CARACTERIZACION Y ESTUDIO DE MEDIDAS E INTEGRALES DIFUSAS A PARTIR DE PROBABILIDADES.

Luis Miguel de Campos Ibáñez

Departamento de Ciencias de la Computación
e Inteligencia Artificial

Memoria que, para optar al grado de doctor, presenta el licenciado en Ciencias Matemáticas D. Luis Miguel de Campos Ibáñez.

Director

Profesor Dr. D. Manuel Jorge Bolaños Carmona

V° B°

Director

Profesor Dr. D. Miguel Delgado Calvo-Flores

V° B°

UNIVERSIDAD DE GRANADA. Mayo de 1987

Quiero manifestar mi agradecimiento al Dr. D. Jorge Bolaños Carmona y al Dr. D. Miguel Delgado Calvo-Flores, por su labor de dirección y apoyo constante, asi como al Dr. D. Serafin Moral Callejón y al Prof. D. Antonio González Muñoz por sus valiosas sugerencias, a los demás miembros del Grupo de trabajo en Razonamiento Aproximado de la Universidad de Granada, a los compañeros del Colegio Universitario de Jaén y a todos los que de algún modo han contribuido a la realización de esta memoria .

INDICE.

CAPITULO IV

– INTRODUCCION GENERAL –

El concepto de subconjunto difuso fué definido con el objeto de modelizar aquellas situaciones en que los límites de un subconjunto del referencial dado aparecen como mal definidos, por la vaguedad inherente a las condiciones que se imponen a sus elementos o por el relativo cumplimiento de éstas.

Como es sabido, tal noción se basa en la relativización de la función característica y encuentra su fundamento en la lógica multivaluada. La riqueza del concepto permite una diversidad de enfoques matemáticos, como el propiamente conjuntista, el funcional, a través de la función de pertenencia, o el informacional, entendido como el análisis del grado de determinación de un elemento o subconjunto, este último más conectado con la teoría de la decisión o las teorías de la medida.

En todo caso, al ser generalizada la noción clásica de conjunto, se pierde la idea de que éste es una simple agregación de elementos. Coherentemente con esto, puede cuestionarse también la aditividad de las medidas, propia de las teorías clásicas. Aunque la investigación sobre medidas no aditivas comenzó antes de la definición de subconjunto difuso, y dió lugar a diversas líneas con aplicaciones a la física matemática (Choquet [11]), al cálculo de probabilidades (Dempster [16]) o a la estadística matemática (Huber [28]), es la Teoría de subconjuntos difusos la que presta una interpretación semántica y aplicaciones a un extenso campo de problemas de modelización matemática de la información. En suma, las medidas no aditivas encajan conceptualmente con la Teoría de subconjuntos difusos, dando lugar a un pujante desarrollo de las medidas difusas.

Una medida difusa puede entenderse, a su vez, desde diferentes puntos de vista:

a) Como una medida del grado de importancia o relevancia de los subconjuntos de un referencial.

b) Como una evaluación del grado de determinación de un elemento desconocido, mediante la asignación a cada conjunto de una medida que refleje la posible pertenencia de tal elemento al conjunto.

c) Como una valoración del grado en que, globalmente, los elementos de un conjunto verifican ciertas propiedades.

Si adoptamos la definición de Sugeno [52], el campo de las medidas difusas es extraordinariamente amplio; así el estudio de tales medidas se ha realizado, por lo general, a través de la consideración de clases particulares de medidas difusas, como las probabilidades, las posibilidades (Zadeh [62]), las λ-medidas (Sugeno [52]), las evidencias (Shafer [46]), las medidas descomponibles (Weber [55]), etcétera. Aunque se tienen resultados globales, se hace necesaria una sistematización de la clase general de las medidas difusas. Uno de los mecanismos que se han mostrado más fértiles para el análisis de algunos tipos concretos ha sido la expresión de las medidas difusas en términos de probabilidades; ello explica el desarrollo de la Teoría de la evidencia de entre las diversas modalidades de medidas difusas. Una metodología de estudio general se presenta en esta memoria.

El propio concepto de subconjunto difuso requiere también de mecanismos de evaluación que permitan establecer medidas sobre clases de subconjuntos difusos de un referencial, sobre la base del conocimiento de la medida de los subconjuntos ordinarios (crisp) del mismo. La integración difusa provee tal requerimiento. Así, conectadas íntimamente con las medidas difusas, aparecen las integrales difusas; basta pensar que las primeras definiciones de medidas e integral difusas se dieron simultáneamente (Sugeno [52]).

La extensión de medidas a subconjuntos difusos que permiten las integrales difusas generaliza a éstos las interpretaciones antes mencionadas, de modo que puede hablarse de grados de importancia, inclusión de un elemento desconocido y verificación de condiciones de un subconjunto difuso.

La integración difusa tiene tratamientos diversos; por ejemplo, Sugeno utiliza los operadores máximo y mínimo al poner en cuestión la aditividad, y tales conectivos resultan ser elementos destacados de las clases de t-conormas y t-normas. Durante algún tiempo se entendió a la integral de Sugeno como la natural en el contexto difuso, hasta el punto de denominarla, simplemente, integral difusa. Otros tipos de integrales fueron definidos más tarde (Kruse [36] o Weber [55]), a partir de tipos concretos de medidas difusas. Así mismo, la integral de Sugeno ha sido generalizada mediante la modificación de los operadores binarios utilizados (Suárez [51]).

Recientemente algunos autores (Nguyen [41] o Weber [56]) han sugerido la aplicación del operador funcional de Choquet [11] como alternativa general a las integrales conocidas; la definición de esperanza monótona en base al citado funcional (Bolaños, Lamata y Moral [5]) y el estudio de sus propiedades, viene a cubrir esta necesidad.

En su trabajo introductor, Sugeno siguió la idea de establecer un paralelismo entre su integral y la integral clásica de Lebesgue, lo que contribuyó a la mencionada asunción de aquella como un operador natural en el campo difuso. Sin embargo, se carece de un análisis global del papel de las diferentes integrales difusas en el contexto general de las medidas difusas, por lo que pretendemos aquí realizar una contribución en este sentido. Las caracterizaciones de la esperanza monótona y la integral de Sugeno que se llevan a cabo en la presente memoria nos permiten, entre otras cosas, establecer un paralelismo formal entre estos dos funcionales, que corrobora a ambos como alternativas para la extensión de medidas difusas. La integración clásica ocupa así su lugar como caso particular de la esperanza monótona para medidas aditivas, correspondiente a la integral de Sugeno sobre posibilidades. De este modo se obtiene indirectamente un enfoque original del paralelismo entre probabilidades y posibilidades.

En su conjunto, este trabajo ha sido planteado para el caso de referencial finito, dado que, por una parte, solo para él tienen sentido muchas de las definiciones y resultados, por otra, permite focalizar el estudio sobre los aspectos que nos interesan, evitando problemas de medibilidad y continuidad que no aportarían nada a nuestros objetivos, y finalmente, la práctica totalidad de las aplicaciones al campo de la modelización del conocimiento quedan cubiertas por el caso finito.

Como objetivos fundamentales nos proponemos los siguientes:

1) Caracterización y estudio de las integrales difusas generales (esperanza monótona e integral de Sugeno), y análisis de sus posibles generalizaciones, de las relaciones existentes entre ellas y de sus propiedades más relevantes.

2) Caracterización de la clase general de las medidas difusas a partir de probabilidades.

3) Establecimiento de métricas en el conjunto de las medidas difusas, que permiten cuantificar relaciones y

características básicas de las mismas, como la semejanza entre medidas, o la falta de aditividad y/o la información que proporciona cada una de ellas.

4) Definición y estudio de relaciones de inclusión y operaciones de conjunción y disyunción entre medidas difusas, y justificación de las mismas.

Para la consecución de tales objetivos, la metodología empleada puede resumirse en las siguientes etapas:

a) Estudio del comportamiento de la esperanza monótona sobre funciones no negativas del referencial, fijando los elementos relevantes para el valor de aquella y, a partir de tales elementos:

b) Caracterización de la esperanza monótona y la integral de Sugeno, sus relaciones con las probabilidades y posibilidades, sus semejanzas formales y sus posibles extensiones, y

c) Caracterización de medidas difusas a partir de conjuntos ordenados de medidas de probabilidad.

En esta última encuentran su fundamento:

d) La definición de distancias entre medidas difusas, lo que permite establecer índices de semejanza entre dos medidas, de falta de aditividad y de calidad de la información de una medida, y

e) La definición en términos de las caracterizaciones de relaciones de inclusión y métodos de combinación de medidas difusas.

Aunque se analiza con detalle en lo que sigue, conviene señalar que los resultados obtenidos corresponden a un amplio espectro de problemas de interés y actualidad en el campo que nos ocupa. Así, si se piensa en una medida difusa como reflejo del grado de importancia de un subconjunto, son importantes los problemas de determinar el conjunto de los posibles mecanismos de integración entre la medida y la función de pertenencia de un subconjunto difuso (o, en general, de valores de variables cualesquiera) y de la adecuación de la elección del funcional en cada caso. Si la medida se entiende en el contexto de la detección de un elemento desconocido, es fundamental evaluar el grado de información que proporciona. Finalmente, tanto en la interpretación anterior como en la valoración del grado en que se verifican determinadas propiedades, se requiere un suficiente análisis de las características fundamentales de una medida difusa, como la aditividad, precisión, etcétera, así como de las relaciones de semejanza, inclusión o combinación entre diversas medidas. Creemos que el presente trabajo facilita en alguna medida la resolución de tales problemas.

De acuerdo con la metodología empleada, la memoria se estructura en cuatro capítulos.

En el primero se recogen las definiciones y resultados conocidos sobre medidas e integrales difusas que posibilitan el desarrollo de los siguientes, constituidos básicamente por aportaciones originales.

El segundo capítulo constituye la clave metodológica del trabajo. En él se obtienen las caracterizaciones de medidas e integrales difusas mencionadas.

El capítulo tercero se dedica al estudio de métricas en el conjunto de medidas difusas, y a la definición de diversos índices de no aditividad e información, cuyas propiedades y relaciones también se analizan. Los resultados se ejemplifican en el estudio completo de un caso de interés.

Finalmente, el capítulo cuarto aborda la inclusión, conjunción y disyunción de medidas, y el estudio comparativo de nuestras definiciones con algunas otras que han sido establecidas para el caso particular de evidencias.

Elemento fundamental de todo trabajo de investigación, que complementa a la obtención de resultados sobre problemas previos, es el planteamiento de nuevas vías de trabajo. Entre las que se originan en esta memoria cabe destacar:

a) El estudio más detallado de la adecuación de las integrales difusas a diferentes contextos y problemas de aplicación.

b) La definición y sistematización de nuevas generalizaciones de integrales difusas.

c) El análisis de los índices propuestos u otros que puedan ser de interés, a partir de métricas distintas de las aquí utilizadas.

d) La profundización en el estudio de los métodos de combinación de medidas difusas, de modo que puedan obtenerse definiciones o algoritmos que subsanen las dificultades teóricas y prácticas aún existentes.

e) La generalización de cuantos resultados sea posible al caso de referenciales infinitos, si bien presumiblemente los problemas que se plantearían serían de índole bien diferente a los aquí estudiados.

- CAPITULO I -

1.0. INTRODUCCION.

Este capítulo está dedicado a la presentación de las definiciones y resultados conocidos en que se basa el resto de la memoria. Fundamentalmente se trata aquí de definir las medidas e integrales difusas, y exponer los diferentes tipos de unas y otras que se pueden considerar, así como las relaciones entre ellos.

En el caso de las medidas, se define a éstas como valoraciones monótonas y acotadas de subconjuntos del referencial. Esta definición no plantea dificultades en el caso de referencial finito, puesto que las diferentes alternativas que pueden encontrarse en la literatura especializada difieren básicamente en las propiedades de continuidad que se exigen. En todo caso la que adoptamos es la definición más general en nuestro contexto.

Se aborda también en el presente capítulo la definición y estudio de algunos conceptos inevitables en cualquier exposición sobre medidas difusas. Destacan entre ellos los siguientes:

- La dualidad, que asocia a cada medida difusa otra deducible de ella, lo que impone que cualquier análisis deba asumir parejas de medidas, y no medidas individuales. Sin embargo ello no plantea problemas dado que puede considerarse que dos medidas duales contienen la misma información.
- Las normas y conormas triangulares como conectivos que cumplen una doble función en el campo de las medidas e integrales difusas. De una parte, determinan la conexión entre los valores que toma una medida sobre diferentes subconjuntos, definiendo así algunos tipos importantes de medidas. Por otro lado juegan un papel decisivo en la definición de las integrales difusas.

- La clasificación, en lo que hasta el momento se conoce, de las medidas difusas en diferentes tipos o clases, así como las relaciones existentes entre éstas. Dado que el campo de las medidas difusas es muy amplio, y que algunos de los tipos han sido definidos para abordar problemas particulares, es imprescindible tener siempre presente cuales de las propiedades y resultados que obtengamos tienen carácter general, y cuales se refieren solo a clases particulares de medidas.

Por lo que respecta a las integrales difusas, se recogen aquí las definiciones y propiedades de los dos funcionales que cabe considerar como generales, sin entrar en el análisis de aquellos otros definidos para tipos particulares de medidas. Uno de ellos, la integral de Sugeno, ha sido considerado como el mecanismo natural de conexión entre medidas difusas y valores de funciones valuadas en el intervalo unidad, para la extensión de medidas difusas a subconjuntos difusos. El otro, la esperanza monótona, proviene de una idea muy anterior que ha sido recientemente reivindicada en el campo de las medidas difusas.

De cualquier modo, no se trata de realizar aquí una exposición exhaustiva de los conocimientos sobre medidas e integrales difusas disponibles al comienzo de nuestro trabajo, sino de aquellos que resultarán relevantes para el mismo.

En el apartado 1.1 se dan las definiciones de medida difusa, dualidad, normas y conormas, de algunos tipos fundamentales de medidas, y las propiedades básicas de estas nociones.

Una clase de medidas difusas especialmente interesante y que ha sido estudiada con cierta profundidad es la de las medidas generadas por una asignación básica de probabilidad (evidencias). A la exposición de los elementos fundamentales de la Teoría de la evidencia se dedica el apartado 1.2.

Por las razones antes apuntadas, la clasificación y ordenación de los tipos de medidas difusas nos será imprescindible. En el apartado 1.3 se definen de forma más exhaustiva los diferentes tipos de medidas difusas, y se exponen las relaciones entre ellos.

Partiendo de la idea, que se verá confirmada en el capítulo siguiente, de que la integral de Sugeno y la esperanza monótona deben coexistir como métodos alternativos de integración, la definición y propiedades conocidas de ambos funcionales se recogen en el apartado 1.4.

1.1. DEFINICIONES GENERALES SOBRE MEDIDAS DIFUSAS.

En este apartado expondremos brevemente los conceptos fundamentales sobre las medidas difusas que vamos a utilizar, que son las definidas por Sugeno [52]. Existen otras formulaciones diferentes de medidas difusas como las de Butnariu [10], Klement [32], Klement, Lowen y Schwyhla [34] o Höhle [26] que no estudiaremos.

Un desarrollo más detallado de la teoría de medidas difusas de Sugeno puede encontrarse en Banon [1] y Dubois y Prade [17].

Sea un conjunto finito $X=\{x_1,x_2,..,x_n\}$ y sea $\mathcal{A}\subseteq\mathcal{P}(X)$ un álgebra.

<u>Definición 1.1</u>

Sea $g:\mathcal{A}\longrightarrow[0,1]$. Se dice que g es una medida difusa sobre $(X,\mathcal{A})$ si y solo si se cumple

1.- $g(\emptyset)=0$, $g(X)=1$.

2.- Si $A,B\in\mathcal{A}$, $A\subseteq B$, entonces $g(A)\leq g(B)$.

Una medida difusa es una función de conjunto normalizada y monótona. Se puede considerar como una generalización del concepto de probabilidad, sustituyendo la condición de aditividad por la más débil de monotonía.

Para evitar problemas de medibilidad, y puesto que X es un conjunto finito, consideraremos siempre como álgebra $\mathcal{A}=\mathcal{P}(X)$; por tanto las medidas difusas que manejaremos estarán siempre definidas sobre $(X,\mathcal{P}(X))$, y para abreviar, hablaremos de medidas difusas definidas sobre X.

<u>Nota:</u> La tercera condición exigida por Sugeno en la definición de medidas difusas (continuidad de la medida para sucesiones monótonas) siempre se verifica en conjuntos finitos,

y por este motivo la omitimos.

Notaremos al conjunto de las medidas difusas definidas sobre X mediante $\mathfrak{M}(X)$ o simplemente $\mathfrak{M}$, si no hay dudas sobre cual es el referencial escogido.

Definición 1.2

Sea $g \in \mathfrak{M}(X)$. Se llama medida difusa dual de g a la g^* definida por

$$g^*(A) = 1 - g(\overline{A}) \quad \forall A \subseteq X,$$

donde $\overline{A}$ representa el complementario del subconjunto A.

Puesto que obviamente también g es la medida difusa dual de g^*, hablaremos de parejas de medidas difusas duales, notándolas (g, g^*).

El concepto de dualidad es de gran importancia ya que permite obtener dos representaciones alternativas de una información. Así pues consideraremos que una medida difusa y su dual contienen la misma información, solo que codificada de diferente modo.

Definición 1.3

Una medida difusa g se dice autodual si coincide con su medida dual:

$$g(A) = g^*(A) \quad \forall A \subseteq X.$$

Los ejemplos más sencillos de medidas autoduales son las medidas de probabilidad. De este modo el conjunto PR de todas las probabilidades definidas sobre X es un subconjunto de $\mathfrak{M}$.

Las t-normas y t-conormas son útiles en el estudio de las medidas difusas, y sobre todo básicas para la formulación de ciertas integrales difusas que recogemos en este capítulo, así como para el desarrollo de algunos apartados del capítulo II de esta memoria. Por estos motivos exponemos ahora las definiciones más importantes relativas a t-normas y t-conormas. Un estudio más detallado puede encontrarse en Schweizer y Sklar [48] y [49] o en Dubois y Prade [18].

<u>Definición 1.4</u>

Una aplicación $\perp : [0,1] \times [0,1] \longrightarrow [0,1]$ se dice que es una conorma triangular (t-conorma) si y solo si cumple:

1.- Propiedades de frontera

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

2.- Monotonía

$$\text{si } a \leq b \text{ y } c \leq d \text{ entonces } a \perp c \leq b \perp d \quad \forall a,b,c,d \in [0,1]$$

3.- Conmutatividad

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

4.- Asociatividad

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

Una aplicación $* : [0,1] \times [0,1] \longrightarrow [0,1]$ se dice que es una norma triangular (t-norma) si y solo si cumple las propiedades 2,3 y 4 anteriores, y además

1'.- $0*0=0; \quad 1*a=a*1=a \quad \forall a \in [0,1].$

<u>Definición 1.5</u>

Si $\perp$ es una t-conorma, entonces la t-norma $\perp'$ definida por

$$a \perp' b = 1 - ((1-a) \perp (1-b))$$

se llama t-norma dual de $\perp$.

Si $*$ es una t-norma, entonces la t-conorma $*'$ definida por

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

se llama t-conorma dual de $*$.

$\perp'$ y $\perp$, $*$ y $*'$ son parejas de norma y conorma duales.

Las principales t-conormas son:

-operador máximo: $a \vee b = \max(a,b)$

-suma probabilística: $a+b-ab$

-suma acotada: $\min(1,a+b)$

Sus respectivas t-normas duales son:

-operador mínimo: $a \wedge b = \min(a,b)$

-operador producto: ab

-operador de Lukasiewicz: $\max(0,a+b-1)$

Puede comprobarse fácilmente que

$\max(0,a+b-1) \leq ab \leq \min(a,b) \leq \max(a,b) \leq a+b-ab \leq \min(1,a+b)$ $\forall a,b \in [0,1]$.

<u>Definición 1.6</u>

Se dice que la medida difusa g está basada en la t-conorma $\perp$ si y solo si

$$\forall A,B \subseteq X,\quad A \cap B = \emptyset \Rightarrow g(A \cup B) = g(A) \perp g(B).$$

Se dice que la medida difusa g' está basada en la t-norma $*$ si y solo si

$$\forall A,B \subseteq X,\quad A \cup B = X \Rightarrow g'(A \cap B) = g'(A) * g'(B).$$

Se trata de expresar que el valor de medida de $A \cup B$ (A y B disjuntos) o el de $A \cap B$ (A y B exhaustivos) solo depende de los valores de medida de A y de B a través de la conorma $\perp$ y la norma $*$ respectivamente.

Es fácil comprobar que si una medida g está basada en la conorma $\perp$ (respectivamente norma $*$) entonces su medida dual está basada en la norma dual $\perp'$ (respectivamente conorma dual $*'$).

Veamos algunos ejemplos de medidas difusas basadas en normas y conormas:

Ejemplo 1.1

Las medidas de probabilidad están basadas, a la vez, en la conorma

$$a \perp b = \min(1, a+b)$$

y en su norma dual

$$a * b = \max(0, a+b-1).$$

Ejemplo 1.2

Las medidas difusas basadas en la t-conorma del máximo verifican

$$g(A \cup B) = \max(g(A), g(B)) \quad \forall A, B \subseteq X$$

no limitándose esta relación al caso de conjuntos disjuntos.

Estas medidas, introducidas por Zadeh [62] se denominan medidas de posibilidad.

Si Π es una medida de posibilidad podemos definir la distribución de posibilidad asociada como una aplicación

$$\pi : X \longrightarrow [0,1]$$

$$\pi(x) = \Pi(\{x\}) \quad \forall x \in X.$$

Es obvio que

$$\Pi(A) = \sup_{x \in A} \pi(x) \quad \forall A \subseteq X$$

y las medidas de posibilidad quedan caracterizadas conociendo la distribución de posibilidad asociada.

Nota: Si se considera un referencial infinito, las medidas de posibilidad no pueden en general considerarse medidas difusas (ver Puri y Ralescu [42]).

Para más detalles sobre las medidas de posibilidad, consultar Zadeh [62], Nguyen [40] y [41], y Moral [38].

Ejemplo 1.3

Las medidas difusas basadas en la t-norma del mínimo verifican

$$g'(A\cap B)=\min(g'(A),g'(B))\ \forall A,B\subseteq X.$$

Se denominan medidas de necesidad, y son las medidas duales de las posibilidades.

Ejemplo 1.4

Sugeno [52] definió una clase de medidas difusas, denominadas λ-medidas, caracterizadas por la propiedad

$$\forall A,B\subseteq X,\ \ A\cap B=\emptyset\ \Rightarrow\ g_\lambda(A\cup B)=g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A)g_\lambda(B),$$

siendo $\lambda>-1$.

Estas medidas están basadas en la t-conorma

$$a\perp b=\min(1,a+b+\lambda ab).$$

Si $\lambda\in(-1,0]$, la medida dual de g_λ es otra λ-medida g_μ, con $\mu\in[0,+\infty)$ (en concreto $\mu=-\lambda/(1+\lambda)$) y recíprocamente, o sea

$$\forall A\subseteq X\ \ g_{-\lambda/(1+\lambda)}(A)=1-g_\lambda(\overline{A}).$$

1.2. TEORIA DE LA EVIDENCIA.

Una clase especialmente importante de medidas difusas son las generadas por una evidencia. En este apartado repasaremos las definiciones y propiedades más importantes de la Teoría de la evidencia, fundamentada en los trabajos de Dempster [16] y Shafer [46].

El concepto fundamental es el de asignación básica de probabilidad (a.b.p.).

<u>Definición 1.7</u>

Una aplicación $m : \mathcal{P}(X) \longrightarrow [0,1]$ es una a.b.p. si y solo si cumple

1.- $m(\emptyset)=0$.

2.- $\sum_{A\subseteq X} m(A)=1$.

En palabras del propio Shafer: "m(A) mide la masa de creencia confinada en A, pero que se puede mover hacia cualquier punto de A".

<u>Definición 1.8</u>

Dada una a.b.p. m sobre X, $A\subseteq X$ es un elemento o conjunto focal si y solo si $m(A)>0$.

Las a.b.p. no son funciones de conjunto monótonas, por lo que no son medidas difusas. Pero siempre tienen asociadas una pareja de medidas difusas duales.

<u>Definición 1.9</u>

La medida de creencia asociada a una a.b.p. m es una aplicación $bel : \mathcal{P}(X) \longrightarrow [0,1]$ definida por

$$bel(A)= \sum_{B\subseteq A} m(B).$$

Definición 1.10

La medida de plausibilidad asociada a una a.b.p. m es una aplicación Pl $:\mathcal{P}(X)\longrightarrow[0,1]$ definida por

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

Es fácil comprobar que bel y Pl son medidas difusas duales, y además

$$bel(A)\leq Pl(A) \quad \forall A\subseteq X.$$

La relación que liga las a.b.p. y las medidas asociadas es biunívoca, ya que hay una única asignación básica de probabilidad que genere a bel y Pl, que es

$$m(A)=\sum_{B\subseteq A} (-1)^{|A-B|}bel(B) \quad \forall A\subseteq X,$$

donde $|.|$ representa el cardinal de los subconjuntos.

Por tanto una evidencia puede representarse tanto por su a.b.p. como por sus medidas de creencia y plausibilidad asociadas.

Las medidas difusas bel y Pl se caracterizan por ser capacidades de orden infinito (ver Choquet [11]):

Proposición 1.1

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

1.- $bel(\emptyset)=0$, $bel(X)=1$.

2.- $\forall m\in\mathbb{N}$, $\forall A_i\subseteq X$, $i\in\{1,2,..,m\}$

$$bel(\bigcup_{i=1}^{m} A_i)\geq \sum_{\substack{I\subseteq\{1,..,m\}\\ I\neq\emptyset}} (-1)^{|I|+1}bel(\bigcap_{i\in I} A_i).$$

Proposición 1.2

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

1.- $Pl(\emptyset)=0$, $Pl(X)=1$.

2.- $\forall m\in\mathbb{N}$, $\forall A_i\subseteq X$, $i\in\{1,2,..,m\}$

$$Pl(\bigcap_{i=1}^{m} A_i)\leq \sum_{\substack{I\subseteq\{1,..,m\}\\ I\neq\emptyset}} (-1)^{|I|+1} Pl(\bigcup_{i\in I} A_i).$$

Veamos algunos ejemplos importantes de evidencias:

Ejemplo 1.5

Una medida de probabilidad P sobre X es una medida de plausibilidad y de creencia simultáneamente. Los únicos elementos focales de su asignación básica de probabilidad pueden ser los conjuntos unitarios, y

$$m(\{x\})=P(\{x\})=p(x)\ \forall x\in X$$

siendo $p(\cdot)$ la distribución de probabilidad asociada a la medida P. Recíprocamente, toda evidencia cuyos únicos posibles elementos focales sean conjuntos unitarios tiene medidas de plausibilidad y creencia que coinciden con una misma medida de probabilidad.

Ejemplo 1.6

Las medidas de necesidad y de posibilidad son medidas de creencia y de plausibilidad respectivamente. Su asignación básica de probabilidad m se caracteriza por tener sus elementos focales anidados, es decir, existen $A_i\subseteq X$, $i=1,2,..,n$, tales que

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

Si Π es una medida de posibilidad con distribución asociada π, de forma que

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

entonces la a.b.p. asociada a Π, m_π , es

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

Es inmediato comprobar que sus elementos focales están anidados.

Recíprocamente, toda a.b.p. cuyos elementos focales estén anidados tiene asociadas medidas de creencia y de plausibilidad que son medidas de necesidad y posibilidad, respectivamente.

Ejemplo 1.7

La a.b.p. dada por

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

es apropiada cuando no se posee información alguna sobre el fenómeno a estudiar. Esta representación de la ignorancia es marcádamente diferente de la usual en teoría de la probabilidad, donde la ignorancia se caracteriza por la distribución de probabilidad uniforme sobre X.

Las medidas de creencia y plausibilidad asociadas, que llamaremos ignorancia inferior y superior respectivamente, son

$$bel_o(A)=\begin{cases} 1 & \text{si } A=X \\ 0 & \text{si } A\neq X \end{cases} \qquad Pl_o(A)=\begin{cases} 0 & \text{si } A=\emptyset \\ 1 & \text{si } A\neq\emptyset \end{cases}$$

Puede comprobarse que Pl_o es una posibilidad con distribución asociada $\pi(x)=1\ \forall x\in X$.

1.3. CLASES DE MEDIDAS DIFUSAS.

En este apartado haremos una clasificación de las medidas difusas en diferentes clases en orden descendente de inclusión, y por tanto con más propiedades en cada paso. Un esquema de este tipo puede encontrarse en Banon [1], aunque nosotros seguiremos el propuesto por Lamata [37], más completo que el de Banon, y basado en el concepto de dualidad entre medidas difusas.

Puesto que toda medida difusa g tiene una medida dual g^*, y esta correspondencia es biunívoca, consideraremos las medidas por pares duales, aunque a veces nos refiramos solo a uno de los elementos del par.

<u>1. Parejas de medidas ordenadas.</u>

La familia PO es la formada por las parejas duales (g,g^*), de modo que $g(A) \leq g^*(A)$ $\forall A \subseteq X$ (abreviadamente $g \leq g^*$).

En adelante hablaremos de g como la medida inferior y de g^* diremos que es la medida superior. En algunos casos se dice que g codifica la información de forma pesimista, mientras que g^* lo hace de modo optimista.

Desde otro punto de vista, podemos decir que el valor de medida de cualquier subconjunto A de X no está bien determinado, y puede oscilar en el intervalo $[g(A),g^*(A)]$. Volveremos sobre esta interpretación en el capítulo IV, al estudiar una relación de inclusión entre medidas difusas.

<u>2. Parejas de medidas aditivo-coherentes.</u>

La familia AC es la formada por todas las parejas de medidas ordenadas (g,g^*) que verifican

$$\forall A,B\subseteq X,\quad A\cap B=\emptyset \quad \begin{cases} g^{*}(A\cup B)\leq g^{*}(A)+g^{*}(B) \\ g(A\cup B)\geq g(A)+g(B) \end{cases}$$

Este tipo de medidas aparece en Huber [29]. Es inmediato comprobar que las únicas medidas aditivo-coherentes que son autoduales son las probabilidades.

3. Parejas de medidas que acotan una probabilidad.

La familia PA es la formada por las parejas de medidas duales (g,g^{*}) que verifican que existe una probabilidad P tal que

$$g(A)\leq P(A)\leq g^{*}(A) \quad \forall A\subseteq X.$$

En Huber [29] se puede encontrar una caracterización de este tipo de medidas.

Esta familia y la anterior no coinciden, pero AC∩PA contendrá a las familias que consideremos posteriormente.

Es evidente que las únicas medidas autoduales de PA son las probabilidades.

4. Medidas representables.

La familia MR es la formada por las parejas duales (g,g^{*}) para las que existe una familia $\mathcal{P}_o$ de medidas de probabilidad tal que

$$\forall A\subseteq X \quad \begin{cases} g(A)=\inf\limits_{P\in\mathcal{P}_o} P(A) \\ g^{*}(A)=\sup\limits_{P\in\mathcal{P}_o} P(A) \end{cases}$$

Estas medidas, introducidas por Dempster [16] han sido caracterizadas por Wolf [57] (ver Huber [29]) y Giles [23] desde puntos de vista diferentes.

La familia MR está contenida en la intersección de AC y PA pero no coincide con ella.

5. Capacidades de orden dos.

La familia C2 es la formada por las parejas duales (g,g^*) tales que

$$\forall A,B\subseteq X \quad \begin{cases} g(A\cup B)\geq g(A)+g(B)-g(A\cap B) \\ g^*(A\cup B)\leq g^*(A)+g^*(B)-g^*(A\cap B) \end{cases}$$

es decir, g y g^* son respectivamente capacidades monótona y alternante de orden dos (ver Choquet [11]).

En Huber [29] se da una demostración de que C2⊂MR. En el capítulo IV de esta memoria, donde estudiaremos en detalle este tipo de medidas, daremos otra demostración de este resultado.

6. Evidencias.

La familia EV es la formada por parejas (bel,Pl) de medidas de creencia y plausibilidad asociadas a una misma a.b.p..

Puesto que bel y Pl son capacidades de orden infinito, es evidente que EV⊂C2.

7. Evidencias consonantes.

La familia EC es la formada por parejas (N,Π) de medidas de necesidad y posibilidad duales.

