:

$$\sup_{i\varepsilon I}\{h_i\} = \sup_{i\varepsilon I} h_i^+ - \inf_{i\varepsilon I} h_i^- \qquad \forall\, I \subset\{1,\ldots,p\}$$

$$\inf_{i\varepsilon I}\{h_i\} = \inf_{i\varepsilon I} h_i^+ - \sup_{i\varepsilon I} h_i^- \qquad \forall\, I \subset\{1,\ldots,p\}$$

por tanto, para la esperanza con respecto a Pl(.), se tiene:

$$E_{Pl}(\inf_{i\varepsilon\{1,.,p\}} h_i) = E_{Pl}(\inf_{i\varepsilon\{1,.,p\}} h_i) + E_{Pl}(-\sup_{i\varepsilon\{1,.,p\}} h_i)$$

$$\leqslant \sum_{I\subset\{1,.,p\}} (-1)^{|I|+1} E_{Pl}(\sup_{i\varepsilon I} h_i^+) - E_{Bel}(\sup_{i\varepsilon\{1,.,p\}} h_i^-)$$

$$\leqslant \sum_{I\subset\{1,.,p\}} (-1)^{|I|+1} E_{Pl}(\sup_{i\varepsilon I} h_i^+) - (\sum_{I\subset\{1,.,p\}} (-1)^{|I|+1} E_{Bel}(\inf_{i\varepsilon I} h_i^-)$$

$$= \sum_{I\subset\{1,.,p\}} (-1)^{|I|+1} E_{Pl}(\sup_{i\varepsilon I} h_i^+) + E_{Pl}(-\inf_{i\varepsilon I} h_i^-)$$

$$= \sum_{I\subset\{1,.,p\}} (-1)^{|I|+1} E_{Pl}(\sup_{i\varepsilon I} h_i)$$

la demostración para medidas de creencia es análoga a la anterior.

Sabemos que cuando dos variables aleatorias f y g son - independientes,

$$E(f.g) = E(f).E(g)$$

siendo, E(.) la esperanza matemática clásica.

En el siguiente teorema, demostramos una propiedad análoga para las esperanzas inferiores y superiores, cuando --- existe independencia fuerte.

Proposición 3.17

Sea m una evidencia bidimensional en $X_1 \times X_2$. Entonces las siguientes condiciones son equivalentes.

1) m_1 y m_2 son fuertemente independientes.

2) $I^*(h_1.h_2/m) = I^*(h_1/m).I^*(h_2/m), \quad \forall\, h_i : X_i \rightarrow R_o^+, \ i=1,2$

3) $I_*(h_1.h_2/m) = I_*(h_1/m).I_*(h_2/m), \quad \forall\, h_i : X_i \rightarrow R_o^+, \ i=1,2$

Demostración

Probaremos 1⇔2. La demostración 1⇔3, es análoga a la anterior.

1⇒2. Bajo la hipótesis de independencia fuerte:

$$I^*(h_1,h_2) = \sum_{A \subset X_1 \times X_2} m(A) \sup_{(x_1,x_2) \varepsilon A} h_1(x_1).h_2(x_2)$$

$$= \sum_{A_1 \times A_2 \subset X_1 \times X_2} m_1(A_1).m_2(A_2) \sup_{\substack{x_1 \varepsilon A_1 \\ x_2 \varepsilon A_2}} h_1(x_1).h_2(x_2)$$

teniendo en cuenta que, si $h_i \geqslant 0$ y $h_2 \geqslant 0$, entonces:

$$\sup_{\substack{x_1 \varepsilon A_1 \\ x_2 \varepsilon A_2}} h_1(x_1).h_2(x_2) = (\sup_{x_1 \varepsilon A_1} h_1(x_1)) (\sup_{x_2 \varepsilon A_2} h_2(x_2))$$

obtenemos:

$$I^*(h_1.h_2/m) = (\sum_{A_1 \subset X_1} m_1(A_1) \sup_{x_1 \varepsilon A_1} h_1(x_1))(\sum_{A_2 \subset X_2} m_2(A_2) \sup_{x_2 \varepsilon A_2} h_2(x_2))$$

$$= I^*(h_1/m_1).I^*(h_2/m_2).$$

2⇒1. Proberamos en primer lugar que existe independencia, para lo cual, se ha de verificar que:

$$Pl(A_1 \times A_2) = Pl_1(A_1).Pl_2(A_2) \qquad \forall\, A_1 \subset X_1,\ A_2 \subset X_2.$$

Esta igualdad se deduce del comportamiento de 2), tomando:

$$h_i = I(A_i) \qquad i=1,2$$

y teniendo encuenta que:

$$I^*(I_{A_i}/m_i) = \sum_{B_i \cap A_i \neq \phi} m_i(B_i) = Pl_i(A_i)$$

$$I^*(I_{A_1}.I_{A_2}/m) = I^*(I_{A_1 \times A_2}/m) = Pl(A_1 \times A_2)$$

sean ahora

$$h_1(x_1) = \begin{cases} 1 & \text{si} \quad x_1 \varepsilon A_1 \\ 1/2 & \text{si} \quad x_1 \not\varepsilon A_1 \end{cases}$$

$$h_2(x_2) = \begin{cases} 1 & \text{si} \quad x_2 \varepsilon A_2 \\ 1/2 & \text{si} \quad x_2 \not\varepsilon A_2 \end{cases}$$

aplicando 2), se obtiene:

$$I^*(h_1.h_2/m) = I^*(h_1/m_1).I^*(h_2/m_2)$$

es decir,

$$1/4\ Pl(X_1 \times X_2) + 1/4\ Pl(A_1 \times X_2 \cup X_1 \times A_2) + 1/2\ Pl(A_1 \times A_2)$$

$$= (1/2\ Pl_1(X_1) + 1/2\ Pl_1(A_1))(1/2\ Pl_2(X_2) + 1/2 Pl_2(A_2)$$

lo que es equivalente a,

$$1/4 + 1/4\ Pl(A_1 \times X_2 \cup X_1 \times A_2) + 1/2\ Pl(A_1 \times A_2)$$

$1/4 + 1/4\, Pl_1(A_1) + 1/4\, Pl_2(A_2) + 1/4\, Pl_1(A_1).Pl_2(A_2) =$

simplificando y teniendo encuenta la independencia, obtenemos definitivamente:

$$Pl(A_1\times X_2 \cup X_1\times A_2) = Pl_1(A_1) + Pl_2(A_2) - Pl(A_1\times A_2) \qquad (32)$$

Por último, vamos a comprobar, que si m no se concentra en los rectángulos, entonces, no se verifica (32).

Supongamos que existe $B\subset X_1\times X_2$ tal que $B\neq P_1(B)\times P_2(B)$, pero $m(B)\neq 0$. Sea $(x_1,x_2)\in P_1(B)\times P_2(B)-B$ y hagamos $A_1=\{x_1\}$, --- $A_2=\{x_2\}$. En éstas condiciones:

$$Pl(A_1\times X_2\cup X_1\times A_2) = \sum_{\substack{x_1\in P_1(A)\\ x_2\in P_2(A)}} m(A) = \sum_{x_1\in P_1(A)} m(A) + \sum_{x_2\in P_2(A)} m(A) - \sum_{\substack{x_1\in P_1(A)\\ x_2\in P_2(A)}} m(A)$$

$$= Pl_1(A_1) + Pl_2(A_2) - \sum_{\substack{x_1\in P_1(A)\\ x_2\in P_2(A)}} m(A)$$

Puesto que, $(x_1,x_2)\in A$ supone $x_1\in P_1(A)$ y $x_2\in P_2(A)$ y, por otra parte, existe B con $m(B)>0$ y $(x_1,x_2)\notin B$, se tiene:

$$\sum_{\substack{x_1\in P_1(A)\\ x_2\in P_2(A)}} m(A) > \sum_{(x_1,x_2)\in A} m(A) = Pl(A_1\times A_2)$$

con lo que:

$$Pl(A_1\times X_2\cup X_1\times A_2) < Pl_1(A_1) + Pl_2(A_2) - Pl(A_1\times A_2) \;;$$

en contra de (32). Por tanto, la evidencia se concentra en los rectángulos y existe independencia fuerte.

<u>Corolario</u>

En las condiciones de la proposición anterior, si $h_1\geqslant 0$,

$h_2 \geq 0$ y existe independencia fuerte, entonces:

a) $E_{Pl}(h_1.h_2) = E_{Pl_1}(h_1).E_{Bel_2}(h_2)$

b) $E_{Bel}(h_1.h_2) = E_{Bel_1}(h_1).E_{Pl_2}(h_2)$

Demostración

a) $E_{Pl}(h_1.h_2) = -E_{Bel}(h_1(-h_2)) = -E_{Bel_1}(h_1).E_{Bel_2}(-h_2)$

$= E_{Bel_1}(h_1).E_{Pl_2}(h_2)$

b) Es análoga a la anterior.

2.2. LA ESPERANZA MONOTONA RESPECTO A LAS MEDIDAS DE POSIBILIDAD Y NECESIDAD

Es inmediato que todas las propiedades de la esperanza monótona con respecto a una evidencia, son directamente --- trasladables al caso de medidas de posibilidad y necesidad. En éste apartado, nos vamos a dedicar a obtener, una expresión de la esperanza monótona, para el caso particular de - éstas últimas medidas, que nos permitirá calcular las integrales de un modo sencillo.

Sea X un conjunto finito, h una aplicación de X en R^+, π una distribución de posibilidad en X y Π y N las medidas de posibilidad y necesidad asociadas a la misma. En éstas - condiciones, si notamos $X=\{x_1,\ldots,x_n\}$ con:

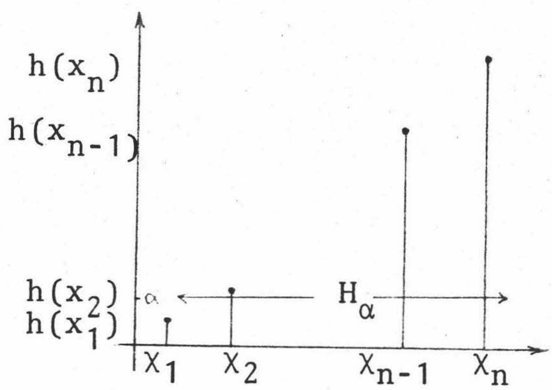
$$h_1(x_1) \leq \ldots \leq h_n(x_n),$$

se tiene:

$$I^*(h/m_\pi) = E_\Pi(h) = \int_0^\infty \Pi(H_\alpha)d\alpha$$

donde:

$$H_\alpha=\{x_i \in X/h(x_i) \geq \alpha\}$$



entonces:

$$E_\Pi(h) = \int_0^{h(x_1)} \Pi(\{x_1,\ldots,x_n\})d\alpha + \int_{h(x_1)}^{h(x_2)} \Pi(\{x_2,\ldots,x_n\})d\alpha + \ldots +$$

$$+ \int_{h(x_{n-1})}^{h(x_n)} \Pi(\{x_n\})d\alpha$$

$$= \Pi(\{x_1,..,x_n\})h(x_1) + \Pi(\{x_2,\ldots,x_n\})(h(x_2)-h(x_1)) +$$

$$+\ldots\ldots+ \Pi(\{x_{n-1},x_n\})(h(x_{n-1})-h(x_{n-2})) +$$

$$+ \Pi(\{x_n\})(h(x_n)-h(x_{n-1}))$$

$$= h(x_1)[\Pi(\{x_1,\ldots,x_n\}) - \Pi(\{x_2,\ldots,x_n\})] +$$

$$+ h(x_2)[\Pi(\{x_2,\ldots,x_n\}) - \Pi(\{x_3,\ldots,x_n\})] +$$

$$+ h(x_{n-1})[\Pi(\{x_{n-1},x_n\}) - \Pi(\{x_n\})] +$$

$$+ h(x_n)\ \Pi(\{x_n\})$$

$$= \sum_{i=1}^{n} h(x_i)[\sup_{i\leqslant j\leqslant n} \pi(x_j) - \sup_{i+1\leqslant j\leqslant n} \pi(x_j)]$$

De forma análoga, si N es la medida de necesidad dual de Π,

$$I_*(h/m_\pi) = E_N(h) = \int_0^\infty N(H_\alpha)d\alpha =$$

$$= h(x_1)\ N(\{x_1,\ldots,x_n\}) + [h(x_2)-h(x_1)]N(\{x_2,..,x_n\})$$

$$+\ldots+ [h(x_{n-1})-h(x_{n-2})]N(\{x_{n-1},x_n\}) +$$

$$+ [h(x_n)-h(x_{n-1})]N(\{x_n\}) =$$

$$\sum_{i=1}^{n} h(x_i)\ [\,(N(\{x_i,\dots,x_n\})-N(\{x_{i+1},\dots,x_n\})]\ =$$

$$\sum_{i=1}^{n} h(x_i)\ [\,1-\Pi(\{x_1,\dots,x_{i-1}\})-1+\Pi(\{x_1,\dots,x_i\})]\ =$$

$$\sum_{i=1}^{n} h(x_i)\ (\sup_{1\ j\ i} \pi(x_j)-\sup_{1\ j\ i-1} \pi(x_j))$$

en resumen obtenemos:

$$I^*(h/m_\pi)\ =\sum_{i=1}^{n}\ h(x_i)\ \ (\sup_{1\ j\ n} \pi(x_j)-\sup_{1\ j\ n} \pi(x_j))$$

$$I_*(h/m_\pi)\ =\sum_{i=1}^{n}\ h(x_i)\ \ (\sup_{1\ j\ i} \pi(x_j)-\sup_{1\ j\ i-1} \pi(x_j))$$

2.3. LA ESPERANZA MONOTONA CON RESPECTO A UNA λ-MEDIDA

El objetivo de éste apartado es identico al del anterior, significando que las propiedades generales de la esperanza monótona con respecto a una λ-medida se pueden obtener particularizando las de las integrales superior e inferior.

Sea g_λ una λ-medida y $h:X \to R_o^+$. Supongamos que se nota $X=\{x_1,\dots,x_n\}$ con $h(x_1) \leqslant h(x_2) \leqslant \dots \leqslant h(x_n)$.
En éstas condiciones:

$$E_{g_\lambda}(h)\ =\ \int_o^{\infty} g_\lambda(H_\alpha)\,d\alpha\ =$$

$$=\int_o^{h(x_1)} d\alpha+\int_{h(x_1)}^{h(x_2)} g_\lambda(x_2,\dots,x_n)\,d\alpha+\dots+\int_{h(x_{n-1})}^{h(x_n)} g_\lambda(x_n)\,d\alpha$$

$$=\sum_{i=1}^{n}[\,h(x_i)-h(x_{i-1})]\,g_\lambda(x_i,x_{i+1},\dots,x_n)$$

donde hemos supuesto $h(x_o)=0$.

Operando en la expresión anterior, obtenemos:

$$E_{g_\lambda}(h) = \sum_{i=1}^{n-1} h(x_i)[g_\lambda(x_i,..,x_n)-g_\lambda(x_{i+1},..,x_n)]+h(x_n)g_\lambda(x_n)$$

$$= \sum_{i=1}^{n-1} h(x_i)[g_\lambda(x_i)+g_\lambda(x_{i+1},..,x_n)+\lambda g_\lambda(x_i)g_\lambda(x_{i+1},..,x_n)$$

$$-g_\lambda(x_{i+1},..,x_n)]+h(x_n)g_\lambda(x_n)$$

$$= \sum_{i=1}^{n} h(x_i)[g_\lambda(x_i)(1+\lambda g_\lambda(x_{i+1},...,x_n))]$$

en resumen:

$$E_{g_\lambda}(h) = \sum_{i=1}^{n} h(x_i)\{g_\lambda(x_i)[1+\lambda g_\lambda(\{x_{i+1},...,x_n\})]\}$$

Como consecuencia inmediata de ésta expresión se obtiene que; cuando $\lambda=0$ (g_o es una medida de probabilidad)

$$E_{g_o}(h) = \sum_{i=1}^{n} h(x_i)g_o(x_i)$$

- Cuando $\lambda>0$, (g_λ es una medida de creencia)

$$E_{g_\lambda}(h) > \sum_{i=1}^{n} h(x_i)g_\lambda(x_i)$$

- Cuando $-1 < \lambda < 0$, (g_λ es una medida de plausibilidad)

$$E_{g_\lambda}(h) < \sum_{i=1}^{n} h(x_i)g_\lambda(x_i)$$

- CAPITULO IV -

Problemas de decisión con información general. Criterios de decisión.

0.-INTRODUCCION.

Todos, de un modo u otro, tenemos formada una idea intuitiva de lo que es un proceso de decisión, ya que, en la vida real, a menudo se presentan situaciones de éste tipo, que conllevan la elección de una alternativa en función de unos objetivos que ya debemos tener previstos de antemano.

La Teoria de la Decisión como rama de la Matemática - aplicada, pretende la construcción de Modelos Matemáticos que nos permitan, bajo unas hipótesis, poder adoptar nuestras decisiones de un modo racional, de forma que nos aseguren lo más posible, que la alternativa que elegimos o DECISION sea la que mejor nos conduzca a conseguir los objetivos prefijados.

Los elementos fundamentales de un proceso de decisión son:

1.-DECISOR. Individuo (fisico, jurídico,...) que debe tomar la decisión.

2.-AMBIENTE. Conjunto de circunstancias que influyen sobre el problema y que no pueden ser controladas - por el decisor. Habitualmente, estas variables no controladas tienen varias posibilidades que se denominan ESTADOS DE LA NATURALEZA. Una hipótesis básica en Teoria de la Decisión es que en un momento determinado la Naturaleza se encuentra en un estado bien definido que puede ser conocido o no con certeza por el decisor.

3.- ALTERNATIVAS. Son las diferentes posibilidades que se ofrecen al decisor para que éste haga su decisión. Se denominan tambien acciones.

4.- OBJETIVOS. Son aquellas metas que el decisor pretende conseguir mediante sus acciones. Estos objetivos han de estar especificados de antemano por el decisor.

5.- CONSECUENCIAS. La interacción de una alternativa d con un estado de la naturaleza w proporcionará al decisor una consecuencia o resultado c(d,w). Cuando c(d,w) es numérica se suele llamar RECOMPENSA.

6.- DURACION. A menudo, los procesos de decisión transcurren en el tiempo de un modo etápico,en el sentido que no solo se requiere una decisión en un instante sino que se necesitan una serie de ellas en momentos sucesivos.

Los procesos de decisión se clasifican de acuerdo con las variantes de sus distintos elementos (ver S.Rios 1977). Dado el tema de ésta memoria, nosotros solo nos ocuparemos de las ditintas modalidades en que el verdadero estado de laNaturaleza puede manifestarse al decisor representadas por el grado de conocimiento que éste posee sobre aquel. Clasicamente se considera que puede haber un AMBIENTE DE CERTIDUMBRE.- El estado actual de la Naturaleza es perfectamente conocido con lo que cada acción nos dará una consecuencia perfectamente conocida y previsible.

RIESGO.- Se conocen los estados, pero la información disponible sobre el verdadero se reduce a una ley de probabilidad conocida por el decisor. Asi pues, para cada acción,las consecuencias solo pueden ser previstas en términos de la mencionada ley de probabilidad.

INCERTIDUMBRE.- O bien no tenemos idea de la existencia de algun estado o bien no conocemos la forma de presentarse los mismos.

Normalmente, se supone éste segundo caso, ya que el primero no tiene mucho sentido, pues la modelización de cualquier situación solo puede emplear los elementos conocidos. Atendiendo al grado de información que se posea dentro del desconocimiento supuesto, se distingue entre:

Incertidumbre TOTAL.- No se conoce absolutamente nada de la forma o ley en que se presentan los estados.

Incertidumbre PARCIAL.- Se sabe que los estados de la naturalezase presentan segun una entre varias leyes de probabilidad conocidas, pero no se sabe concretamente cual de ella es.

En general el ambiente se caracteriza mediante la información que se supone que el decisor posee sobre los estados. Clasicamente se admite que, o existe una falta de información total o el conocimiento disponible sobre el estado de la naturaleza puede materializarse mediante distribuciones de probabilidad (considerando la certidumbre total como un caso particular de distribuciones degeneradas).

Es obvio que existen muchas situaciones en las que la información disponible es tal que resulta practicamente imposible aceptar la existencia de una distribución de probabilidad que la represente. Supongamos por ejemplo la información que, sobre el tiempo, suelen proporcionar unas personas a otras: " hace frio ", " la temperatura es agradable " o " no llueve demasiado ",...

Como ya hemos dicho con anterioridad, la Teoria de la Evidencia de Shafer en combinación con la Teoria de Subconjuntos Difusos, proporciona herramientas para modelizar situacionescomo la descrita en párrafo precedente.

El objetivo básico de éste capitulo es analizar problemas de decisión en ambiente general, es decir suponiendo que la información sobre los estados no es necesariamente, o la ignorancia o una distribución de probabilidad.

Comenzaremos planteando rigurosamente el problema a considerar y los distintos elementos que forman parte del mismo.

Al igual que en capítulos anteriores supondremos que tanto el conjunto de estados como el de acciones son finitos. Con ello tratamos de evitar problemas de medibilidad y convergencia completamente ajenos al tema de nuestro estudio.

Vamos a admitir que la información disponible sobre el verdadero estado de la naturaleza es una evidencia representada por una asignaciónbásica de probabilidad M, con lo que se engloban los casos clásicos de riesgo e incertidumbre.

En estas condiciones se establece una aplicación g
$g : R^n \longrightarrow R^2$ tal que para cada $\tilde{r} \varepsilon R^n$ $g(r)=\{I_*(r/M), I^*(r/M)\}$
que permite asociar a cada decisión d la imagen de su vector de recompensas. La segunda parte del apartado II está dedicada al estudio de las propiedades de g y la justificaciónde su empleo para la construcción de reglas de decisión razonables.

Existen dos formas de abordar éste último problema. En primer lugar está la obtención de criterios a partir de relaciones de orden (preferencias) construidas directamente sobre R^2. En segundo lugar se tiene la actuación mediante funciones valor $v : R^2 \longrightarrow R$. Como es bien conocido estos dos caminos no son independientes, si bien se consideran separados por resaltar la filosofia subyacente en cada uno de ellos.

El apartado III está dedicado al estudio de los criterios que se obtienen al trabajar directamente en R^2. Basicamente, se analizan la dominancia segun el cuadrante positivo y las reglas de tipo lexicográfico.

El apartado IV recoge el empleo de las funciones valor contemplando tres casos particulares de gran interés: $v(x_1,x_2)= \max(x_1,x_2)$, $v(x_1,x_2)= \min(x_1,x_2)$, $v(x_1,x_2)= c_1x_1+ c_2x_2$.

El modelo propuesto generaliza los ambientes de riesgo e incertidumbre, y algunos de los criterios de decisión más usuales en Teoria Clásica (el de la esperanza matemática, el de Wald o el de Hurwicz), pueden obtenerse como casos particulares de los aquí obtenidos. A ésto se dedica el apartado V.

Para terminar indicaremos que son dos las ideas que han conformado la organización y contenido de éste capítulo: en primer lugar, la de presentar un modelo de decisión más general que los considerados por la teoria clásica y en segundo lugar la de indicar que herramientas pueden emplearse para conseguir criterios de decisión razonables y que forma adoptan éstos bajo determinadas hipótesis de comportamiento del decisor. Es obvio que existen puntos de vista alternativos sobre como abordar el problema y aspectos del mismo que no han sido considerados aquí. El apartado VI recoge sugerencias en relación con éstos temas y futuras vias de investigación.

<u>NOTA</u>. A lo largo de todo éste capítulo empleamos repetidas veces los términos " riesgo ", " actitud del decisor frente al riesgo " , "optimismo " y " pesimismo ". Queremos

destacar que, salvo en los modelos particulares de la Teoria de la Decisión Clásica, éstos términos se emplean con un significado intuitivo y no formal. Así, por ejemplo, para nosotros un decisor pesimista será aquel que pretenda siempre asegurar una ganancia, aunque ésta sea baja, mientras que el calificativo de optimista se aplicará a aquel que admite la posibilidad de ganancias bajas siempre que, simultaneamente, exista la posibilidad de ganancias altas.

1.- PLANTEAMIENTO DEL PROBLEMA. RESULTADOS BASICOS

A lo largo de todo el capítulo consideraremos un problema de decisión con los siguientes elementos:

1.- Un conjunto de estados de la Naturaleza que supondremos finito

$$\Omega = \{w_1, w_2, \dots, w_n\}$$

2.- Un conjunto de acciones que tambien será finito

$$\mathcal{D} = \{d_1, d_2, \dots, d_m\}$$

3.- Una función $r : \mathcal{D} \times \Omega \longrightarrow R$ tal que a cada decisión d_i y a cada estado w_j hace corresponder un número real

$$(d_i, w_j) \longrightarrow r(d_i, w_j) \equiv r_{ij}$$

siendo r_{ij} la recompensa que percibe el decisor por haber elegido la decisión d_i cuando se ha presentado el estado w_j. r se puede representar mediante una matriz que llamaremos MATRIZ DE RECOMPENSAS, con la que trabajaremos a partir de ahora.