8. Medidas de tipo crisp.

La familia CR es la formada por parejas duales de medidas de necesidad y posibilidad que únicamente toman los valores 0 y 1. Están asociadas a a.b.p. con un único elemento focal A_o. Son por tanto del tipo

$$g(A)=\begin{cases} 1 & \text{si } A_o\subseteq A \\ 0 & \text{en otro caso} \end{cases} \qquad g^*(A)=\begin{cases} 1 & \text{si } A_o\cap A\neq\emptyset \\ 0 & \text{en otro caso} \end{cases}$$

La distribución de posibilidad asociada es

$$\pi_{A_o}(x)=\begin{cases} 1 & \text{si } x\in A_o \\ 0 & \text{en otro caso} \end{cases}$$

9. Medidas de Sugeno.

La familia $G\lambda$ está formada por pares (g_λ, g_μ) de λ-medidas de Sugeno, con $\lambda\in[0,+\infty)$ y $\mu=-\lambda/(1+\lambda)\in(-1,0]$.

Banon [1] probó que las λ-medidas son evidencias ($G\lambda\subset EV$).

10. Medidas de probabilidad.

Sea PR el conjunto de medidas de probabilidad. Es obvio que las probabilidades son λ-medidas de Sugeno para $\lambda=0$, y por tanto $PR\subset G\lambda$.

11. Medidas de Dirac o probabilidades degeneradas.

La familia DI está formada por las probabilidades degeneradas en algún punto de X

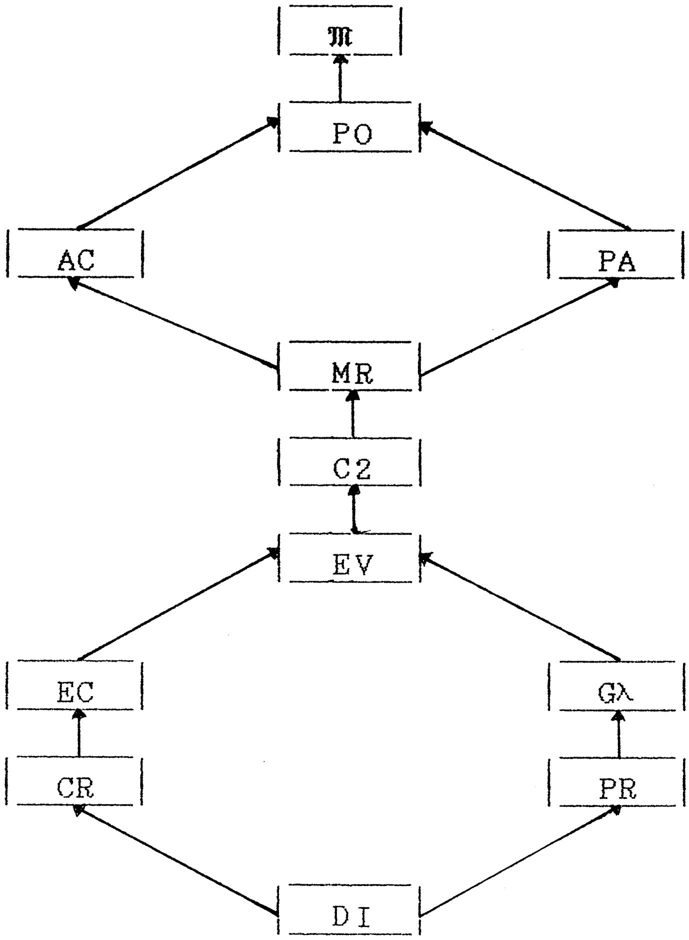
$$P^{x_o}(A)=\begin{cases} 1 & \text{si } x_o\in A \\ 0 & \text{si } x_o\notin A \end{cases} \qquad \forall x_o\in X.$$

con distribución de probabilidad asociada

$$p^{x_o}(x)=\begin{cases} 1 & \text{si } x=x_o \\ 0 & \text{si } x\neq x_o \end{cases}$$

Es obvio que $DI=PR\cap CR$, es decir, las probabilidades degeneradas son las únicas medidas que son a la vez posibilidades y probabilidades.

El diagrama siguiente, extraido de Lamata [37], ilustra gráficamente las relaciones entre las diferentes familias de medidas. El sentido de las flechas indica inclusión.


$\mathfrak{M}$
PO
AC
PA
MR
C2
EV
EC
Gλ
CR
PR
DI

1.4. INTEGRALES DIFUSAS.

Cuando sobre un conjunto X se dispone de una medida difusa, es interesante contar con herramientas que permitan resumir la información proporcionada por una función en un solo valor, que sería una especie de promedio de la función, en términos de la medida subyacente. Tales herramientas son las integrales difusas.

En este apartado estudiaremos dos tipos de integrales difusas: la integral de Sugeno y sus generalizaciones (integrales t-normadas) y la esperanza monótona.

Existen otros tipos de integrales difusas que no consideraremos en esta memoria, fundamentalmente porque solo están definidas para tipos particulares de medidas difusas de Sugeno. En esa línea están los trabajos de Kruse [35] y [36] para λ-medidas, Zi-Xiao [63] para posibilidades, y Weber [55] para medidas descomponibles.

1.4.1. LA INTEGRAL DE SUGENO. GENERALIZACIONES.

Sugeno [52] dió la siguiente definición de integral difusa:

<u>Definición 1.11</u>

Sea g una medida difusa definida sobre X, y

$$h: X \longrightarrow [0,1]$$

La integral difusa de h respecto de g es

$$\oint h\circ g = S_g(h) = \sup_{F\subseteq X} \left(\inf_{x\in F} h(x) \wedge g(F)\right)$$

El propio Sugeno probó que la integral anterior puede escribirse como

$$S_g(h)=\sup_{\alpha\in[0,1]}(\alpha\wedge g(H_\alpha))$$

donde $H_\alpha=\{x\in X/\ h(x)\geq\alpha\}$, y $\wedge$ denota el operador mínimo, como en lo sucesivo se utilizará $\vee$ para el máximo.

La integral de Sugeno ha sido profundamente estudiada por diversos autores desde diferentes puntos de vista, y utilizada para múltiples aplicaciones (estadística, Kandel [31], teoría de la información, Batle y Trillas [2], diagnóstico médico, Vila y Delgado [53], Gupta et al. [24],...). En el plano teórico, podemos destacar los trabajos de Batle y Trillas [2], Bolaños [4], y Ralescu y Adams [43], donde se dan diversas caracterizaciones de la integral de Sugeno, y los de Wang [54] y Ralescu [44] que estudian propiedades de convergencia.

Algunas de las propiedades más importantes de la integral de Sugeno son:

1.- Si $a\in[0,1]$, entonces $S_g(a)=a$.

2.- $h(x)\leq h'(x)\ \forall x\in X \Rightarrow S_g(h)\leq S_g(h')$.

3.- $g(A)\leq g'(A)\ \forall A\subseteq X \Rightarrow S_g(h)\leq S_{g'}(h)\quad \forall h:X\longrightarrow[0,1]$.

4.- Si $a\in[0,1]$ y $h:X\longrightarrow[0,1]$

$$S_g(a\vee h)=a\vee S_g(h)$$

$$S_g(a\wedge h)=a\wedge S_g(h)$$

5.- $S_g(h\vee h')\geq S_g(h)\vee S_g(h')\quad \forall h,h':X\longrightarrow[0,1]$.

6.- $S_g(h\wedge h')\leq S_g(h)\wedge S_g(h')\quad \forall h,h':X\longrightarrow[0,1]$.

7.- $S_g(I_A)=g(A)$, siendo I_A la función característica del subconjunto A de X.

8.- Si P es una medida de probabilidad, $\forall h:X\longrightarrow[0,1]$

$$|S_P(h)-\int_X h\,dP|\leq\frac{1}{4}$$

siendo $\int_X h\,dP$ la esperanza matemática de la función h respecto de la probabilidad P.

9.- Si $X=\{x_1, x_2, .., x_n\}$, y se cumple

$$h(x_1) \le h(x_2) \le \ldots \le h(x_n)$$

entonces

$$S_g(h) = \bigvee_{i=1}^{n} (h(x_i) \wedge g(A_i))$$

donde $A_i = \{x_i, x_{i+1}, \ldots, x_n\}$, $i=1, .., n$.

Todas estas propiedades están demostradas en Sugeno [52].

Suárez [51], debilitando los conceptos de t-norma y t-conorma (no exigiendo las propiedades asociativa y conmutativa a estos operadores) ha generalizado la integral de Sugeno, introduciendo las llamadas integrales seminormadas y semiconormadas. Nuestro interés se centra en las integrales seminormadas, cuando se emplea no una t-seminorma sino una norma triangular, a las que llamaremos integrales normadas.

Definición 1.12

Sea g una medida difusa definida sobre X, $*$ una t-norma y

$$h: X \longrightarrow [0,1]$$

La integral normada de h respecto a g es

$$IN_g(h) = \sup_{\alpha \in [0,1]} (\alpha * g(H_\alpha))$$

siendo $H_\alpha = \{x \in X / \ h(x) \ge \alpha\}$.

La integral de Sugeno es un caso particular de esta integral, cuando la t-norma $*$ coincide con el mínimo.

Las integrales normadas verifican la mayoría de las propiedades que cumple la integral de Sugeno. Destacamos las siguientes:

1.- Si $a \in [0,1]$, entonces $IN_g(a)=a$.

2.- $h(x) \le h'(x) \ \forall x \in X \Rightarrow IN_g(h) \le IN_g(h')$.

3.- $g(A)\leq g'(A)\ \forall A\subseteq X \Rightarrow IN_g(h)\leq IN_{g'}(h)\quad \forall h:X\longrightarrow[0,1]$.

4.- Si $a\in[0,1]$ y $h:X\longrightarrow[0,1]$

$$IN_g(a\vee h)=a\vee IN_g(h)$$

$$IN_g(a*h)=a*IN_g(h)$$

5.- $IN_g(h\vee h')\geq IN_g(h)\vee IN_g(h')\quad \forall h,h':X\longrightarrow[0,1]$.

6.- $IN_g(I_A)=g(A)$, siendo I_A la función característica del subconjunto A de X.

7.- Si $X=\{x_1,x_2,..,x_n\}$, y se cumple

$$h(x_1)\leq h(x_2)\leq...\leq h(x_n)$$

entonces

$$IN_g(h)=\bigvee_{i=1}^{n}(h(x_i)*g(A_i))$$

donde $A_i=\{x_i,\ x_{i+1},...,x_n\}$, $i=1,..,n$.

La demostraciones de estas propiedades pueden encontrarse en Suárez [51].

1.4.2. LA ESPERANZA MONOTONA.

La integral que vamos a describir ahora fue introducida en el año 1953 por Choquet [11], junto a la teoría de capacidades. Posteriormente ha sido utilizada por Huber [28] y [29] y Beran [3] en aspectos teóricos del cálculo de probabilidades y estadística matemática. En el marco de la teoría de subconjuntos difusos, Nguyen [41] la ha empleado sobre medidas de posibilidad, y más recientemente ha sido extendida por Bolaños, Lamata y Moral [5] (a quienes se debe el término esperanza monótona) para medidas difusas cualesquiera y funciones positivas; una generalización a funciones cualesquiera ha sido estudiada por Lamata [37].

En cuanto a sus aplicaciones, podemos destacar los trabajos de Bolaños, Lamata y Moral [6] y [7] sobre teoría de la decisión y extensión de medidas difusas, y de Bolaños y de Campos [8] sobre diagnóstico médico.

Sea X un conjunto finito y h una aplicación

$$h : X \longrightarrow \mathbb{R}_o^+$$

Si P es una medida de probabilidad sobre X, es conocido que

$$E_P(h) = \sum_{x \in X} p(x)h(x) = \int_o^\infty P([h \geq x])dx.$$

Basándonos en esta expresión de la esperanza matemática, introducimos la definición de esperanza monótona respecto a una medida difusa.

<u>Definición 1.13</u>

Sea g una medida difusa definida sobre X, y $h: X \longrightarrow \mathbb{R}_o^+$.

La esperanza monótona de h respecto a la medida g es

$$E_g(h) = \int_o^\infty g(H_\alpha)d\alpha$$

donde $H_\alpha = \{x \in X / \ h(x) \geq \alpha\}$, conjuntos que por adecuación al contexto difuso denominaremos α-cortes de h.

La esperanza monótona siempre existe y es finita cualesquiera que sean g y h, y es obvio que constituye una generalización de la esperanza matemática, reduciéndose a ésta cuando la medida difusa empleada sea una probabilidad.

Las propiedades más importantes de la esperanza monótona son:

1.- Si $a \in \mathbb{R}_o^+$ entonces $E_g(a) = a$.

2.- $h(x) \leq h'(x) \ \forall x \in X \Rightarrow E_g(h) \leq E_g(h')$.

3.- $g(A) \leq g'(A) \ \forall A \subseteq X \Rightarrow E_g(h) \leq E_{g'}(h) \quad \forall h: X \longrightarrow \mathbb{R}_o^+$.

4.- $E_g(I_A) = g(A)$, siendo I_A la función característica de $A \subseteq X$.

5.- Si $c,b\in\mathbb{R}_o^+$ entonces $E_g(c+bh)=c+bE_g(h)$ $\forall h:X\longrightarrow\mathbb{R}_o^+$.

6.- Para cualquier medida difusa g y $h:X\longrightarrow[0,1]$

$$|S_g(h)-E_g(h)|\leq \frac{1}{4}$$

generalizándose la cota de Sugeno.

7.- La esperanza monótona no es un operador lineal en general; solo lo es para medidas difusas que sean probabilidades.

8.- Si $\{g_i\}$, $i=1,..,r$ son medidas difusas sobre X y $\lambda_i\in[0,1]$, $i=1,..,r$ son tales que $\sum_{i=1}^{r}\lambda_i=1$, entonces

$$E_g(h)=\sum_{i=1}^{r}\lambda_i E_{g_i}(h) \quad \forall h:X\longrightarrow\mathbb{R}_o^+$$

siendo $g(A)=\sum_{i=1}^{r}\lambda_i g_i(A)$ $\forall A\subseteq X$.

La demostración de estas propiedades puede encontrarse en Bolaños, Lamata y Moral [5] y [6].

Dempster [16] introdujo dos conceptos de esperanza para las asignaciones básicas de probabilidad que Smets [50] ha caracterizado de la siguiente forma:

Definición 1.14

Sea $h:X\longrightarrow\mathbb{R}$ una función cualquiera y m una a.b.p. sobre X. Las integrales superior e inferior de h respecto a m se definen como

$$I^*(h/m)=\sum_{A\subseteq X} m(A)\ (\sup_{x\in A} h(x))$$

$$I_*(h/m)=\sum_{A\subseteq X} m(A)\ (\inf_{x\in A} h(x))$$

Estas integrales también son una generalización de la esperanza matemática, pues para evidencias de tipo probabilístico

$$I^*(h/m)=I_*(h/m)=\int_X hdP$$

siendo P la probabilidad obtenida a partir de m.

Bolaños, Lamata y Moral [5] han probado que estas integrales son un caso particular de la esperanza monótona. Concretamente, si m es una a.b.p. con medidas de creencia y plausibilidad bel y Pl respectivamente, y $h:X \longrightarrow \mathbb{R}_o^+$, entonces

$$I^*(h/m)=E_{Pl}(h)$$

$$I_*(h/m)=E_{bel}(h)$$

Por último indiquemos que la extensión de la esperanza monótona a funciones cualesquiera puede hacerse de dos formas:

Dado $h:X \longrightarrow \mathbb{R}$, como X es finito existirá $\min_{x \in X} h(x)=c$. La función definida mediante $h'(x)=h(x)-c$ es no negativa y podemos calcular su esperanza monótona. Entonces, para h, definimos

$$E_g(h)=E_g(h')+c=E_g(h-c)+c.$$

Esta definición está justificada por la linealidad de $E_g(\cdot)$ para constantes aditivas.

Otra posibilidad es proceder de un modo más parecido al clásico, descomponiendo $h:X \longrightarrow \mathbb{R}$ como $h=h^+-h^-$, siendo

$$h^+=\max(h,0),\ h^-=\max(-h,0)$$

Entonces definimos

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

donde g^* es la medida dual de g.

Se puede probar (ver Lamata [37]) que ambas extensiones son equivalentes y conservan todas las propiedades de la esperanza monótona para funciones no negativas.

- CAPITULO II -

2.0. INTRODUCCION.

Junto a algunos resultados destacables en relación con la caracterización de la esperanza monótona y la integral de Sugeno, en este capítulo se prueban las bases metodológicas de todo el trabajo. La idea clave es la consideración de las ordenaciones posibles entre los elementos de un referencial finito para la caracterización de cada medida difusa por un conjunto de medidas de probabilidad.

Aunque el orden natural de estudio debiera situar en primer lugar a las medidas difusas, por nuestra parte analizamos antes la esperanza monótona, dado que nuestros resultados básicos sobre medidas difusas surgen precisamente de este análisis. En efecto, del hecho de que los subconjuntos cuya medida es relevante para el valor de la esperanza monótona (y para la integral de Sugeno) formen una sucesión monótona, y se distingan entre si por la inclusión de uno o varios elementos, se deduce una ordenación entre éstos, dependiente de la función que se pretende integrar. Ello permite, en un segundo paso, entender que la medida difusa funciona, cara a la obtención de la integral, como una probabilidad, y que de esta forma, de una medida se generan n! probabilidades, correspondientes a las ordenaciones posibles de los elementos del referencial.

Esa es la línea de razonamiento de la caracterización de medidas difusas por probabilidades. A partir de ella, se obtiene una serie de interesantes resultados con referencia a las medidas, sus tipos, las integrales, y las relaciones entre todos estos conceptos.

Sucintamente, cabe destacar entre los resultados originales de este capítulo a:

- Las caracterizaciones de la esperanza monótona y de la integral de Sugeno como funcionales que verifican determinadas propiedades.
- La caracterización de una medida difusa por un conjunto de n! probabilidades.
- La demostración del paralelismo formal entre los dos tipos de integrales, y entre posibilidades y probabilidades en el contexto de la integración difusa.

El apartado 2.1 se dedica al estudio y caracterización de la esperanza monótona, introduciendo el concepto de equiordenación de funciones, y analizando las relaciones de la esperanza monótona con la esperanza matemática y la forma en que aquella posee propiedades que relajan la linealidad de ésta.

La caracterización de las medidas difusas y su relación con la dualidad son el objeto del apartado 2.2.

En el apartado 2.3 se caracteriza la integral de Sugeno, se pone de manifiesto el paralelismo antes mencionado con la esperanza monótona (jugando la F-aditividad el papel de la aditividad), y entre probabilidades y posibilidades, de las que se obtiene una caracterización directa. Estos resultados permiten también probar que la integral de Sugeno sobre posibilidades es el correlato de la integral clásica sobre probabilidades.

En el apartado 2.4 se prueba una caracterización de medidas difusas a partir de posibilidades, y se demuestra que los conjuntos de probabilidades y posibilidades asociados a una misma medida difusa son óptimamente consistentes y mutuamente deducibles, lo que puede interpretarse como la existencia de dos mecanismos de codificación para una información dada por una medida difusa.

Finalmente, en el apartado 2.5 se estudian las posibles generalizaciones de la esperanza monótona y la integral de Sugeno a la luz de sus caracterizaciones respectivas, y se muestra como la clave de la posible utilización de diferentes t-normas o t-conormas es la distributividad entre ellas.

2.1. CARACTERIZACION DE LA ESPERANZA MONOTONA.

Como hemos visto en el capítulo anterior, la esperanza monótona es una generalización a medidas difusas de la esperanza matemática para probabilidades. La propiedad fundamental de la esperanza matemática es la linealidad, y esta no se conserva en la esperanza monótona más que en el caso particular en que la medida difusa sea una probabilidad. La proposición siguiente puede encontrarse en Bolaños, Lamata y Moral [5].

Proposición 2.1

La condición necesaria y suficiente para que se verifique:

$$\forall\ h_1,\ h_2\ : X \longrightarrow \mathbb{R}_0^+,\ E_g(h_1+h_2) = E_g(h_1)+E_g(h_2),$$

es que g sea una medida de probabilidad.

Demostración:

La condición suficiente es inmediata, ya que sobre medidas de probabilidad la esperanza monótona coincide con la esperanza matemática.

La condición necesaria exige probar que cualesquiera que sean $A, B \in P(X)$, con $A \cap B = \emptyset$, se cumple que $g(A \cup B) = g(A)+g(B)$.

Si I_A e I_B son las funciones características de A y B respectivamente, se tiene , por la hipótesis

$$g(A\cup B)=E_g(I_{A\cup B})=E_g(I_A+I_B)=E_g(I_A)+E_g(I_B)=g(A)+g(B),$$

lo que prueba que g es una medida de probabilidad.#

Cuando la medida difusa que se considera no es una probabilidad, la linealidad se pierde al no verificarse la aditividad. No obstante, y por tratarse de una extensión, cabe cuestionarse si la esperanza monótona posee alguna propiedad más débil en relación con la aditividad, siquiera para alguna clase particular de funciones. La proposición siguiente proporciona una expresión de la esperanza monótona que nos permitirá analizar esta cuestión.

En lo sucesivo notaremos, para cualquier función no negativa $h : X \longrightarrow \mathbb{R}_0^+$, $h(x_i)=h_i$, $i=1,..,n$.

Proposición 2.2

Si los valores de una función no negativa h están ordenados de la forma $h_1 \leq h_2 \leq ... \leq h_n$, la esperanza monótona de h respecto de la medida difusa g puede expresarse como

$$E_g(h)=h_1+\sum_{i=2}^{n}(h_i-h_{i-1})g(A_i),$$

o, alternativamente, como

$$E_g(h) = \sum_{i=1}^{n-1} h_i(g(A_i)-g(A_{i+1}))+h_n g(A_n) \ , \qquad (1)$$

donde $A_i = \{x_i, x_{i+1},...,x_n\}$, $i=1,..n$.

Demostración:

Los α-cortes de la función h son:

$$H_\alpha = \begin{cases} \emptyset \quad \text{si } \alpha > h_n \\ \{x_n\}=A_n \quad \text{si } h_{n-1} < \alpha \leq h_n \\ \dots\dots\dots\dots\dots\dots \\ \{x_i,..,x_n\}=A_i \quad \text{si } h_{i-1} < \alpha \leq h_i \\ \dots\dots\dots\dots\dots\dots \\ \{x_1,..,x_n\}=A_1=X \quad \text{si } \alpha \leq h_1 \end{cases}$$

Por tanto

$$E_g(h)=\int_0^{+\infty} g(H_\alpha)d\alpha = h_1 g(X)+\sum_{i=2}^{n}(h_i-h_{i-1})g(A_i),$$

expresión que si desarrollamos y reagrupamos proporciona

$$E_g(h) = \sum_{i=1}^{n-1} h_i(g(A_i)-g(A_{i+1}))+h_n g(A_n).\#$$

La expresión (1) nos permite considerar el hecho de que dos funciones con igual ordenación de sus valores dan lugar a los mismos conjuntos A_i (en cuya configuración solo influye ese orden), y la esperanza monótona de su suma será la suma de las esperanzas monótonas de ambas, puesto que la función suma ordena sus valores del mismo modo. Por tanto la esperanza monótona puede considerarse como un funcional lineal para aquellas funciones que ordenen sus valores de igual forma.

Es necesario formular con más rigor y precisión esta idea. Para ello definimos a continuación una relación binaria entre funciones no negativas, que llamaremos equiordenación, y que será de gran utilidad en las caracterizaciones de las integrales difusas que pretendemos obtener.

<u>Definición 2.1</u>

Dadas dos funciones no negativas h y h', diremos que h está equiordenada con h', y lo notaremos por $h \simeq h'$, si y solo si h es constante o, para cualquier pareja $x_i, x_j \in X$ tales que $h_i < h_j$, se cumple que $h'_i \leq h'_j$.

Esta definición pretende reflejar la idea de que dos funciones tienen sus valores ordenados de la misma forma, bien entendido que dos elementos en los que una función tome el mismo valor pueden ser considerados en cualquier orden.

Las propiedades de esta relación son las siguientes:

1.- $\simeq$ es reflexiva.

2.- $\simeq$ es simétrica:

En efecto, si $h \simeq h'$, y h' es constante, se cumple obviamente la condición $h' \simeq h$. Si $h \simeq h'$ y para una pareja $x_i, x_j \in X$ se cumple $h'_i < h'_j$, ha de ser necesariamente $h_i \leq h_j$, ya que si por el contrario $h_i > h_j$, se verificaría $h'_i \geq h'_j$, contra la condición de partida.

Por tanto, por la simetría de la relación, en lugar de la afirmación "h está equiordenada con h'", se dirá que "h y h' están equiordenadas".

3.- $\simeq$ no es transitiva:

Ello se debe a que la definición no está basada en desigualdades estrictas, y esto permite que, por ejemplo, una función constante esté equiordenada con cualquier otra.

En consecuencia no se trata de una relación de equivalencia.

4.- $\simeq$ es compatible con cualquier operación binaria monótona $*$ entre números reales positivos, y en particular con cualquier norma o conorma triangular , es decir, se cumple que

$$\text{si } h \simeq h'' \text{ y } h' \simeq h'', \text{ entonces } h*h' \simeq h''.$$

En efecto, si $h*h'$ no es constante, y se tiene $(h*h')_i<(h*h')_j$, ha de verificarse que $h_i''\leq h_j''$: si suponemos, por el contrario, que $h_i''>h_j''$, al ser $h''\simeq h'$ y $h''\simeq h$ se tendría $h_i\geq h_j$ y $h_i'\geq h_j'$, y por la monotonía de la operación $*$, $h_i*h_i'\geq h_j*h_j'$, lo que contradice la afirmación inicial.

A partir del concepto de equiordenación, la aditividad de la esperanza monótona para funciones que ordenen sus valores del mismo modo puede formalizarse como se ve en la siguiente proposición.

<u>Proposición 2.3</u>

Sean h y h' dos funciones no negativas, y sea g una medida difusa. Si h y h' están equiordenadas, se verifica

$$E_g(h+h')=E_g(h)+E_g(h').$$

<u>Demostración:</u>

Supongamos por comodidad que $h_1\leq h_2\leq\ldots\leq h_n$. Como $h\simeq h'$, entonces también $h_1'\leq h_2'\leq\ldots\leq h_n'$, y por tanto

$$h_1+h_1'\leq h_2+h_2'\leq\ldots\leq h_n+h_n'.$$

En caso de que la ordenación de los valores de h fuese distinta, bastaría modificar convenientemente la notación de los elementos de X.

En estas condiciones, y siendo $A_i=\{x_i,x_{i+1},\ldots,x_n\}$, i=1,..,n, la proposición 2.2 nos permite escribir

$$E_g(h+h')=\sum_{i=1}^{n-1}(h_i+h'_i)(g(A_i)-g(A_{i+1}))+(h_n+h'_n)g(A_n)=$$

$$=\sum_{i=1}^{n-1}h_i(g(A_i)-g(A_{i+1}))+h_ng(A_n)+\sum_{i=1}^{n-1}h'_i(g(A_i)-g(A_{i+1}))+h'_ng(A_n)=$$

$$=E_g(h)+E_g(h').\#$$

Así pues, aunque la esperanza monótona no es un funcional aditivo (lo que es coherente con la pérdida de aditividad de las medidas difusas), si conserva una aditividad para funciones equiordenadas, que denominaremos "equiaditividad" o "aditividad ordenada", y que permitiría hablar de una "equilinealidad" o "linealidad ordenada".

Si las propiedades que caracterizan la esperanza matemática son bien conocidas, al relajar la aditividad por la aditividad ordenada, surge la cuestión de qué propiedades son suficientes para caracterizar al funcional esperanza monótona. El siguiente teorema proporciona la respuesta.

<u>Teorema 2.1</u>

Sea E un operador funcional, definido para funciones no negativas, verificando:

(a) Equiaditividad: si $h\simeq h'$ entonces $E(h+h')=E(h)+E(h')$.

(b) Monotonía: si $\forall\ x\in X$, $h(x)\leq h'(x)$ $(h\leq h')$, entonces $E(h)\leq E(h')$.

(c) Normalización: notando por 1_X la función constantemente igual a la unidad, $E(1_X)=1$.

(d) Homogeneidad: $\forall\ a>0$, $E(ah)=aE(h)$.

Entonces existe una única medida difusa g tal que E es la esperanza monótona respecto de g.

Demostración:

Definimos la valoración $g(A)=E(I_A)$, siendo I_A la función característica de A. Comprobaremos en primer lugar que g es una medida difusa:

1) $g(X)=E(I_X)=1$, por la condición (c).

$g(\emptyset)=E(I_\emptyset)=E(0_X)=0$, por el siguiente razonamiento:

Como 0_X es una función constante, para cualquier h se tiene $0_X \simeq h$, y por la condición (a)

$$E(h)=E(h+0_X)=E(h)+E(0_X),$$

de donde $E(0_X)=0$.

2) Si $A\subseteq B$, $g(A)\leq g(B)$. En efecto, de $A\subseteq B$ se deduce que $I_A(x)\leq I_B(x)$ $\forall x\in X$, y por la condición (b),

$$g(A)=E(I_A)\leq E(I_B)=g(B).$$

Por tanto g es una medida difusa.

Veamos ahora que para cualquier función no negativa h,

$$E(h)=E_g(h)$$

Consideremos una función no negativa h cualquiera, y supongamos que verifica $h_1\leq h_2\leq\ldots\leq h_n$, lo cual, como de costumbre, no supone ninguna restricción. Entonces h puede expresarse del siguiente modo:

$$h(x)=h_1 I_{A_1}(x)+\sum_{i=2}^{n}(h_i-h_{i-1})I_{A_i}(x),$$

donde $A_i=\{x_i,\ x_{i+1},\ldots,x_n\}$, $i=1,..,n$.

Notando $f_i(x)=(h_i-h_{i-1})I_{A_i}(x)$, $i=2,..,n$, y $f_1(x)=h_1 I_{A_1}(x)$,

$$h(x)=\sum_{i=1}^{n}f_i(x).$$

Probaremos ahora que las funciones f_i están equiordenadas:

Si $i=1$, f_1 es constante y por tanto $f_1\simeq f_j$, $j=2,..n$.

Para $i\geq 2$, si $f_i(x_k)<f_i(x_l)$, por la construcción de f_i ha de ser $k<i\leq l$. Ahora los dos casos posibles para f_i y f_j son $i<j$ e $i\geq j$.

En el primero, $i<j \Rightarrow k<j \Rightarrow f_j(x_k)=0\leq f_j(x_l)$.

En el segundo, $i\geq j \Rightarrow j\leq l \Rightarrow f_j(x_l)=h_j-h_{j-1}\geq f_j(x_k)$.

Por tanto $f_i\simeq f_j$, $i,j=1,..,n$.

Aplicando repetidamente que la relación de equiordenación es compatible con la suma (que es una operación monótona), y empleando (a) y (d), tenemos

$$E(h)=E(\sum_{i=1}^{n} f_i)=\sum_{i=1}^{n} E(f_i)=E(h_1 I_{A_1})+\sum_{i=2}^{n} E((h_i-h_{i-1})I_{A_i})=$$

$$=h_1E(I_{A_1})+\sum_{i=2}^{n} (h_i-h_{i-1})E(I_{A_i})=h_1+\sum_{i=2}^{n} (h_i-h_{i-1})g(A_i)=E_g(h),$$

por la proposición 2.2.

Finalmente, la unicidad de g es evidente por su propia definición.#

Cualquier funcional que sea equiaditivo, monótono, normalizado y homogéneo es la esperanza monótona respecto de una medida difusa.

Tenemos una caracterización de la esperanza monótona, pues naturalmente el recíproco del teorema anterior también es cierto: (b), (c), y (d) son propiedades de la esperanza monótona que estudiamos en el Capítulo I, y la condición (a) se verifica por la proposición 2.3.

La siguiente cuestión que nos planteamos es ver si las condiciones enunciadas en este teorema de caracterización son redundantes, o las mínimas para obtener tal caracterización. Para resolver esta cuestión veremos algunos ejemplos que prueban la necesidad de las condiciones (a), (b) y (c).

En cada uno de ellos se considera un funcional E que verifica tres de las cuatro condiciones, y se prueba que la otra no es deducible.

Ejemplo 2.1

Sea $E(h)=\sqrt[n]{\prod_{i=1}^{n} h_i}$

Consideremos dos funciones h y h' tales que $h \leq h'$. Entonces $h_i \leq h'_i$, $i=1,..,n$, y en consecuencia $\prod_1^n h_i \leq \prod_1^n h'_i$, y $E(h) \leq E(h')$. Por tanto E cumple (b).

E cumple (c) porque $E(1_X)=\sqrt[n]{1}=1$.

Sea $a>0$. Se tiene $E(ah)=\sqrt[n]{\prod_{i=1}^{n} ah_i} = a\sqrt[n]{\prod_{i=1}^{n} h_i} = aE(h)$, con lo que E cumple (d).

Sin embargo E no verifica (a):

Si $n=2$, las funciones h y h' definidas por

$$h_1=0.5,\ h_2=0.6$$
$$h'_1=0.4,\ h'_2=0.5$$

están obviamente equiordenadas, pero $E(h)=\sqrt{0.3}$, $E(h')=\sqrt{0.2}$, y $E(h+h')=\sqrt{.99} \neq \sqrt{0.3}+\sqrt{0.2}$.

Ejemplo 2.2

Para $n=2$, sea $E(h)=h_1+\max(h_1,h_2)-\min(h_1,h_2)$, lo que equivale a

$$E(h)=\begin{cases} 2h_1-h_2 & \text{si } h_1 \geq h_2 \\ h_2 & \text{si } h_1 \leq h_2 \end{cases}$$

E verifica la condición (a), dado que si $h \simeq h'$, entonces

$$h_1 \geq h_2 \text{ y } h'_1 \geq h'_2 \text{ ó } h_1 \leq h_2 \text{ y } h'_1 \leq h'_2.$$

En el primer caso

$E(h)=2h_1-h_2$, $E(h')=2h'_1-h'_2$ y $E(h+h')=2(h_1+h'_1)-(h_2+h'_2)=E(h)+E(h')$.

En el segundo caso

$E(h)=h_2$, $E(h')=h'_2$ y $E(h+h')=h_2+h'_2=E(h)+E(h')$.

Es inmediato comprobar que E verifica las condiciones (c) y (d). En cambio no cumple la condición (b), pues por ejemplo, para h y h' definidas por

$$h_1=1,\ h_2=0$$
$$h'_1=1,\ h'_2=1$$

es $h\leq h'$ y $E(h)=2$, $E(h')=1$, por lo que $E(h)>E(h')$.

Ejemplo 2.3

Sea $E(h)=\sum_{i=1}^{n}\lambda_i h_i$, donde $\lambda_i\geq 0$, $i=1,..,n$, y $\sum_{i=1}^{n}\lambda_i>1$.

Es obvio que E satisface las condiciones (a), (b) y (d), mientras que no cumple la condición (c).

Así pues, las condiciones (a), (b) y (c) del teorema 2.1 son imprescindibles para caracterizar la esperanza monótona. Probaremos ahora que la condición (d) de homogeneidad no es imprescindible, puesto que se deduce de las anteriores.

Proposición 2.4

Todo operador funcional que sea equiaditivo, monótono y normalizado, sobre funciones no negativas, es también homogéneo.

Demostración:

Hay que probar que $E(rh)=rE(h)\ \forall\ r\in\mathbb{R}^+$. Distinguimos varios casos:

1) Si $r\in\mathbb{N}$, entonces

$$E(rh)=E(h+\overset{r}{\ldots}+h)=E(h)+\overset{r}{\ldots}+E(h)=rE(h),$$

puesto que $h\simeq h$ y E es equiaditivo.

Como $E(h)=E(r(\frac{1}{r}h))=rE(\frac{1}{r}h)$, se verifica

$$E(\tfrac{1}{r}h)=\tfrac{1}{r}E(h).$$

2) Si $r\in\mathbb{Q}^+$, $r=\frac{p}{q}$, donde $p,q\in\mathbb{N}$. Entonces, visto el caso 1), se cumple

$$E(rh)=E(\tfrac{p}{q}\ h)=E(p\ \tfrac{1}{q}\ h)=p\ E(\tfrac{1}{q}\ h)=p\ \tfrac{1}{q}\ E(h)=rE(h).$$

Por tanto de la condición de equiaditividad se deduce la homogeneidad para racionales positivos.

3) Si $r\in\mathbb{R}^+$, r es límite de una sucesión monótona no decreciente $\{r_n\}$ de racionales positivos. Basta entonces probar que

$$E(rh)=\lim_{n\to\infty} E(r_n h) \qquad (2)$$

para tener probado el resultado:

$$E(rh)=\lim_{n\to\infty} E(r_n h)=\lim_{n\to\infty} (r_n E(h))=(\lim_{n\to\infty} r_n)E(h)=rE(h).$$

La condición de monotonía es ahora necesaria, junto a la equiaditividad para probar la igualdad (2), como sigue:

Si tomamos $c=\max_{x\in X} h(x)$, que excepto en el caso trivial en que h es constantemente cero, es estrictamente positivo, podemos escribir

$$\forall\ \varepsilon>0\ \exists\ n_o\in\mathbb{N}\ /\ n\geq n_o \quad |r-r_n|=r-r_n<\frac{\varepsilon}{c}$$

por ser r el límite de la sucesión no decreciente $\{r_n\}$; entonces

$$rh(x)-r_n h(x)=(r-r_n)h(x)<\frac{\varepsilon}{c}\ h(x)\leq\varepsilon\ ,\ \forall x\in X.$$

Luego

$$\forall x\in X \quad rh(x)<r_n h(x)+\varepsilon,$$

y aplicando la monotonía y equiaditividad de E tenemos

$$E(rh(x))\leq E(r_n h(x)+\varepsilon)=E(r_n h(x))+E(\varepsilon).$$

Por otra parte

$$E(\varepsilon)=E(\varepsilon 1)=\varepsilon E(1)=\varepsilon 1=\varepsilon,$$

en virtud de la condición de normalización, y siempre que ε sea un número racional, lo que no supone restricción alguna, ya que $\mathbb{Q}^+$ es un conjunto denso en $\mathbb{R}^+$.

De $E(rh)\leq E(r_n h)+\varepsilon$ se deduce que $|E(rh)-E(r_n h)|\leq\varepsilon$, lo que prueba la validez de la expresión (2).#

Así pues, el teorema de caracterización de la esperanza monótona adopta una forma más refinada:

Teorema 2.2

Un operador funcional E definido para funciones no negativas es

(a) equiaditivo

(b) monótono

(c) normalizado

si, y solamente si existe una única medida difusa g tal que E es la esperanza monótona respecto de la medida g.

2.2. CARACTERIZACION DE MEDIDAS DIFUSAS POR PROBABILIDADES.

Hemos visto como la esperanza monótona conserva la aditividad con carácter general solo para funciones igualmente ordenadas. Cabe analizar entonces por qué el orden de los valores de la función resulta decisivo.

Si recordamos la expresión (1) de la esperanza monótona, obtenida en la proposición 2.2:

$$E_g(h) = \sum_{i=1}^{n-1} h_i(g(A_i)-g(A_{i+1}))+h_n g(A_n)$$

observamos que realmente $E_g(h)$ es un promedio de los valores de h ponderado por ciertos pesos, que son:

$$\begin{cases} p_i=g(A_i)-g(A_{i+1}), & i=1,..,n-1 \\ p_n=g(A_n). & \end{cases}$$

Puesto que

$$\sum_{i=1}^{n} p_i=g(A_1)=g(X)=1 \text{ y } p_i\geq 0, \quad i=1,..,n,$$

podemos interpretar los valores p_i como los de una función de probabilidad. Entonces $E_g(h)$ equivale a la esperanza matemática de h con respecto a esa distribución de probabilidad.

Pero, ¿ de qué dependen los valores p_i?. Obviamente de la medida g, pero también de los conjuntos A_i, que a su vez dependen de h únicamente en la ordenación del conjunto X que determina a través de sus valores.

Podemos pues, enunciar:

<u>Proposición 2.5</u>

La esperanza monótona de una función no negativa h respecto de una medida difusa g coincide con la esperanza matemática de h respecto de una probabilidad que solo depende de g y de la ordenación de los valores de h.

De este modo se obtiene una respuesta a la cuestión planteada al inicio de este apartado. La aditividad de la esperanza monótona para funciones equiordenadas h y h' se explica porque las esperanzas monótonas de h, h', y h+h' son esperanzas matemáticas sobre la misma distribución de probabilidad, lo que de ninguna manera puede garantizarse cuando las dos funciones no inducen el mismo orden.

Además, puesto que la probabilidad empleada en cada caso solo depende, dada g, de la ordenación que la función a integrar establece sobre los elementos de X, el número de distribuciones de probabilidad que, como máximo, podrían aparecer coincide con n!, que es el número de ordenaciones o permutaciones posibles de un conjunto de n elementos.

Desde el punto de vista de la esperanza monótona, y para una función concreta h, el papel de la medida difusa g queda reducido a la determinación de la probabilidad para la que la esperanza monótona es la esperanza matemática de h. En general, de la información contenida en g es posible deducir un número n! de tales probabilidades. Entonces tiene sentido asociar a cada medida difusa esas n! probabilidades, y plantearse si la información contenida en éstas a su vez determina la medida g.

Estudiemos en primer lugar como podemos asociar a cada medida difusa un conjunto de n! probabilidades.

Como hemos comentado, cada una de estas probabilidades se obtiene a partir de una ordenación (o permutación) de los elementos de X; por ejemplo para la ordenación

$$\{x_1, x_2, x_3, \ldots, x_n\},$$

la probabilidad asociada es

$$P_{\{1,2,\ldots,n\}}(x_1)=g(\{x_1\})$$
$$P_{\{1,2,\ldots,n\}}(x_2)=g(\{x_1,x_2\})-g(\{x_1\})$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$P_{\{1,2,..,n\}}(x_i)=g(\{x_1,x_2,..,x_i\})-g(\{x_1,x_2,..,x_{i-1}\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$P_{\{1,2,..,n\}}(x_n)=g(\{x_1,x_2,..,x_n\})-g(\{x_1,x_2,..,x_{n-1}\}).$$

En general, las ordenaciones posibles de los elementos de X vienen dadas por las permutaciones de un conjunto de n elementos, que constituyen el grupo S_n.

Definición 2.2

Llamaremos probabilidades asociadas a una medida difusa g a las P_σ dadas por las funciones de probabilidad

$$p_\sigma(x_{\sigma(1)})=p_{\sigma\sigma(1)}=g(\{x_{\sigma(1)}\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$p_\sigma(x_{\sigma(i)})=p_{\sigma\sigma(i)}=$$

$$=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(i)}\})-g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(i-1)}\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$p_\sigma(x_{\sigma(n)})=p_{\sigma\sigma(n)}=1-g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(n-1)}\}),$$

para cada $\sigma=(\sigma(1),\sigma(2),..,\sigma(n))\in S_n$.

Así, para cada medida difusa g se tiene un único conjunto de n! medidas de probabilidad. El conocimiento de las probabilidades asociadas a g no permite, sin saber también de que permutación proviene cada una de ellas, reconstruir la medida g. Esto se debe a que diferentes medidas difusas pueden generar el mismo conjunto de probabilidades asociadas, como se ve en el siguiente ejemplo.

Ejemplo 2.4

Consideremos, sobre $X=\{x_1,x_2\}$ las medidas difusas g y g' definidas por

A	g(A)	g'(A)
$\{x_1\}$	0.4	0.5
$\{x_2\}$	0.5	0.6
$\{x_1, x_2\}$	1	1

Las probabilidades asociadas a g son:

para la permutación (1,2), $p_{(1,2)}(x_1)=0.4$, $p_{(1,2)}(x_2)=0.6$

para la permutación (2,1), $p_{(2,1)}(x_2)=0.5$, $p_{(2,1)}(x_2)=0.5$.

Las probabilidades asociadas a g' son:

para $\sigma=(1,2)$, $p'_{(1,2)}(x_1)=0.5$, $p'_{(1,2)}(x_2)=0.5$

para $\sigma=(2,1)$, $p'_{(2,1)}(x_1)=0.4$, $p'_{(2,1)}(x_2)=0.6$,

de modo que g y g' tienen asociadas las mismas probabilidades.

Como puede comprobarse en este sencillo ejemplo, el orden desempeña un papel decisivo en la determinación de la medida a partir de sus probabilidades asociadas.

Proposición 2.6

El conjunto de probabilidades asociadas a una medida difusa determina a ésta de forma única si se conoce la permutación correspondiente a cada probabilidad.

Demostración:

Bastará probar que si dos medidas difusas g y g' tienen el mismo conjunto de probabilidades asociadas, y éstas corresponden a las mismas permutaciones, ambas medidas coinciden, es decir,

$$P_\sigma = P'_\sigma \ \forall \sigma \in S_n \Rightarrow g = g'$$

Consideremos un subconjunto cualquiera $A \subseteq X$. Siempre existe una permutación que coloca los elementos de A en los primeros lugares, esto es, si

$$A = \{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$$

existe $\sigma_o \in S_n$ tal que

$$\sigma_o(1)=i_1$$
$$\cdots\cdots$$
$$\sigma_o(k)=i_k$$

En esas condiciones

$$g(A)=g(\{x_{i_1},..,x_{i_k}\})=g(\{x_{i_1},...,x_{i_k}\})-g(\{x_{i_1},..,x_{i_{k-1}}\})+$$
$$+g(\{x_{i_1},..,x_{i_{k-1}}\})-...-g(\{x_{i_1},x_{i_2}\})+g(\{x_{i_1},x_{i_2}\})-g(\{x_{i_1}\})+$$
$$+g(\{x_{i_1}\})=p_{\sigma_o}(x_{i_k})+...+p_{\sigma_o}(x_{i_1})=\sum_{j=1}^{k} p_{\sigma_o}(x_{i_j})=P_{\sigma_o}(A).$$

Analogamente obtenemos $g'(A)=P'_{\sigma_o}(A)$.

Por tanto, de $P_{\sigma_o}(A)=P'_{\sigma_o}(A)$, se deduce $g(A)=g'(A)$.#

El proceso seguido en la demostración indica además la forma concreta en que puede obtenerse g a partir de sus probabilidades asociadas.

Esta proposición nos permite interpretar cualquier medida difusa en términos de sus probabilidades asociadas (que no tienen por qué ser diferentes entre si), hecho que será fundamental en el desarrollo de los capítulos posteriores.

En este sentido, como acabamos de ver, la ordenación es básica, y no podemos prescindir de ella, salvo en tipos especiales de medidas difusas, como se verá en el capítulo IV. Puede ocurrir que las mismas n! probabilidades ordenadas de diferente modo proporcionen distintas medidas difusas. Quizás el caso más notable sea el de una medida difusa y su medida dual. Antes de exponerlo en la proposición siguiente necesitamos una definición.

Definición 2.3

Diremos que dos permutaciones $\sigma,\sigma^*\in S_n$ son duales, o que σ^* es dual de σ, si se verifica

$$\sigma^*(i)=\sigma(n-i+1),\ i=1,..,n.$$

Proposición 2.7

Condición necesaria y suficiente para que dos medidas difusas g y g^* sean duales es que tengan asociadas las mismas n! probabilidades correspondientes a permutaciones duales, es decir, que $P_\sigma=P^*_{\sigma^*}$ si σ y σ^* son duales.

Demostración:

Condición necesaria:

Por comodidad consideremos las permutaciones duales

$$\sigma=(1,2,..,n-1,n) \text{ y } \sigma^*=(n,n-1,..,2,1),$$

y en general procederíamos de forma análoga.

$p_\sigma(x_1)=g(\{x_1\})=1-g^*(\{x_2,..,x_n\})=p^*_{\sigma^*}(x_1)$.

. . . .

$p_\sigma(x_i)=g(\{x_1,..,x_i\})-g(\{x_1,..,x_{i-1}\})=$

$=1-g^*(\{x_{i+1},..,x_n\})-1+g^*(\{x_i,..,x_n\})=p^*_{\sigma^*}(x_i)$.

. . . .

$p_\sigma(x_n)=1-g(\{x_1,..,x_{n-1}\})=g^*(\{x_n\})=p^*_{\sigma^*}(x_n)$.

Condición suficiente:

Basta probar que $g(A)+g^*(\bar{A})=1\ \forall A\subseteq X$.

Por comodidad, y sin que suponga restricción, sean

$$A=\{x_1,x_2,..,x_i\} \text{ y } \bar{A}=\{x_{i+1},..,x_n\}).$$

Consideremos la permutación $\sigma=(1,2,..,n)$, y su dual $\sigma^*=(n,n-1,..,1)$.

$$g(A)=g(\{x_1,..,x_i\})=\sum_{j=1}^{i}(g(\{x_1,..,x_j\})-g(\{x_1,..,x_{j-1}\}))=$$

$$=\sum_{j=1}^{i}p_\sigma(x_j)=P_\sigma(A).$$

$$g^*(\bar{A})=g^*(\{x_{i+1},..,x_n\})=1-(g^*(\{x_1,..,x_n\})-g^*(\{x_{i+1},..,x_n\}))=$$

$$=1-\sum_{j=1}^{i}(g^*(\{x_j,..,x_n\})-g^*(\{x_{j+1},..,x_n\}))=1-\sum_{j=1}^{i}p_{\sigma^*}^*(x_j)=$$

$$=1-P_{\sigma^*}^*(A).$$

Por la hipótesis, $P_\sigma(A)=P_{\sigma^*}^*(A)$, por lo que $g(A)+g^*(\bar{A})=1$.#

Si aceptamos que una medida difusa y su dual contienen la misma información, pero codificada de forma diferente, el resultado anterior podría interpretarse diciendo que las n! probabilidades asociadas contienen la información de la medida difusa, y los diferentes órdenes en que aparezcan son diferentes codificaciones de la misma.

Como consecuencia inmediata de la proposición anterior se caracteriza la autodualidad de una medida difusa en términos de sus probabilidades asociadas:

Corolario 2.1

Una medida difusa g es autodual si y solamente si las probabilidades asociadas a cada permutación y a su dual coinciden.

Otro caso interesante es aquel en el que la medida difusa sea una probabilidad; se trata del caso extremo en que las probabilidades asociadas coinciden entre si:

Proposición 2.8

Una medida difusa g es una medida de probabilidad si y solamente si sus n! probabilidades asociadas son idénticas.

Demostración:

La condición necesaria es obvia por la aditividad de las medidas de probabilidad, y la definición de probabilidades asociadas a una medida difusa.

Para probar la condición suficiente, notaremos P a la probabilidad asociada única. Dado cualquier subconjunto $A\subseteq X$, al igual que en la demostración de la proposición 2.6, podemos deducir que $g(A)=P_{\sigma_o}(A)$ para alguna permutación $\sigma_o \in S_n$.

Puesto que para cualquier permutación es $P(A)=P_\sigma(A)$ $\forall A \subseteq X$, obviamente $g(A)=P(A)$, y g es una probabilidad.#

2.1. CARACTERIZACION DE LA INTEGRAL DE SUGENO.

El funcional más ampliamente utilizado hasta el momento sobre medidas difusas es la integral de Sugeno. En este apartado estudiaremos una caracterización de la misma similar a la del apartado 2.1 para la esperanza monótona, y estableceremos el paralelismo existente entre esos dos funcionales.

Como hemos visto en el capítulo I, la integral de Sugeno de una función $h:X\longrightarrow[0,1]$ respecto de una medida difusa g puede expresarse como

$$\int h\circ g = S_g(h)=\bigvee_{i=1}^{n}(h_i\wedge g(A_i)), \qquad (3)$$

supuesto que los valores de h están ordenados de la forma

$$h_1\leq h_2\leq\ldots\leq h_n,$$

y como anteriormente $A_i=\{x_i, x_{i+1},\ldots,x_n\}$, $i=1,\ldots,n$.

Recordemos también que entre las propiedades que verifica la integral de Sugeno se encuentran:

1) Si $h(x)\leq h'(x)$ $\forall x\in X$, entonces $S_g(h)\leq S_g(h')$.

2) $S_g(1_X)=1$.

3) $\forall$ $a\in[0,1]$, $S_g(a\wedge h)=a\wedge S_g(h)$.

Por nuestra parte, y respecto de la relación de equiordenación, anteriormente definida, probaremos otra propiedad que se mostrará esencial para la caracterización de S_g:

<u>Proposición 2.9</u>

Sean dos funciones $h,h':X\longrightarrow[0,1]$, y sea g una medida difusa. Si h y h' están equiordenadas, se verifica

$$S_g(h\vee h')=S_g(h)\vee S_g(h').$$

<u>Demostración:</u>

Supongamos, como de costumbre, que se verifica

$$h_1\leq h_2\leq\ldots\leq h_n.$$

Como $h \simeq h'$, entonces también

$$h'_1 \leq h'_2 \leq \ldots \leq h'_n,$$

y por tanto

$$h_1 \vee h'_1 \leq h_2 \vee h'_2 \leq \ldots \leq h_n \vee h'_n.$$

En esas condiciones

$$S_g(h \vee h') = \bigvee_{i=1}^{n} ((h_i \vee h'_i) \wedge g(A_i)) = \bigvee_{i=1}^{n} ((h_i \wedge g(A_i)) \vee (h'_i \wedge g(A_i))) =$$

$$= \bigvee_{i=1}^{n} (h_i \wedge g(A_i)) \vee \bigvee_{i=1}^{n} (h'_i \wedge g(A_i)) = S_g(h) \vee S_g(h'),$$

utilizando la distributividad del operador mínimo respecto del máximo.#

Puede entonces considerarse una "equi-F-aditividad" o "F-aditividad ordenada" de la integral de Sugeno, por analogía con la equiaditividad de la esperanza monótona.

Observamos que las propiedades básicas de la integral de Sugeno son iguales o semejantes a las de la esperanza monótona, reemplazando la suma y el producto por el máximo y el mínimo. Resulta entonces lógico pensar en una caracterización de la integral de Sugeno similar a la establecida en el teorema 2.1 para la esperanza monótona.

Teorema 2.3

Sea E un operador funcional no negativo, definido para funciones valuadas en el intervalo unidad, verificando:

(a') Si $h \simeq h'$, entonces $E(h \vee h') = E(h) \vee E(h')$.

(b) Si $h \leq h'$, entonces $E(h) \leq E(h')$.

(c) $E(1_X) = 1$.

(d') $\forall a \in (0,1]$, $E(a \wedge h) = a \wedge E(h)$.

Entonces, y solo entonces, existe una única medida difusa g tal que E es la integral de Sugeno respecto de g.

Demostración:

Condición suficiente:

Se deduce inmediatamente de las propiedades de la integral de Sugeno que acabamos de exponer.

Condición necesaria:

Supongamos ahora que E verifica (a'), (b), (c) y (d'). Como en el teorema 2.1 definimos la valoración

$$g(A)=E(I_A),\ \forall A\subseteq X,$$

y probaremos que es una medida difusa:

La monotonía de g se deduce directamente de la condición (b), y de (c) se concluye que $g(X)=1$. Resta ver que $g(\emptyset)=0$; para ello hay que considerar que $\forall a\in(0,1]$, se cumple

$$E(a)=E(a\wedge 1_X)=a\wedge E(1_X)=a\wedge 1=a,$$

por las condiciones (c) y (d'). En consecuencia, utilizando (a'),

$$\forall a\in(0,1],\ \ a=E(a)=E(a\vee 0_X)=E(a)\vee E(0_X)=a\vee E(0_X),$$

y ello implica $E(0_X)\le a$ $\forall a\in(0,1]$, pudiendo afirmarse entonces que $g(\emptyset)=E(0_X)=0$, al ser E no negativo por hipótesis.

Hay que demostrar ahora que coincide con la integral de Sugeno sobre g.

Sea $h:X\longrightarrow[0,1]$, y supongamos, sin pérdida de generalidad que $h_1\le h_2\le\ldots\le h_n$. Entonces h puede expresarse del siguiente modo:

$$h(x)=\bigvee_{i=1}^{n}(h_i\wedge I_{A_i}(x)),$$

donde $A_i=\{x_i,\ x_{i+1},\ldots,x_n\}$, $i=1,\ldots,n$.

Notando $t_i(x)=h_i\wedge I_{A_i}(x)$, $i=1,\ldots,n$, $h(x)=\bigvee_{i=1}^{n} t_i(x)$.

Probaremos que las funciones t_i están equiordenadas:

Si $t_i(x_k)<t_i(x_l)$, ha de ser, por la construcción de t_i, $k<i\le l$. Los dos casos posibles para t_i y t_j son $i<j$ e $i\ge j$.

En el primero, $i<j \Rightarrow k<j \Rightarrow t_j(x_k)=0\le t_j(x_l)$.

En el segundo, $i\geq j \Rightarrow l\geq j \Rightarrow t_j(x_l)=h_j\geq t_j(x_k)$.

En cualquier caso se verifica por tanto $t_i\simeq t_j$, $i,j=1,..,n$.

Aplicando repetidamente que la relación de equiordenación es compatible con el máximo (pues es una operación monótona), y empleando (a') y (d'), tenemos

$$E(h)=E(\bigvee_{i=1}^{n} t_i)=\bigvee_{i=1}^{n} E(t_i)=\bigvee_{i=1}^{n} E(h_i\wedge I_{A_i})=\bigvee_{i=1}^{n} (h_i\wedge E(I_{A_i}))=$$

$$=\bigvee_{i=1}^{n} (h_i\wedge g(A_i)),$$

que coincide con la expresión (3) de la integral de Sugeno. Por último, la unicidad de g es evidente.#

Así pues, el paralelismo entre los dos tipos de integrales difusas es total, responden al mismo esquema formal, y difieren únicamente en los operadores empleados: suma y producto en la esperanza monótona, y máximo y mínimo en la integral de Sugeno. Sin embargo, las propiedades matemáticas de estos operadores confieren a ambos funcionales un carácter particular, que los hace útiles en diferentes contextos.

Una de las diferencias que pueden encontrarse es que, mientras la esperanza monótona tiene sentido para funciones no negativas cualesquiera, la integral de Sugeno pierde todo su significado para funciones valuadas fuera del intervalo unidad. Así, si se tiene en cuenta que la medida de cualquier subconjunto de X no supera en ningún caso la unidad, el hecho de que una función tomara valores por encima de 1 sería irrelevante para el valor de la integral de Sugeno. Ciertos intentos de modificar la definición para extender este funcional a $\mathbb{R}_0^+$ (Kandel [30]) han sido fuertemente criticados (ver Klement y Ralescu [34]).

Igual que en el apartado 2.1, nos planteamos ahora el problema de averiguar si las condiciones del teorema 2.3 son mínimas o hay alguna redundante y es posible refinar dicho teorema. En este caso no puede suprimirse ninguna: consideraremos situaciones en las que, del cumplimiento de tres de las condiciones, no se deduce la restante.

Ejemplo 2.5

Consideremos $X=\{x_1, x_2\}$ y el funcional

$$E(h)=(h_1 \vee h_2) \wedge (2(h_1 \wedge h_2)-(h_1 \wedge h_2)^2).$$

Puede comprobarse que E verifica las condiciones (b), (c) y (d'). Sin embargo no cumple (a'):

Para las funciones h y h' dadas por

$$h_1=0, \quad h_2=0.8$$
$$h'_1=0.2, \quad h'_2=0.3$$

se cumple que $h \simeq h'$ y en cambio

$$E(h \vee h')=0.36 \neq 0 \vee 0.3=E(h) \vee E(h').$$

Nota: El caso en que no se verifica (b) puede estudiarse teóricamente. De la demostración del teorema 2.3 se deduce que el cumplimiento de las condiciones (a'), (c) y (d') obliga al funcional E a adoptar la forma

$$E(h)=\bigvee_{i=1}^{n} (h_i \wedge E(I_{A_i})),$$

y por tanto, conociendo $E(I_A)$ $\forall A \subseteq X$ queda determinado E. No hay inconveniente para considerar la existencia de un funcional para el que no se verifique la monotonía sin más que asignar, para algunos subconjuntos A y B de X, con $A \subseteq B$, valores de E que verifiquen $E(I_A) > E(I_B)$.

Ejemplo 2.6

El funcional

$$E(h)=(\bigwedge_{x \in X} h(x)) \wedge 0.5$$

obviamente verifica las condiciones (a'), (b) y (d'), y sin embargo $E(1_X)=0.5$, y no se cumple (c).

Ejemplo 2.7

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

$$E(h)=\begin{cases} h_1\vee 0.7h_2 & \text{si } h_1\leq h_2 \\ h_2\vee 0.4h_1 & \text{si } h_1\geq h_2 \end{cases}$$

Puede comprobarse que E cumple (a'), (b) y (c), y en cambio no verifica (d'). Para una función h con valores

$$h_1=0.4,\ h_2=0.9$$

y tomando a=0.5, se tiene

$$a\wedge E(h)=0.5 \neq 0.4=E(a\wedge h).$$

Así pues, en el caso de la integral de Sugeno ninguna de las condiciones es redundante, a diferencia de lo que ocurre con la esperanza monótona. En el apartado 2.5, cuando tratemos el tema de definir otras integrales difusas que respondan al mismo esquema formal de propiedades, pero empleando otros operadores, daremos una explicación de esta diferencia entre S_g y E_g.

Existen analogías entre ambos funcionales también en otros sentidos. Estudiaremos como la integral de Sugeno para el caso de las medidas de posibilidad se comporta en forma similar a la esperanza monótona en el caso de medidas de probabilidad.

Como hemos repetidamente expuesto en anteriores apartados, si una medida difusa es una probabilidad P, la esperanza monótona respecto de ella no es más que la esperanza matemática, y puede escribirse

$$E_P(h)=\sum_{i=1}^{n} p_i h_i . \qquad (4)$$

La proposición siguiente da una expresión análoga para S_g cuando la medida difusa sea una posibilidad, resultado que puede encontrarse en Vila y Delgado [53], y para el que hemos adaptado la demostración a nuestro contexto.

Proposición 2.10

Sea Π una medida de posibilidad, y sea h una función valuada en el intervalo unidad. La integral de Sugeno de h respecto de Π es

$$S_\pi(h)=\bigvee_{i=1}^{n}(\pi_i\wedge h_i), \qquad (5)$$

donde $\pi_i=\Pi(\{x_i\})$, $i=1,..,n$.

Demostración:

Supongamos que h verifica que $h_1\leq h_2\leq\ldots\leq h_n$. Entonces

$$S_\pi(h)=\bigvee_{i=1}^{n}(h_i\wedge\Pi(\{x_i,x_{i+1},\ldots,x_n\}))=\bigvee_{i=1}^{n}(h_i\wedge\bigvee_{j=i}^{n}\Pi(\{x_j\})=$$

$$=\bigvee_{i=1}^{n}\bigvee_{j=i}^{n}(h_i\wedge\pi_j).$$

Como

$$h_i\wedge\pi_i\leq\bigvee_{j=i}^{n}(h_i\wedge\pi_j),\quad i=1,..,n,$$

entonces

$$\bigvee_{i=1}^{n}(h_i\wedge\pi_i)\leq\bigvee_{i=1}^{n}\bigvee_{j=i}^{n}(h_i\wedge\pi_j)=S_\pi(h).$$

Si suponemos que

$$\bigvee_{i=1}^{n}(h_i\wedge\pi_i)<S_\pi(h),$$

entonces existen i_o, j_o, con $j_o\geq i_o$, tales que

$$h_i\wedge\pi_i<h_{i_o}\wedge\pi_{j_o},\ i=1,..,n.$$

Pero, para $i=j_o$ tenemos

$$h_{j_o}\wedge\pi_{j_o}<h_{i_o}\wedge\pi_{j_o},$$

lo cual es imposible porque $h_{j_o}\geq h_{i_o}$.

Por tanto no puede darse la desigualdad estricta, y debe ocurrir que

$$S_\pi(h)=\bigvee_{i=1}^{n}(\pi_i\wedge h_i).\ \#$$

Como se observa, las expresiones (4) y (5) son formalmente iguales, lo que confirma la consideración de la integral de Sugeno como una suerte de esperanza (que se ha dado en llamar esperanza difusa), sobre todo en el caso de posibilidades.

Probamos ahora dos proposiciones que caracterizan los casos particulares de medidas de probabilidad y de posibilidad.

Proposición 2.11

Si en el teorema 2.2 sustituimos la condición (a) de equiaditividad por la más restrictiva de aditividad:

$$E(h+h')=E(h)+E(h'),\ \forall h,h':X \longrightarrow \mathbb{R}_o^+,$$

conservando las restantes condiciones, entonces, y solo entonces E es la esperanza matemática respecto de una medida de probabilidad.