$\mathcal{D} \backslash \Omega$	w_1	w_2 ...	w_j ...	w_n
d_1	r_{11}	r_{12}...	r_{1j}...	r_{1n}
d_2	r_{21}	r_{22}...	r_{2j}...	r_{2n}
			⋮	
d_i	r_{i1}	r_{i2}...	r_{ij}...	r_{in}
			⋮	
d_m	r_{m1}	r_{m2}...	r_{mj}...	r_{mn}

Sobre la forma de construir recompensas que valoren - las consecuencias de las acciones puede consultarse, por - ejemplo, Fishburn (1970, 1973,...) y Keeney y Raiffa (1976)

En estas condiciones, la recompensa que podrá obtener el decisor al efectuar una elección será uno cualquiera de los valores del vector n-dimensional asociado a la misma,

$$d_i \longrightarrow [r_{i1}, r_{i2}, \ldots, r_{ij}, \ldots, r_{in}] \equiv \tilde{r}_i \varepsilon R^n$$

y el valor concreto de la recompensa dependerá del es tado de la naturaleza que se presente.

4.- Una cierta información I sobre el verdadero estado de la Naturaleza. Como hemos dicho, vamos a suponer que ésta información no es, necesariamente, de tipo probabilístico sino que nuestro conocimiento viene expresado mediante una evidencia, materializada en una asignación básica de proba bilidad M sobre .

5.- Una relación de preferencias del decisor. Puesto que - hemos admitido que las consecuencias están medidas en términos de recompensas, no supone falta alguna de generalidad considerar que entre las consecuencias de las acciones ésta relación de preferencia se representa por el orden natural de R ($r_{ij} \succsim r_{kp} \Longleftrightarrow r_{ij} \geqq r_{kp}$).

¿Cual será en un momento determinado la mejor decisión . En lógica nuestra contestación seria; aquella cuyo vector de recompensas asociado fuese mayor. Esta que,en teoria es una contestación sencilla, en la práctica no lo es tanto ya que pasa por definir relaciones de orden en R^n. De hecho todo criterio de decisión supone una actuación de éste tipo.

La teoria de la decisión clásica ofrece numerosos ejemplos dē criterios que suponen ordenar directamente R^n (por ejemplo el lexicográfico o los que usan la dominancia). Una forma indirecta de abordar el problema es aplicar cada vector n-dimensional r_i en un valor $r_i \in R$ y luego tomar la decisión de acuerdo con el orden natural de R. De hecho, la mayor parte de los criterios clásicos de decisión consisten en construir una aplicación $f:R^n \longrightarrow R$ (denominada función valor por Keeney y Raiffa (1976)) con

$$f(\tilde{r}_i) = f\left[r_{i1},\ldots,r_{in}\right] = r_i \in R$$

y considerar que la acción d_i es más preferida que la d_k si y solo si r_i es mayor o igual que r_k

$$d_i \succsim d_k \quad \text{ó} \quad \left[r_{i1},\ldots,r_{in}\right] \succsim \left[r_{k1},\ldots,r_{kn}\right]$$

si y solo si

$$r_i \geqq r_k$$

Separando la preferencia estricta de la indiferencia, se dice que d_i es más preferida que d_k si y solo si $r_i > r_k$

y que d_i es indiferente a d_k si y solo si $r_i=r_k$.

De acuerdo con ésto, la decisión optima d^* será aquella tal que:

$$f(d^*)= \max_i f(d_i)$$

La aplicación f se construye teniendo en cuenta los valores de las recompensas, la información disponible sobre el estado de la Naturaleza y la actitud del decisor frente al riesgo.

El criterio de Wald ofrece un ejemplo típico de ésta forma de actuar. En un ambiente de incertidumbre total y supuesto que el decisor tiene aversión al riesgo, tiene sentido construir f tal que

$$f\left[r_{i1},\ldots,r_{in}\right] = \min\left[r_{i1},\ldots,r_{in}\right]$$

con lo que

$$d_i \gtrsim d_k \Longleftrightarrow \min\left[r_{i1},\ldots,r_{in}\right] \geq \min\left[r_{k1},\ldots,r_{kn}\right]$$

y por tanto, la decisión optima d_e será aquella tal que

$$\min_j \left[r_{ej}\right] = \max_i \min_j \left[r_{ij}\right]$$

Nosotros tambien vamos a emplear la misma idea. Consideraremos que una solución para el problema de decisión será encontrar un orden total en el conjunto $\mathfrak{D}$ construido teniendo en cuenta la función r, la información I y la actitud del decisor frente al riesgo.

De acuerdo con nuestras hipótesis, la información I es una evidencia materializada en una Asignación Básica de Probabilidad M sobre Ω. De acuerdo con los resultados del capítulo anterior podemos establecer una aplicación $g:R^n \longrightarrow R^2$ tal que:

$$g(\tilde{r}) = \left[I_*(\tilde{r}/M) \ , \ I^*(\tilde{r}/M) \right] \qquad (33)$$

siendo $I_*(r/M)$ e $I^*(r/M)$ la integral inferior y superior, respectivamente, de r con respecto a M. Observese que g se construye a partir de r e I pero sin tener en cuenta de una forma específica la actitud del decisor ante el riesgo.

Mediante ésta g, el problema de ordenar las filas de la matriz de recompensas se transforma en el de ordenar elementos de R^2. Si bien en éste espacio no existe tampoco un orden natural, es obvio, que su reducida dimensión lo hace más cómodo para trabajar.

En éstas condiciones y supuesto que se acepta el uso de g tenemos dos formas de obtener criterios de decisión:

1.- Establecer una función valor $v:R^2 \longrightarrow R$ y considerar:

$$\{a,b\} \succsim \{a',b'\} \iff v(a,b) \geq v(a',b')$$

2.- Trabajar directamente en R^2.

Es obvio que éstos dos caminos no son independientes, ya que cualquier aplicación de R^2 en R establece un orden total no estricto en R^2 y, bajo determinadas condiciones, un orden total no estricto de R^2 se puede representar mediante una aplicación real valuada.

Es inmediato que cualquier orden establecido en R^2, en particular mediante h, deberá reflejar la actitud del decisor frente al riesgo.

Observese que la introducción de g permite realizar en dos pasos la construcción de la aplicación f. Antes de pasar al análisis de los criterios de decisión que se obtienen por medio de los caminos anteriormente incluidos, y al obje to de justificar éstos, vamos a estudiar algunas propiedades interesantes de g.

Estudio de la aplicación g.

De acuerdo con (33)

$$g(\tilde{r}_i) = \{I_*(\tilde{r}_i/M) \ , \ I^*(\tilde{r}_i/M)\}$$

para cada decisión d_i.

Si tenemos en cuenta las definiciones de integral superior e inferior (ver Cap-3º) el par $g(\tilde{r}_i)$ tiene las siguientes caracteristicas:

1).- $I_*(\tilde{r}_i/M) \leqq I^*(\tilde{r}_i/M) \qquad \forall\, i$

2).- Su primera componente, $I_*(\tilde{r}_i/M)$, le asigna al vector $\tilde{r}_i$ y por tanto a la decisión correspondiente d_i, un valor igual al de la esperanza matemática de ese mismo vector con respecto a la medida de probabilidad que obtendriamos cuando toda la masa de evidencia m(A), relativa al conjunto focal A de Ω, se la adjudicaramos a aquel estado de la Naturaleza, perteneciente a A, $w_k \varepsilon A c \Omega$,en el cual se verifique que el vector de recompensas restringido a A, alcance su mínimo.

El resto de los elementos de A quedarian con probabilidad nula, y, en el caso de que un estado de la Naturaleza - fuese a la vez de mínima utilidad para más de un conjunto - focal, la probabilidad que le asignariamos sería la suma de las evidencias correspondientes a dichos conjuntos focales.

Así pues, $I_*(\tilde{r}_i/M)$ puede interpretarse como una evaluaciónpesimista de la recompensa esperada segun la información disponible.

3).-El valor $I^*(\tilde{r}_i/M)$ es la esperanza matemática de $\tilde{r}_i$ con respecto a la medida de probabilidad, que obtendriamos al a signar la evidencia relativa a A, al estado de la Naturaleza $w_k \varepsilon A$, cuya recompensa fuese mayor. Si hubiese un estado que al igual que en el caso de la integral inferior perteneciese a dos o más conjuntos focales, y tal que en él se alcanzase el máximo de las recompensas, su probabilidad asociada sería la suma de las evidencias relativas a los conjuntos supuestos.

$I^*(\tilde{r}_i/M)$ puede entonces considerarse como una evaluación optimista de la recompensa esperada según la información disponible.

Tendremos que hacer notar que las distribuciones de probabilidad empleadas en la interpretación de éstas dos integrales no son fijas para Ω, sino que dependen de i, es decir, de la decisión d_i, a cuyo vector se le vaya a aplicar g. Obtenemos asi dos distribuciones de probabilidad diferentes asociadas a una misma acción d_i, y que cambian al variar ésta.

Los vectores $\tilde{r}_i$, juegan un doble papel a la hora de construir g: Por una parte,son usados para encontrar las disribuciones de proabbilidad superior e inferior asociadas a Ω y por otra parte, serán la función a partir de la cual y teniendo en cuenta las dos distribuciones de probabilidad anteriores, nos serviran para calcular las esperanzas.

De acuerdo con éstas consideraciones, es posible considerar $g(\tilde{r}_i)$ como un par de elementos o como un intervalo. No hay diferencia sustancial entre ambas interpretaciones, ya que la recompensa esperada por el decisor al tomar la decisión d_i, tiene como cotas inferior y superior los valores $I_*(\tilde{r}_i/M)$ e $I^*(\tilde{r}_i/M)$ respectivamente. Asi pues

$$g(\tilde{r}_i) = \left[I_*(\tilde{r}_i/M) \ , \ I^*(\tilde{r}_i/M) \right]$$

puede considerarse como un " intervalo de riesgo ".

Como posteriormente veremos, si nuestra elección se hace en base a utilizar la primera componente de g, estaremos utilizando un criterio pesimista, si por el contrario nos - fijamos unicamente en el extremo superior, nuestro critério será optimista.

Para ilustrar éstas caracteristicas de g proponemos el siguiente ejemplo.

Ejemplo .4.1

Supongamos que se tiene:

- $\mathcal{D} = \{d_1, d_2\}$
- $\Omega = \{w_1, w_2, w_3\}$
- Matriz de recompensas

	w_1	w_2	w_3
d_1	40	25	15
d_2	25	40	30

- Asignación básica de probabilidad

$$M(A) = \begin{cases} 0.3 & A = \{w_1\} \\ 0.4 & A = \{w_2, w_3\} \\ 0.3 & A = \{w_1, w_3\} \\ 0 & \text{en otro caso} \end{cases}$$

Es inmediato que:

$$d_1 \left(\begin{array}{l} I_*(\tilde{r}_i/M) = \sum_{A \subset U} M(A) \inf_{a \in A} r_{1j} = 22.5 \\ \\ I^*(\tilde{r}_i/M) = \sum_{A \subset U} M(A) \sup_{a \in A} r_{1j} = 34 \end{array} \right.$$

$$d_2 \left\{ \begin{array}{l} I_*(\tilde{r}_i/M) = \sum\limits_{A \subset U} M(A) \inf\limits_{a \in A} r_{2j} = 27 \\ \\ I^*(\tilde{r}_i/M) = \sum\limits_{A \subset U} M(A) \sup\limits_{a \in A} r_{2j} = 32.5 \end{array} \right.$$

Vamos a ver que éstos valores han de coincidir con los calculados como esperanzas con respecto a las distribucio - nes de probabilidad mínima y máxima relativas a d_1 y a d_2 - respectivamente.

d_1

	w_2	w_3	w_1
r	25	15	40
			0.3
M	0.4		
		0.3	

	w_2	w_3	w_1
r	25	15	40
P_*	o	0.7	0.3
P^*	0.4	0	0.6

$$E(\tilde{r}_i/P_*) = 22.5$$

$$E(\tilde{r}_i/P^*) = 34$$

d_2

	w_2	w_3	w_1
r	40	30	25
			0.3
M	0.4		
	0.3		

	w_2	w_3	w_1
r	40	30	25
P_*	0	0.4	0.6
P^*	0.4	0.3	0.3

$$E(\tilde{r}_2/P_*)= 27$$

$$E(\tilde{r}_2/P^*)= 32.5$$

De acuerdo con las propiedades de I_* e I^* que hemos -- probado en el capítulo anterior, dependiendo del tipo de -- evidencia que se considere, tendremos, algunos casos particulares que daran lugar a situaciones interesantes.

1).- Si la evidencia es la ignorancia, representada -- por

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

las integrales tendrian como valor.

$$I_*(\tilde{r}_i/m_o)= \min\ [r_{i1},r_{i2},\ldots,r_{in}]$$

$$I^*(\tilde{r}_i/m_o)= \text{máx}\ [r_{i1},r_{i2},\ldots,r_{in}]$$

con lo que

$$g(\tilde{r}_i)= [\min_j r_{ij}\ ,\ \text{máx}_j\ r_{ij}]$$

Desde un punto de vista formal, éste resultado se justifica por el hecho de que en ausencia de información la -- probabilidad inferior le asigna toda la masa de evidencia a $\min_j \tilde{r}_{ij}$ mientras que la probabilidad superior se la asigna a $\text{máx}_j\ \tilde{r}_{ij}$.

Desde un punto de vista intuitivo resulta obvio que, - en ausencia de información $\min_j r_{ij}$ y $\text{máx}_j\ r_{ij}$ son, respectivamente , la mínima y la máxima recompensa esperada por el decisor cuando toma la decisión d_i.

2).-Si la evidencia es de tipo probabilistico, repre - sentada por

$$m_p(A)=\begin{cases} p(w_j) & \text{si } A=\{w_j\} \quad j=1,\dots,n \\ 0 & \text{en otro caso} \end{cases}$$

es obvio que nos encontramos en un ambiente de riesgo, con- distribución $\{p(w_j) \quad j=1,\dots,n\}$ sobre Ω.

En éstas condiciones

$$I_*(\tilde{r}_i/m_p) = I^*(\tilde{r}_i/m_p) = \sum_{j=1}^{n} r_{ij}p(w_j)$$

de modo que, como en la Teoria Bayesiana clásica, hemos de trabajar con la esperanza matemática,

Es inmediato que la certidumbre puede considerarse - como un caso particular de evidencia probabilistica con la distribución P concentrada en un solo punto.

La ignorancia y la evidencia probabilística represen - tan casos extremos en el sentido siguiente: la ignorancia - produce para cada decisión, el mayor intervalo de riesgo posible, mientras que una evidencia probabilística reduce éste a un punto. Para cualquier evidencia comprendida entre - las dos (según la relación de inclusión definida en Cap I) tendremos una situación intermedia. Asi, dos evidencias con la propiedad de ser una más fina que la otra, nos proporcionarán, para cada decisión, intervalos que tambien estarán - ligados por la relación de inclusión. De acuerdo con la proposición 3.15 del capítulo III, se tiene:

3).- Si m_1 y m_2 son asignaciones básicas de probabilidad sobre Ω (representando sendas informaciones) tales que $m_1 \subset m_2$, entonces para cualquier decisión $d_i \varepsilon \mathcal{D}$, se verifica

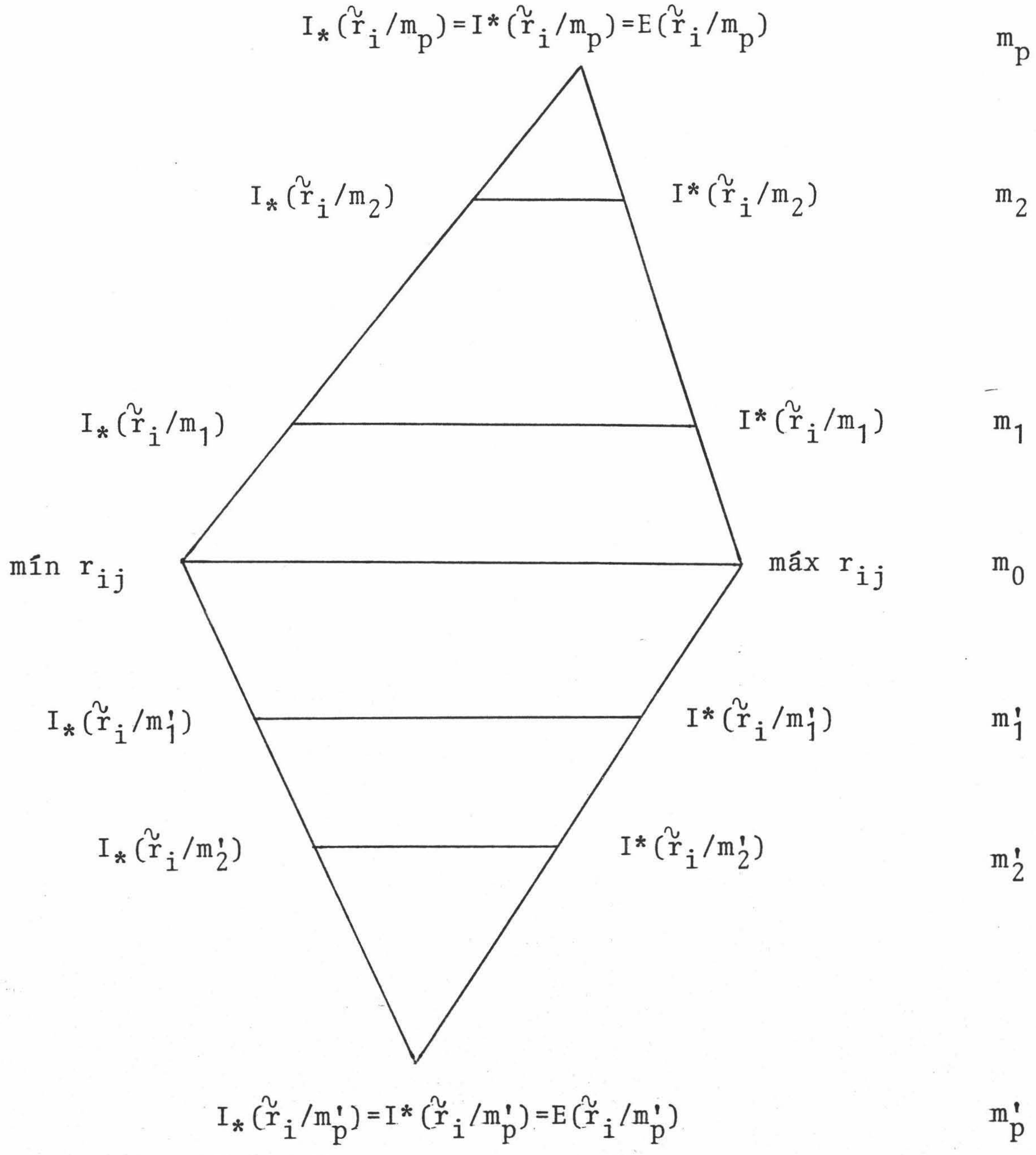
$$[I_*(\tilde{r}_i/m_2)\ ,\ I^*(\tilde{r}_i/m_2)]\ c\ [I_*(\tilde{r}_i/m_1)\ ,\ I^*(\tilde{r}_i/m_1)]$$

Si m_1 está contenida en m_2, entonces los elementos focales de m_2 están más atomizados que los de m_1 y, en consecuencia, al asignar la masa de evidencia a los elementos -- $w_j \varepsilon \Omega$, habrá más estados de la Naturaleza con probabilidad no nula en el caso de m_2 que en el caso de m_1. De ésta forma, cuanto mayor sea la información de que dispongamos, -- más próximas estarán las distribuciones de probabilidad superior e inferior y por tanto tambien las esperanzas de --- cualquier $\tilde{r} \varepsilon R^n$, que, como sabemos, coinciden con la inte -- gralsuperior e inferior respecto de la evidencia.

Desde un punto de vista intuitivo, éste resultado se - justifica por el hecho de que, cuanto mejor sea la información disponible sobre el verdadero estado de la Naturaleza, menor debe ser el intervalo de riesgo para el decisor.

Hay que hacer notar que cualqüier información debe contener a la ignorancia (proporcionando un intervalo de riesgo menor), mientras que dos evidencias probabilísticas son siempre no comparables (y, en general, proporcionarán dos esperanzas distintas) si bien ambas son de la misma "calidad" en términos de la evidencia proporcionada.

El gráfico siguiente resume estos resultados:



$I_*(\tilde{r}_i/m_p) = I^*(\tilde{r}_i/m_p) = E(\tilde{r}_i/m_p)$
m_p
$I_*(\tilde{r}_i/m_2)$
$I^*(\tilde{r}_i/m_2)$
m_2
$I_*(\tilde{r}_i/m_1)$
$I^*(\tilde{r}_i/m_1)$
m_1
mín r_{ij}
máx r_{ij}
m_0
$I_*(\tilde{r}_i/m_1')$
$I^*(\tilde{r}_i/m_1')$
m_1'
$I_*(\tilde{r}_i/m_2')$
$I^*(\tilde{r}_i/m_2')$
m_2'
$I_*(\tilde{r}_i/m_p') = I^*(\tilde{r}_i/m_p') = E(\tilde{r}_i/m_p')$
m_p'

Ejemplo 4.2

Sea $\mathcal{D} = \{d_1, d_2, d_3\}$

$\Omega = \{w_1, w_2, w_3\}$

con la siguiente matriz de recompensas.

	w_1	w_2	w_3
d_1	500	2500	5000
d_2	4000	2500	1500
d_3	2500	1500	3000

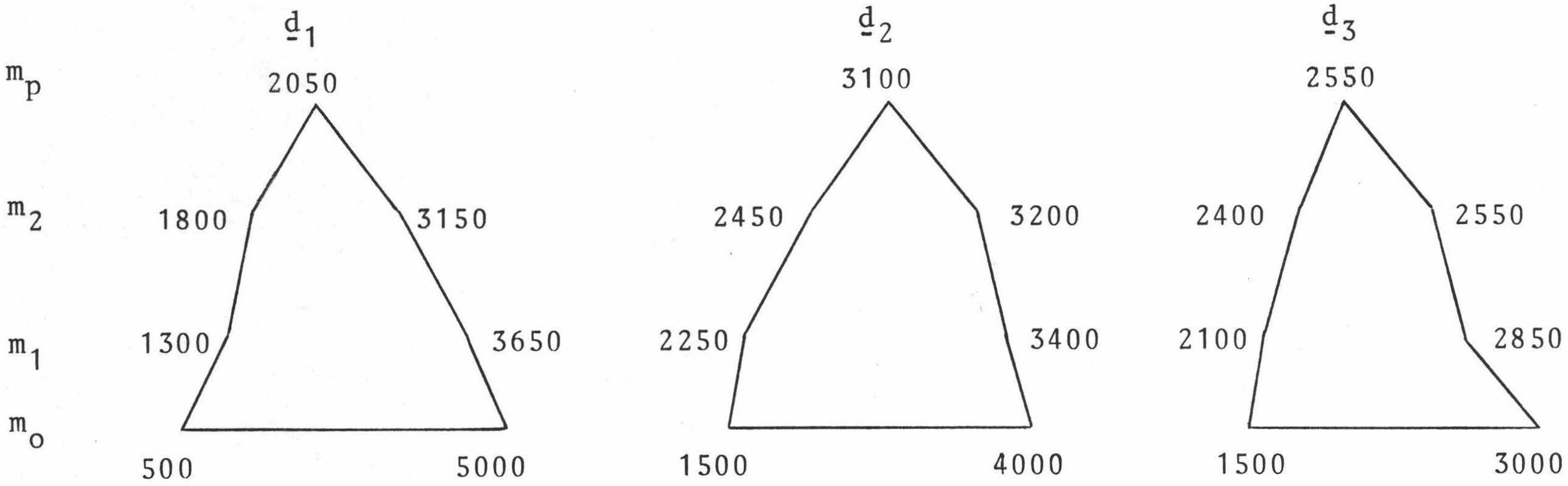
supongamos que se tienen las siguientes evidencias:

$$m_o(A) = \begin{cases} 1 & \text{si } A = \Omega \\ 0 & \text{'' } A \neq \Omega \end{cases} \qquad m_1(A) = \begin{cases} 0.3 & \text{si } A = \{w_1\} \\ 0.4 & \text{'' } A = \{w_2, w_3\} \\ 0.3 & \text{'' } A = \{w_1, w_3\} \end{cases}$$

$$m_2(A) = \begin{cases} 0.3 & \text{si } A = \{w_1\} \\ 0.2 & \text{'' } A = \{w_2\} \\ 0.2 & \text{'' } A = \{w_3\} \\ 0.3 & \text{'' } A = \{w_1, w_3\} \end{cases} \qquad m_p(A) = \begin{cases} 0.6 & \text{si } A = \{w_1\} \\ 0.1 & \text{'' } A = \{w_2\} \\ 0.3 & \text{'' } A = \{w_3\} \end{cases}$$

Es fácil comprobar que $m_p \subset m_1 \subset m_2 \subset m_o$.