Demostración:

Condición necesaria:

Puesto que podemos descomponer h de la forma

$$h(x)=\sum_{i=1}^{n} h_i I_{\{x_i\}}(x),$$

entonces

$$E(h)=\sum_{i=1}^{n} E(h_i I_{\{x_i\}})=\sum_{i=1}^{n} h_i E(I_{\{x_i\}}),$$

y notando $p_i=E(I_{\{x_i\}})$, como

$$\sum_{i=1}^{n} p_i=\sum_{i=1}^{n} E(I_{\{x_i\}})=E(\sum_{i=1}^{n} I_{\{x_i\}})=E(1_X)=1,$$

$$E(h)=\sum_{i=1}^{n} p_i h_i$$

es la esperanza matemática de h respecto de la probabilidad $P=(p_1,p_2,..,p_n)$.

La condición suficiente es evidente.#

Proposición 2.12

Si en el teorema 2.3, en lugar de la condición (a') de equi-F-aditividad exigimos que se verifique la más restrictiva condición de F-aditividad:

$$E(h\vee h')=E(h)\vee E(h'),\ \forall h,h':X\longrightarrow[0,1],$$

conservando las restantes condiciones, entonces y solo entonces E es la integral de Sugeno respecto de una medida de posibilidad.

Demostración:

Condición necesaria:

Puesto que $I_A\vee I_B=I_{A\cup B}$ $\forall A,B\subseteq X$, y dado que g se define como $g(A)=E(I_A)$ $\forall A\subseteq X$, entonces

$$g(A\cup B)=E(I_{A\cup B})=E(I_A\vee I_B)=E(I_A)\vee E(I_B)=g(A)\vee g(B),$$

y la medida difusa g es una posibilidad.

Condición suficiente:

Si g es una medida de posibilidad, por la proposición 2.10,

$$S_g(h)=\bigvee_{i=1}^{n}(h_i\wedge g(\{x_i\})).$$

Luego

$$S_g(h\vee h')=\bigvee_{i=1}^{n}((h_i\vee h_i')\wedge g(\{x_i\}))=$$

$$=\bigvee_{i=1}^{n}((h_i\wedge g(\{x_i\}))\vee(h_i'\wedge g(\{x_i\})))=$$

$$=\bigvee_{i=1}^{n}(h_i\wedge g(\{x_i\}))\vee\bigvee_{i=1}^{n}(h_i'\wedge g(\{x_i\}))=$$

$$=S_g(h)\vee S_g(h').\#$$

Es claro después de estos resultados que la esperanza matemática se extiende a medidas difusas relajando la aditividad, y obteniéndose un funcional equiaditivo que es la esperanza monótona. También es evidente que la integral de Sugeno para medidas difusas generales es una relajación de la integral de Sugeno para posibilidades, sustituyendo la F-aditividad por la más débil equi-F-aditividad.

Frente a la idea, expuesta por el propio Sugeno y generalmente asumida, de que el funcional equiparable a la integral de Sugeno es la esperanza matemática clásica, creemos haber probado que esto es afirmable solo para el caso particular de las medidas de posibilidad, mientras que en general, para cualquier medida difusa, el paralelismo debe establecerse entre la integral de Sugeno y la esperanza monótona.

Tenemos por tanto dos tipos de integrales difusas formalmente iguales, pero conceptualmente diferentes. La esperanza monótona responde a un planteamiento "aditivo", de promedio clásico, mientras que la integral de Sugeno se adecua a un esquema de tipo "máx-min".

Un tema en el que no entramos es el de qué tipo de funcional utilizar en cada caso. Suponemos que, dependiendo de las características del problema que se plantee, será más adecuado uno u otro método de integración difusa.

2.4. CARACTERIZACION DE MEDIDAS DIFUSAS POR POSIBILIDADES.

En el apartado 2.2 vimos que a una medida difusa cualquiera g podía asociársele un conjunto de $n!$ probabilidades ordenadas que, además de determinar a la medida, eran las probabilidades con las que se efectuaría el cálculo de la esperanza monótona de cualquier función respecto de g. Así mismo hemos analizado como el comportamiento de las medidas de posibilidad para la integral de Sugeno es comparable al de las probabilidades en la esperanza monótona. Resulta entonces lógico plantear si se puede obtener una caracterización de las medidas difusas en términos de posibilidades, siguiendo una metodología similar a la empleada en el caso de las probabilidades. Veremos que tal caracterización es posible.

Comenzamos definiendo las posibilidades asociadas a una medida difusa:

<u>Definición 2.4</u>

LLamaremos posibilidades asociadas a una medida difusa g a las Π_σ dadas por las distribuciones de posibilidad

$$\pi_{\sigma\sigma(1)}=\Pi_\sigma(\{x_{\sigma(1)}\})=g(\{x_{\sigma(1)}\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\pi_{\sigma\sigma(i)}=\Pi_\sigma(\{x_{\sigma(i)}\})=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(i)}\})$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\pi_{\sigma\sigma(n)}=\Pi_\sigma(\{x_{\sigma(n)}\})=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(n)}\})=1,$$

para cada $\sigma=(\sigma(1),\sigma(2),..,\sigma(n))\in S_n$.

La proposición siguiente muestra como el conocimiento de las posibilidades asociadas determina la medida de la que provienen, como ocurría con las probabilidades asociadas, y en las mismas condiciones que entonces.

Proposición 2.13

El conjunto de posibilidades asociadas a una medida difusa determina a esta de forma única si se conoce la permutación correspondiente a cada posibilidad.

Demostración:

Bastará probar que si dos medidas g y g' son tales que $\Pi_\sigma=\Pi'_\sigma$ $\forall\sigma\in S_n$, entonces g y g' coinciden.

Dado cualquier $A\subseteq X$, existe una permutación $\sigma_o\in S_n$ que coloca en los primeros lugares los elementos de A. Entonces, teniendo en cuenta como se definen las posibilidades asociadas, es obvio que $g(A)=\Pi_{\sigma_o}(A)$ y $g'(A)=\Pi'_{\sigma_o}(A)$.

Por tanto $g(A)=g'(A)$ $\forall A\subseteq X$. #

El resultado siguiente reafirma la analogía entre el papel que juegan probabilidades y posibilidades en los dos funcionales que venimos estudiando.

Proposición 2.14

La integral de Sugeno de cualquier función h valuada en el intervalo unidad respecto de una medida difusa g cualquiera coincide con la obtenida respecto de la medida de posibilidad Π_σ asociada a g, y correspondiente a la permutación determinada por el orden de los valores de h.

Demostración:

Supuesta la habitual ordenación de los valores de h

$$h_1\leq h_2\leq\ldots\leq h_n$$

correspondiente a la permutación $\sigma=(n,n-1,..,2,1)$, y notando $A_i=\{x_i, x_{i+1},\ldots,x_n\}$, i=1,..,n, la expresión

$$S_g(h)=\bigvee_{i=1}^{n}(h_i\wedge g(A_i)),$$

permite concluir que

$$S_g(h)=S_{\pi_\sigma}(h),$$

sin más que tener en cuenta que $g(A_i)=\Pi_\sigma(\{x_i\})$, $i=1,..,n$.#

Para cada medida difusa tenemos caracterizaciones alternativas en términos de n! probabilidades y n! posibilidades. Puede establecerse una relación entre la posibilidad y la probabilidad asociadas, para una misma permutación.

Proposición 2.15

Las posibilidades asociadas a una medida difusa determinan sus probabilidades asociadas y recíprocamente, según las expresiones

$$\begin{cases} p_{\sigma\sigma(1)}=\pi_{\sigma\sigma(1)} \\ p_{\sigma\sigma(i)}=\pi_{\sigma\sigma(i)}-\pi_{\sigma\sigma(i-1)}, \ i=2,..,n \end{cases}$$

y
$$\pi_{\sigma\sigma(i)}=p_{\sigma\sigma(1)}+...+p_{\sigma\sigma(i)}, \quad i=1,..,n, \ \forall\sigma\in S_n.$$

Demostración:

Se deduce inmediatamente de las definiciones de P_σ y Π_σ.#

La relación entre posibilidades y probabilidades asociadas a una medida difusa no se limita a obtener unas a partir de otras, sino que como es presumible por ser codificaciones distintas de una misma información, tienen una coherencia que se manifiesta en la verificación de una propiedad de consistencia.

La noción de consistencia entre posibilidades y probabilidades ha sido objeto de numerosos trabajos, como los de Zadeh [62], Dubois y Prade [18], Moral [38], Höhle [27] y Natvig [39]. Por nuestra parte adoptamos la definición de Dubois y Prade, que afirma la consistencia cuando la medida de probabilidad no supera en ningún caso a la de posibilidad. Así, podemos enunciar:

Proposición 2.16

Las medidas de probabilidad y posibilidad asociadas a una medida difusa g, y correspondientes a una misma permutación $\sigma \in S_n$ verifican

$$P_\sigma(A) \le \Pi_\sigma(A), \ \forall A \subseteq X, \ \forall \sigma \in S_n,$$

es decir, son consistentes.

Demostración:

Sin pérdida de generalidad, consideremos la permutación $\sigma=(1,2,..,n-1,n)$. En ese caso

$$\pi_{\sigma i}=p_{\sigma 1}+p_{\sigma 2}+...+p_{\sigma i}, \ i=1,..,n.$$

Por la construcción de Π_σ es evidente que

$$\pi_{\sigma 1} \le \pi_{\sigma 2} \le ... \le \pi_{\sigma n}.$$

Sea A cualquier subconjunto de X. Se verifica

$$\Pi_\sigma(A)= \max_{x_j \in A} \pi_{\sigma j}=\pi_{\sigma j_o},$$

donde $j_o=\max \{j \in \{1,2..,n\} / \ x_j \in A\}$.

Por tanto

$$\Pi_\sigma(A)=\pi_{\sigma j_o} = \sum_{i=1}^{j_o} p_{\sigma i} = \sum_{x_i \in A} p_{\sigma i} + \sum_{\substack{i \le j_o \\ x_i \notin A}} p_{\sigma i} \ge \sum_{x_i \in A} p_{\sigma i} = P_\sigma(A). \#$$

Además, podemos afirmar que las probabilidades y posibilidades asociadas para la misma permutación presentan la mayor consistencia posible, en el sentido de la siguiente proposición:

Proposición 2.17

Si Π_σ y P_σ son la posibilidad y probabilidad asociadas a una medida difusa g para la permutación $\sigma \in S_n$, entonces no existe una probabilidad P que verifique

$$P_\sigma(A) \le P(A) \le \Pi_\sigma(A) \ \forall A \subseteq X.$$

Demostración:

Como de costumbre probemos el resultado para la permutación $\sigma=(1,2,..,n)$ por comodidad.

Supongamos que existe una probabilidad P en esas condiciones. Puesto que

$$\Pi_\sigma(\{x_1,..,x_i\})=\pi_{\sigma i}=p_{\sigma 1}+p_{\sigma 2}+...+p_{\sigma i}=P_\sigma(\{x_1,..,x_i\}),$$

entonces debe ocurrir que

$$P(\{x_1,..,x_i\})=P_\sigma(\{x_1,..,x_i\}),\quad i=1,..,n.$$

Luego se verifica

$$P(\{x_1,..,x_i\})-P(\{x_1,..,x_{i-1}\})=P_\sigma(\{x_1,..,x_i\})-P_\sigma(\{x_1,..,x_{i-1}\}),$$

y por tanto

$$P(\{x_i\})=P_\sigma(\{x_i\}),\quad i=1,..,n,$$

y P coincide con P_σ.#

No se puede, en general, hacer la misma afirmación para posibilidades: hay casos en que existe una medida de posibilidad Π distinta de Π_σ tal que

$$P_\sigma(A)\leq\Pi(A)\leq\Pi_\sigma(A),\quad \forall A\subseteq X.$$

2.5. OTRAS INTEGRALES DIFUSAS.

En los apartados anteriores hemos caracterizado dos tipos de integrales difusas, la integral de Sugeno y la esperanza monótona, como funcionales que verificaban una serie de propiedades. Además veíamos que las propiedades que cumplían ambas integrales eran formalmente iguales, variando únicamente los operadores empleados. Pretendemos ahora, en este apartado, indagar si es posible definir otras integrales difusas que respondan al mismo modelo, pero utilizando otros operadores.

En general el problema es analizar si tiene sentido la definición de funcionales E, que para funciones $h:X \longrightarrow Y$, con $Y=\mathbb{R}_o^+$ o $Y=[0,1]$, verifiquen:

($a^{\perp}$) Si $h \simeq h'$, entonces $E(h \perp h')=E(h) \perp E(h')$.

(b) Si $h \leq h'$, entonces $E(h) \leq E(h')$.

(c) $E(1_X)=1$.

(d^{*}) $\forall a \in Y-\{0\}$, $E(a*h)=a*E(h)$,

donde $\perp$ y $*$ son operaciones binarias en Y.

Se plantea inmediatamente la cuestión de qué tipo de condiciones deben verificar los operadores $\perp$ y $*$.

Como primera condición, y dado que se trata de generalizar la esperanza monótona y la integral de Sugeno a clases de funcionales de similares caracterizaciones, nos restringimos a tipos de operadores binarios que incluyan suma y producto por un lado, y máximo y mínimo por otro. El máximo y el mínimo son, respectivamente, una conorma y una norma triangular en $[0,1]$, mientras que la suma es una "ley de composición" (en el sentido de Schweizer y Sklar [49]) en $\mathbb{R}_o^+$, y el producto verifica ser una t-norma cuando se restringe a $[0,1]$, y no tiene estructura de ley de composición en $\mathbb{R}_o^+$. En consecuencia vamos a considerar operaciones $\perp, * : Y\times Y \longrightarrow Y$, que verifiquen:

1.- $\perp$ y $*$ son asociativas.

2.- $\perp$ y $*$ son conmutativas.

3.- $\perp$ y $*$ son monótonas.

4.- $a*1=a$, y $a*0=0$ $\forall a\in Y$.

5.- $0\perp 0=0$.

Para el caso $Y=[0,1]$, exigiremos además la condición

6.- $a\perp 1=1$, y $a\perp 0=a$ $\forall a\in Y$.

En este último caso, las anteriores propiedades determinan justamente las normas y conormas triangulares.

Para determinar la clase de funcionales que pudieran generarse mediante operadores binarios que verifiquen las anteriores propiedades, empezamos sustituyendo la condición (d^*) por (d) y (d') en las integrales difusas conocidas, para posteriormente considerar la sustitución de (a) y (a') por ($a^{\perp}$).

<u>Proposición 2.18</u>

Condición necesaria para que un funcional E satisfaga (a), (b), (c) y (d^*) es que la operación $*$ sea el producto o coincida con éste para todos los pares (a,b), con $a\in\mathbb{R}_o^+$ y $b\in\{E(I_A)/A\subseteq X\}$.

<u>Demostración:</u>

Consideremos un valor $a\in\mathbb{R}_o^+$, un subconjunto A de X, y la función $h:X\longrightarrow\mathbb{R}_o^+$ definida por

$$h(x)=\begin{cases} a & \text{si } x\in A \\ 0 & \text{si } x\notin A. \end{cases}$$

Por una parte, h puede expresarse como

$$h(x)=a*I_A(x) \quad \forall x\in X,$$

de donde

$$E(h)=E(a*I_A)=a*E(I_A).$$

Por otro lado es obviamente

$$h(x)=aI_A(x) \quad \forall x\in X,$$

y por la proposición 2.4, dado que se verifican (a), (b) y (c), también se cumple (d), por lo que

$$E(h)=E(aI_A)=aE(I_A).$$

Es inmediato entonces que

$$a*E(I_A)=aE(I_A),$$

y * coincide con el producto en todos los pares que pueden interesarnos.#

Obsérvese que para cualquier par no incluido entre los anteriores, los valores resultantes de la operación * no tienen ninguna relevancia en el cálculo de E, por lo que en la práctica, el producto es la única operación binaria compatible con (a), (b), (c) y (d^*).

Como consecuencia inmediata de esta proposición, podemos afirmar que la esperanza monótona no puede generalizarse empleando otro operador distinto del producto.

Por su parte, la integral de Sugeno resulta un caso particular de una familia de funcionales obtenidos mediante la sustitución del mínimo en (d') por cualquier t-norma. Tales funcionales han sido denominados integrales difusas t-normadas por Suárez [51] (ver también Weber [56]). El siguiente teorema caracteriza este tipo de integrales.

<u>Teorema 2.4</u>

Sea E un funcional no negativo definido para funciones valuadas en el intervalo unidad, verificando:

(a') Si $h\simeq h'$, entonces $E(h\vee h')=E(h)\vee E(h')$.

(b) Si $h\leq h'$, entonces $E(h)\leq E(h')$,

(c) $E(1_X)=1$.

(d^*) $\forall a\in(0,1]$, $E(a*h)=a*E(h)$,

donde * es una norma triangular.

Entonces y solo entonces, existe una única medida difusa g tal que E es la integral difusa t-normada respecto de g correspondiente a la norma *.

Demostración:

Condición necesaria:

De forma similar a la empleada en el teorema 2.3, puede probarse que la valoración $g(A)=E(I_A)$ $\forall A\subseteq X$, es una medida difusa (basta cambiar el operador mínimo de entonces por la norma *).

Suponiendo que la función h verifica $h_1\leq h_2\leq\ldots\leq h_n$, podemos expresar h como

$$h(x)=\bigvee_{i=1}^{n}(h_i * I_{\{x_i,\ldots,x_n\}}(x))=\bigvee_{i=1}^{n}t_i(x).$$

En la demostración del teorema 2.3 está probado que

$$t_i\simeq t_j,\quad i,j=1,\ldots,n,$$

por procedimientos igualmente aplicables ahora. Por tanto, empleando repetidamente (a') se tiene

$$E(h)=\bigvee_{i=1}^{n}E(h_i * I_{\{x_i,\ldots,x_n\}}),$$

y la condición (d^*) nos permite afirmar

$$E(h)=\bigvee_{i=1}^{n}(h_i * E(I_{\{x_i,\ldots,x_n\}}))=\bigvee_{i=1}^{n}(h_i * g(\{x_i,\ldots,x_n\})),$$

que es justamente la expresión de la integral difusa t-normada correspondiente a *.

Condición suficiente:

Las condiciones (b), (c) y (d^*) son propiedades bién conocidas de esta integral. La condición (a') se prueba de forma idéntica a la empleada en la demostración de la proposición 2.9, sin más que sustituir nuevamente $\wedge$ por * y tener en cuenta que cualquier t-norma es distributiva respecto del máximo.#

Si se compara este teorema con la proposición 2.18, se aprecia que no es posible generalizar la esperanza monótona mediante la sustitución del producto por otros operadores, mientras que si es factible obtener una generalización de la integral de Sugeno al cambiar el mínimo por cualquier t-norma. Este hecho se debe a la verificación o no de la propiedad distributiva entre los operadores implicados, como se pone de manifiesto en la siguiente proposición.

<u>Proposición 2.19</u>

Una condición necesaria para poder definir un funcional E que verifique ($a^{\perp}$), (b), (c) y (d^{*}) es que la operación $*$ sea distributiva respecto de la operación $\perp$, al menos en todos los casos relevantes para el cálculo de E (que son los del tipo $(a*c)\perp(b*c)=(a\perp b)*c$, con $a,b\in Y$ y $c\in\langle E(I_A)/A\subseteq X\rangle$).

<u>Demostración:</u>

Consideremos los valores $a,b\in Y$, un subconjunto A de X y las funciones $h,h':X\longrightarrow Y$ definidas por

$$h(x)=\begin{cases} a & \text{si } x\in A \\ 0 & \text{en otro caso}\end{cases} \qquad h'(x)=\begin{cases} b & \text{si } x\in A \\ 0 & \text{en otro caso}\end{cases}$$

Evidentemente h y h' están equiordenadas, y la función $h\perp h'$ es

$$(h\perp h')(x)=\begin{cases} a\perp b & \text{si } x\in A \\ 0\perp 0=0 & \text{en otro caso}\end{cases}$$

Como h,h' y $h\perp h'$ se pueden escribir también mediante $h(x)=a*I_A(x)$, $h'(x)=b*I_A(x)$, $(h\perp h')(x)=(a\perp b)*I_A(x)$, $\forall x\in X$, entonces, por (d^{*})

$$E(h)=a*E(I_A),\ E(h')=b*E(I_A),\ E(h\perp h')=(a\perp b)*E(I_A).$$

Como debe cumplirse ($a^{\perp}$),

$$(a*E(I_A))\perp(b*E(I_A))=(a\perp b)*E(I_A),$$

y $*$ es distributiva respecto de $\perp$ en los valores relevantes para el cálculo de E.#

A partir de este resultado es obvio por qué la esperanza monótona no es generalizable y la integral de Sugeno si lo es:

1) En el primer caso la operación $\perp$, definida en $\mathbb{R}_o^+$, es la suma, y el único operador distributivo con respecto a la suma es, de entre los que consideramos, el producto.

2) En el segundo caso, el operador $\perp$ es el máximo, y toda t-norma es distributiva con respecto a él.

Fijadas respectivamente la suma y el máximo, las posibilidades de generalización están aclaradas. Cabe plantearse la sustitución de estos operadores para obtener una familia de funcionales más extensa.

En caso de considerar como conjunto Y el intervalo [0,1], la sustitución del máximo por otras t-conormas ha sido abordada por Suárez [51], quien al encontrar serias dificultades (como el hecho de que para una función de valor constante, el funcional no respetara ese valor), desechó la utilidad de tales generalizaciones. Puede darse una justificación teórica de esta situación, a través de los siguientes resultados:

<u>Proposición 2.20</u>

Condición necesaria para que exista alguna t-norma distributiva respecto de una t-conorma dada $\perp$ es que ésta sea idempotente, es decir, que $c\perp c=c \ \forall c\in[0,1]$.

<u>Demostración:</u>

Si existe alguna norma $*$ tal que

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

entonces , si tomamos a=b=1, puesto que $1\perp 1=1$ y $1*c=c \ \ \forall c\in[0,1]$, tenemos que

$$c=c\perp c \ \forall c\in[0,1].\#$$

Por tanto, combinando este resultado y la proposición 2.19, se concluye que la generalización de las integrales t-normadas solo puede realizarse mediante t-conormas idempotentes. Pero:

Proposición 2.21

La única t-conorma idempotente es el máximo.

Demostración:

Sea $\perp$ una t-conorma idempotente, y consideremos dos valores $x,y\in[0,1]$, tales que $x\leq y$. Por las propiedades de las t-conormas,

$$x\leq y \text{ e } y\leq y \Rightarrow x\perp y\leq y\perp y=y \Rightarrow x\perp y\leq y,$$

$$0\leq x \text{ e } y\leq y \Rightarrow y=0\perp y\leq x\perp y \Rightarrow y\leq x\perp y.$$

Por tanto

$$x\perp y=x\vee y \ \forall x,y\in[0,1].\#$$

Puede concluirse por tanto que si se desean unas propiedades razonables, la integral de Sugeno no puede generalizarse más que a las integrales t-normadas, y el operador máximo es insustituible.

Resulta importante remarcar el diferente papel que desempeñan el máximo y el mínimo en la integral de Sugeno, que desde otro punto de vista (relación entre medidas difusas y ordinarias, sobre subconjuntos difusos), fué también señalado por Bolaños [4]. Esta distinción no proviene del hecho de que el máximo sea una conorma y el mínimo una norma, puesto que es conocido que la integral de Sugeno puede definirse intercambiando el papel de ambos operadores (ver Suárez [51]). La diferencia proviene de cual sea el operador usado para conectar valores de la función y medidas de los conjuntos, y cual el empleado para globalizar los resultados así obtenidos. Un desarrollo paralelo al que hemos llevado a cabo podría hacerse intercambiando los papeles de $\vee$ y $\wedge$; las generalizaciones factibles serían las resultantes de sustituir el máximo por otra conorma cualquiera, conservando el mínimo, lo que daría lugar a las integrales t-conormadas de Suárez. La condición clave sería en este caso la distributividad de cualquier t-conorma respecto del mínimo, que a su vez es la única t-norma idempotente.

En caso de considerar como conjunto Y a $\mathbb{R}_o^+$, la familia de operadores a considerar es muy extensa y compleja, por lo que una exploración exhaustiva de los posibles funcionales válidos, resultantes de reemplazar la suma, sobrepasa los objetivos de este trabajo, y será objeto de estudios posteriores.

- CAPITULO III -

3.0. INTRODUCCION.

La caracterización de una medida difusa mediante probabilidades dada en el capítulo anterior, permite identificarla con n! puntos en el espacio real de n dimensiones. Dado que existe una relación biunívoca entre los puntos correspondientes a dos medidas, dada en términos de la permutación correspondiente, pueden establecerse métricas en el conjunto de las medidas difusas en función de las distancias que se definan entre sus probabilidades asociadas. La posibilidad de establecer así una cuantificación del mayor o menor parecido entre medidas o conjuntos de medidas tiene un extraordinario interés desde el punto de vista de la modelización de la información, ya se entienda ésta como generada objetivamente por una medida imprecisa, o por una opinión subjetiva. Cabría incluso configurar mecanismos de ponderación o consenso entre diferentes medidas.

A partir del establecimiento de una distancia entre medidas difusas, nos planteamos en este capítulo la definición de diversos índices basados en las distancias que una medida difusa tenga con respecto a ciertas medidas destacadas. De esta forma se construyen:

a) Un índice de no aditividad en términos de la distancia a la probabilidad más cercana.

b) Indices de información, mediante las distancias a la medida de ignorancia o, inversamente, a las medidas de Dirac, que contienen respectivamente la mínima y la máxima información posible sobre la determinación de un elemento.

El apartado 3.1 está dedicado al establecimiento de las métricas, tomando en particular una definición derivada de la distancia euclídea, si bien cualquier distancia entre probabilidades sería válida. Se prueba la coherencia de la definición de distancia con la dualidad de medidas y se establece la forma en que podría ser utilizada la caracterización por posibilidades alternativamente a la dada por probabilidades para la definición de distancia.

En el apartado 3.2 se aborda y resuelve el problema de determinar la probabilidad más próxima a una medida difusa dada, tema clave en el apartado siguiente. Esta idea se generaliza a un conjunto de medidas y se prueba la coincidencia de nuestro planteamiento con algunas definiciones sugeridas en el caso de medidas de evidencia; así, para estas medidas la probabilidad más cercana resulta ser , como parece intuitivo, el resultado de la equidistribución de masas de evidencia.

En estas condiciones es posible plantearse la cuantificación del cumplimiento de la propiedad de aditividad por parte de una medida difusa. El apartado 3.3 se dedica a este tema, obteniéndose un índice de no aditividad que es mínimo (nulo) para probabilidades y máximo para las medidas de ignorancia.

Partiendo de la interpretación según la cual una medida difusa reflejaría el grado de certidumbre sobre la pertenencia de un elemento fijo y desconocido a cada subconjunto, el apartado 3.4 se dedica a la definición y estudio de índices que reflejan el nivel de información que cada medida posee en la determinación de dicho elemento desconocido. A este respecto, hay que resaltar que la idea clásica de entropía no es válida en el contexto difuso, debido a que la mínima información no viene dada aquí por la medida uniforme, sino por la de ignorancia. Se establecen índices de certidumbre e incertidumbre, que difieren según se menciona más arriba, en tener como referencia la medida de ignorancia o las degeneradas en un punto; en este último caso, considerar la distancia a la más cercana o incluir las distancias a todas ellas establece un nuevo matiz.

También en el apartado 3.4 se analizan las relaciones de nuestros índices entre si y con una medida de entropía para medidas difusas definida previamente, estudiando estas relaciones especialmente en el caso de probabilidades, donde se plasman muy claramente las diferencias existentes.

Finalmente, en el apartado 3.5 se realiza un estudio completo para un tipo particular de medidas difusas, que resulta especialmente interesante por las características de éstas y que muestra un funcionamiento adecuado de los índices definidos.

3.1. DISTANCIAS ENTRE MEDIDAS DIFUSAS.

Pretendemos en este apartado construir herramientas que nos permitan establecer comparaciones entre diferentes medidas difusas, en el sentido de poder cuantificar la semejanza o diferencia entre ellas; ello permitirá establecer interesantes índices y promedios.

La importancia que tiene el poder evaluar la distancia o el parecido entre dos medidas difusas es, a nuestro juicio, grande. Si pensamos que una medida difusa puede modelizar una información o un estado de opinión de una persona, podríamos entonces cuantificar en qué medida dos informaciones, o dos estados de opinión son parecidos o no.

Puesto que sabemos (proposición 2.6 del capítulo II) que podemos caracterizar una medida difusa g definida sobre un referencial finito $X=\{x_1,x_2,..,x_n\}$ como un conjunto de $n!$ probabilidades, cada una de ellas asociada a una determinada permutación de S_n, el procedimiento que emplearemos será el de definir una distancia entre probabilidades y obtener la distancia entre medidas difusas a través de sus probabilidades asociadas.

Para definir una distancia entre probabilidades, podemos pensar en éstas como puntos de $\mathbb{R}^n$, y emplear cualquier distancia usual de ese espacio. Recordemos cuales son las distancias más empleadas sobre $\mathbb{R}^n$:

Si $x,y \in \mathbb{R}^n$, $x=(x_1,x_2,..,x_n)$, $y=(y_1,y_2,..,y_n)$

a) $$d_q(x,y)=\sqrt[q]{\sum_{i=1}^{n} |x_i-y_i|^q} \qquad q\geq 2$$

es la distancia de Minkowski. Para $q=2$, d_2 es la distancia euclídea.

b) $$d_v(x,y)=\sum_{i=1}^{n} |x_i-y_i|$$

es la distancia del valor absoluto.

c) $$d_m(x,y)= \max_{1\le i\le n} |x_i-y_i|$$

es la distancia del máximo.

Cualquiera de ellas puede usarse para establecer una métrica en el conjunto de probabilidades. Otras muchas definiciones de distancia entre probabilidades pueden encontrarse en la literatura (ver por ejemplo Huber [29] y Burbea y Rao [9]), algunas de ellas definidas en contextos particulares, relacionados con problemas de convergencia y robustez.

Puesto que resulta conveniente, a efectos comparativos, que la distancia tome valores en el intervalo [0,1], tal normalización exige conocer donde se alcanza la distancia máxima (la mínima es obvio que se alcanza cuando las dos probabilidades coincidan) y que valor toma.

Lema 3.1

Para cualesquiera $x,y\in\mathbb{R}^n$ tales que $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = 1$, $x_i, y_i \ge 0$, i=1,..,n, se verifica:

(a) $d_q(x,y)\le \sqrt[q]{2}$ $\forall q\ge 2$.

(b) $d_v(x,y)\le 2$.

(c) $d_m(x,y)\le 1$,

y en todos los casos se alcanzan esas cotas.

Demostración:

(a) $\sum_{i=1}^{n} |x_i-y_i|^q \le \sum_{i=1}^{n} (x_i-y_i)^2$, puesto que $|x_i-y_i|\le 1$ $\forall i$.

$$\sum_{i=1}^{n} (x_i-y_i)^2 = \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} x_i y_i \le \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} y_i^2 \le 1+1=2,$$

puesto que $0\le x_i\le 1 \Rightarrow x_i^2\le x_i \Rightarrow \sum_{i=1}^{n} x_i^2 \le \sum_{i=1}^{n} x_i = 1$.

Luego $d_q(x,y)\leq \sqrt[q]{2}$.

(b) $d_v(x,y)=\sum_{i=1}^{n} |x_i-y_i|\leq \sum_{i=1}^{n} (x_i+y_i)\leq 2.$

(c) Es evidente, pues $|x_i-y_i|\leq 1\ \forall i$.

En cualquier caso, es obvio que si

$$\begin{cases} x_i=0 & \forall i\neq j \\ x_j=1 & \end{cases} \qquad \begin{cases} y_i=0 & \forall i\neq k \\ y_k=1 & \end{cases}$$

se alcanzan las cotas anteriores.#

Definición 3.1

Dadas dos medidas de probabilidad P y P' sobre X, definimos las siguientes distancias entre probabilidades:

(a) $$s_q(P,P')=\sqrt[q]{\frac{1}{2}\sum_{i=1}^{n} |p_i-p'_i|^q} \qquad q\geq 2$$

(b) $$s_v(P,P')=\frac{1}{2}\sum_{i=1}^{n} |p_i-p'_i|$$

(c) $$s_m(P,P')=\max_{1\leq i\leq n} |p_i-p'_i|,$$

donde $p_i=p(x_i)$ $\forall i\in\{1,2,..,n\}$.

Ahora, dadas dos medidas difusas g y g', podemos definir la distancia entre ellas basándonos en las distancias entre sus probabilidades asociadas, para una misma permutación $\sigma\in S_n$. Por coherencia, si hemos elegido una distancia concreta para probabilidades, debemos mantenerla, si bien ello no es indispensable.