Calculando losintervalos de riesgo para cada decisión obtenemos:



$\underline{d}_1$
$\underline{d}_2$
$\underline{d}_3$
m_p
2050
3100
2550
m_2
1800
3150
2450
3200
2400
2550
m_1
1300
3650
2250
3400
2100
2850
m_o
500
5000
1500
4000
1500
3000

Consideremos ahora la evidencia probabilística dada - por:

$$m_{p'}(A) = \begin{cases} 0.1 & A = \{w_1\} \\ 0.2 & A = \{w_2\} \\ 0.7 & A = \{w_3\} \end{cases}$$

para la cual

$$E(\tilde{r}_1/p') = 4050 \qquad E(\tilde{r}_2/p') = 1950 \qquad E(\tilde{r}_3/p') = 2650$$

que no coinciden con las correspondientes a m_p.

4).- Sean $\tilde{r}$, $\tilde{r}' \in R^n$. Si $\tilde{r} \geq \tilde{r}'$ $(r_i \geq r_i' \quad \forall i)$ se cumple

$$\begin{aligned} I_*(r/m) &\geqq I^*(r/m) \\ I_*(r/m) &\geqq I^*(r/m) \end{aligned} \qquad (34)$$

cualquiera que sea la evidencia m y reciprocamente.

En términos de la teoría clásica de la Decisión ésta - propiedad se enuncia diciendo, que si $\tilde{r}_i$ domina a $\tilde{r}_j$, entonces los extremos superior e inferior del intervalo de riesgo de d_i son mayores que los correspondientes de d_j cualquiera que sea la información disponible sobre el verdadero estado de la Naturaleza.

Esta propiedad se justifica e interpreta de modo simple tanto desde un punto de vista formal como intuitivo. -

<u>Ejemplo</u> 4.3

Supongamos $\Omega = \{w_1, w_2, w_3\}$

con la asignación básica de probabilidad:

$$m(A)=\begin{cases} 0.3 & A=\{w_1\} \\ 0.4 & A=\{w_2,w_3\} \\ 0.3 & A=\{w_1,w_3\} \end{cases}$$

consideremos ahora los vectores:

$$r_1=(4o,25,15) \geqq (35,1o,15)=r_2$$

Es inmediato que:

$$I_*(r_1/m)= 22.5 \qquad 19=I_*(r_2/m)$$

$$I^*(r_1/m)= 34 \qquad 26.5=I^*(r_2/m)$$

Hay que destacar que puede encontrarse evidencias m y vectores $\tilde{r}$ y $\tilde{r}'$ para los que se cumpla (34) y sin embargo $\tilde{r} \not\geqq \tilde{r}'$ tal como presenta el siguiente ejemplo.

Ejemplo 4.4

Sea $\Omega=\{w_1,w_2,w_3,w_4,w_5\}$

la asignación básica de probabilidad dada por:

$$m(A)=\begin{cases} 0.2 & A=\{w_1\} \\ 0.2 & A=\{w_2,w_3\} \\ 0.5 & A=\{w_1,w_3\} \\ 0.1 & A=\{w_4,w_5\} \end{cases}$$

y los vectores

$$r_1 = (1,2,3,4,5)$$

$$r_2 = (5,4,3,2,1)$$

Es inmediato comprobar que:

$$I_*(r_1/m) = 1.5 \qquad 3.2 = I_*(r_2/m)$$

$$I^*(r_1/m) = 2.8 \qquad 4.5 = I^*(r_2/m)$$

Observese que la mayor parte de la evidencia (un total de 0.9) se concentra en $\{w_1, w_2, w_3\}$, conjunto en el que r_2 - tiene componentes mayores que r_1. El hecho de que la amplitud de los intervalos de riesgo de r_1 y r_2 sea igual es consecuencia de que los valores en r_1 y en r_2 son los mismos.

De acuerdo con éste ejemplo podemos dar una versión -- más débil de la propiedad 4, que solo puede expresarse en - términos intuitivos.

Si r y r' son tales que la masa de evidencia proporcionada por una cierta m se concentra sobre conjuntos en los - que r tiene mayores valores que r', entonces:

$$I_*(r/m) \geqq I_*(r'/m)$$

e

$$I^*(r/m) \geqq I^*(r'/m)$$

2.- CRITERIOS DE DECISION BASADOS EN EL INTERVALO DE RIESGO.

Es obvio que cualquier relación de orden establecida - en R^2 de acuerdo con las preferencias del decisor debe permitir la comparación de elementos del conjunto

$$\{g(\tilde{r}_i) \quad i=1,\ldots,m\}$$

y por lo tanto de $\mathfrak{D}$, lo cual, nos puede conducir a la construcción de reglas de decisión. Nosotros nos vamos a limitar al análisis del orden "natural" (dominancia) y del orden lexicográfico en R^2, los cuales, a nuestro juicio generan formas de actuar bastante razonables dadas las caracteristicas de el modelo.

Criterio de dominancia.

Definición 4.1.- Sean $d_i, d_k \varepsilon \mathfrak{D}$. Diremos que d_i domina (debilmente a d_k según g y notaremos $d_i \succsim_g d_k$ si

$$\begin{aligned} I_*(r_i/M) &\geqq I_*(r_k/M) \\ I^*(r_i/M) &\geqq I^*(r_k/M) \end{aligned} \qquad (35)$$

Cuando al menos alguna de las desigualdades anteriores es estricta, diremos que d_i domina estrictamente según g a d_k.

De acuerdo con las propiedades de g es razonable aceptar que una decisión que domine a todas es una decisión optima. En particular, si las domina estrictamente es la única decisión optima.

En las figuras 1 y 2 representamos las dos posibilidades de dominancia estricta de d_i sobre d_k:

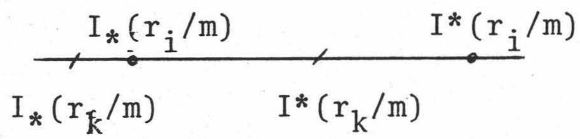


fig 1

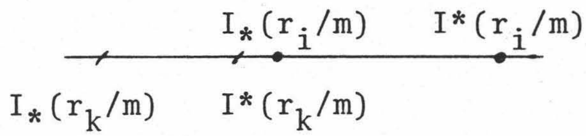


fig 2

Vemos que en ambos casos, no hay condición alguna sobre la amplitud del intervalo de riesgo.

Según hemos probado anteriormente se tiene:

$$\tilde{r}_i \geq \tilde{r}_j \Longrightarrow d_i \succsim_g d_k \qquad (36)$$

lo que nos indica que $\succsim_g$ reproduce sobre R^2 la dominancia de R^n, la cual, a su vez, puede interpretarse como la relación de orden básica del problema.

Ahora bien, $\succsim_g$ no es una relación de orden total, y por tanto, no tiene porqué existir un elemento maximal en $\mathfrak{D}$. Puesto que tambien parece razonable aceptar que una decisión dominada según g, no debe ser escogida por el decisor, tiene sentido, introducir el concepto de eficiencia u optimalidad Pareto.

Definición.4.2

Diremos que $d_i \in \mathfrak{D}$ es eficiente según g, si no existe $d_k \in \mathfrak{D}$ tal que $d_k \succsim_g d_i$.

Notamos $E_g(\subset \mathfrak{D})$ el conjunto de acciones eficientes. Si $E_g=\{d^*\}$ netonces d^* es decisión óptima única.

Es obvio que el decisor debe centrar su atención en E_g eliminando a priori todas las acciones de $\mathfrak{D}-E_g$. Ahora bien, por construcción, dos acciones de E_g, son no comparables, - salvo que tengan el mismo intervalo de riesgo. De éste modo, para un decisor coherente con la dominancia como critério, - son igualmente aceptables, debiendo recurrir a formas de selección adicionales, si se quiere individualizar una decisión. Cualquiera de los critérios que veremos posteriormente puede servir para éste fin.

Recordando que cuanto "mejor" es la información disponible, menor es el intervalo de riesgo, y que, para una decisión dada, las probabilidades superior e inferior dependen -

del vector que se esté integrando, resulta intuitivamente aceptable que, en cualquier caso, una decisión con intervalo de riesgo pequeño es siempre deseable para un decisor que pretenda asegurar su ganancia (optimista), ya que en éste caso la información disponible permite predecir bastante precisamente el valor de la recompensa que puede esperarse al tomar tal decisión. Por el contrario,una decisión con intervalo de riesgo grande será deseable para un decisor capaz de exponerse a algunas pérdidas con tal de tener la oportunidad de obtener mayores beneficios (Optimista).

De acuerdo con ésto, la amplitud de $[I_*(./M) , I^*(./M)]$ puede emplearse para decidir dentro de E_g, dependiendo de la actitud del decisor frente al riesgo.

Proposición .4.1

Si $\underline{d}, \overline{d} \in E_g$ son tales que:

$$I^*(\underline{d}/M) - I_*(\underline{d}/M) = \min_{d \in E_g} \{ I^*(d/M) - I_*(d/M) \} \tag{37}$$

$$I^*(\overline{d}/M) - I_*(\overline{d}/M) = \max_{d \in E_g} \{ I^*(d/M) - I_*(d/M) \} \tag{38}$$

entonces:

$$I^*(\overline{d}/M) = \max_{d \in \mathfrak{D}} I^*(d/M) \tag{39}$$

$$I_*(\underline{d}/M) = \max_{d \in \mathfrak{D}} I_*(d/M) \tag{40}$$

Demostración.

Sea d un elemento arbitrario de E_g. Por la definición de éste conjunto tiene que ser:

$$I^*(\underline{d}/M) \geqq I^*(d/M) \quad e \quad I_*(\underline{d}/M) \leqq I_*(d/M) \qquad (41)$$

ó

$$I^*(\underline{d}/M) \leqq I^*(d/M) \quad e \quad I_*(\underline{d}/M) \geqq I_*(d/M) \qquad (42)$$

$$I^*(\overline{d}/M) \geqq I^*(d/M) \quad e \quad I_*(\overline{d}/M) \leqq I_*(d/M) \qquad (43)$$

ó

$$I^*(\overline{d}/M) \leqq I^*(d/M) \quad e \quad I_*(\overline{d}/M) \geqq I_*(d/M) \qquad (44)$$

No pueden darse (41) y (44) ya que entonces no se cumplirian (37) y (38) Así pues, necesariamente tienen que verificarse (42) y (43) o lo que es lo mismo:

$$I^*(\overline{d}/M) = \max_{d \varepsilon E_g} I^*(d/M)$$

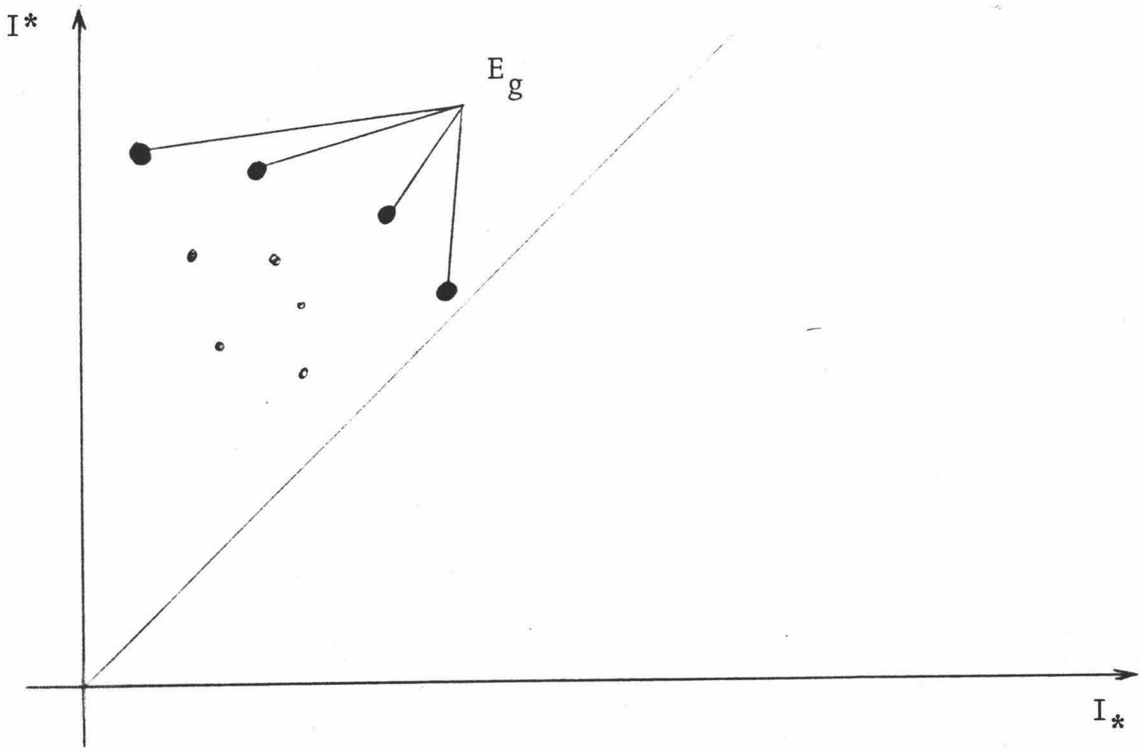
$$I_*(\underline{d}/M) = \max_{d \varepsilon E_g} I_*(d/M)$$

De acuerdo con las propiedades de E_g (37) y (38) quedan probadas.