Definición 3.2

Sean g y g' medidas difusas definidas sobre X, con probabilidades asociadas P_σ y P'_σ ,$\sigma\in S_n$ respectivamente. Definimos las siguientes distancias entre g y g':

(a) $S_q(g,g')=\sqrt[q]{1/n!\sum_{\sigma\in S_n} s_q(P_\sigma,P'_\sigma)^q}=\sqrt[q]{1/(2n!)\sum_{\sigma\in S_n}\sum_{i=1}^{n}|p_{\sigma i}-p'_{\sigma i}|^q}$

con $q\geq 2$.

(b) $S_v(g,g')=1/n!\sum_{\sigma\in S_n} s_v(P_\sigma,P'_\sigma)=1/(2n!)\sum_{\sigma\in S_n}\sum_{i=1}^{n}|p_{\sigma i}-p'_{\sigma i}|$.

(c) $S_m(g,g')=\max_{\sigma\in S_n}\{s_m(P_\sigma,P'_\sigma)\}=\max_{\sigma\in S_n}\max_{1\leq i\leq n}|p_{\sigma i}-p'_{\sigma i}|$,

donde $p_{\sigma i}=p_\sigma(x_i)$, $p'_{\sigma i}=p'_\sigma(x_i)$ $\forall\sigma\in S_n$ $\forall i\in\{1,2,..,n\}$.

Algunas expresiones aparecen divididas por n! con objeto de que las distancias varíen entre cero y uno. Dado que las probabilidades son casos particulares de medidas difusas, las distancias de la definición 3.2 son extensiones de s_q, s_v y s_m.

Probaremos que efectivamente, S_q, S_v y S_m poseen las propiedades de una distancia.

Proposición 3.1

S_m, S_v y S_q, $q\geq 2$, establecen métricas en el conjunto $\mathfrak{M}$ de las medidas difusas definidas sobre X, y en cualquiera de los casos, su máximo valor posible es la unidad.

Demostración:

Notemos por S_o a cualquiera de las hipotéticas distancias. Si $g=g'$ entonces es obvio que $S_o(g,g')=0$; y si $S_o(g,g')=0$ entonces $s_o(P_\sigma,P'_\sigma)=0$ $\forall\sigma\in S_n$, y por tanto $P_\sigma=P'_\sigma$ $\forall\sigma\in S_n$. Como sabemos que las probabilidades (con su orden) determinan unívocamente la medida, deducimos que $g=g'$.

La simetría se verifica evidentemente.

Para probar la desigualdad triangular basta con tener en cuenta que una medida difusa puede considerarse como un punto del espacio $\mathbb{R}^m$, con $m=n\cdot n!$

$$g=(p_{\sigma_1 1},..,p_{\sigma_1 n},p_{\sigma_2 1},..,p_{\sigma_2 n},....,p_{\sigma_{n!}1},..,p_{\sigma_{n!}n}),$$

notando $\sigma_1,\sigma_2,..,\sigma_{n!}$ a cada una de las n! permutaciones de S_n, y que si prescindimos de las constantes de normalización (que no son relevantes a estos efectos), tanto S_q como S_v y S_m son respectivamente las distancias de Minkowski, del valor absoluto y del máximo en el espacio $\mathbb{R}^m$.

La máxima distancia, en cualquier caso, se alcanzará para aquellas medidas difusas cuyas probabilidades asociadas sean todas degeneradas en un punto (medidas de Dirac), de modo que para una misma permutación σ, P_σ y P'_σ se focalicen en puntos distintos.

Esta máxima distancia se alcanza, por ejemplo, si tanto g como g' son distintas probabilidades degeneradas, pues es sabido (proposición 2.8 del Capítulo II) que las probabilidades asociadas a una medida difusa que es ella misma una probabilidad coinciden con ésta última. También se alcanza la distancia máxima si g y g' son las medidas de ignorancia superior e inferior Pl_o y bel_o.#

Otra interesante propiedad de las distancias entre medidas difusas es:

Proposición 3.2

Para cualquiera de las distancias S_q, S_v y S_m se verifica

$$S_\circ(g,g')=S_\circ(g^*,g'^*)$$

donde g^* y g'^* son las medidas duales de g y g' respectivamente.

Demostración:

Puesto que una medida difusa y su dual tienen asociadas las mismas n! probabilidades, aunque con órdenes duales (proposición 2.7 del Capítulo II), en las fórmulas que definen las distancias aparecen las mismas probabilidades para g y g' que para g^* y g'^*.#

Por tanto una medida difusa dista de otra en la misma cuantía en que distan sus medidas duales, lo cual resulta lógico desde el punto de vista de que medidas duales codifican la misma información de diferente modo.

Teniendo en cuenta que las probabilidades son medidas autoduales, se verifica:

Corolario 3.1

Para cualquiera de las distancias S_q, S_v y S_m

$$S_o(g,P)=S_o(g^*,P)$$

para cualquier medida difusa g y cualquier probabilidad P.

Una medida difusa y su dual distan lo mismo de cualquier probabilidad, y también de cualquier medida autodual.

Puesto que tenemos diversas alternativas para la distancia entre medidas difusas, se hacen necesarios algunos comentarios sobre nuestra preferencia por alguna de ellas.

Considerando que trabajamos con distribuciones de probabilidad, que la aditividad o su pérdida en las medidas difusas será uno de los aspectos que analizaremos, y que es deseable que todas las probabilidades asociadas influyan en la evaluación de las distancias, pensamos que puede descartarse S_m, en la que solo la mayor distancia entre probabilidades es relevante, además de estar basada en una filosofía no aditiva.

La distancia S_v plantea la dificultad operativa, bien conocida, de manejar el operador valor absoluto, además de poseer un poder de discriminación menor que S_q.

Finalmente, utilizaremos S_2 de entre todas las del tipo S_q por simplicidad, teniendo en cuenta que por las razones referidas con respecto al valor absoluto, las potencias de exponente par son más operativas. No obstante las demás distancias son válidas y puede resultar interesante emplearlas en algún caso.

Antes de continuar, conviene hacer una reflexión. Hemos basado nuestras definiciones de distancias entre medidas en el hecho de que una medida difusa es un conjunto ordenado de probabilidades. Otra perspectiva muy diferente que podemos adoptar es basarse en que también una medida difusa cualquiera es un conjunto ordenado de n! posibilidades. Entonces podemos definir distancias entre posibilidades y emplear éstas para definir la distancia entre medidas difusas.

Nuevamente, si identificamos una posibilidad con un punto de $\mathbb{R}^n$, cualquier distancia definida sobre ese espacio es válida, pero parece que por las características de las medidas de posibilidad, lo mejor es emplear la distancia del máximo (que hemos rechazado en el enfoque probabilista).

<u>Definición 3.3</u>

Si Π y Π' son medidas de posibilidad sobre X, definimos la distancia entre Π y Π' por

$$d(\Pi,\Pi')=\max_{1\leq i\leq n}|\pi_i-\pi'_i|$$

donde $\pi_i=\pi(x_i)$ y $\pi'_i=\pi'(x_i)$ $\forall i\in\{1,2,..,n\}$.

Como antes, la distancia entre posibilidades nos permite definir una distancia entre medidas difusas.

<u>Definición 3.4</u>

Si g y g' son medidas difusas sobre X, con posibilidades asociadas Π_σ y Π'_σ , $\sigma\in S_n$ respectivamente, definimos

$$D(g,g')=\max_{\sigma\in S_n} d(\Pi_\sigma\ ,\Pi'_\sigma)=\max_{\sigma\in S_n}\ \max_{1\leq i\leq n}|\pi_{\sigma i}-\pi'_{\sigma i}|$$

Proposición 3.3

D es una distancia en el conjunto $\mathfrak{M}$ de las medidas difusas definidas sobre X.

La demostración es análoga a la de la proposición 3.1.

Una dificultad que aparece aquí, pero no en el enfoque anterior se debe a que, para una posibilidad, las n! posibilidades asociadas no coinciden todas con la dada. Entonces podría ocurrir que D(Π,Π') no fuese una extensión de d(Π,Π'). Afortunadamente no es así.

Proposición 3.4

Si Π y Π' son posibilidades, entonces

$$D(\Pi,\Pi')=d(\Pi,\Pi')$$

Demostración:

Como ya sabemos, para la permutación $\sigma=(\sigma(1),\sigma(2),..,\sigma(n))$

$$\pi_{\sigma\sigma(i)}=\Pi(\{x_{\sigma(1)},..,x_{\sigma(i)}\})=\bigvee_{j=1}^{i}\Pi(\{x_{\sigma(j)}\})=\bigvee_{j=1}^{i}\pi_{\sigma(j)}.$$

Como siempre existe una permutación $\sigma_i\in S_n$ tal que $\sigma_i(1)=i$, y por tanto

$$\pi_{\sigma_i\sigma_i(1)}=\pi_{\sigma_i(1)}=\pi_i$$

entonces

$$D(\Pi,\Pi')\geq\max_{1\leq i\leq n}|\pi_i-\pi'_i|=d(\Pi,\Pi').$$

Por otra parte, puesto que para cualquier permutación σ

$$|\pi_{\sigma\sigma(i)}-\pi'_{\sigma\sigma(i)}|=|\bigvee_{j=1}^{i}\pi_{\sigma(j)}-\bigvee_{j=1}^{i}\pi'_{\sigma(j)}|\leq\bigvee_{j=1}^{i}|\pi_{\sigma(j)}-\pi'_{\sigma(j)}|$$

entonces

$$d(\Pi_\sigma,\Pi'_\sigma)=\bigvee_{i=1}^{n}|\pi_{\sigma\sigma(i)}-\pi'_{\sigma\sigma(i)}|\leq\bigvee_{j=1}^{n}|\pi_{\sigma(j)}-\pi'_{\sigma(j)}|=d(\Pi,\Pi').$$

Luego

$$D(\Pi,\Pi')\leq d(\Pi,\Pi')$$

y se verifica la igualdad.#

De todas formas en el desarrollo posterior solo emplearemos la vía de las probabilidades con la distancia S_2, a la que por simplificar notaremos solo por S. Un estudio paralelo basado en posibilidades y la distancia D puede ser objeto de trabajos futuros.

3.2. PROBABILIDAD MAS CERCANA A UNA MEDIDA DIFUSA.

Puesto que disponemos ya de una forma de medir el parecido entre medidas difusas (por medio de la distancia S), podemos plantearnos hallar la probabilidad más próxima a una medida difusa dada. Resolver este problema puede resultar interesante desde dos puntos de vista:

En primer lugar nos permite aproximar de la mejor manera posible una medida difusa por una probabilidad. Si tal aproximación es suficientemente buena, se puede sustituir la medida difusa por su probabilidad más próxima, en caso de que sea necesario aplicar las herramientas de la teoría de la probabilidad, más potentes que las de la teoría de medidas difusas.

En segundo lugar, si evaluamos la distancia entre una medida y la medida de probabilidad más próxima, disponemos de un valor que cuantifica la desviación de esa medida difusa respecto de un comportamiento aditivo, es decir, podemos construir un índice de la falta de aditividad de una medida difusa, que es una de sus características fundamentales.

En algunos problemas (y cuando los valores de la medida difusa representan grados de creencia) se habla de medidas optimistas y pesimistas, situando la neutralidad en las probabilidades. El índice que definiremos también puede dar idea del grado de optimismo o pesimismo de una medida difusa.

El problema de cuantificar la falta de aditividad de una medida difusa es análogo al que se plantea la teoría de subconjuntos difusos de medir la difusividad de los mismos, es decir, como de difusos son, entendiendo ésto como lejanía de un comportamiento nítido (entropía difusa).

Así pues, dada una medida difusa g, el problema que queremos resolver es el de encontrar la probabilidad P_g que minimiza la distancia de g a una probabilidad, es decir, encontrar P_g tal que

$$S(g,P_g)= \min_{P\in PR} S(g,P).$$

Proposición 3.5

Dada una medida difusa g, con probabilidades asociadas P_σ, $\sigma\in S_n$, la probabilidad más próxima a dicha medida, P_g, es el promedio de las probabilidades asociadas a g

$$P_g(x)= \frac{1}{n!}\sum_{\sigma\in S_n} P_\sigma(x) \quad \forall x\in X.$$

Demostración:

Teniendo en cuenta como se define la distancia S, debemos resolver

$$\min_{P\in PR} \sqrt{1/(2n!)\sum_{\sigma}\sum_{i=1}^{n}(p_{\sigma i}-p_i)^2}\ .$$

Como la raiz cuadrada y las constantes no influyen para minimizar, el problema se reduce a obtener:

$$\min_{P\in PR}\sum_{\sigma}\sum_{i=1}^{n}(p_{\sigma i}-p_i)^2= \min_{P\in PR} f(p_1,p_2,..,p_n).$$

Si derivamos parcialmente

$$\frac{\partial f}{\partial p_i} = -2\sum_{\sigma}(p_{\sigma i}-p_i)=0 \Longleftrightarrow \sum_{\sigma} p_{\sigma i}=n!p_i, \text{ o sea}$$

$$p_i=1/n!\sum_{\sigma} p_{\sigma i} \quad \forall i\in\{1,2,..,n\},$$

es el único posible mínimo. Para asegurarnos, construimos el hessiano H con las derivadas parciales segundas

$$\frac{\partial^2 f}{\partial p_j \partial p_i}=0;\ \frac{\partial^2 f}{\partial^2 p_i} = 2n! \quad \forall i,j\in\{1,2,..,n\}$$

$$H=\begin{bmatrix} 2n! & 0.\ldots.0 \\ 0 & 2n!\ldots0 \\ \vdots & \vdots \\ 0 & 0\ldots2n! \end{bmatrix}$$

Como H es una matriz obviamente definida positiva,

$$P_g(x)=1/n!\sum_{\sigma} P_{\sigma}(x)$$

es la probabilidad más próxima a la medida difusa g.#

La probabilidad más cercana a una medida difusa es , pues, el centro de gravedad de sus probabilidades asociadas.

Estudiemos los casos particulares de evidencias y posibilidades.

Proposición 3.6

Si g es una medida de creencia o plausibilidad con asignación básica de probabilidad m, la probabilidad más próxima a g es la resultante de equidistribuir la masa de evidencia de cada subconjunto del referencial entre sus elementos, es decir

$$P_g(x)=\sum_{x\in A}\frac{m(A)}{|A|}$$

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

Demostración:

Supondremos que g es una creencia. En caso de que g fuese una plausibilidad obtendríamos el resultado por dualidad.

Puesto que g puede expresarse como

$$g(B)=\sum_{A\subseteq B} m(A)$$

y dada una permutación σ, si k_i es tal que $\sigma(k_i)=i$ entonces

$$p_{\sigma}(x_i)=g(\{x_{\sigma(1)},\ldots,x_{\sigma(k_i)}\})-g(\{x_{\sigma(1)},\ldots,x_{\sigma(k_i-1)}\})=$$

$$=\sum_{\substack{A\subseteq\{x_{\sigma(k)}/k\leq k_i\} \\ x_i\in A}} m(A)$$

Luego

$$P_g(x_i)=1/n!\sum_{\sigma} P_\sigma(x_i)=1/n!\sum_{\sigma}\sum_{\substack{A\subseteq\{x_{\sigma(k)}/k\leq k_i\}\\ x_i\in A}} m(A).$$

¿ Cuántas veces aparece m(A) en esa expresión, para cada $x_i\in A$?. Tantas como permutaciones haya donde los $|A|$ elementos de A ocupen los $|A|$ primeros lugares en la ordenación generada por la permutación. Como hay tantas permutaciones en las que x_i es el primero de los puntos de A que aparecen como aquellas en que sea otro punto de A el que aparece primero, entonces el número total es $n!/|A|$. Luego

$$P_g(x_i)=\frac{1}{n!}\,\frac{n!}{|A|}\sum_{x_i\in A} m(A)=\sum_{x_i\in A}\frac{m(A)}{|A|}.\#$$

Esta probabilidad coincide con la que, de una forma un tanto empírica, proponen Dubois y Prade [19] como probabilidad asociada a una evidencia, esto es, como la forma más razonable de "probabilizar" a ésta.

Nuestra formulación provee además un mecanismo de interpretación de los funcionales definibles sobre una evidencia. En efecto, dada una evidencia m, como se ha recogido en el Capítulo I, pueden definirse dos funcionales, denominados integral superior e integral inferior, de la forma:

$$I^*(h/m)=\sum_{A\subseteq X}(m(A)\max_{x_i\in A} h(x_i))$$

$$I_*(h/m)=\sum_{A\subseteq X}(m(A)\min_{x_i\in A} h(x_i))$$

que son un caso particular de la esperanza monótona, correspondientes a las medidas de plausibilidad y creencia respectivamente. Pues bien, la esperanza matemática con respecto a P_g proporciona el funcional

$$E_{P_g}(h)=\sum_{i=1}^{n} h(x_i)P_g(x_i)=\sum_{i=1}^{n} h(x_i)\Big(\sum_{x_i\in A}\frac{m(A)}{|A|}\Big)=$$

$$= \sum_{A\subseteq X} m(A)\left(\frac{1}{|A|}\sum_{x_i\in A} h(x_i)\right),$$

que puede considerarse intermedio entre las integrales superior e inferior, e interpretarse como una valoración "neutra" entre el "optimismo" de la primera y el "pesimismo" de la segunda.

Si la probabilidad P_g más próxima a g resulta de especial interés en el caso en que g proviene de una evidencia, otro tanto ocurre cuando g es una medida de posibilidad. En efecto, en el trabajo anteriormente citado, Dubois y Prade desarrollan un mecanismo de asignación de una probabilidad P a una medida de posibilidad Π dada, y reciprocamente; su resultado, dado por:

$$p_i = \frac{1}{i}\pi_i - \sum_{j=i+1}^{n} \frac{1}{j(j-1)}\pi_j \qquad i=1,..,n$$

con $\pi_1 \geq \pi_2 \geq ... \geq \pi_n$

también coincide con la probabilidad P_π más próxima a Π, puesto que Dubois y Prade probaron que P coincide con la probabilidad que se asociaría a Π por el hecho de ser ésta una medida de plausibilidad.

Volviendo al caso general, la asignación de una probabilidad a una medida difusa dada mediante el mecanismo de la mínima distancia verifica una interesante propiedad de continuidad, recogida en la siguiente proposición.

Proposición 3.7

Dotando al conjunto $\mathfrak{M}$ de medidas difusas definidas sobre X de la topología inducida por la distancia S, y al conjunto PR de las probabilidades definidas sobre X de la topología inducida por la distancia s, la aplicación que a cada medida difusa le hace corresponder la probabilidad más próxima a ella

$$\mathfrak{M} \longrightarrow PR$$

$$g \longrightarrow P_g$$

es continua.

Demostración:

Podemos identificar $\mathfrak{M}$ y PR con subconjuntos de $\mathbb{R}^{nn!}$ y $\mathbb{R}^{n}$ respectivamente. Además S y s son, salvo constantes, las distancias euclídeas en estos espacios. En esas condiciones, la aplicación que a

$$g=(p_{\sigma_1 1},\ldots,p_{\sigma_1 n},\ldots,p_{\sigma_2 1},\ldots,p_{\sigma_2 n},\ldots\ldots,p_{\sigma_{n!} 1},\ldots,p_{\sigma_{n!} n})$$

le hace corresponder

$$P_g=(1/n!\sum_{\sigma\in S_n} p_{\sigma 1},1/n!\sum_{\sigma\in S_n} p_{\sigma 2},\ldots,1/n!\sum_{\sigma\in S_n} p_{\sigma n}),$$

es obviamente continua.#

Por tanto medidas muy cercanas llevan asociadas probabilidades más próximas también muy cercanas.

Veamos algunos ejemplos sencillos del cálculo de la probabilidad más próxima a una medida difusa y de la distancia mínima entre ellas.

Ejemplo 3.1

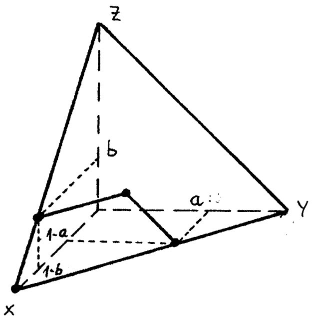
Consideremos el referencial $X=\{x_1,x_2,x_3\}$, y una medida de posibilidad Π definida por

$\pi(x_1)=1$, $\pi(x_2)=a$, $\pi(x_3)=b$, siendo $a\geq b$.

Las 6 probabilidades asociadas son

σ	$p_{\sigma 1}$	$p_{\sigma 2}$	$p_{\sigma 3}$
(1,2,3)	1	0	0
(1,3,2)	1	0	0
(2,1,3)	1-a	a	0
(2,3,1)	1-a	a	0
(3,1,2)	1-b	0	b
(3,2,1)	1-a	a-b	b

Cuando, como en este caso, el número n de elementos del referencial es tres, podemos representar la medida difusa representando sus probabilidades asociadas como puntos de $\mathbb{R}^3$ pertenecientes al hiperplano x+y+z=1 ($x=p_{\sigma 1}$, $y=p_{\sigma 2}$, $z=p_{\sigma 3}$)



Representación gráfica de Π.

Esta representación nos permitirá visualizar e interpretar geométricamente muchos resultados y conceptos. Desgraciadamente para $n\geq 4$ la representación gráfica ya no es posible, aunque si la interpretación geométrica.

La probabilidad más cercana a Π es P_π definida por

$$p_\pi(x_1)=1-\frac{1}{2}a-\frac{1}{6}b$$

$$p_\pi(x_2)=\frac{1}{2}a-\frac{1}{6}b$$

$$p_\pi(x_3)=\frac{1}{3}b$$

La mínima distancia, $S(\Pi,P_\pi)$ es

$$S(\Pi,P_\pi)=\sqrt{\frac{a^2+b^2}{4}-\frac{ab}{6}}$$

Ejemplo 3.2

Sea de nuevo $X=\{x_1,x_2,x_3\}$ y m la a.b.p. asociada a una evidencia. Si notamos $m_i=m(\{x_i\})$, $m_{ij}=m(\{x_i,x_j\})$ y $m_{123}=m(X)$, las probabilidades asociadas a la creencia

$$bel(A)=\sum_{B\subseteq A} m(B)$$

son

σ	$P_{\sigma 1}$	$P_{\sigma 2}$	$P_{\sigma 3}$
(1,2,3)	m_1	m_2+m_{12}	$1-m_1-m_2-m_{12}$
(1,3,2)	m_1	$1-m_1-m_3-m_{13}$	m_3+m_{13}
(2,1,3)	m_1+m_{12}	m_2	$1-m_1-m_2-m_{12}$
(2,3,1)	$1-m_2-m_3-m_{23}$	m_2	m_3+m_{23}
(3,1,2)	m_1+m_{13}	$1-m_1-m_3-m_{13}$	m_3
(3,2,1)	$1-m_2-m_3-m_{23}$	m_2+m_{23}	m_3

La probabilidad más próxima es P_m definida por

$$p_m(x_1)=m_1+\frac{1}{2}(m_{12}+m_{13})+\frac{1}{3}m_{123}$$

$$p_m(x_2)=m_2+\frac{1}{2}(m_{12}+m_{23})+\frac{1}{3}m_{123}$$

$$p_m(x_3)=m_3+\frac{1}{2}(m_{13}+m_{23})+\frac{1}{3}m_{123}$$

La mínima distancia es

$$S(m,P_m)=\sqrt{\frac{1}{4}(m_{12}^2+m_{13}^2+m_{23}^2)+\frac{1}{3}m_{123}^2+\frac{1}{3}m_{123}(m_{12}+m_{13}+m_{23})+\frac{1}{12}(m_{12}m_{13}+m_{12}m_{23}+m_{13}m_{23})}$$

Una vez resuelto el problema de encontrar la probabilidad más próxima a una medida difusa dada, inmediatamente podemos pensar en hallar la probabilidad más cercana a un conjunto finito de medidas difusas, o más generalmente, la medida difusa g_G más cercana a un conjunto $G=\{g_1,g_2,\ldots,g_m\}$ de medidas difusas. Esta medida podría considerarse en cierto modo como una medida de compromiso entre las iniciales, una medida promedio. Así, tendríamos un acuerdo entre opiniones o informaciones, y la distancia entre la medida hallada y las iniciales daría idea del grado de conflicto entre esas opiniones o informaciones.

Naturalmente primero hay que escoger la manera de medir la distancia de una medida difusa a un conjunto de medidas.

Nosotros hemos elegido, para mantener la coherencia con la elección de S, la siguiente:

Definición 3.5

Dadas medidas difusas $g_0, g_1, \dots, g_m$, definimos la distancia de g_0 al conjunto $G=\{g_1, g_2, \dots, g_m\}$ mediante

$$S_c(g_0, G) = \sqrt{\sum_{i=1}^{m} S^2(g_0, g_i)}$$

El término distancia se usa aquí en un sentido no riguroso, dado que no es una verdadera distancia, aunque esta nomenclatura es habitualmente utilizada por abuso de lenguaje al establecer valoraciones de la discrepancia entre puntos y conjuntos.

Proposición 3.8

Dado el conjunto de medidas difusas $G=\{g_1, g_2, \dots, g_m\}$, la probabilidad P_G más próxima a G es

$$p_G(x) = \frac{1}{m} \sum_{i=1}^{m} p_{g_i}(x) = \frac{1}{m} \frac{1}{n!} \sum_{i=1}^{m} \sum_{\sigma} p^i_{\sigma}(x) \quad \forall x \in X$$

donde P^i_{σ} es la probabilidad asociada a la medida g_i para la permutación $\sigma \in S_n$, $i=1, \dots, m$.

Demostración:

Hay que minimizar la función

$$f_G(P) = f_G(p_1, \dots, p_n) = \sum_{i=1}^{m} \sum_{\sigma} \sum_{j=1}^{n} (p^i_{\sigma j} - p_j)^2$$

Si derivamos parcialmente

$$\frac{\partial f_G}{\partial p_j} = -2 \sum_{i=1}^{m} \sum_{\sigma} (p^i_{\sigma j} - p_j) = 0 \iff p_j = \frac{1}{m} \frac{1}{n!} \sum_{i=1}^{m} \sum_{\sigma} p^i_{\sigma j}$$

Si calculamos el hessiano H_G obtenemos

$$H_G = 2mn!\, I_n$$

donde I_n es la matriz unidad de orden n, de modo que H_G es una matriz definida positiva.#

Así pues, la probabilidad más próxima a un conjunto de medidas difusas se obtiene promediando las mn! probabilidades asociadas a esas medidas (o promediando las m probabilidades más próximas a cada medida g_i).

Proposición 3.9

Dado un conjunto de medidas difusas $G=\{g_1,g_2,...,g_m\}$, la medida difusa más próxima a G es

$$g_G(A)=\frac{1}{m}\sum_{i=1}^{m} g_i(A) \quad \forall A\subseteq X$$

es decir, la combinación convexa uniforme de las medidas que están en G.

Demostración:

Notando como es habitual P_σ y P^i_σ a las probabilidades asociadas a g y g_i respectivamente, debemos minimizar

$$f(g)=\sum_{i=1}^{m}\sum_{\sigma}\sum_{j=1}^{n}(p^i_{\sigma j}-p_{\sigma j})^2.$$

$$\frac{\partial f}{\partial p_{\sigma j}} = -2\sum_{i=1}^{m}(p^i_{\sigma j}-p_{\sigma j})=0 \Leftrightarrow P_{\sigma j}=\frac{1}{m}\sum_{i=1}^{m}P^i_{\sigma j}$$

El hessiano en este caso es

$$H=2mI_{nn!}$$

que es definido positivo. El mínimo se alcanza pues para la medida difusa g_G con probabilidades asociadas

$$P_\sigma=\frac{1}{m}\sum_{i=1}^{m}P^i_\sigma$$

y puede comprobarse facilmente que esas probabilidades corresponden a la medida difusa

$$g_G=\frac{1}{m}\sum_{i=1}^{m}g_i. \quad \#$$

Como para el caso de la probabilidad más próxima a una medida difusa dada, la medida difusa más próxima a un conjunto de medidas posee interesantes propiedades en relación con la esperanza monótona.

Proposición 3.10

Si g_a es la medida difusa más próxima al conjunto de medidas difusas $G=\{g_1,g_2,..,g_m\}$, para cualquier función $h:X \longrightarrow \mathbb{R}_o^+$ se verifica

$$E_{g_a}(h)=\frac{1}{m}\sum_{i=1}^{m} E_{g_i}(h)$$

es decir, la esperanza monótona respecto a g_a es el promedio de las esperanzas monótonas respecto a las medidas difusas de G.

Demostración:

Puesto que g_a es una combinación convexa de las medidas g_i, i=1,..,m, una propiedad de la esperanza monótona estudiada en el Capítulo I garantiza el resultado.#

Diversas familias de medidas difusas, como las ordenadas, las aditivo-coherentes, las que acotan una probabilidad, las representables, las capacidades de orden dos y las evidencias, son cerradas frente a combinaciones convexas (ver por ejemplo Lamata [37]). Por tanto podemos afirmar que la medida difusa más próxima a un conjunto de medidas difusas de una de esas familias también es de la misma familia.

Las familias de evidencias consonantes y de evidencias de tipo crisp no son convexas, pero su clausura convexa es la familia de evidencias. Así, la medida difusa más próxima a un conjunto de evidencias consonantes o de medidas de tipo crisp es una evidencia.

3.3. INDICE DE NO ADITIVIDAD DE MEDIDAS DIFUSAS.

Puesto que el problema de encontrar la probabilidad P_g más próxima a una medida difusa dada g tiene solución, y podemos calcular la distancia de la medida a dicha probabilidad, tal distancia proporciona una cuantificación de lo que se aparta g de una filosofía probabilista, de un comportamiento aditivo.

Una de las formas de evaluar cómo de difuso es un determinado subconjunto difuso de un referencial es medir de alguna manera su lejanía de un subconjunto crisp apropiado. En nuestro caso el método es semejante; para ver cuan difusa es una medida difusa (entendiendo que difuso aquí significa una ausencia de aditividad), podemos medir la distancia a la probabilidad más cercana.

Puesto que todo subconjunto difuso normalizado induce una medida de posibilidad y recíprocamente, hay que aclarar que la difusividad como subconjunto y la falta de aditividad como medida no son el mismo concepto (y por eso nuestro índice de no aditividad, restringido a posibilidades no tiene que verificar ninguna de las axiómaticas existentes para los índices de difusividad (ver por ejemplo De Luca y Termini [14] y [15])). Quizás la mínima distancia de una posibilidad a medidas difusas de tipo crisp pudiera interpretarse como un índice de difusividad, aunque no entramos más en ese tema.

Partimos de una medida difusa g, que tiene por probabilidades asociadas P_σ $\sigma \in S_n$. La probabilidad más próxima es

$$P_g = \frac{1}{n!} \sum_{\sigma} P_\sigma$$

y la distancia de g a P_g es

$$S(g,P_g)=\sqrt{1/(2n!)\sum_{\sigma}\sum_{i=1}^{n}(p_{\sigma i}-p_{gi})^2}$$

que también puede escribirse

$$S(g,P_g)=\sqrt{1/(2n!^3)\sum_{\sigma\in S_n}\sum_{i=1}^{n}\Big(\sum_{\tau\in S_n}(p_{\sigma i}-p_{\tau i})\Big)^2}$$

Como deseamos que nuestro índice de no aditividad varíe entre cero y uno, veamos que valores extremos puede tomar $S(g,P_g)$.

Lema 3.2

$S(g,P_g)$ toma valor mínimo igual a cero si y solo si g es una probabilidad, y cuando g es una medida de ignorancia alcanza valor máximo igual a $\sqrt{\dfrac{n-1}{2n}}$.

Demostración:

La primera afirmación es evidente, puesto que S es una distancia.

De la expresión de la distancia de g a P_g puede deducirse que ésta será máxima para aquella medida que tenga asociadas probabilidades degeneradas, todas las posibles y en igual número (como le ocurre a las medidas de ignorancia), en cuyo caso

$$S(g,P_g)=\sqrt{1/(2n!^3)\Big[(n-1)!\ (n!-(n-1)!)^2+(n-1)!^2(n!-(n-1)!)\Big]n}=$$

$$=\sqrt{\frac{n-1}{2n}}\ .\#$$

Definición 3.6

Se define el índice de no aditividad de una medida difusa g, con probabilidades asociadas P_σ, $\sigma\in S_n$ como

$$A(g)=\sqrt{\frac{2n}{n-1}}\ S(g,P_g)=\sqrt{\frac{n}{n-1}\frac{1}{n!^3}\sum_{i=1}^{n}\sum_{\sigma\in S_n}\Big(\sum_{\tau\in S_n}(p_{\sigma i}-p_{\tau i})\Big)^2}$$

Proposición 3.11

El índice $A(\bullet)$ verifica

(a) $A(g)\in[0,1]\ \forall g\in\mathfrak{M}$.