Teniendo en cuenta éste resultado, el decidir dentro de E_g por medio de la longitud del intervalo de riesgo equivale a emplear $I_*(/M)$ o $I^*(/M)$ como función valor (Keeney y Raiffa (1976)), lo que corresponde, a lo que posteriormente denominaremos critério pesimista y critério optimista, respectivamente.

Si representamos gráficamente la imagen de $\{\tilde{r}_i \; i=1,..,n\}$ según g, obtenemos una nube de puntos en el plano, situados todos ellos por encima de la diagonal principal, ya que, en cualquier caso, la integral superior siempre es mayor o igual que la inferior. Es inmediato, que E_g está formado por las decisiones representadas por los puntos de la nube situados más arriba y más a la derecha. Por definición, $\underline{d}$

está asociada al punto de E_g más cercano (en distancia euclidea) a la diagonal principal y, $\overline{d}$ al más lejano, o lo -- que es lo mismo, $\underline{d}$ es el más cercano al eje de abcisas y $\overline{d}$ al de ordenadas, lo que demuestra graficamente la proposición anterior.



Hay que destacar por último que, puesto que $\precsim_g$ es una relación de orden bastante "natural", podemos limitar la -- aplicación de cualquier critério de decisión al conjunto E_g

Dicho de otro modo, la relación de dominancia no permite obtener una decisión optima, salvo en casos particulares pero proporciona un método para eliminar "malas acciones".

<u>Ejemplo</u> .4.5

Consideremos un problema con:

- Matriz de recompensas:

	w_1	w_2	w_3
d_1	40	25	15
d_2	25	15	30
d_3	35	10	15

-Asignación básica de probabilidad:

$$M(A) = \begin{cases} 0.3 & A = \{w_1\} \\ 0.4 & A = \{w_2, w_3\} \\ 0.3 & A = \{w_1, w_3\} \end{cases}$$

Puesto que r_3 está dominado por r_1, podemos eliminar d_3

Por otra parte:

$$I_*(d_1/M) = 22.5 \qquad I^*(d_1/M) = 34$$

$$I_*(d_2/M) = 21 \qquad I^*(d_2/M) = 28.5$$

Es inmediato que $d_1 \succ_g d_2$ de modo que d_1 es en éste -caso la decisión optima.

Supongamos ahora un problema como el anterior pero --con una acción más, d_4, con vector de recompensas:

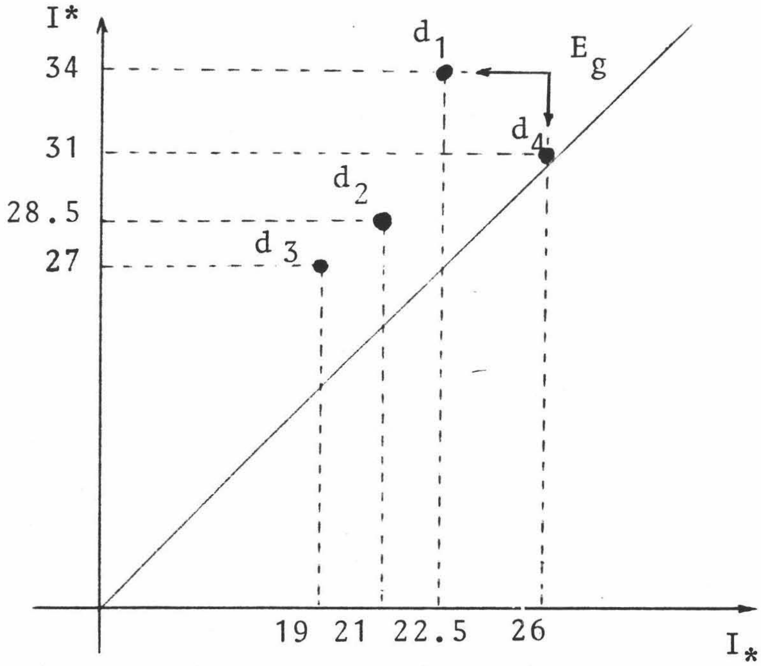
$$r_4 = (\ 35,\ 20,\ 25\ )$$

para la cual:

$$I_*(\tilde{r}_4/M) = 26 \qquad I^*(\tilde{r}_4/M) = 31$$

Así pues, d_4 no es comparable con d_1 mediante $\gtrsim_g$ y --

por tanto $E_g = \{d_1, d_4\}$. Para optar por una de éstas dos acciones, el decisor deberá emplear algún critério adicional; atendiendo, por ejemplo, a la amplitud del intervalo de riesgo, d_4 es preferible a d_1 para un decisor pesimista, mientras que ocurre al contrario para uno optimista.



Nota

De acuerdo con los resultados de Yu (1974), $\gtrsim_g$ no es otra cosa que la dominancia con respecto al cuadrante positivo de R^2. Todos los desarrollos anteriores pueden repetirse empleando estructuras de dominación (conos) más generales.

Orden lexicográfico.

Según la postura del decisor frente al riesgo consideraremos dos órdenes, el directo y el inverso, que esencialmente consiste en aplicar el método lexicográfico comenzando por la primera componente o por la segunda del intervalo de riesgo respectivamente.

A) Para un decisor pesimista (tiene aversión al riesgo), lo más importante es que la cota inferior de sus recompensas esperadas ($I_*(./M)$) sea lo mayor posible. De éste modo establecerá la siguiente relación de preferencias:

$$d_i \succ_L d_k \iff I_*(\tilde{r}_i/M) > I_*(\tilde{r}_k/M) \quad \text{ó}$$

$$I_*(\tilde{r}_i/M) = I_*(\tilde{r}_k/M) \quad \text{e} \quad I^*(\tilde{r}_i/M) > I^*(\tilde{r}_k/M)$$

$$d_i \sim_L d_k \iff I_*(\tilde{r}_i/M) = I_*(\tilde{r}_k/M) \quad \text{e} \quad I^*(\tilde{r}_i/M) = I^*(\tilde{r}_k/M)$$

Puede comprobarse que ésta relación es total. El criterio de decisión correspondiente será denominado LEXICOGRAFICO DIRECTO, ya que comienza por la primera componente del intervalo de riesgo.

B) Si un decisor es optimista, valorará fundamentalmente el extremo superior del intervalo de riesgo ya que, cuanto mayor sea éste, mayor podrá ser la recompensa a obtener. De éste modo, establecerá la siguiente relación de preferencias:

$$d_i \succ_{⅂} d_k \iff I^*(\tilde{r}_i/M) > I^*(\tilde{r}_k/M) \quad \text{ó}$$

$$I^*(\tilde{r}_i/M) = I^*(\tilde{r}_k/M) \quad \text{e} \quad I_*(\tilde{r}_i/M) > I_*(\tilde{r}_k/M)$$

$$d_i \sim_{⅂} d_k \iff I^*(\tilde{r}_i/M) = I^*(\tilde{r}_k/M) \quad \text{e} \quad I_*(\tilde{r}_i/M) = I_*(\tilde{r}_k/M)$$

Necesariamente se tiene una relación de orden total. - Al critério de decisiónasociado lo denominaremos LEXICOGRAFICO INVERSO, ya que, ahora se comienza por comparar la segunda componente del intervalo de riesgo.

Es inmediato comprobar que:

$$d_i \succsim_g d_k \Longrightarrow d_i \succsim_L d_k \quad ; \quad d_i \succsim_{\neg} d_k$$

<u>Ejemplo</u> .4.6

Consideremos un problema con matriz de recompensas:

	w_1	w_2	w_3
d_1	50	40	20
d_2	45	55	30
d_3	40	30	25

y supongamos que la asignación básica de probabilidad es:

$$M(A) = \begin{cases} 0.3 & A = \{w_1\} \\ 0.4 & A = \{w_1, w_2\} \\ 0.3 & A = \{w_1, w_3\} \end{cases}$$

es inmediato que:

$$\left.\begin{array}{l} g(\tilde{r}_1) = (\ 37,\ 50\ ) \\ g(\tilde{r}_2) = (\ 40.5,\ 49\ ) \\ g(\tilde{r}_3) = (\ 31.5,\ 40\ ) \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} d_1 \succsim_{\neg} d_2 \succ_{\neg} d_3 \\ \\ d_2 \succ_L d_1 \succ_L d_3 \end{array}\right.$$

Observese que, r_3 está dominada por r_1 y r_2 y así d_3 - está dominada según g por d_2 y d_1, de manera que, tambien - es la peor decisión según los ordenes lexicográficos. Por - definición, d_1, $d_2 \in E_g$, de modo que, el empleo de $\succsim_L$ ó $\succsim_1$ permite individualizar una decisión (óptima) en éste conjunto.

Si consideramos el caso del ejemplo 3.5, resulta:

$$d_4 \succ_L d_1 \quad ; \quad d_1 \succ_1 d_4$$

lo que permite individualizar una decisión (óptima) dentro del correspondiente E_g.

Nota

Es posible obtener una versión modificada de éstos critérios de decisión empleando la variante del orden lexicográfico que contempla la existencia de niveles de aspiración. (ver Keeney y Raiffa (1976)).

3.- CRITERIOS DE DECISION BASADOS EN UNA FUNCION VALOR

Según hemos comentado anteriormente, una forma alternativa de establecer un criterio de decisión es construir una función valor(Keeney y Raiffa (1976))

$$v : R^2 \longrightarrow R$$

y asegurar

$$d_i \succsim d_k \quad \text{si y solo si} \quad v\,[g(\tilde{r}_i)] \geqq v\,[g(\tilde{r}_k)]$$

Esta forma de actuar, no es independiente de la analizada en el capítulo anterior tal y como ya apuntamos, (ver Fishburn 1970, 1973). Posteriormente, volveremos a insis--

tir sobre el tema.

De acuerdo con las propiedades de g, y de los desarrollos de los dos apartados anteriores, el problema puede --- plantease en términos de construir una función valor para - un problema biatributo con las siguientes caracteristicas:

1) Los dos atributos no son contradictorios, de modo - que, pueden crecer simultaneamente.

2) El decisor aspira a conseguir el más alto nivel en los dos atributos (elemento máximal en $\gtrsim_g$), si bien no -- tiene por que existir una decisión en la que se satisfaga - éste requerimiento.