(b) $A(g)=0$ si y solo si g es una probabilidad.

(c) $A(g)=1$ si g es una medida de ignorancia.

(d) Si g^* es la medida difusa dual de g, entonces $A(g)=A(g^*)$.

Demostración:

(a), (b) y (c) son evidentes por el lema anterior; (d) se deduce de que g y g^* tienen las mismas probabilidades asociadas.#

Otra interesante propiedad de $A(\bullet)$ es la siguiente:

Proposición 3.12

La aplicación

$$A:\mathfrak{M}\longrightarrow[0,1]$$
$$g\longrightarrow A(g)$$

es continua respecto de la topología inducida en $\mathfrak{M}$ por la distancia S (y la topología usual en $[0,1]$).

Demostración:

Si notamos $k=\sqrt{\dfrac{2n}{n-1}}$, podemos decir que

$$A(g)=kS(g,P_g).$$

Como S es una distancia en $\mathfrak{M}$

$$S(g,P_g)\le S(g,g')+S(g',P_g)\le S(g,g')+S(g',P_{g'})+S(P_{g'},P_g)$$

y por tanto

$$A(g)-A(g')=k(S(g,P_g)-S(g',P_{g'}))\le k(S(g,g')+S(P_g\ ,P_{g'}))\ \forall g,g'\in\mathfrak{M}.$$

Razonando analogamente obtenemos

$$A(g')-A(g)\le k(S(g,g')+S(P_g\ ,P_{g'}))$$

y de este modo

$$|A(g)-A(g')|\le k(S(g,g')+S(P_g\ ,P_{g'}))\ \forall g,g'\in\mathfrak{M}.$$

Puesto que S restringido a probabilidades coincide con s, y la aplicación que asocia P_g a g es continua

$$\exists\ \delta'>0 \text{ tal que } S(g,g')<\delta \Rightarrow S(P_g, P_{g'})<\frac{\varepsilon}{2k}$$

Si elegimos $\delta=\min(\delta', \frac{\varepsilon}{2k})$, entonces, si $S(g,g')<\delta$

$$|A(g)-A(g')|\leq k(S(g,g')+S(P_g, P_{g'}))<k(\delta+\frac{\varepsilon}{2k})\leq k(\frac{\varepsilon}{2k} + \frac{\varepsilon}{2k})=\varepsilon,$$

para cualquier $\varepsilon>0$, y por tanto $A(\bullet)$ es continua.#

Así pues, medidas difusas próximas entre sí muestran una falta de aditividad parecida, como era de esperar.

Antes de proseguir queremos hacer algunos comentarios sobre la relación entre aditividad y autodualidad en medidas difusas. Es conocido que si una medida es aditiva (probabilidad) es autodual; el recíproco no es cierto: hay medidas autoduales que no son aditivas. La autodualidad es desde nuestro punto de vista, y hablando libremente, una casualidad en las medidas de probabilidad: una medida difusa por el hecho de ser autodual no tiene que ser ni más ni menos aditiva que otra. Se puede dar la situación extrema de una medida autodual con índice de no aditividad máximo, como es el caso, en $X=\{x_1,x_2,x_3\}$ de la medida difusa g definida por

$$g(\{x_1\})=g(\{x_2\})=g(\{x_3\})=0$$
$$g(\{x_1,x_2\})=g(\{x_1,x_3\})=g(\{x_2,x_3\})=1$$

cuyas probabilidades asociadas son las mismas que las de las medidas de ignorancia, como se puede fácilmente comprobar.

<u>Nota:</u> Sin embargo, para algunas familias particulares de medidas difusas (como las aditivo-coherentes y las que acotan una probabilidad) la aditividad y la autodualidad son equivalentes. En tales casos, el valor $S(g,g^*)$, que podría considerarse un índice adecuado de autodualidad, sería también de no aditividad.

No es fácil obtener expresiones concretas de $A(g)$ para diferentes tipos de medidas difusas, debido a la dificultad de calcular de forma teórica las n! probabilidades asociadas. Para n=3 si hemos visto (ejemplos 3.1 y 3.2) expresiones generales de $S(g,P_g)$ para posibilidades y evidencias, que normalizadas (multiplicando por $\sqrt{3}$) proporcionan el índice A(g); con esta información empírica, A(g) parece tener un comportamiento adecuado como índice de no aditividad. Para evidencias de tipo crisp si es posible obtener una expresión general.

Proposición 3.13

Sea g una evidencia de tipo crisp focalizada en el conjunto B, de modo que $|B|=p$ (y $|X|=n$). Entonces

$$A(g)=\sqrt{\frac{n(p-1)}{(n-1)p}}$$

Demostración:

Sea g una medida crisp inferior; en caso de considerar la medida superior, el resultado se obtiene por dualidad.

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

Las probabilidades asociadas a g son probabilidades degeneradas en cada uno de los elementos de B, n!/p de cada tipo. Por tanto

$$A(g)=\sqrt{\frac{n}{n-1}\frac{1}{n!}\left[\frac{n!}{p}(1-\frac{1}{p})^2+(n!-\frac{n!}{p})(\frac{1}{p})^2\right]p} =$$

$$=\sqrt{\frac{n}{n-1}\frac{1}{n!}n!\,\frac{p-1}{p}} = \sqrt{\frac{n(p-1)}{p(n-1)}} ,$$

ya que la probabilidad más próxima a g es

$$P_g(x)=\begin{cases} 1/p & \text{si } x\in B \\ 0 & \text{si } x\notin B \end{cases} \qquad \#$$

El índice de no aditividad de una evidencia crisp solo depende del número de elementos del referencial y del número de elementos del foco.

Corolario 3.2

Sean g_B y g_C dos evidencias crisp focalizadas en los subconjuntos B y C de X, respectivamente.

Si $|B| \leq |C|$ entonces $A(g_B) \leq A(g_C)$.

Este resultado es muy lógico: cuanto mayor es el conjunto donde se concentra la evidencia, menos aditiva resulta ésta.

3.4. INDICES DE INFORMACION PARA MEDIDAS DIFUSAS.

En este apartado el problema que planteamos es el siguiente: supongamos que en una medida difusa g, para cualquier subconjunto A de X, el valor $g(A)$ representa el grado de creencia en que un elemento desconocido x_o del referencial X está en el subconjunto A; esta idea está presente en el origen de las más importantes aportaciones a la teoría de medidas difusas (como la propia definición de medida difusa de Sugeno [52], la definición de medidas de posibilidad de Zadeh [62] o la interpretación en términos de medidas difusas de las medidas de evidencia de Shafer [46]).

Solo interpretaremos (en este apartado) el significado de una medida difusa en el sentido anterior. Todo lo que sigue no tendría sentido en caso de que g modelizara, por ejemplo, grados de importancia de subconjuntos del referencial.

En este contexto, es de enorme interés tratar de estudiar las medidas difusas desde el punto de vista de la información que aportan para la determinación de ese elemento x_o desconocido. Evidentemente será preferible una medida difusa a otra, si la primera nos proporciona una información que nos permita elegir con más certeza cual es el elemento buscado que la segunda.

Las medidas difusas modelizan la información disponible, pero si en último extremo hay que tomar una decisión sobre cual es el elemento desconocido, será preferible aquella medida que lo delimite mejor. Por supuesto hay que suponer ciertas las diferentes informaciones que tengamos. Surgen entonces problemas de compatibilidad y de combinación entre las diferentes informaciones en los que, de momento, no entramos. Nuestro objetivo ahora es dar un índice de la calidad de una medida difusa para delimitar el elemento x_o.

Las probabilidades proporcionan informaciones precisas, pero dispersas, mientras que las posibilidades ofrecen informaciones imprecisas, pero totalmente coherentes, siendo ambas situaciones extremas. En medio se situan las restantes medidas difusas que reflejan en general informaciones, ni totalmente precisas, ni completamente coherentes.

Pensamos que, pese a la precisión que poseen las medidas de probabilidad, no pueden en general considerarse como las más informativas, y así, el índice A(g) definido en el apartado anterior, puesto que solo refleja parecido a una probabilidad, no puede entenderse como un índice de información. Es posible imaginar situaciones en que, frente a una medida difusa muy próxima a la probabilidad uniforme, tengamos otra medida que se aproxime (aunque menos) a otra probabilidad más informativa, y puede que sea preferible esta última desde el punto de vista de la determinación de x_o. Veamos un ejemplo concreto:

Ejemplo 3.3

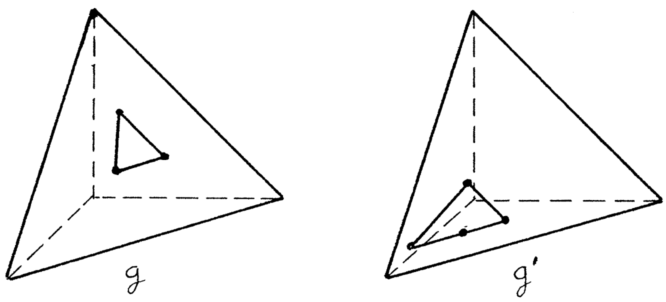
Sea $X=\{x_1,x_2,x_3\}$ y sean g y g' medidas difusas sobre X definidas por

A	g(A)	g'(A)
$\{x_1\}$	1/3	0.8
$\{x_2\}$	1/3	0.2
$\{x_3\}$	1/3	0.1
$\{x_1,x_2\}$	2/3	0.9
$\{x_1,x_3\}$	2/3	0.9
$\{x_2,x_3\}$	5/6	0.4

Si calculamos sus probabilidades asociadas,

σ	$p_{\sigma 1}$	$p_{\sigma 2}$	$p_{\sigma 3}$	$p'_{\sigma 1}$	$p'_{\sigma 2}$	$p'_{\sigma 3}$
(1,2,3)	1/3	1/3	1/3	0.8	0.1	0.1
(1,3,2)	1/3	1/3	1/3	0.8	0.1	0.1
(2,1,3)	1/3	1/3	1/3	0.7	0.2	0.1
(2,3,1)	1/6	1/3	1/2	0.6	0.2	0.2
(3,1,2)	1/3	1/3	1/3	0.8	0.1	0.1
(3,2,1)	1/6	1/2	1/3	0.6	0.3	0.1

La representación geométrica de g y g' es



Las probabilidades más próximas son

$$P_g\ (0.2\widehat{7},\ 0.36\widehat{1},\ 0.36\widehat{1})$$

que es muy parecida a la probabilidad uniforme, y

$$P_{g'}\ (0.71\widehat{6},\ 0.1\widehat{6},\ 0.11\widehat{6})$$

que es razonablemente cercana a una probabilidad degenerada. El índice de no aditividad de cada medida es

$$A(g)=0.144\ ,\ A(g')=0.15$$

Luego g es algo más parecida a una probabilidad. En cambio creemos que g' es preferible desde el punto de vista de la información, ya que ésta se dirige hacia un elemento concreto (x_1 en este caso) más claramente que g.

Las medidas difusas más informativas no cabe duda que son las probabilidades P^i degeneradas en cada punto x_i (medidas de Dirac)

$$P^i(x_i)=1 \ , \ P^i(x_j)=0 \ \forall j\neq i$$

puesto que sin ningún género de dudas indican el punto x_i como el único posible, no producen incertidumbre alguna.

En un contexto probabilístico tradicional se considera que la representación de la mayor incertidumbre es la probabilidad uniforme. En el campo más general de las medidas difusas, pensamos que esta idea no debe mantenerse, pues una distribución uniforme supone que hay tantas razones para pensar que el elemento x_o es x_i como que es x_j, para cualquier i y j, y esto supone, o disponer de una información previa en tal sentido, o hacer uso del principio de la razón insuficiente.

Una forma más adecuada de reflejar un estado de verdadera ignorancia es decir que cualquier elemento es posible con grado máximo, y necesario con grado cero, lo cual se traduce en las medidas difusas de ignorancia superior e inferior

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

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

con asignación básica de probabilidad

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

3.4.1. MEDIDAS DE ENTROPIA.

En Yager [58] y [59] se da una definición de entropía para evidencias que generaliza la noción de entropía de Shannon. La definición siguiente es una caracterización de la misma dada por el propio Yager.

Definición 3.7

Sea una evidencia definida sobre X, con a.b.p. m y medida de plausibilidad asociada Pl. Se define la entropía de dicha evidencia como

$$E(m) = -\sum_{A\subseteq X} m(A)\ln(PL(A))$$

donde ln representa la función logaritmo neperiano.

En el mismo trabajo pueden encontrarse las propiedades que verifica E(m).

Esta entropía presenta un serio inconveniente, debido a que tanto las evidencias posibilísticas (y en particular la ignorancia Pl_o) como las medidas de Dirac tienen entropía nula, y la incertidumbre es máxima cuando dispongamos de Pl_o y nula cuando tengamos P^i. Por tanto E(m) no es una buena medida de incertidumbre.

Para obviar este problema, el propio Yager ([59] y [60]) introduce la medida de especificidad asociada a una evidencia (ver también Higashi y Klir [25] y Dubois y Prade [20]).

Definición 3.8

Sea una evidencia en X representada por la a.b.p. m. Se define la especificidad de esa evidencia mediante

$$Sp(m) = \sum_{\substack{A\subseteq X \\ A\neq\emptyset}} \frac{m(A)}{|A|}$$

Entre otras propiedades, que pueden encontrarse en las referencias citadas, destacamos aquí que la especificidad alcanza valor mínimo ($1/|X|$) para la ignorancia m_o, y es máxima e igual a uno para probabilidades.

La entropía mide la inconsistencia de una evidencia, y la especificidad mide la precisión de la misma: si la evidencia se concentra en conjuntos de cardinal pequeño la especificidad será alta, y si lo hace en conjuntos de cardinal grande, ésta será baja. Según Yager, la combinación de ambas medidas proporciona una idea de la calidad de la evidencia.

Moral [38], en esa misma línea, mejora los trabajos de Yager construyendo índices que resuman en uno las medidas de entropía y de especificidad.

La principal limitación que, desde nuestro punto de vista, presentan estas formulaciones es que son específicas para evidencias, y no se pueden extender a medidas difusas más generales.

En una línea no muy divergente de la seguida por Yager, pero aplicable a casos más generales, podemos hacer uso de nuestra caracterización por conjuntos ordenados de n! probabilidades de medidas difusas cualesquiera, y definir una entropía para las mismas.

Definición 3.9

Sea g una medida difusa cualquiera definida sobre el referencial X, y sean P_σ, $\sigma\in S_n$ sus probabilidades asociadas. Definimos la entropía de g como

$$En(g)= \frac{1}{n!}\sum_{\sigma\in S_n} e(P_\sigma),$$

donde e(·) representa la entropía de Shannon para probabilidades

$$e(P)= -\sum_{i=1}^{n} p(x_i)\ln(p(x_i)).$$

Proposición 3.14

Si P es una probabilidad, entonces En(P)=e(P).

Demostración:

Si P es una probabilidad, las n! probabilidades asociadas coinciden con P: $P_\sigma=P\ \forall\sigma\in S_n$. Por tanto

$$En(P)= \frac{1}{n!}\sum_{\sigma\in S_n} e(P)=e(P).\#$$

Así pues, En(•) generaliza la noción de entropía de Shannon, aunque no la entropía de Yager para evidencias, como puede comprobarse fácilmente. También es evidente que una medida difusa y su dual tienen la misma entropía. Así mismo, puede verificarse que En(·) es una función continua respecto de la topología inducida en $\mathfrak{M}$ por la distancia S.

<u>Proposición 3.15</u>

Sea g una medida difusa y P_σ, $\sigma\in S_n$ sus probabilidades asociadas. Entonces En(g)=0 si y solamente si todas las probabilidades asociadas son medidas de Dirac.

<u>Demostración:</u>

Puesto que e(P)=0 si y solo si $P=P^i$ para algún $i\in\{1,2,..,n\}$, entonces, y puesto que $e(P)\geq 0$ cualquiera que sea la probabilidad P,

$$En(g)=0 \iff e(p_\sigma)=0 \ \forall\sigma\in S_n$$

y el resultado es ya evidente.#

No es fácil fijar las clases de medidas difusas con entropía nula, pero es posible dar una caracterización de las mismas si nos restringimos a las capacidades de orden dos, que constituyen una amplia clase de medidas difusas con interesantes propiedades y que estudiaremos más extensamente en el capítulo siguiente.

<u>Proposición 3.16</u>

Si g es una capacidad de orden dos, entonces

$$En(g)=0 \iff g \text{ es una evidencia de tipo crisp.}$$

<u>Demostración:</u>

La condición suficiente es evidente, puesto que toda evidencia de tipo crisp tiene asociadas probabilidades degeneradas, lo que por la proposición anterior equivale a que tenga entropía nula.

Para la condición necesaria, de nuevo por la proposición 3.15 basta probar que cualquier capacidad de orden dos g que tenga todas sus probabilidades asociadas degeneradas, lo que evidentemente equivale a que se verifique

$$g(A)=0 \text{ ó } 1 \ \forall A\subseteq X,$$

es de tipo crisp.

Haremos la demostración para capacidades inferiores de orden dos, y puesto que g y g^* tienen asociadas las mismas probabilidades, el resultado se cumplirá también para capacidades superiores de orden dos.

Vamos a probar que existe $C\subseteq X$ tal que

$$\begin{array}{l} \forall A\supseteq C \quad g(A)=1 \\ \text{y} \\ \forall A\not\supseteq C \quad g(A)=0 \end{array}$$

Supongamos que no fuese cierto. Entonces

$$\forall C\subseteq X \quad \begin{array}{l} \exists A\supseteq C \text{ tal que } g(A)=0 \\ \text{ó} \\ \exists A\not\supseteq C \text{ tal que } g(A)=1 \end{array}$$

Sea $C\neq X$ tal que $g(C)=1$ (si $\forall C\neq X$, $g(C)=0$ entonces g es la medida de ignorancia bel_0, que es de tipo crisp). No puede existir un subconjunto $A\supseteq C$ de modo que $g(A)=0$, luego $\exists A_1\not\supseteq C$ tal que $g(A_1)=1$. Puesto que

$$g(A_1\cup C)+g(A_1\cap C)\geq g(A_1)+g(C)=2$$

entonces $g(A_1\cap C)=1$ y $|A_1\cap C|<|C|$, puesto que $A_1\not\supseteq C$.

Dado $A_1\cap C$, $\exists A_2\not\supseteq A_1\cap C$ y $g(A_2)=1$. De nuevo

$$g(A_2\cup(A_1\cap C))+g(A_2\cap A_1\cap C)\geq g(A_2)+g(A_1\cap C)=2$$

y por tanto $g(A_2\cap A_1\cap C)=1$, y $|A_2\cap A_1\cap C|<|A_1\cap C|$.

Si continuamos este proceso podemos llegar, en a lo sumo $|C|-1$ pasos a

$$g(A_{i-1}\cap\ldots\cap A_1\cap C)=g(\{x_k\})=1$$

y otra vez $\exists A_i\not\supseteq\{x_k\}$ tal que $g(A_i)=1$. Como $A_i\cap\{x_k\}=\emptyset$ obtenemos

$$1=g(A_i\cup\{x_k\})+g(\emptyset)\geq g(A_i)+g(\{x_k\})=2,$$

lo cual es absurdo, y por tanto g es de tipo crisp.#

Así, nuestra entropía solo es nula para evidencias crisp, que son las únicas capacidades completamente coherentes (pues son posibilidades) y al mismo tiempo representan una información con certidumbre total, aunque con dispersión variable (máxima en la ignorancia y nula en las medidas de Dirac).

En este sentido nuestra entropía se comporta mejor que la de Yager, pues ésta se anula también para posibilidades cualesquiera (más aún, para todas las evidencias cuyos elementos focales se corten dos a dos).

De todos modos $En(\cdot)$ presenta el inconveniente de no discriminar entre las diferentes medidas crisp. Así, como le ocurre a $E(\cdot)$, tanto la ignorancia como las probabilidades degeneradas tienen entropía nula. Para superar esta dificultad, y de modo análogo a lo que hace Yager mediante la especificidad, se podría complementar nuestra entropía con el índice de no aditividad $A(\cdot)$ definido en el apartado anterior, para obtener una pareja de valores que, conjuntamente, midan la "calidad" de una medida difusa.

En las siguientes secciones se presentan otras alternativas totalmente diferentes para definir índices de incertidumbre para medidas difusas, que se basan en la distancia entre medidas definida en el apartado 3.1. Ello es posible si se tiene en cuenta que las medidas que proporcionan más incertidumbre son las de ignorancia bel_o y Pl_o, y las que proporcionan menos son las medidas de Dirac. Así, caben varias opciones que pasamos a analizar a continuación.

Los elementos que manejaremos son: una medida difusa g, sus probabilidades asociadas P_σ, $\sigma \in S_n$, la probabilidad más próxima a g, P_g, las medidas de ignorancia bel_o y Pl_o, la probabilidad uniforme P_u y las medidas de Dirac P^i, $i=1,..,n$.

Una primera opción sería calcular la distancia de P_g a la probabilidad uniforme P_u, y una segunda hallar la distancia de P_g a las probabilidades P^i. Sin embargo rechazamos ambos métodos porque para calcular la distancia resumen la información de g en una única probabilidad P_g, con evidente pérdida de información.

Otra posibilidad es emplear la distancia de la medida g a P_u. Tampoco es aceptable este método porque presupone que la medida menos informativa es la probabilidad uniforme, lo cual no es cierto en el campo de las medidas difusas.

Aún quedan dos alternativas que consideramos válidas:

a) Medir la distancia entre g y una medida de ignorancia, lo que proporciona un <u>indice de certidumbre:</u> a mayor distancia, menos incertidumbre genera g.

b) Medir la distancia entre g y las medidas de Dirac, lo que produce un <u>indice de incertidumbre:</u> a mayor distancia, más incertidumbre proporciona la medida difusa g.

Estudiemos por separado estos dos métodos.

3.4.2. INDICE DE CERTIDUMBRE.

Medir la distancia de una medida difusa a una medida de ignorancia presenta el problema de que, puesto que disponemos de dos codificaciones de la ignorancia, bel_o y Pl_o, cabría condiderar dos distancias para una misma medida; ¿ Cual debemos elegir ?. Parece lógico pensar que aquella codificación de la ignorancia que se adecue más a las características de la medida difusa en cuestión. Puesto que toda medida difusa g tiene una codificación dual g^*, y bel_o y Pl_o son medidas duales, creemos que las medidas pueden clasificarse en dos categorías: las más próximas a bel_o (de tipo pesimista) y las más cercanas a Pl_o (de tipo optimista). Así, elegiremos la medida de ignorancia que se parezca más a g.

Definición 3.10

Sea g una medida difusa sobre X y sean bel_o y Pl_o las medidas de ignorancia. Definimos el índice de certidumbre de g como

$$C(g)=\min(S(g,bel_o),S(g,Pl_o))$$

siendo S la distancia entre medidas habitual.

Estudiemos a continuación algunas propiedades de $C(\cdot)$.

Proposición 3.17

El índice de certidumbre $C(\cdot)$ verifica:

(a) $C(g)=0$ si y solamente si $g=bel_o$ o $g=Pl_o$.

(b) $C(g)=\min(S(g,bel_o),S(g^*,bel_o))=\min(S(g^*,Pl_o),S(g,Pl_o))$.

(c) $C(g)=C(g^*)$.

Demostración:

(a) es evidente por ser S una distancia.

(b) y (c) se deducen de que la distancia entre dos medidas y sus duales es la misma.#

Proposición 3.18

$C(\cdot)$ es una función continua entre $\mathfrak{M}$, con la topología inducida por la distancia S, y $\mathbb{R}$ con la topología usual.

Demostración:

Pretendemos probar que, dado $\varepsilon>0$ existe $\delta>0$ tal que si $S(g,g')<\delta$ entonces $|C(g)-C(g')|<\varepsilon$.

Supongamos que $C(g)\geq C(g')$. Entonces

$$S(g,bel_o)\geq \min(S(g',bel_o),S(g',Pl_o))$$

$$S(g,Pl_o)\geq \min(S(g',bel_o),S(g',Pl_o))$$

Si $\min(S(g',bel_o),S(g',Pl_o))=S(g',bel_o)$, entonces

$|C(g)-C(g')|=C(g)-C(g')=\min(S(g,bel_o),S(g,Pl_o))-S(g',bel_o)\leq$

$\leq S(g,bel_o)-S(g',bel_o)\leq S(g,g')$

verificándose la última desigualdad en virtud de la desigualdad triangular.

Si $\min(S(g',bel_o),S(g',Pl_o))=S(g',Pl_o)$, entonces

$|C(g)-C(g')|=C(g)-C(g')=\min(S(g,bel_o),S(g,Pl_o))-S(g',Pl_o)\leq$

$\leq S(g,Pl_o)-S(g',Pl_o)\leq S(g,g')$.

Si $C(g)\leq C(g')$ se procedería de manera semejante.

En cualquier caso

$$|C(g)-C(g')|\leq S(g,g')$$

y por tanto $C(\cdot)$ es continua. De hecho hemos probado que $C(\cdot)$ es lipschitziana.#

Vamos ahora a estudiar en qué casos $C(g)$ coincide con la distancia de g a Pl_o y en qué otros es la distancia de g a bel_o.

<u>Proposición 3.19</u>

Si g es una medida difusa verificando

$$g(\{x_i\})+g(\overline{\{x_i\}})+g(\{x_j\})+g(\overline{\{x_j\}})\leq 2 \text{ (respectivamente } \geq)$$

para cualesquiera $x_i,x_j\in X$, con $i\neq j$, entonces

$$S(g,bel_o)\leq S(g,Pl_o) \text{ (respectivamente } \geq).$$

<u>Demostración:</u>

Tomemos dos permutaciones con orden dual. Por comodidad

$$\sigma=(n,n-1,..,2,1) \text{ y } \sigma^*=(1,2,..,n-1,n).$$

Notamos por $P_{Pl\tau}$ y $P_{bel\tau}$ a las probabilidades asociadas a Pl_o y bel_o respectivamente, para la permutación $\tau\in S_n$. Entonces

$$P_{Pl\sigma}=P^n, \quad P_{Pl\sigma^*}=P^1$$

$$P_{bel\sigma}=P^1,\quad P_{bel\sigma^*}=P^n$$

Por tanto

$$\sum_{i=1}^{n}(P_{\sigma i}-P_{Pl\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^*i}-P_{Pl\sigma^*i})^2=(1-g(\{x_n\}))^2+$$

$$+\sum_{i=2}^{n-1}(g(\{x_i,..,x_n\})-g(\{x_{i+1},..,x_n\}))^2+(1-g(\{x_2,..,x_n\}))^2+$$

$$+(1-g(\{x_1\}))^2+\sum_{i=2}^{n-1}(g(\{x_1,..,x_i\})-g(\{x_1,..,x_{i-1}\}))^2+$$

$$+(1-g(\{x_1,..,x_{n-1}\}))^2.$$

También

$$\sum_{i=1}^{n}(P_{\sigma i}-P_{bel\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^*i}-P_{bel\sigma^*i})^2=$$

$$=g(\{x_n\})^2+\sum_{i=2}^{n-1}(g(\{x_i,..,x_n\})-g(\{x_{i+1},..,x_n\}))^2+$$

$$+g(\{x_2,..,x_n\})^2+g(\{x_1\})^2+$$

$$+\sum_{i=2}^{n-1}(g(\{x_1,..,x_i\})-g(\{x_1,..,x_{i-1}\}))^2+g(\{x_1,..,x_{n-1}\}))^2.$$

Por tanto

$$\left[\sum_{i=1}^{n}(P_{\sigma i}-P_{Pl\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^*i}-P_{Pl\sigma^*i})^2\right]-$$

$$-\left[\sum_{i=1}^{n}(P_{\sigma i}-P_{bel\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^*i}-P_{bel\sigma^*i})^2\right]=$$

$$=(1-g(\{x_n\}))^2+(1-g(\{x_2,..,x_n\}))^2+(1-g(\{x_1\}))^2+$$

$$+(1-g(\{x_1,..,x_{n-1}\}))^2-\Big[g(\{x_n\})^2+g(\{x_2,..,x_n\})^2+g(\{x_1\})^2+$$

$$+g(\{x_1,..,x_{n-1}\})^2\Big]=$$

$$=4-2\Big[g(\{x_n\})+g(\{x_2,..,x_n\})+g(\{x_1\})+g(\{x_1,..,x_{n-1}\})\Big]\geq 0 \iff$$

$$\iff g(\{x_n\})+g(\{x_2,..,x_n\})+g(\{x_1\})+g(\{x_1,..,x_{n-1}\})\leq 2.$$

Si se verifica esa condición para todo $x_i,x_j\in X$, entonces sumamos en todas las parejas de permutaciones y se obtiene el resultado.#

La proposición anterior es una condición suficiente. Veamos una condición necesaria y suficiente:

Proposición 3.20

Sea g una medida difusa. Entonces

$$S(g,bel_o)\le S(g,Pl_o) \text{ (respectivamente } \ge)$$

si y solamente si

$$\sum_{i=1}^{n}\left[g(\{x_i\})+g(\overline{\{x_i\}})\right]\le n \text{ (respectivamente } \ge).$$

Demostración:

En la proposición anterior hemos visto que, para parejas de permutaciones duales σ y σ^*, de modo que $\sigma(1)=h$ y $\sigma(n)=k$

$$\left[\sum_{i=1}^{n}(P_{\sigma i}-P_{Pl\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^* i}-P_{Pl\sigma^* i})^2\right]-$$

$$-\left[\sum_{i=1}^{n}(P_{\sigma i}-P_{bel\sigma i})^2+\sum_{i=1}^{n}(P_{\sigma^* i}-P_{bel\sigma^* i})^2\right]\ge 0 \qquad (1)$$

si y solo si

$$g(\{x_h\})+g(\overline{\{x_h\}})+g(\{x_k\})+g(\overline{\{x_k\}})\le 2.$$

Permutaciones tales que $\sigma(1)=h$ y $\sigma(n)=k$ hay $(n-2)!$, y para todas ellas la expresión (1) es la misma. Luego

$$\sum_{\sigma}\sum_{i=1}^{n}(P_{\sigma i}-P_{Pl\sigma i})^2-\sum_{\sigma}\sum_{i=1}^{n}(P_{\sigma i}-P_{bel\sigma i})^2=$$

$$=\sum_{\substack{h,k\\h\ne k}}(n-2)!\left[4-2(g(\{x_h\})+g(\overline{\{x_h\}})+g(\{x_k\})+g(\overline{\{x_k\}}))\right]\ge 0 \Leftrightarrow$$

$$\Leftrightarrow \sum_{\substack{h,k\\h\ne k}}4(n-2)!\ge\sum_{\substack{h,k\\h\ne k}}2(n-2)!\left[(g(\{x_h\})+g(\overline{\{x_h\}})+g(\{x_k\})+g(\overline{\{x_k\}}))\right]$$

y puesto que posibilidades de combinar h y k distintos hay $\binom{n}{2}=\frac{n(n-1)}{2}$, lo anterior ocurre si y solo si

$$\frac{n(n-1)}{2}\,4(n-2)!=2n!\ge 2(n-2)!\sum_{\substack{h,k\\h\ne k}}\left[(g(\{x_h\})+g(\overline{\{x_h\}})+g(\{x_k\})+g(\overline{\{x_k\}}))\right]$$

$$\Leftrightarrow n(n-1)\ge\sum_{\substack{h,k\\h\ne k}}\left[(g(\{x_h\})+g(\overline{\{x_h\}})+g(\{x_k\})+g(\overline{\{x_k\}}))\right].$$

Como cada h se puede combinar con n-1 k distintos, los sumandos están repetidos n-1 veces cada uno, o sea que

$$S(g,Pl_o)\geq S(g,bel_o) \iff n\geq \sum_{i=1}^{n} (g(\{x_i\})+g(\overline{\{x_i\}})).\#$$

Corolario 3.3

Si g es una medida difusa tal que $g(A)\leq g^*(A)$ $\forall A\subseteq X$, entonces

$$S(g,bel_o)\leq S(g,Pl_o).$$

Demostración:

Basta considerar que si

$$g(A)\leq g^*(A)=1-g(\overline{A}) \Rightarrow g(A)+g(\overline{A})\leq 1 \ \forall A\subseteq X \Rightarrow g(\{x_i\})+g(\overline{\{x_i\}})\leq 1 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} (g(\{x_i\})+g(\overline{\{x_i\}}))\leq n.\#$$

Así pues, las medidas ordenadas inferiores se parecen más a bel_o y las ordenadas superiores a Pl_o, lo cual resulta lógico, ya que se comparan codificaciones del mismo tipo.