Así pues, parece razonable exigir que v sea creciente en sus dos argumentos, es decir;

$$(x_1,x_2) \geqq (y_1,y_2) \Longrightarrow v(x_1,x_2) \geqq v(y_1,y_2)$$

de tal manera que:

$$d_i \gtrsim_g d_k \Longrightarrow v\left[g(\tilde{r}_i) \right] \geqq v\left[g(\tilde{r}_k) \right]$$

Existen multiples formas de construir una función valor con éstas características, cada una de las cuales co - rresponderá a un conjunto fiferente de hipótesis, sobre el sistema de preferencias del decisor (ver Keeney y Raiffa - (1976) ó Chankong y Haimes (1983), para un estudio detallado del modelo biatributo). Como en el apartado anterior, - nos vamos a limitar a estudiar tres casos particulares de - interés:

1) $v^*:R^2 \longrightarrow R$ tal que $v^*(x_1,x_2)=$ máx (x_1,x_2)

2) $v_*:R^2 \longrightarrow R$ tal que $v_*(x_1,x_2)=$ mín (x_1,x_2)

3) $v_c:R^2 \longrightarrow R$ tal que $v_c(x_1,x_2)= c_1x_1+c_2x_2 \; (c_1,c_2)\in R_2^+$

Criterio optimista

Mediante el empleo de v^*, se llega a la siguiente relación de orden en $\mathfrak{D}$.

$$d_i \succsim^* d_k \iff v^*(g(\tilde{r}_i)) \geqq v^*(g(\tilde{r}_k))$$

$$\iff \text{máx } \{I_*(\tilde{r}_i/M) \ , \ I^*(\tilde{r}_i/M)\} \geqq$$

$$\text{máx } \{I_*(\tilde{r}_k/M) \ , \ I^*(\tilde{r}_k/M)\}$$

$$\iff I^*(\tilde{r}_i/M) \geqq I^*(\tilde{r}_k/M)$$

ya que $I^*(\tilde{r}/M) \geqq I_*(\tilde{r}/M) \quad \forall \tilde{r} \varepsilon R^n$.

Este criterio lleva a tomar la decisión en base unicamente a la integral superior, y por tanto, lo emplearan personas optimistas o arriesgadas. Es propio de un decisor que asigne toda la evidencia que se tiene sobre un subconjunto $A c$, a aquel estado de la Naturaleza $w_j \varepsilon A$ cuya recompensa sea mayor, comportamiento obviamente optimista, ya que supone colocarse en la circunstancia más favorable.

Es inmediato que:

1.a) $d_i \succsim_g d_k \Longrightarrow d_i \succsim^* d_k$

1.b) $d_i >^* d_k \iff d_i \succ_1 d_k$, definiendo las preferencias estrictas, en términos de desigualdades estrictas. En caso de que

$$I^*(\tilde{r}_i/M) = I^*(\tilde{r}_k/M) \text{ se tiene } d_i \sim^* d_k$$

y habrá que acudir a criterios de selección adicionales si se quiere individualizar una decisión.

Una posibilidad, es recurrir a la comparación de

$I_*(r_i/M)$ e $I_*(r_k/M)$, reobteniendo, de éste modo, el orden - lexicográfico inverso.

1.c) Este criterio no tiene en cuenta cuestión alguna relativa a las tasas marginales de sustitución.

Criterio pesimista

Mediante v_* se llega a la siguiente relación de orden en $\mathfrak{D}$:

$$d_i \succsim_* d_k \iff v_*(g(\tilde{r}_i)) \geq v_*(g(\tilde{r}_k))$$

$$\iff \text{mín}\{ I_*(\tilde{r}_i/M) \ , \ I^*(\tilde{r}_i/M) \} \geq$$

$$\text{mín}\{ I_*(\tilde{r}_k/M) \ , \ I^*(\tilde{r}_k/M) \}$$

$$\iff I_*(\tilde{r}_i/M) \geq I_*(\tilde{r}_k/M)$$

ya que $I^*(\tilde{r}/M) \geq I_*(\tilde{r}/M) \quad \forall \tilde{r} \varepsilon R^n$.

Llamaremos pesimista a éste criterio, porque selecciona una decisión, comparando los valores de las integrales - inferiores de las distintas acciones, con lo cual tratamos de elegir aquella decisión cuyo riesgo inferior esperado -- sea máximo.

El decisor se pone en el caso más desfavorable para el suponiendo que para cada decisión d_i, i=1,...,m, toda la masa de evidencia del conjunto $A \varepsilon \Omega$ se asignará a aquel $w_j \varepsilon A$, cuya recompensa para cada d_i sea menor.

Como en el caso optimista es inmediato que:

2.a) $d_i \succsim_g d_k \iff d_i \succsim_* d_k$

2.b) $d_i \succ_* d_k \iff d_i \succ_L d_k$

definiendo nuevamente las preferencias estrictas en térmi-- nos de desigualdades estrictas. Si

$$I_*(\tilde{r}_i/M) = I_*(\tilde{r}_k/M)$$

entonces d_i y d_k son indiferentes, y si se desea escoger u- na de ellas hay que recurrir a criterios adicionales. Una - posibilidad es comparar $I^*(\tilde{r}_i/M)$ con $I^*(\tilde{r}_k/M)$, reobteniendo en éste caso, el orden lexicográfico directo.

2.c) Este criterio tampoco tiene en cuenta, cuestión - alguna relativa a las tasas marginales de sustitución.

Nota

De acuerdo con la proposición 3.1, si se emplea la longitud del intervalo de riesgo, como un criterio de selección dentro de E_g, se obtienen los que aquí denominamos criteri- os pesimista y optimista. Este resultado es intuitivamente justificable ya que un decisor que opte por decisiones con pequeño intervalo, está buscando la seguridad en la predic- ción de la recompensa esperada, (no tiene afición al ries- go), mientras que, uno que busque una decisión con un gran intervalo de riesgo está apostando por la obtención de re-- compensas elevadas.

Criterio lineal

Consideremos que un decisor acepta el empleo de una -- función valor lineal:

$$v_c(x_1,x_2) = c_1x_1+c_2x_2 \ : \ R^2 \longrightarrow R$$

Es bien conocido (ver Keeney y Raiffa (1976) ó Chan-kong y Haimes (1973)) que se trata de un individuo para el cual, las tasas marginales de sustitución son constantes e iguales a c_1/c_2 (x_2 sobre x_1) y c_2/c_1 (x_1 sobre x_2).

Por otra parte, y por las características del modelo tiene que ser $c_1 \geqq 0$, $c_2 \geqq 0$.

Es obvio que, cualquiera que sea $c \geq 0$, la relación de orden inducida por v_c es identica a la inducida por

$$v_{c/k} = c_{1/k} x_1 + c_{2/k} x_2$$

siendo k, una constante positiva arbitraria. Esto nos permite trabajar con funciones valor lineales normalizadas. Existen, tres formas típicas de realizar ésta normalización:

1) $k=c_1$
con lo que se obtiene

$$v_t(x_1,x_2)=x_1+tx_2 \qquad t \varepsilon R^+$$

donde t es la tasa marginal de sustitución de x_1 sobre x_2.

2) $k=c_2$
obteniendose

$$v_r(x_1,x_2)=rx_1+x_2 \qquad r \varepsilon R^+$$

siendo ahora r la tasa marginal de sustitución de x_2 sobre x_1.

3) $k=c_1+c_2$
que nos proporciona

$$v_\alpha(x_1,x_2)=(1-\alpha)x_1+\alpha x_2 \qquad \alpha \varepsilon \left[0,1\right]$$

el coeficiente α mide la importancia relativa que el deci--

sor da a cada uno de los atributos. En nuestro modelo y dadas las características de g, nos parece más interesante una formulación de éste tipo, motivo por el cual nos vamos a centrar en ésta forma de la función valor.

A través de g, v_α establece la siguiente relación de preferencias en $\mathfrak{D}$:

$$d_1 \gtrsim_\alpha d_2 \iff v_\alpha(g(\tilde{r}_i)) \geq v_\alpha(g(\tilde{r}_k))$$

$$\iff \alpha I^*(\tilde{r}_i/M)+(1-\alpha)I_*(\tilde{r}_i/M) \geqq$$

$$\alpha I^*(\tilde{r}_k/M)+(1-\alpha)I_*(\tilde{r}_k/M)$$

Si para una decisión fija d_i, hacemos variar α de 0 a 1, obtenemos, todos los puntos del intervalo que tiene por extremos $I_*(\tilde{r}_i/M)$, $I^*(\tilde{r}_i/M)$ y en particular:

$$\alpha=0 \longrightarrow v_o\left[I_*(\tilde{r}_i/M) , I^*(\tilde{r}_i/M)\right] = I_*(\tilde{r}_i/M)$$

que nos reproduce el criterio pesimista

$$\alpha=1 \longrightarrow v_1\left[I_*(\tilde{r}_i/M) , I^*(\tilde{r}_i/M)\right] = I^*(\tilde{r}_i/M)$$

que nos proporciona el criterio optimista.

En general, cuanto mayor es α , mayor es la importancia que el decisor concede a $I^*(./M)$ y reciprocamente. Así, α puede considerarse como una medida de su grado de optimismo. Un decisor neutro ante el riesgo deberá tomar $\alpha=1/2$. En éstas condiciones parece razonable admitir que el valor

$$1/2\ I^*(./M) + 1/2\ I_*(./M)$$

juega en el ambiente de información general, el mismo papel que la esperanza matemática en el ambiente de riesgo clásico.

Son inmediatas las siguientes propiedades:

1) Si $d_i \gtrsim_g d_k$ entonces $d_i \gtrsim_\alpha d_k \quad \forall \alpha \varepsilon [0,1]$

2) Las tasas marginales de sustitución son $\alpha/1-\alpha$ y $1-\alpha/\alpha$, de modo que, por las características del modelo, éstas tasas marginales reflejan el optimísmo o pesimísmo del decisor y son fáciles de medir atendiendo a éste caracter.

3) Dadas $d_i, d_k \in \mathfrak{D}$ el valor de α para el cual $d_i \sim_\alpha d_k$ se obtiene a partir de la ecuación:

$$\alpha \left| I^*(\tilde{r}_i/M) - I^*(\tilde{r}_k/M) \right| = (1-\alpha) \left| I_*(\tilde{r}_i/M) - I_*(\tilde{r}_k/M) \right|$$

donde como es lógico, juegan un papel crucial, las tasas marginales de sustitución.

De acuerdo con la propiedad 1, la ecuación anterior solo tiene solución si $d_i, d_k \varepsilon E_g$, hecho intuitivamente obvio por las propiedades de éstos conjuntos.

Si notamos $\bar{\alpha}$ el valor de la indiferencia entre d_i y d_k entonces para todo $\lambda < \bar{\alpha}$ se tomará la decisión más pesimista de las dos (menor intervalo de risego), mientras que, para $\lambda > \bar{\alpha}$ se optará por la más optimista (mayor intervalo de riesgo).

Ejemplo.4.7

Consideremos nuevamente el problema del ejemplo 3.5 en el cual:

$$g(\tilde{r}_1) = (22.5, 34)$$

$$g(\tilde{r}_2) = (21, 28.5)$$

$$g(\tilde{r}_3) = (\ 19,\ 27\ )$$

$$g(\tilde{r}_4) = (\ 26,\ 31\ )$$

Puesto que $E_g = \{d_1, d_4\}$, solo hemos de fijar nuestra atención en éstas dos decisiones.

Es inmediato comprobar, que el valor de α para el cual son indiferentes es:

$$\overline{\alpha} = 6/13$$

y como d_1 es más optimista que d_4

$$\lambda > \overline{\alpha} \Longrightarrow d_1 \succ_\lambda d_4 \quad ; \quad \lambda < \overline{\alpha} \Longrightarrow d_4 \succ_\lambda d_1$$

Sea, por ejemplo, $\lambda = 0.6$. Por definición:

$$v_{0.6}(g(\tilde{r}_1)) = 0.6 \times 34 + 0.4 \times 22.5 = 29.4$$

$$v_{0.6}(g(\tilde{r}_4)) = 0.6 \times 31 + 0.4 \times 26 = 28$$

tal y como cabia esperar.

4.- OBTENCION DE ALGUNOS CRITERIOS DE DECISION CLASICOS

Según sabemos, los ambientes de riesgo e incertidumbre se presentan mediante evidencias particulares (probabilística e ignorancia, respectivamente).

A continuación vamos a comprobar, como, en éstos casos, nuestra formulación reproduce algunos de los criterios de decisión más usuales.

Evidencia probabilística

Según hemos indicado repetidas veces, si M es una evidencia probabilística asociada a una distribución de probabilidad P entonces:

$$I_*(\tilde{r}/M) = I^*(\tilde{r}/M) = E[\tilde{r}/p] \qquad \forall\, r \in R$$

Así pues, g es en realidad una función de R^n en R que puede emplearse como función valor, dado lo cual, los criterios anteriormente descritos, se reducen al de la esperanza que es el criterio clásico en ambiente de riesgo.