Para medidas que no son ordenadas, pensar en ellas como de tipo inferior o superior depende del valor de la suma

$$\sum_{i=1}^{n} (g(\{x_i\})+g(\overline{\{x_i\}}))$$

La simplicidad del cómputo de esta suma permite obtener una primera aproximación sobre el carácter pesimista u optimista de cualquier medida difusa, y que cabe interpretar como una relajación de la idea de super o subaditividad. Esto puede resultar interesante en términos prácticos.

Por lo que se refiere al índice de certidumbre, se tiene, en resumen:

$$C(g)=\begin{cases} S(g,bel_o) & \text{sii} \quad \sum_{i=1}^{n} (g(\{x_i\})+g(\overline{\{x_i\}}))\leq n \\ S(g,Pl_o) & \text{sii} \quad \sum_{i=1}^{n} (g(\{x_i\})+g(\overline{\{x_i\}}))\geq n \end{cases}$$

Veamos un ejemplo de como funciona $C(\cdot)$.

Ejemplo 3.4

Sea Π una medida de posibilidad sobre $X=\{x_1,x_2,x_3\}$, definida por

$$\pi(x_1)=1,\ \pi(x_2)=a,\ \pi(x_3)=b,\ \text{con } a\geq b.$$

Entonces $C(\Pi)=S(\Pi,Pl_o)$,

$$C(\Pi)=\sqrt{\frac{6a^2+4b^2-10a-6b-2ab+8}{12}}$$

En caso de que a=b, entonces

$$C(\Pi)=\sqrt{2/3}\ (1-a).$$

La calidad de $C(\cdot)$ como índice de información es solo parcial. En efecto, resulta muy indicativo de falta de información sobre la determinación de un elemento del referencial cuando $C(\cdot)$ es bajo. Para valores altos, sin embargo, se presenta el inconveniente de que la medida en cuestión podría ser muy informativa (cercana a una medida de Dirac), o tratarse de una medida atípica y poco informativa; de hecho el máximo de $C(\cdot)$ no se alcanza para las medidas de Dirac. La inadecuación parece provenir de que no todas las medidas difusas responden a una interpretación informativa, sino que pueden reflejar otras muchas orientaciones (como la de evaluar el grado de importancia de subconjuntos del referencial o las opiniones subjetivas de un individuo sobre la verificación de una cualidad).

Por ejemplo, una medida como la ya mencionada:

$$g(\{x_1\})=g(\{x_2\})=g(\{x_3\})=0$$

$$g(\{x_1,x_2\})=g(\{x_1,x_3\})=g(\{x_2,x_3\})=1$$

da valor máximo de $C(\cdot)$, pero resulta en absoluto informativa; parece responder más bien a la evaluación de los subconjuntos que a la identificación de un elemento.

Parece natural complementar este índice con otro cuyas características se orienten en sentido inverso; el índice de incertidumbre que a continuación se estudia resulta valioso a este respecto.

3.4.3. INDICE DE INCERTIDUMBRE.

Si pensamos que las medidas difusas que menos incertidumbre provocan son las probabilidades degeneradas P^i, $i=1,..,n$, parece razonable emplear la distancia entre medidas, para una medida difusa g y P^i para definir un índice de incertidumbre de g. Como disponemos de n probabilidades de este tipo, podemos elegir como índice de incertidumbre la menor de las distancias entre g y P^i, $i=1,..,n$.

Definición 3.11

Sea g una medida difusa sobre X. Definimos el índice de incertidumbre de g mediante

$$I(g)= \min_{1\leq i\leq n} S(g,P^i).$$

Si $I(g)$ toma valores próximos a cero, entonces g es muy parecida a una probabilidad degenerada, y por tanto tiene poca incertidumbre. En cambio si $I(g)$ toma valores altos, g es poco parecida a ninguna probabilidad P^i, e interpretamos que genera mucha incertidumbre.

El índice de incertidumbre se puede expresar, desarrollando la expresión $S(g,P^i)$, como

$$I(g)= \min_{1\leq i\leq n} \sqrt{1/(2n!)\ \sum_{\sigma} \sum_{j=1}^{n} (p_{\sigma j}-p^i_j)^2} =$$

$$= \min_{1\leq i\leq n} \sqrt{1/(2n!)\ \sum_{\sigma} \left(\sum_{j=1}^{n} p^2_{\sigma j}+1-2p_{\sigma i} \right)} =$$

$$= \sqrt{1/(2n!)\ (\sum_{\sigma} \sum_{j=1}^{n} p_{\sigma j}^{2} + n! - 2 \max_{1\le i\le n} \sum_{\sigma} p_{\sigma i})}$$

En la proposición siguiente enunciamos algunas propiedades de $I(\cdot)$.

Proposición 3.21

El índice $I(\cdot)$ verifica:

(a) $I(g)=0$ si y solamente si $g=P^{i}$ para algún $i\in\{1,2,..,n\}$.

(b) $I(g)$ es máximo e igual a $\sqrt{(n-1)/n}$ si g es una medida de ignorancia.

(c) $I(g)=I(g^{*})$.

Demostración:

(a) Es evidente por ser S una distancia en $\mathfrak{M}$.

(b) Vista la expresión desarrollada de $I(g)$, el máximo se alcanzará cuando las probabilidades asociadas a la medida sean degeneradas, y por razones de simetría, habrá el mismo número de ellas, como es el caso de las medidas de ignorancia bel_o y Pl_o. El valor alcanzado es

$$\sqrt{1/(2n!)\ (n!+n!-2(n-1)!)} = \sqrt{\frac{n-1}{n}}\ .$$

(c) Se deduce por ser las probabilidades medidas autoduales.#

Pequeñas variaciones de g provocan variaciones también pequeñas en su incertidumbre:

Proposición 3.22

El índice $I(\cdot)$ es una función continua, de $\mathfrak{M}$ con la topología inducida por S, a $\mathbb{R}$ con la topología usual.

Demostración:

Supongamos que, dadas dos medidas difusas g y g',

$$I(g)\ge I(g').$$

Entonces si

$$I(g')=\min_{1\le i\le n} S(g',P^i)=S(g',P^k),$$

$$|I(g)-I(g')|=I(g)-I(g')=\min_{1\le i\le n} S(g,P^i)-S(g',P^k)\le$$

$$\le S(g,P^k)-S(g',P^k)\le S(g,g'),$$

por verificarse la desigualdad triangular para S. Por tanto I(•) no solo es continua, sono que además es lipschitziana.#

Conviene, para que la incertidumbre varíe en [0,1], dividir I(•) por su valor máximo y definir un índice de incertidumbre normalizado

$$In(g)=\sqrt{\frac{n}{n-1}}\; I(g)$$

que obviamente conserva todas las propiedades de I(•).

Veamos que expresión adopta I(•) para medidas difusas de tipo crisp.

<u>Proposición 3.23</u>

Sea g una medida difusa de tipo crisp focalizada en el conjunto B de cardinal q. Entonces

$$I(g)=\sqrt{\frac{q-1}{q}}$$

<u>Demostración:</u>

g tiene asociadas n!/q probabilidades degeneradas en cada punto perteneciente a B. Luego

$$I(g)=\sqrt{1/(2n!)\;(n!+n!-2n!/q)}=\sqrt{\frac{q-1}{q}}\;.\#$$

La incertidumbre que genera una medida crisp solo depende del cardinal del conjunto sobre el que está focalizada.

<u>Proposición 3.24</u>

Sea g una medida difusa, y sea P_g la probabilidad promedio de las asociadas a g. Entonces

$$I(g)=S(g,P^k)$$

donde $p_g(x_k) = \max_{1\leq i\leq n} p_g(x_i)$.

Demostración:

Al desarrollar la expresión de $I(\cdot)$ ya vimos que el mínimo se alcanzaba donde fuese máxima la suma $\sum_{\sigma} p_{\sigma i}$.

Como $\sum_{\sigma} p_{\sigma i} = n!\,p_g(x_i)$, el resultado es evidente.#

Así pues, la probabilidad degenerada de la que g dista menos es la que concentra toda la masa de probabilidad en el mismo elemento en que es mayor la probabilidad más próxima a la medida g.

Nota: Hay cierta analogía entre lo anterior y lo que le ocurre a la medida de especificidad de Yager para posibilidades. En Dubois y Prade [21] puede verse que

$$Sp(\Pi) = \max_{1\leq i\leq n} p_{\pi}(x_i).$$

Veamos un ejemplo de como funciona el índice de incertidumbre:

Ejemplo 3.5

Sea Π una posibilidad sobre $X=\{x_1,x_2,x_3\}$ definida por

$$\pi(x_1)=1,\ \pi(x_2)=a,\ \pi(x_3)=b,\ \text{con } a\geq b.$$

Entonces $I(\Pi)=S(\Pi,P^1)$,

$$I(\Pi) = \sqrt{\frac{6a^2+4b^2-2ab}{12}}$$

$$In(\Pi) = \sqrt{\frac{6a^2+4b^2-2ab}{8}}$$

Si $a=b$, entonces

$$I(\Pi) = \sqrt{2/3}\ a,\quad In(\Pi)=a,$$

y en este caso $C(\Pi)+I(\Pi)=\sqrt{2/3}$.

Un inconveniente que presenta el índice $I(\cdot)$ es que solo utiliza el mínimo de las distancias de g a las diferentes P^i, sin tener en cuenta las restantes distancias; es decir, mide la incertidumbre sobre cual es el elemento desconocido con referencia solo al elemento al que la medida otorga más oportunidades de ser el buscado. Puede así ocurrir, cuando deseemos comparar dos medidas difusas g y g' por el grado de incertidumbre que posean, que las distancias de g a las probabilidades P^i sean todas menores que las distancias de g' a las P^i. Si g está más próxima a todas las P^i que g', no queda claro cual proporciona menos incertidumbre (estar cerca a la vez de todos los elementos de información máxima es tan poco informativo como estar simultáneamente lejos de todos).

Veamos un ejemplo:

<u>Ejemplo 3.6</u>

Sean Π y Π' dos posibilidades sobre $X=\{x_1,x_2,x_3\}$ definidas por

$$\pi(x_1)=1,\ \pi(x_2)=0,\ \pi(x_3)=1$$

$$\pi'(x_1)=1,\ \pi'(x_2)=0.1,\ \pi'(x_3)=1$$

Si calculamos la distancia de Π y Π' a P^i, $i=1,2,3$,

$$S(\Pi,P^1)=0.701 \qquad S(\Pi',P^1)=0.697$$

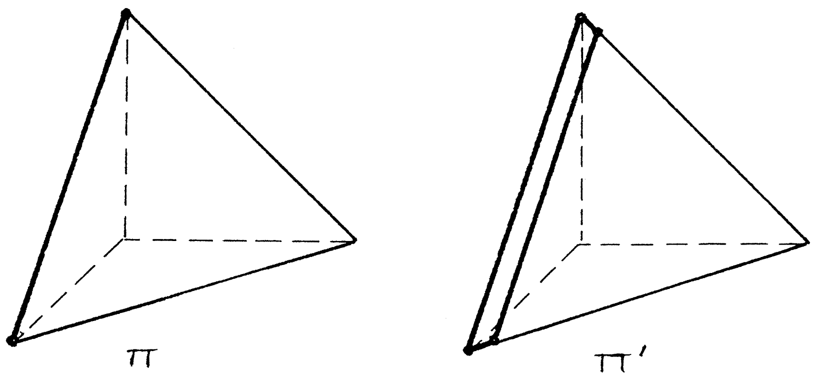
$$S(\Pi,P^2)=1 \qquad S(\Pi',P^2)=0.967$$

$$S(\Pi,P^3)=0.701 \qquad S(\Pi',P^3)=0.967$$

Por tanto

$$I(\Pi')=0.697\leq 0.701=I(\Pi)$$

y con este criterio Π' proporciona menos incertidumbre que Π, cuando parece que debe ser al contrario, puesto que Π otorga posibilidad cero a x_2 mientras que Π' le concede posibilidad 0.1. Gráficamente, Π y Π' son



Creemos que esto ocurre porque solo tenemos en cuenta la distancia de Π y Π' a P^1 o P^3, que es efectivamente menor para Π'. Pero también es menor la distancia de Π' a P^2 que la de Π.

Una forma de solventar esta dificultad es relativizar el índice para que, con la misma filosofía, utilice todas las distancias $S(g,P^i)$.

Proponemos como índice corregido de incertidumbre

$$Ic(g)= \min_{1\leq i\leq n} \sqrt{n\, S^2(g,P^i)/\sum_{j=1}^{n} S^2(g,P^j)}$$

que convenientemente desarrollado es

$$Ic(g)= \sqrt{\frac{\sum_{\sigma}\sum_{j=1}^{n} p_{\sigma j}^2 + n!(1-2\max_{1\leq i\leq n} p_{gi})}{\sum_{\sigma}\sum_{j=1}^{n} p_{\sigma j}^2 + n!(1-\frac{2}{n})}}$$

donde $p_{gi}=\frac{1}{n!}\sum_{\sigma} p_{\sigma i}$ es la probabilidad más próxima a g.

Puede comprobarse que $Ic(\cdot)$ verifica las mismas propiedades que $I(\cdot)$ o $In(\cdot)$ (vale cero solo para probabilidades degeneradas, alcanza valor máximo igual a uno para las medidas de ignorancia, coincide para medidas duales y varía de forma continua).

En el ejemplo anterior, si empleamos $Ic(\cdot)$ obtenemos

$$Ic(\Pi)=0.866, \quad Ic(\Pi')=0.874$$

y ahora ya Π proporciona menos incertidumbre que Π', como cabía esperar.

3.4.4. RELACIONES ENTRE LOS INDICES Y LA ENTROPIA.

En las secciones anteriores hemos definido y estudiado índices de certidumbre e incertidumbre para medidas difusas. Nos proponemos ahora analizar si hay alguna relación entre ellos, y para el caso particular de medidas de probabilidad, estudiar su comportamiento y compararlos con la entropía de Shannon.

En cuanto a la relación entre los diversos índices, desgraciadamente hay que decir que no son equivalentes: aunque en la mayoría de los casos, dadas dos medidas difusas, los tipos de índices señalan a una misma medida como preferible desde el punto de vista de la incertidumbre, hay casos en que no ocurre esto.

Veamos algunos ejemplos:

Ejemplo 3.7

Sean Π y Π' como en el ejemplo 3.6. Ya vimos que

$$I(\Pi)=0.701\geq 0.697=I(\Pi')$$

$$Ic(\Pi)=0.866\leq 0.874=Ic(\Pi')$$

luego $I(\cdot)$ e $Ic(\cdot)$ no son equivalentes. Si calculamos el índice de certidumbre para Π y Π', obtenemos

$$C(\Pi)=0.577\geq 0.519=C(\Pi')$$

y con este índice es preferible Π a Π'. Por tanto tampoco son equivalentes $I(\cdot)$ y $C(\cdot)$.

Ejemplo 3.8

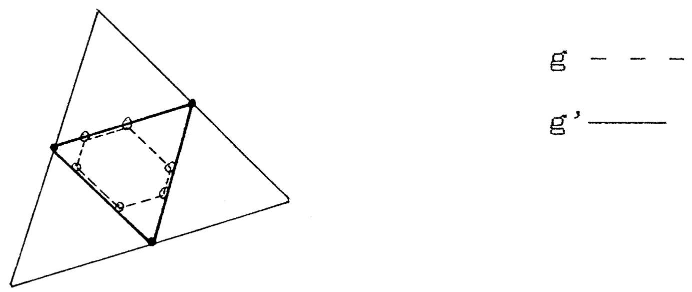
Sean g y g' dos medidas difusas definidas sobre $X=\{x_1, x_2, x_3\}$ por

A	$g(A)$	$g'(A)$
$\{x_1\}$	0.3	0
$\{x_2\}$	0.1	0
$\{x_3\}$	0.1	0
$\{x_1, x_2\}$	0.5	0.5
$\{x_1, x_3\}$	0.5	0.5
$\{x_2, x_3\}$	0.5	0.5

Puede comprobarse que g es una creencia y g' una capacidad inferior de orden dos. Las probabilidades asociadas a g y g' son:

σ	$p_{\sigma 1}$	$p_{\sigma 2}$	$p_{\sigma 3}$	$p'_{\sigma 1}$	$p'_{\sigma 2}$	$p'_{\sigma 3}$
(1,2,3)	0.3	0.2	0.5	0	0.5	0.5
(1,3,2)	0.3	0.5	0.2	0	0.5	0.5
(2,1,3)	0.4	0.1	0.5	0.5	0	0.5
(2,3,1)	0.5	0.1	0.4	0.5	0	0.5
(3,1,2)	0.4	0.5	0.1	0.5	0.5	0
(3,2,1)	0.5	0.4	0.1	0.5	0.5	0

Si representamos esas medidas obtenemos



Entonces

$$C(g)=0.451\leq 0.5=C(g')$$

y g' es preferible a g. En cambio

$$Ic(g)=0.905\leq 1=Ic(g')$$

y ahora es g es preferible a g'. De modo que tampoco son equivalentes $C(\cdot)$ e $Ic(\cdot)$.

No es de extrañar esta falta de equivalencia entre los índices, pues ya se ha comentado que $C(\cdot)$ no funciona bien para valores altos (que no tienen por qué indicar certidumbre elevada) mientras que, valores elevados de $I(\cdot)$ o $Ic(\cdot)$ no necesariamente indican mucha incertidumbre, dado que no alcanzan su máximo solo para medidas de ignorancia. De todos modos, al ser complementarios en su comportamiento, creemos que conjuntamente proporcionan una idea ajustada sobre la certidumbre o incertidumbre de una medida difusa. En efecto, un valor bajo de $I(\cdot)$ o $Ic(\cdot)$ indica que la medida es buena desde el punto de vista informativo (cercana a una medida de Dirac); un valor relativamente alto de estos índices, complementado con un valor bajo de $C(\cdot)$ indica poca calidad informativa de la medida (cercana a la ignorancia). Finalmente, si ambos tipos de índices dan resultados simultáneamente elevados, cabe pensar en que se trate de una medida atípica, probablemente no diseñada desde un punto de vista informativo, y sería necesario analizar su calidad de información por otros criterios.

Veamos ahora que forma adoptan los diferentes índices para el caso de medidas de probabilidad, que resulta particularmente ilustrativo.

<u>Proposición 3.25</u>

Sea P una medida de probabilidad. Entonces

(a) $I(P)=\sqrt{\frac{1}{2}\left(1-2\max_{1\le j\le n} p_j+\sum_{j=1}^{n} p_j^2\right)}$

(b) $Ic(P)=\sqrt{\dfrac{1+\sum_{j=1}^{n} p_j^2-2\max_{1\le j\le n} p_j}{1+\sum_{j=1}^{n} p_j^2-\frac{2}{n}}}$

(c) $C(P)=\sqrt{\frac{1}{2}\left(1+\sum_{j=1}^{n} p_j^2-\frac{2}{n}\right)}$

Demostración:

(a) $S(P,P^i)=s(P,P^i)=\sqrt{\frac{1}{2}\sum_{j=1}^{n}(p_j-p_j^i)^2} = \sqrt{\frac{1}{2}\left(\sum_{j=1}^{n}p_j^2+1-2p_i\right)}$,

luego

$$I(P)=\min_{1\le i\le n} S(P,P^i)=\sqrt{\frac{1}{2}\left(1-2\max_{1\le j\le n} p_j+\sum_{j=1}^{n}p_j^2\right)}$$

(b) Recordando la expresión desarrollada de Ic(•), y teniendo en cuenta que para una probabilidad (g=P) tenemos que $P_\sigma=P$ $\forall\sigma\in S_n$, y $P_g=P$, entonces

$$Ic(P)=\sqrt{\frac{\sum_{j=1}^{n}p_j^2+1-2\max_{1\le j\le n}p_j}{\sum_{j=1}^{n}p_j^2+1-\frac{2}{n}}}$$

(c) $C(P)=\min(S(P,bel_o),S(P,Pl_o))=S(P,Pl_o)=$

$$=\sqrt{1/(2n!)\sum_{i=1}^{n}\left((n-1)!\sum_{\substack{j=1\\ i\neq j}}^{n}p_j^2+(p_i-1)^2\right)} =$$

$$=\sqrt{1/(2n)\sum_{i=1}^{n}\left(\sum_{j=1}^{n}p_j^2+1-2p_i\right)} = \sqrt{1/(2n)\left(n\sum_{j=1}^{n}p_i^2+n-2\right)} =$$

$$=\sqrt{\frac{1}{2}\left(1+\sum_{j=1}^{n}p_j^2-\frac{2}{n}\right)}\ .\#$$

Corolario 3.4

Si P es una medida de probabilidad, entonces

$$Ic(P)=\frac{I(P)}{C(P)}$$

Así pues, Ic(•) para probabilidades es una combinación de los índices de incertidumbre y certidumbre, lo cual no ocurre para medidas difusas generales.

Si disponemos de dos probabilidades P y P', y deseamos saber cual de las dos es preferible desde el punto de vista de la determinación del elemento desconocido del referencial, podría utilizarse cualquiera de los diversos índices. Puesto que, un criterio basado en el índice de certidumbre podría contradecir a otro basado en el de incertidumbre, es necesario caracterizar los fundamentos de ambos, que resultan especialmente claros en el caso de las probabilidades.

1.- Para $C(\cdot)$,

$$C(P)\geq C(P') \iff \sum_{j=1}^{n}(p_j^2-p_j'^2)\geq 0 \iff \sum_{j=1}^{n} p_j^2 \geq \sum_{j=1}^{n} p_j'^2.$$

Es claro pues que la suma de cuadrados de las probabilidades elementales es decisiva en el índice de certidumbre.

2.- Para $I(\cdot)$,

$$I(P)\leq I(P') \iff \sum_{j=1}^{n}(p_j^2-p_j'^2)\leq 2(\max_{1\leq j\leq n} p_j - \max_{1\leq j\leq n} p_j').$$

En consecuencia, para el índice de incertidumbre, la probabilidad elemental más alta juega también un papel determinante.

3.- Para $Ic(\cdot)$,

$$C(P)\geq C(P') \text{ e } I(P)\leq I(P') \Rightarrow Ic(P)\leq Ic(P'),$$

pero el recíproco no es cierto en general, aunque si se verifica

$$Ic(P)\leq Ic(P') \Rightarrow I(P)\leq I(P') \text{ ó } C(P)\geq C(P').$$

Entonces $Ic(\cdot)$ puede entenderse como un compendio entre los índices anteriores, y podría utilizarse en caso de conflicto entre ellos.

Ninguna de estas formas de decidir que probabilidad P o P' es preferible coincide en general (aunque si en muchos casos) con el sistema que proporciona la entropía de Shannon:

$$e(P)\leq e(P') \iff -\sum_{i=1}^{n} p_i \ln(p_i) \leq -\sum_{i=1}^{n} p_i' \ln(p_i').$$

Veamos algunos ejemplos que ilustren por qué ocurre así:

Ejemplo 3.9

Sean P=(0.6, 0.3, 0.1) y P'=(0.6, 0.4, 0). Entonces

$$e(P)=0.898 \text{ y } e(P')=0.673$$

y puesto que P' tiene menos entropía, es preferible a P. También es así para C(•), pues

$$C(P)=0.629 \text{ y } C(P')=0.653$$

En cambio

$$I(P)=0.36 \text{ e } I(P')=0.4$$

y con este índice es mejor P que P', lo mismo que ocurre para Ic(•), ya que

$$Ic(P)=0.572 \text{ e } Ic(P')=0.612.$$

En líneas generales I(•) prima aquellas probabilidades que establezcan una mayor diferencia entre un determinado p_i y los restantes. Así P(0.6, 0.3, 0.1) es preferida a P'(0.6, 0.4, 0) porque aunque el valor máximo en ambas es 0.6, P distancia más ese valor de los restantes (0.3 y 0.1) que P' (0.4 y 0).

En cambio C(•) no se fija en el valor máximo de cada probabilidad, discrimina solamente por la suma de cuadrados.

Ic(•) es una mezcla de ambos.

Para la entropía de Shannon es preferible eliminar completamente a x_3 como resultado posible, aún a expensas de hacer x_2 más probable. En cambio para I(•) es preferible mantener la mayor diferencia entre x_1 y los elementos restantes, aún a riesgo de discriminar peor entre los elementos menos probables.

Otro ejemplo que resulta aparentemente contradictorio con el anterior es el siguiente:

Ejemplo 3.10

Sean P=(0.91, 0.05, 0.04) y P'=(0.9, 0.1, 0).

$e(P')=0.325\leq 0.364=e(P)$, y P' es mejor que P.

$C(P')=0.759\leq 0.763=C(P)$, y P es mejor que P'.

$I(P')=0.1\leq 0.296=I(P)$, y P' es mejor que P.

$Ic(P')=0.131\leq 0.388=Ic(P)$, y P' es mejor que P.

La contradicción es solo relativa, y existe una importante diferencia entre ambos ejemplos: mientras en el primero se mantiene constante la probabilidad máxima, en el segundo hay una ligera disminución de probabilidad máxima en P' con respecto a P. Entonces C(•) prefiere aquí la probabilidad no nula en x_3 porque el aumento de probabilidad en x_1 y x_3 supera en términos de suma de cuadrados a la disminución de la probabilidad en x_2. Por su parte, I(•) designa a P' como preferible al no compensar el aumento en la probabilidad máxima (0.91 por 0.90) la diferencia en términos de suma de cuadrados (que aquí actua en sentido contrario que en C(•)). Finalmente, e(•) continua designando como preferible a la distribución que descarta x_3.

Es necesario destacar, por último, que los índices propuestos son estrechamente dependientes del criterio elegido para establecer la distancia, que ha sido el de la métrica euclídea.

Otra alternativas, como las basadas en el valor absoluto o en el máximo, darían lugar a la formulación de índices con características diferentes, si bien la idea de definir estos índices a partir de distancias entre probabilidades asociadas a las medidas, tiene carácter general.

3.5. ESTUDIO COMPLETO DE UN CASO PARTICULAR.

A modo de ejemplo resumen, vamos a aplicar todos los conceptos definidos en este capítulo a un tipo de medidas difusas que aparece en Huber [28], y que responde a la operación de descuento de evidencias (ver Shafer [47]), para el caso particular de probabilidades.

Sea P una probabilidad definida sobre X, y sea $0 \leq \varepsilon \leq 1$. Se define

$$g^*_\varepsilon(A) = \begin{cases} (1-\varepsilon)P(A)+\varepsilon & \text{si } A \neq \emptyset \\ 0 & \text{si } A=\emptyset \end{cases}$$

que es obviamente una medida difusa. Su medida dual es

$$g_\varepsilon(A) = \begin{cases} (1-\varepsilon)P(A) & \text{si } A \neq X \\ 1 & \text{si } A=X \end{cases}$$

y puesto que $g_\varepsilon(A) \leq g^*_\varepsilon(A)\ \forall A \subseteq X$, (g, g^*) es un par ordenado de medidas duales.

Veamos que también son capacidades de orden dos, más concretamente evidencias:

En efecto, si

$$m_\varepsilon(A) = \begin{cases} (1-\varepsilon)p_i & \text{si } A=\{x_i\},\ p_i=P(\{x_i\}) \\ 0 & \text{si } |A| \geq 2,\ A \neq X \\ \varepsilon & \text{si } A=X \end{cases}$$

puesto que

$$\sum_{A \subseteq X} m_\varepsilon(A) = \sum_{i=1}^{n} (1-\varepsilon)p_i + \varepsilon = 1$$

m_ε es una a.b.p.. Sus medidas de creencia y plausibilidad asociadas son

$$\forall A \neq X \quad bel_\varepsilon(A) = \sum_{B \subseteq A} m_\varepsilon(B) = \sum_{x_i \in A} m_\varepsilon(\{x_i\}) = \sum_{x_i \in A} (1-\varepsilon)p_i = (1-\varepsilon)P(A)$$

Por tanto

$$bel_\varepsilon = g_\varepsilon \ , \ Pl_\varepsilon = g^*_\varepsilon$$

y g_ε, g^*_ε son evidencias. Además es fácil comprobar que cualquier evidencia cuyos elementos focales sean solo los conjuntos unitarios y el conjunto total, responde a un esquema de este tipo.

Podemos interpretar estas medidas difusas como probabilidades conocidas inexactamente, fijando el error cometido máximo en la cantidad ε. Obsérvese que g_ε evoluciona desde una probabilidad dada P (ε=0) hasta la ignorancia total (ε=1), pasando por situaciones de error intermedias.

Calculemos las probabilidades asociadas:

Dada la permutación $\sigma=(\sigma(1),\sigma(2),..,\sigma(n))$

para $i\neq 1$ $\quad p_{\sigma\sigma(i)}=g^*_\varepsilon(\{x_{\sigma(1)},..,x_{\sigma(i)}\})-g^*_\varepsilon(\{x_{\sigma(1)},..,x_{\sigma(i-1)}\})$

para $i=1$ $\quad p_{\sigma\sigma(1)}=g^*_\varepsilon(\{x_{\sigma(1)}\})$.

Luego

$$p_{\sigma\sigma(i)}=(1-\varepsilon)P(\{x_{\sigma(i)}\})$$
$$p_{\sigma\sigma(1)}=(1-\varepsilon)P(\{x_{\sigma(1)}\})+\varepsilon$$

o sea, para cada $i\in\{1,2,..,n\}$ hay $(n-1)!$ probabilidades asociadas del tipo

$$p_{\sigma i}=\begin{cases}(1-\varepsilon)p_i & \text{si } \sigma(1)\neq i\\ (1-\varepsilon)p_i+\varepsilon & \text{si } \sigma(1)=i\end{cases}$$

correspondientes a las permutaciones tales que $\sigma(1)=i$.

Una vez conocidas las probabilidades asociadas, podemos hacer un estudio de todos los conceptos desarrollados en este capítulo.

A) Distancia:

Fijada la probabilidad P, para $\varepsilon,\varepsilon'\in[0,1]$ se obtiene

$$S(g_\varepsilon,g_{\varepsilon'})=|\varepsilon-\varepsilon'|C(P)$$

donde $C(\cdot)$ es el índice de certidumbre (mínimo entre las distancias de P a las medidas de ignorancia bel_o y Pl_o).

La distancia entre dos medidas aproximadas de una probabilidad disminuye al disminuir la calidad de la probabilidad y también al decrecer la diferencia entre los errores, como es lógico suponer.

Si ahora suponemos dos probabilidades P y P', fijado $\varepsilon \in [0,1]$ obtenemos

$$S(g_{\varepsilon}, g'_{\varepsilon}) = (1-\varepsilon) S(P,P')$$

resultado también muy lógico: al aumentar el error cometido la distancia va disminuyendo, ya que nos vamos aproximando a la ignorancia en los dos casos.

B) Probabilidad más próxima:

La probabilidad más cercana a g_{ε} es $P_{g_{\varepsilon}}$ definida por

$$P_{g_{\varepsilon}}(A) = (1-\varepsilon) P(A) + |A| \frac{\varepsilon}{n} \quad \forall A \subseteq X$$

que solo coincide con P cuando ésta es la probabilidad uniforme:

$$S(P, P_{g_{\varepsilon}}) = \varepsilon S(P, P_u)$$

C) Indice de no aditividad:

Puede comprobarse que

$$A(g_{\varepsilon}) = \varepsilon$$

La pérdida de aditividad solo depende del error que se comete al aproximar la probabilidad, y aumenta con dicho error.

D) Indice de certidumbre:

$$C(g_{\varepsilon}) = (1-\varepsilon) C(P)$$

y la certidumbre de g_{ε} es la de la probabilidad que la genera, modificada por el factor $1-\varepsilon$. Como es lógico, la certidumbre disminuye al aumentar el error. Además siempre la aproximación g_{ε} de P tiene menos certidumbre que P.

E) Indice de incertidumbre:

Puede comprobarse que

$$I(g_{\varepsilon})=\sqrt{\varepsilon^2 C^2(P)+I^2(P)+\varepsilon(\max_j p_j - \sum_{j=1}^{n} p_j^2)}$$

Si derivamos esa expresión respecto de ε, la derivada es siempre positiva, y puede concluirse que $I(g_{\varepsilon})$ es una función creciente en ε, es decir, la incertidumbre aumenta con el error.