Ignorancia

Es ya conocido que el ambiente de incertidumbre (solo puede asegurarse que el verdadero estado de la Naturaleza - esté en Ω); se representa mediante la ignorancia, es decir $M=m_o$, y entonces:

$$I_*(\tilde{r}/M) = \min_j r_j$$

$$I^*(\tilde{r}/M) = \max_j r_j$$

cualquiera que sea

$$\tilde{r} = (r_1,\dots,r_n) \in R^n$$

Así pues se tiene:

Criterio optimista.- Consitirá en elegir aquella decisión - que nos haga máximo el resultado de la aplicación v*. como

$$v^*\left[\min_{w_j \in \Omega} r_{ij} \, , \, \max_{w_j \in \Omega} r_{ij} \right] = \max_{w_j \in \Omega} r_{ij}$$

entonces se deberá elegir la decisión, para la cual

$$\max_{w_j \in \Omega} r_{ij}$$

sea máximo. Vemos pues, que se obtiene el criterio MAXIMAX de la Teoria clásica.

Criterio pesimísta.- Se trata de maximizar sobre v_*. Puesto que:

$$v_* [\min_j r_{ij} , \max_j r_{ij}] = \min_j r_{ij}$$

se deberá elgir la decisión correspondiente a

$$\max_i \min_j r_{ij}$$

Así pues, estaremos empleando el criterio de Wald (MAXIMIN)

Criterio lineal.- Si se emplea v_α como función valor, se -- trata de encontrar la decisión para la cual:

$$v_\alpha (g(r_i))=\alpha \max_j r_{ij} + (1-\alpha) \min_j r_{ij}$$

sea máximo. Así pues, en éste caso se obtiene el bien conocido criterio de Hurwicz.

5.- NOTAS FINALES

Como ya indicamos en la introducción, el objetivo de - éste capítulo es presentar un modelo de decisión con ambiente general e introducir algunas herramientas para la obtención de criterios de decisión razonables.

Es obvio que existen formas alternativas de abordar - el problema y aspectos del mismo que no han sido tenidos - en cuenta aquí. Sin pretender ser exhaustivos (lo que sería imposible dado el caracter subjetivo de la Teoria de la Decisión), a continuación recogemos algunas cuestiones que consideramos de interés y que esrán objeto de estudios posteriores.

1) La aplicación

$$g : R^n \longrightarrow R^2$$

que constituye la base de todos nuestros criterios, es construye empleando de una forma específica la información disponible sobre el verdadero estado de la Naturaleza, lo que equivale a la adopción de determinados supuestos de comportamiento. En concreto, la asignación de la masa de evidencia varia según el vector que se considere, ya que, para cada elemento focal de M solo se contemplan los valores extremos (mínimo y máximo) de la restricción del mencionado -- vector al citado elemento focal.

Esta forma de repartir la masa de evidencia de cada elemento focal de M supone, implicitamente el admitir que - el decisor para elegir solo considera los valores extremos de las ganancias y la posibilidad de obtenerlos, sin tener en cuenta para nada valores intermedios. Por éste motivo y como hemos podidocomprobar a lo largo del capítulo, se obtiene una estructura especialmente adaptada para la consideración de la actitud del decisor ante el riesgo y la obtención de los correspondientes criterios de decisión.

Son concebibles, obviamente, formas alternativas de - manejar la información disponible (por medio de repartos - distintos de la masa de evidencia) que, por supuesto, conduciran a reglas de decisión diferentes.

2) Mediante g (aceptando su uso de acuerdo con las -

consideraciones del punto anterior), se convierte el problema original en otr biatributo equivalente. Para él y bajo -- ciertas hipótesis de comportamiento, pueden construirse funciones valor, mas generales que las consideradas aquí, que - generaran criterios de decisión distintos de los analizados.

3) Todos los criterios de decisión que hemos descrito - se basan en el orden natural de R, y por ello resultan ser - "crisp". Teniendo en cuenta que la información disponible solo permite obtener "intervalos de riesgo", (en el sentido - de que hay una alta probabilidad de obtener una recompensa - cuyo valor esté contenido en ellos), éstos criterios crisp pueden resultar excesivamente fuertes e incluso inaceptables

Supongamos, por ejemplo, el caso de un decisor optimista, al que se le plantea la elección entre las siguientes decisiones:

$$d_1 \equiv g(r_1) = \{28 \ , \ 29\}$$

$$d_2 \equiv g(r_2) = \{2 \ , \ 29.5\}$$

Es obvio que, aplicando el criterio optimista tal y como está enunciado, $d_2 \succ d_1$. Ahora bien, es facilmente intuible que, aunque el decisor tenga gran afición al riesgo, es dificil que acepte ésta conclusión ya que d_1 parece, en términos generales, "mejor" que d_2. Por éste y otros motivos similares, pensamos que el empleo de relajaciones del orden de R, tales como, "un poco mayor" o "mucho mayor", pueden producir reglas de decisión difusas mas realistas y aplicables.

A título de ejemplo, podemos enunciar la siguiente versión difusa del criterio optimista:

" Si $I^*(\tilde{r}_i/M)$ es mayor que $I^*(\tilde{r}_k/M)$, e

$I_*(\tilde{r}_i/M)$ no es mucho menor que $I_*(\tilde{r}_k/M)$, o si

$I^*(\tilde{r}_i/M)$ es mucho mayor que $I^*(\tilde{r}_k/M)$, entonces d_i es preferida a d_k".

Una forma alternativa de abordar el problema es atender al caracter posibilístico del intervalo de riesgo y considerarlo como un número difuso, acudiendo a técnicas de ordenación de números difusos para la obtención de criterios de decisión.

— BIBLIOGRAFIA —

BIBLIOGRAFIA

BANON, G. (1981)
Distinction between several subsets of fuzzy measures. Fuzzy Sets and Systems 5,325-339.

BELLMAN, R. y L.A. ZADEH (1970)
Decision Making in a fuzzy environment. Management Science, 17 B (4), 141-164

BOLAÑOS, M.J., M.T. LAMATA y S. MORAL (1985)
Decision problems under Shafer's evidence. Comunicacion presentada al I IFSA-Congress. Palma de Mallorca, Julio.

BOLAÑOS, M.J., M.T. LAMATA y S. MORAL (1985)
Alternativas para la extension de medidas difusas. Comunicacion presentada a-la XV Reunion de la SEIO. Gijon.

BOLAÑOS, M.J., M.T. LAMATA y S. MORAL (1985)
La Esperanza Monotona: Una generalizacion de la esperanza probabilistica. Comunicacion presentada a la XV Reunion de la SEIO. Gijon.

BOLAÑOS, M.J., M.T. LAMATA y S. MORAL (1985)
A decision model under general information. BUSEFAL, 24, 56-63.

BUCKLEY, J.J. (1983)
Decision making under risk: A comparison of bayesian and fuzzy sets methods.-Risk Analysis 3, 157-168

CHANKONG, V. y HAIMES, Y. (1983)
Multiobjective decision making. North-Holland.

CHOQUET, G. (1953)
Theory of capacitis. Ann. Inst. Fourier, 5, 131-292.

GAINES, B.R. (1984)
Fundamentals of decision: Probabilistic, possibilistic and other forms of uncertainty in decision analysis. TIMS Studies in Management Sciences, 20, 47 - 65, North-Holland.

DE GROOT, M. (1970)
Optimal Statistical Decisions. McGraw Hill

DELGADO, M. y S. MORAL (1985)
A definition of inclusion for evidences. Fuzzy Mathematics, por aparecer.

DEMPSTER, A.P. (1967)
Upper and lower probabilities induced by a multivalued mapping. Ann. Math. - Stat., 38, 325-339.

DUBOIS, D. y H. PRADE (1980)
Fuzzy sets and systems. Theory and applications. Academic Press.

DUBOIS, D. y H. PRADE (1982)
On several representations of an uncertain body of evidence. En Fuzzy Información and Decision Processes. M.M. Gupta y E. Sanchez (Eds). North-Holland.

DUBOIS, D. y H. PRADE (1983)
Upper and lower possibilities induced by a multivalued mapping. Reports of -- the IFAC Symposium on Fuzzy Information and Knowledge representation. M.M. -- Gupta y E. Sanchez (Eds). Pergamon Press, 147-152.

DUBOIS, D. y H. PRADE (1985)
A set theoretic view of belief functions. Por aparecer en Int. J. of General-Systems.

FISHBURN, P.C. (1970)
Utility theory for decision making. J. Wiley.

GILES, R. (1982)
Foundations for a theory of possibility. En Fuzzy Information and Decision -- Processes. M.M. Gupta y E. Sanchez (Eds). North-Holland.

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, U. (1982)
A mathematical theory of uncertainty. En Fuzzy sets and Possibility Theory. - Recent Developments. R.R. Yager (Ed). Pergamon Press, 344-351.

HUBER, P.J. ((1981)
Robust Statistics. Wiley.

JAYNES, E.T. (1957)
Information theory and statistical mechanics. Phys. Rev. 106, 620-630.

KANDEL, A. (1981)
Fuzzy Expectation and energy states in fuzzy media. Fuzzy Sets and Systems, 6 145-160.

KANDEL, A. (1982)
Fuzzy techniques in pattern reconigtion. J. Wiley & Sons.

KEENY, R.L. y H RAIFA (1976)
Decision with multiple objectives: Preferences and values tradeoffs. J. Wiley

KLEMENT, E.P. y D. RALESCU (1983)
Nonlinearity of the fuzzy integral. Fuzzy sets and Systmes, 11, 309-316.

MORAL, S. (1985)
Informacion difusa. Relaciones entre probabilidad y posibilidad. Tesis Doctoral. Universidad de Granada.

NGUYEN, H. T. (1978)
On conditional possibility distributions. Fuzzy Sets and Systems, 1, 299-310.

NGUYEN, H. T. (1978)
On Random sets and belief functions. J. Math. Anal. and Appl., 65,3, 531-542.

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-Holland, 235-246.

NGUYEN, H.T. (1979)
Some mathematical tools for linguistic probabilities. Fuzzy Sets and Systems- 2, 53-65.

PRADE, H. (1981)
On the link between Dempster's rule of combination of evidence and fuzzy sets intersection. BUSEFAL, 8, 60-64.

PURI, L. y RALESCU, D. (1982)
A possibility measure is not a fuzzy measure. Fuzzy Sets and Systems, 7, 311-313.

RALESCU, D. (1980)
Measures, capacities and optimization with inexact constraints. No publicado.

SCHWEIZER, B. y A. SKLAR (1963)
Associative functions and abstract semi-groups. Publ. Math. Dehecen, 10, 69 - 81.

SHAFER, G. (1976)
A mathematical theory of evidence. Princeton Univ. Press.

SHAFER, G. (1979)
Allocations of probability. Ann. of Prob. 17, 827-839.

SHAFER, G. (1982)
Belief functions and parametric models. Jor. R. Stat. Soc. B, 44, 322-352.

SUGENO, M. (1974)
Theory of fuzzy integrals and its applications. Ph. D. Thesis. Tokyo Inst. of Technology.

TERANO, T. y M. SUGENO (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

THOMAS, M. V. (1979)
A generalized maximum entropy principle. Oper. Res., 27, 1188-1196.

TRIBUS, M. (1969)
Rational Descriptions, decisions and designs. Pergamon Press.

WIERZCHON, S.T. (1982)
On fuzzy measure and fuzzy integral. En Fuzzy Information and Decision Processes. M.M. Gupta y E. Sanchez (Eds). North-Holland.

WOLFERSON, M. y FINE, T. (1982)
Bayes like decision making with upper and lower probabilities. J. Amer. Stat. Ass., 77, 80-88.

YAGER, R.R. (1983)
Entropy and specificity in a mathematical theory of evidence. Int. J. General Systems, 9, 249-260.

YAGER, R.R. (1985)
The entailment principle for Dempster-Shafer granule. Tech. Report MII-512. - Iona College, New Rochelle.

ZADEH, L.A. (1965)
Fuzzy Sets. Information and Control, 8, 338-353.

ZADEH, L.A. (1968)
Probability measures of fuzzy events. J. Math. Anal. and Appl., 23, 421-427.

ZADEH, L.A. (1978)
Fuzzy Sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28.

ZADEH, L.A. (1984)
A simple view of the Dempster-Shafer theory of Evidence. Berkeley Cognitive - Science Report nº 27. California.

DILIGENCIA:

Reunido el Tribunal examinador en el día de la fecha, constituido por:

D. Enrique Trillas Ruiz

D. Francisco Javier Giron Gonzalez-Torre

D. Claudio Alsina i Catalá

D. Concepcion Fernandez Vivas

D. Amparo Vila Miranda

[illegible] -a.

ra juzgar la [illegible]

M. Teresa Lamata Jimenez

Unanimidad

[illegible] de Apto (Cum Laude)

ra que conste, se extiende firmada por los [illegible] ponentes del Tribunal la presente diligencia.

[illegible] 16 Enero de 1986

El Secretario,

El Presidente,

El Vocal,

El Vocal,

El Vocal,