F) Indice de incertidumbre corregido:

De nuevo puede verificarse que

$$Ic(g_{\varepsilon})=\sqrt{\frac{\varepsilon^2 C^2(P)+I^2(P)+\varepsilon(\max_j p_j - \sum_{j=1}^{n} p_j^2)}{\varepsilon^2 C^2(P)+C^2(P)+\varepsilon(\frac{1}{n} - \sum_{j=1}^{n} p_j^2)}}$$

que también es una función creciente en ε: $Ic(g_{\varepsilon})$ aumenta si el error crece.

G) Especificidad:

Puesto que g_{ε} es una evidencia, podemos calcular la especificidad de Yager para esta medida.

$$Sp(g_{\varepsilon})=1-\frac{n-1}{n}\varepsilon$$

que es una función decreciente: la especificidad disminuye conforme aumenta el error.

Las medidas de entropía $E(\cdot)$ y $En(\cdot)$ no adoptan expresiones sencillas, debido al uso de logaritmos en su definición.

Así pues, los diversos conceptos e índices definidos en este capítulo parecen tener un comportamiento bastante adecuado en medidas difusas que respondan a la idea de determinar un elemento desconocido del referencial.

- CAPITULO IV -

4.0. INTRODUCCION.

Si el capítulo anterior se ha dedicado fundamentalmente a la cuantificación de algunas características relevantes de una medida difusa en base a medir la semejanza o discrepancia entre medidas o grupos de medidas a través de una distancia, el presente capítulo aborda las relaciones entre dos medidas difusas que corresponderían, en cierto modo, a la inclusión, conjunción y disyunción en una estructura reticular.

Se parte de la acotación de una medida difusa mediante la definición de dos medidas, que denominaremos asociadas, y que se obtienen maximizando y minimizando sobre el conjunto de probabilidades asociadas. La idea de acotar una medida difusa ha presidido numerosos trabajos en este campo, pero nuestra caracterización proporciona un mecanismo natural.

El estudio de estas medidas asociadas nos conduce a considerar la clase de las capacidades de orden dos como la más adecuada para definir relaciones de inclusión. Sin embargo estas medidas no forman una clase cerrada frente a los mecanismos de combinación que definimos, por lo que se hace necesario considerar una clase más amplia: la de las medidas representables, para las que definimos la conjunción y disyunción.

Sin embargo, no se obtiene una estructura reticular debido a la inexistencia de una medida que modelice la contradicción entre informaciones, esto es, de una "medida vacia".

El apartado 4.1 se dedica a la definición y estudio de las medidas difusas asociadas, que pueden interpretarse como codificaciones extremas, se prueba que el proceso de asignación a una medida de sus medidas asociadas se reitera a partir del primer paso, y que la clase de las medidas asociadas coincide con las capacidades de orden dos. Se completa el apartado con diversas caracterizaciones de estas medidas y su relación con la esperanza monótona.

El apartado 4.2, dedicado a las inclusiones entre medidas, comienza con una crítica a la utilización de las diferencias entre una medida y su dual como medida de la imprecisión; nos parece más adecuado en este sentido la diferencia entre las medidas asociadas, relacionada con la dispersión de las probabilidades correspondientes a cada medida. Se define así una inclusión de medidas a través de tales probabilidades, y se estudian en detalle sus propiedades. También se incluyen en el apartado 4.2 nuevas propiedades sobre las capacidades de orden dos, que muestran como la inclusión se adecua especialmente bien a este tipo de medidas. Por último, definimos otra relación de inclusión (que coincide con la anterior para capacidades) en la clase de las medidas representables.

Para ellas,y ya dentro del apartado 4.3, se definen una conjunción y una disyunción, que se prueba son las operaciones asociadas al orden parcial de la inclusión. En relación con las propiedades que deducimos para estos mecanismos de combinación, resulta interesante señalar:

a) Que para medidas de tipo crisp, la conjunción y disyunción se corresponden con la unión e intersección de los subconjuntos implicados, lo que garantiza una mínima congruencia en nuestras definiciones.

b) Que para medidas de posibilidad, la disyunción se corresponde perfectamente con la unión de los subconjuntos difusos normalizados que éstas representan, no así para la conjunción y la intersección, debido a problemas en relación con la normalización.

c) Que nuestra conjunción no coincide con la derivada de la regla de Dempster y que, aunque puede considerarse más restrictiva en algunos aspectos que ésta, presenta muy buenas propiedades algebraicas y no plantea los problemas de ésta última en relación con la idempotencia.

4.1. MEDIDAS DIFUSAS ASOCIADAS A UNA MEDIDA DADA.

El conocimiento de las probabilidades asociadas a una medida difusa permite obtener dos medidas difusas ordenadas que la acotan. Así, dada una medida difusa cualquiera g, quedan fijadas una medida superior y otra inferior, duales, que representan codificaciones extremas de la información dada por g. En lo que sigue se definen estas medidas a las que caracterizaremos como integrantes de una familia conocida de medidas difusas, y estudiaremos algunas de sus propiedades. Los resultados de este apartado serán de utilidad en los siguientes, en donde definiremos una relación de inclusión y métodos de combinación de medidas difusas.

Sea pues g una medida difusa definida sobre el referencial finito $X=\{x_1,x_2,..,x_n\}$, y sean P_σ $\sigma\in S_n$ sus $n!$ probabilidades asociadas.

<u>Definición 4.1</u>

Las medidas difusas asociadas a g son

$$g^m(A)= \max_{\sigma\in S_n} P_\sigma(A), \quad g_m(A)= \min_{\sigma\in S_n} P_\sigma(A) \quad \forall A\subseteq X.$$

g^m mide un subconjunto como lo haría la probabilidad asociada a g que le asignara mayor peso, mientras que g_m lo mide como la probabilidad que le otorga menor peso. En algún sentido, g^m y g_m son codificaciones extremadamente optimista y pesimista respectivamente de g.

<u>Proposición 4.1</u>

Las medidas g^m y g_m asociadas a una medida difusa g son una pareja de medidas duales ordenadas.

<u>Demostración:</u>

Es evidente.#

Proposición 4.2

Las medidas ordenadas asociadas a una medida difusa y a su medida dual coinciden, es decir

$$(g^*)^m=g^m \quad \text{y} \quad (g^*)_m=g_m.$$

Demostración:

Puesto que las probabilidades asociadas a g y a su dual g^* son las mismas, el resultado es obvio.#

Probemos ahora que las medidas asociadas son cotas de la medida difusa de partida:

Proposición 4.3

Las medidas g^m y g_m asociadas a la medida difusa g verifican

$$g_m(A)\leq g(A)\leq g^m(A) \quad \forall A\subseteq X.$$

Demostración:

Sea $A\subseteq X$ y supongamos que $A=\{x_{i_1},x_{i_2},..,x_{i_r}\}$. Existen permutaciones σ tales que

$$\sigma(1)=i_1,\ \sigma(2)=i_2,..,\ \sigma(r)=i_r.$$

Para cualquiera de esas permutaciones τ

$p_\tau(x_{i_1})=g(\{x_{i_1}\})$

$p_\tau(x_{i_2})=g(\{x_{i_1},x_{i_2}\})-g(\{x_{i_1}\})$

. . . .

$p_\tau(x_{i_r})=g(\{x_{i_1},..,x_{i_r}\})-g(\{x_{i_1},..,x_{i_{r-1}}\})$.

Por tanto

$$P_\tau(A)=\sum_{j=1}^{r} p_\tau(x_{i_j})=g(A),$$

y así

$$g_m(A)\leq g(A)\leq g^m(A).\#$$

Aplicando la proposición 4.2 es evidente que también g^m y

g_m son cotas para la medida g^* dual de g.

Definición 4.2

Denominamos MA al conjunto de medidas difusas que están asociadas a alguna medida difusa, es decir

$$MA=\{g'\in\mathfrak{M} \ / \ \exists g\in\mathfrak{M} \text{ y } g^m=g' \text{ ó } g_m=g'\}.$$

Para dar una caracterización de las medidas de esta clase necesitamos un resultado previo.

Proposición 4.4

Para cualquier medida difusa g se verifica

$$(g_m)_m=g_m \quad \text{y} \quad (g^m)^m=g^m.$$

Demostración:

Basta probar que

$$(g_m)_m\geq g_m \quad \text{y} \quad (g^m)^m\leq g^m,$$

pues las desigualdades inversas se verifican siempre en virtud da la proposición 4.3. Probaremos solo la primera desigualdad; la otra se demuestra de forma análoga.

$$g_m(A)=\min_{\sigma\in S_n} P_\sigma(A)=\min_{\sigma\in S_n} \sum_{x_i\in A} p_\sigma(x_i).$$

Para una permutación σ fija

$$p_\sigma(x_i)=g(B_{\sigma i}\cup\{x_i\})-g(B_{\sigma i})$$

donde $B_{\sigma i}=\{x_{\sigma(1)},x_{\sigma(2)},\ldots,x_{\sigma(h_i-1)}\}$ y $\sigma(h_i)=i$. Luego

$$g_m(A)\leq \sum_{x_i\in A}(g(B_{\sigma i}\cup\{x_i\})-g(B_{\sigma i})) \quad \forall\sigma\in S_n \qquad (1)$$

Si notamos mediante $P_{m\sigma}$ a las probabilidades asociadas a la medida g_m ,

$$(g_m)_m(A)=\min_{\sigma\in S_n} P_{m\sigma}(A)=P_{m\sigma_o}(A)=\sum_{x_i\in A} p_{m\sigma_o}(x_i)=$$

$$=\sum_{x_i\in A}(g_m(B_{\sigma_o i}\cup\{x_i\})-g_m(B_{\sigma_o i})).$$

$$g_m(B_{\sigma_o i}\cup\{x_i\})=\min_{\tau\in S_n} P_\tau(B_{\sigma_o i}\cup\{x_i\})=P_{\tau_o}(B_{\sigma_o i}\cup\{x_i\})=$$

$$= \sum_{x_j \in B_{\sigma_o i} \cup \{x_i\}} p_{\tau_o}(x_j) = \sum_{x_j \in B_{\sigma_o i} \cup \{x_i\}} (g(B_{\tau_o j} \cup \{x_j\}) - g(B_{\tau_o j})).$$

$$g_m(B_{\sigma_o i}) = \min_{\gamma \in S_n} P_\gamma(B_{\sigma_o i}) = P_{\gamma_o}(B_{\sigma_o i}) = \sum_{x_j \in B_{\sigma_o i}} p_{\gamma_o}(x_j) =$$

$$= \sum_{x_j \in B_{\sigma_o i}} (g(B_{\gamma_o j} \cup \{x_j\}) - g(B_{\gamma_o j})) \leq$$

$$\leq \sum_{x_j \in B_{\sigma_o i}} (g(B_{\sigma j} \cup \{x_j\}) - g(B_{\sigma j})) \quad \forall \sigma \in S_n \tag{2}$$

Luego

$$(g_m)_m(A) = \sum_{x_i \in A} \Big[\sum_{x_j \in B_{\sigma_o i} \cup \{x_i\}} (g(B_{\tau_o j} \cup \{x_j\}) - g(B_{\tau_o j})) -$$

$$- \sum_{x_j \in B_{\sigma_o i}} (g(B_{\gamma_o j} \cup \{x_j\}) - g(B_{\gamma_o j})) \Big] =$$

$$= \sum_{x_i \in A} \Big[\sum_{x_j \in B_{\sigma_o i}} (g(B_{\tau_o j} \cup \{x_j\}) - g(B_{\tau_o j})) -$$

$$- \sum_{x_j \in B_{\sigma_o i}} (g(B_{\gamma_o j} \cup \{x_j\}) - g(B_{\gamma_o j})) \Big] +$$

$$+ \sum_{x_i \in A} (g(B_{\tau_o i} \cup \{x_i\}) - g(B_{\tau_o i})) \geq$$

$$\geq \sum_{x_i \in A} (g(B_{\tau_o i} \cup \{x_i\}) - g(B_{\tau_o i})) \geq g_m(A),$$

la primera de las desigualdades se verifica por (2), y la segunda por (1).#

Otra manera de expresar este resultado es decir que las aplicaciones que llevan una medida difusa en sus medidas asociadas inferior (g_m) y superior (g^m) son idempotentes.

Proposición 4.5

La medida difusa g pertenece a la clase MA si y solamente si $g=g^m$ ó $g=g_m$.

Demostración:

La condición suficiente es trivial.

Condición necesaria:

Si g está en la clase MA, entonces $\exists g'\in\mathfrak{M}$ tal que

$$g'^m=g \quad \text{ó} \quad g'_m=g$$

En el primer caso

$$g^m=(g'^m)^m=g'^m=g$$

por la proposición 4.4. El segundo caso es análogo, de modo que la condición necesaria está probada.#

Por este resultado, las medidas difusas asociadas a una dada solo pueden ser aquellas que también estén asociadas a si mismas.

Corolario 4.1

Sea g una medida difusa y sean P_σ $\sigma\in S_n$ sus probabilidades asociadas. Entonces

$$g\in MA \Leftrightarrow \begin{cases} g(A)=\min\limits_{\sigma\in S_n} P_\sigma(A) & \forall A\subseteq X \\ \text{ó} & \\ g(A)=\max\limits_{\sigma\in S_n} P_\sigma(A) & \forall A\subseteq X \end{cases}$$

La característica esencial de una medida difusa de la clase MA es pues que solo depende de las probabilidades asociadas a dicha medida, pero no de las permutaciones que las inducen: podemos regenerar la medida original conociendo solo sus probabilidades asociadas, sin necesidad de saber a que permutación corresponden.

Nota: Tratar de reconstruir la medida original a partir de una única probabilidad, el promedio P_g de las asociadas, no es posible ni para medidas de la clase MA. Hay que restringirse a una clase de medidas aún más pequeña, las medidas de posibilidad (naturalmente también con medidas de probabilidad), para conseguirlo, como vimos en el apartado 3.2 del capítulo anterior.

Aunque hemos dado una caracterización de la clase MA de medidas que se asocian a una dada tomando inferior o superior en el conjunto de las probabilidades asociadas a la misma, aún no podemos identificar esta clase de medidas con ninguna conocida.

En principio, la clase MA está contenida en la clase de las medidas difusas representables (debido al corolario anterior); sin embargo no todas las medidas representables son de la clase MA. El siguiente es un contraejemplo en este sentido:

Ejemplo 4.1

En $X=\{x_1,x_2,x_3,x_4\}$ sea g la medida representable inferior obtenida del conjunto de probabilidades

$\mathscr{E}=\{(0.5,0.3,0.1,0.1),(0.25,0.25,0.25,0.25),(0.3,0.4,0.2,0.1)\}$.

$$g(A)=\min_{P\in\mathscr{E}} P(A)$$

Puede comprobarse (construyendo las 24 probabilidades asociadas a g) que $g_m\neq g$ y $g^m\neq g$; luego por la proposición 4.5 g no pertenece a la clase MA.

En sentido contrario, puede probarse que la clase MA contiene a la clase de las evidencias, pero éstas no son las únicas medidas incluidas en MA, como prueba el siguiente contraejemplo.

Nota: La prueba de que toda evidencia es una medida de MA no se incluye puesto que la proposición siguiente proporciona un resultado más potente.

Ejemplo 4.2

Sea g definida sobre $X=\{x_1,x_2,x_3\}$ por

$$g(\{x_1\})=g(\{x_2\})=g(\{x_3\})=0$$

$$g(\{x_1,x_2\})=g(\{x_1,x_3\})=g(\{x_2,x_3\})=\tfrac{1}{2}$$

Es fácil comprobar que g no es una evidencia, pero en cambio $g_m=g$ y por tanto es de la clase MA.

Así pues se verifica que las inclusiones

$$EV\subset MA\subset MR$$

son estrictas.

En realidad la clase MA coincide con la clase C2 de capacidades de orden dos.

Proposición 4.6

Sea g una medida difusa definida sobre X. g es de la clase MA si y solamente si g es una capacidad de orden dos.

Demostración:

Condición necesaria:

Si g es de la clase MA entonces $g_m=g$ ó $g^m=g$. Vamos a probar que si $g_m=g$ entonces g es una capacidad inferior de orden dos, es decir

$$g(A\cup B)+g(A\cap B)\geq g(A)+g(B) \quad \forall A,B\subseteq X$$

y la demostración de que si $g^m=g$ entonces g es una capacidad superior de orden dos se hace de forma análoga.

$$g(A)=g_m(A)=\min_{\sigma\in S_n} P_\sigma(A)=\min_{\sigma\in S_n} \sum_{x_i\in A} p_\sigma(x_i)\leq \sum_{x_i\in A} p_\sigma(x_i) \quad \forall\sigma\in S_n.$$

Supongamos que

$$A\cap B=\{x_{i_1},\ldots,x_{i_r}\}$$

$$\bar{A}\cap B=\{x_{j_1},\ldots,x_{j_s}\}$$

$$A\cap\bar{B}=\{x_{k_1},\ldots,x_{k_t}\}$$

Consideremos una permutación $\tau\in S_n$ tal que

$$\tau(1)=i_1,\ldots,\tau(r)=i_r$$

$$\tau(r+1)=j_1,\ldots,\tau(r+s)=j_s$$

$$\tau(r+s+1)=k_1,\ldots,\tau(r+s+t)=k_t \;,$$

es decir, una ordenación de los elementos de X de modo que los primeros lugares los ocupan los elementos de $A\cap B$, seguidos de los de $\overline{A}\cap B$ y después los de $A\cap\overline{B}$, y por último los de $\overline{A\cup B}$. Entonces

$$P_\tau(A)=\sum_{x\in A}p_\tau(x)=g(\{x_{i_1}\})+g(\{x_{i_1},x_{i_2}\})-g(\{x_{i_1}\})+\ldots$$

$$\ldots+g(\{x_{i_1},\ldots,x_{i_r}\})-g(\{x_{i_1},\ldots,x_{i_{r-1}}\})+$$

$$+g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s},x_{k_1}\})-g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s}\})+$$

$$\ldots+g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s},x_{k_1},\ldots,x_{k_t}\})-$$

$$-g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s},x_{k_1},\ldots,x_{k_{t-1}}\})=$$

$$=g(\{x_{i_1},\ldots,x_{i_r}\})-g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s}\})+$$

$$+g(\{x_{i_1},\ldots,x_{i_r},x_{j_1},\ldots,x_{j_s},x_{k_1},\ldots,x_{k_t}\})=$$

$$=g(A\cap B)-g(B)+g(A\cup B).$$

Por tanto

$$g(A)\le g(A\cap B)-g(B)+g(A\cup B)$$

y g es una capacidad inferior de orden dos.

Condición suficiente:

Vamos a probar que si g es una capacidad inferior de orden dos, entonces

$$g(A)\le \min_{\sigma\in S_n} P_\sigma(A)=g_m(A) \quad \forall A\subseteq X$$

y por tanto g será de la clase MA. La demostración de que si g es una capacidad superior de orden dos entonces $g(A)\ge g^m(A)$ $\forall A\subseteq X$ es análoga.

Supongamos que $A=\{x_{i_1},x_{i_2},\ldots,x_{i_r}\}$. Entonces

$$P_\sigma(A)=\sum_{j=1}^{r}p_\sigma(x_{i_j})=\sum_{j=1}^{r}\left(g(B_{\sigma i_j}\cup\{x_{i_j}\})-g(B_{\sigma i_j})\right)$$

donde $B_{\sigma i_j}=\{x_{\sigma(1)},x_{\sigma(2)},\ldots,x_{\sigma(h_{i_j}-1)}\}$ y $\sigma(h_{i_j})=i_j$.

Sin pérdida de generalidad, podemos suponer que

$$\sigma^{-1}(i_r)<\sigma^{-1}(i_{r-1})<\ldots<\sigma^{-1}(i_1).$$

Si definimos los conjuntos A_{i_j} mediante

$$A_{i_j}=\{x_{i_j},x_{i_{j+1}},\ldots,x_{i_r}\},\quad j=1,\ldots,r$$

entonces

$$B_{\sigma i_j}\cup A_{i_j}=B_{\sigma i_j}\cup\{x_{i_j}\},\quad B_{\sigma i_j}\cap A_{i_j}=\{x_{i_{j+1}},\ldots,x_{i_r}\}=A_{i_{j+1}},\quad j=1,\ldots,r-1$$

$$B_{\sigma i_r}\cup A_{i_r}=B_{\sigma i_r}\cup\{x_{i_r}\},\quad B_{\sigma i_r}\cap A_{i_r}=\emptyset.$$

Luego

$$\sum_{j=1}^{r} g(B_{\sigma i_j}\cup\{x_{i_j}\})\geq\sum_{j=1}^{r} g(B_{\sigma i_j})+\sum_{j=1}^{r} g(A_{i_j})-\sum_{j=1}^{r-1} g(A_{i_{j+1}})=$$

$$=\sum_{j=1}^{r} g(B_{\sigma i_j})+\sum_{j=1}^{r} g(A_{i_j})-\sum_{j=2}^{r} g(A_{i_j})=\sum_{j=1}^{r} g(B_{\sigma i_j})+g(A_{i_1}).$$

Por tanto

$$g(A_{i_1})=g(A)\leq\sum_{j=1}^{r}(g(B_{\sigma i_j}\cup\{x_{i_j}\})-g(B_{\sigma i_j}))=P_\sigma(A),$$

y así

$$g(A)\leq\min_{\sigma\in S_n}P_\sigma(A)=g_m(A),\quad\forall A\subseteq X.\ \#$$

<u>Corolario 4.2</u>

La medida difusa g es una capacidad de orden dos si y solamente si $g=g^m$ ó $g=g_m$.

La clase MA no es sino la clase C2, y por tanto las capacidades de orden dos se caracterizan por obtenerse como superior o inferior del conjunto de sus probabilidades asociadas.

Las capacidades de orden dos poseen también interesantes propiedades desde el punto de vista de la esperanza monótona. En primer lugar daremos un resultado válido para cualquier medida difusa.

Proposición 4.7

Si P_σ $\sigma\in S_n$ son las probabilidades asociadas a una medida difusa g, se verifica, $\forall h:X\longrightarrow\mathbb{R}_o^+$,

$$\min_{\sigma\in S_n} E_{P_\sigma}(h)\le E_g(h)\le \max_{\sigma\in S_n} E_{P_\sigma}(h).$$

Demostración:

Como hemos probado (proposición 2.5 del capítulo II), la esperanza monótona de una función h respecto de una medida difusa g coincide con la esperanza matemática de h respecto de una de las probabilidades asociadas (la correspondiente a la permutación que ordena los elementos de X igual que la función h), es decir,

$$E_g(h)=E_{P_{\sigma_o}}(h) \quad \text{para alguna } \sigma_o\in S_n.$$

Por tanto

$$\min_{\sigma\in S_n} E_{P_\sigma}(h)\le E_g(h)\le \max_{\sigma\in S_n} E_{P_\sigma}(h).\#$$

En general, nada más puede afirmarse a este respecto. No obstante, para el caso particular de las capacidades de orden dos, necesariamente se alcanza una de las anteriores cotas para el valor de la esperanza monótona, y ello constituye además una caracterización de las medidas de dicha clase.

Proposición 4.8

Condición necesaria y suficiente para que una pareja de medidas duales g y g^* sean capacidades de orden dos inferior y superior respectivamente, es que $\forall h:X\longrightarrow\mathbb{R}_o^+$,

$$E_g(h)=\min_{\sigma\in S_n} E_{P_\sigma}(h) \quad \text{y/o} \quad E_{g^*}(h)=\max_{\sigma\in S_n} E_{P_\sigma}(h).$$

Demostración:

Condición necesaria:

Si g,g^* son capacidades de orden dos, entonces

$$g=\min_{\sigma\in S_n} P_\sigma \quad y \quad g^*=\max_{\sigma\in S_n} P_\sigma ,$$

y es obvio que

$$g\leq P_\sigma\leq g^* \quad \forall\sigma\in S_n$$

Por tanto, por la monotonía de la esperanza monótona

$$E_g(h)\leq E_{P_\sigma}(h)\leq E_{g^*}(h), \quad \forall h:X\longrightarrow \mathbb{R}_o^+ ,$$

de modo que

$$E_g(h)\leq \min_{\sigma\in S_n} E_{P_\sigma}(h)$$

$$E_{g^*}(h)\geq \max_{\sigma\in S_n} E_{P_\sigma}(h)$$

La proposición anterior prueba las desigualdades inversas.

Condición suficiente:

Para cualquier $A\subseteq X$ se verifica, por la hipótesis

$$g(A)=E_g(I_A)=\min_{\sigma\in S_n} E_{P_\sigma}(I_A)=\min_{\sigma\in S_n} P_\sigma(A),$$

y en virtud de la caracterización que proporciona el corolario 4.2, se deduce que g es una capacidad inferior de orden dos. Para g^* se prueba analogamente.#

Los resultados que acabamos de probar permiten establecer de una forma simple otra caracterización, ya conocida, de las capacidades de orden dos (ver Huber [29]).

<u>Proposición 4.9</u>

a) g es una capacidad inferior de orden dos si y solo si

$$E_g(h_1+h_2)\geq E_g(h_1)+E_g(h_2), \quad \forall h_1,h_2:X\longrightarrow \mathbb{R}_o^+ .$$

b) g^* es una capacidad superior de orden dos si y solo si

$$E_{g^*}(h_1+h_2)\leq E_{g^*}(h_1)+E_{g^*}(h_2), \quad \forall h_1,h_2:X\longrightarrow \mathbb{R}_o^+ .$$

<u>Demostración:</u>

a) Condición necesaria:

Si g es una capacidad inferior de orden dos, por la proposición 4.8, $\forall h_1,h_2:X\longrightarrow \mathbb{R}_o^+$

$$E_g(h_1+h_2)=\min_{\sigma\in S_n} E_{P_\sigma}(h_1+h_2)=\min_{\sigma\in S_n}(E_{P_\sigma}(h_1)+E_{P_\sigma}(h_2))\geq$$

$$\geq \min_{\sigma\in S_n} E_{P_\sigma}(h_1)+\min_{\sigma\in S_n} E_{P_\sigma}(h_2)=E_g(h_1)+E_g(h_2).$$

Condición suficiente:

Si $E_g(h_1+h_2)\geq E_g(h_1)+E_g(h_2)$ $\forall h_1,h_2:X\longrightarrow\mathbb{R}_0^+$, tomando $h_1=I_A$ y $h_2=I_B$, se verifica

$$g(A)+g(B)=E_g(I_A)+E_g(I_B)\leq E_g(I_A+I_B)=E_g(I_{A\cup B}+I_{A\cap B})=$$

$$=E_g(I_{A\cup B})+E_g(I_{A\cap B})=g(A\cup B)+g(A\cap B),$$

debido a que las funciones $I_{A\cup B}(x)$ e $I_{A\cap B}(x)$ están equiordenadas, y la esperanza monótona es lineal para este tipo de funciones.

Por tanto g es una capacidad inferior de orden dos.

b) Se prueba de forma análoga.#

4.2. INCLUSION ENTRE MEDIDAS DIFUSAS.

La pérdida de aditividad de las medidas difusas lleva asociado otro fenómeno que consiste en una falta de precisión real en los valores de la medida difusa: si consideramos una pareja de medidas duales ordenadas, $g \leq g^*$, ambas codifican la misma información, pero el valor de medida asignado a cualquier subconjunto A de X es $g(A)$ ó $g^*(A)$; esto puede interpretarse como una indefinición del valor de medida de A, que oscilaría en el intervalo $[g(A), g^*(A)]$. Así, las medidas g y g^* proporcionan las cotas mínima y máxima para la variación del valor de medida de cualquier subconjunto.

Desde este punto de vista, y refiriéndonos solo a medidas ordenadas, la máxima imprecisión correspondería a las medidas de ignorancia (con intervalo de imprecisión $[0,1]$), y la mayor precisión a las medidas autoduales, en las que $g(A)$ y $g^*(A)$ coinciden.

En el caso de medidas difusas cualesquiera, no ordenadas, la interpretación anterior no es tan clara. Cabe recurrir a un enfoque más general, considerando las capacidades de orden dos g_m y g^m asociadas a una medida difusa g, estudiadas en el apartado anterior: el intervalo de imprecisión del valor de la medida g sobre A sería $[g_m(A), g^m(A)]$, y $g(A)$ podría interpretarse como un "valor modal" (en el sentido de considerar el valor de medida de A como un número difuso con moda $g(A)$ y holguras $g^m(A)-g(A)$ y $g(A)-g_m(A)$. Para más detalles referentes a los números difusos consultar, por ejemplo, Dubois y Prade [17] y Delgado y Verdegay [12]).

Este otro criterio coincidiría con el primero cuando la medida considerada fuese una capacidad de orden dos, por lo expuesto en el apartado anterior, y en caso contrario, parece responder mejor a la idea de imprecisión en la evaluación de un subconjunto; en efecto, el intervalo $[g(A),g^{*}(A)]$ depende exclusivamente de la medida de A y de la de su complementario, mientras que el $[g_m(A),g^m(A)]$ depende de todas las evaluaciones probabilísticas de A correspondientes a las posibles ordenaciones que pueden considerarse sobre los elementos del referencial.

El objetivo fundamental de este apartado es relacionar la falta de precisión de las medidas difusas con la dispersión inherente a las probabilidades asociadas a éstas. Para ello vamos a definir una relación de inclusión entre medidas difusas y estudiaremos detalladamente sus propiedades. Dado que la mayor riqueza de éstas se encuentra en las capacidades de orden dos, las dos acepciones de la imprecisión que hemos comentado son equivalentes para nuestros fines.

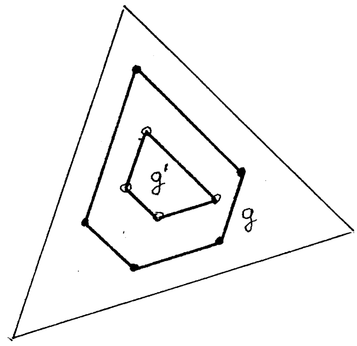
La idea sobre la que se basa la definición de inclusión es la siguiente: cuanto más dispersas sean las probabilidades asociadas a una medida difusa, menos precisión poseerá ésta. Geométricamente, las probabilidades asociadas a una medida difusa son un conjunto de (a lo sumo) n! puntos del hiperplano afín de $\mathbb{R}^n$ de ecuación $X_1+X_2+..+X_n=1$. El cierre convexo de esos puntos proporciona una región cuyos vértices son probabilidades asociadas. Pues bien, intuitivamente parece lógico pensar que si la región asociada a una medida dada queda en el interior de la correspondiente a una segunda medida, la primera presenta menos dispersión que la segunda, y por tanto es más precisa en la determinación de los valores de medida de cada subconjunto. Vamos a formalizar esta idea.

Definición 4.3

Sean g y g' dos medidas difusas definidas sobre el referencial X, y sean respectivamente P_σ y P'_σ , $\sigma\in S_n$ sus probabilidades asociadas. Diremos que la medida g' está contenida o incluida en la medida g, y lo notaremos mediante $g'\subset g$, si todas las probabilidades asociadas a g' son combinaciones convexas de las probabilidades asociadas a g:

$$g'\subset g \iff \forall x\in X,\ \forall\sigma\in S_n \quad p'_\sigma(x)=\sum_{\tau\in S_n}\lambda_{\tau\sigma}p_\tau(x)$$

donde $\lambda_{\tau\sigma}\geq 0\ \forall\tau\in S_n$ y $\sum_{\tau\in S_n}\lambda_{\tau\sigma}=1$.



g' está contenida en g.

Nota: Puesto que las probabilidades asociadas a una medida difusa y a su dual coinciden (salvo en el orden), en realidad estamos definiendo una relación de inclusión en el conjunto de pares de medidas duales, aunque al hablar de ella nos refiramos en algunos casos a uno de los elementos del par.

Estudiemos algunas propiedades elementales de la relación de inclusión:

(a) Es reflexiva: $g\subset g\ \forall g\in\mathfrak{M}$.

(b) Es transitiva: $g\subset g'$ y $g'\subset g'' \Rightarrow g\subset g''\ \forall g,g',g''\in\mathfrak{M}$.

Estas propiedades son evidentes, e indican que la relación definida es un preorden.

(c) Existen elementos máximos, uno de los cuales es la pareja de medidas duales de ignorancia total (bel_o, Pl_o):

En efecto, puesto que las probabilidades asociadas a bel_o y Pl_o son las degeneradas P^i, $i=1,..,n$, obviamente cualquier probabilidad se expresa como combinación convexa de tales probabilidades.

(d) No existe elemento mínimo, pero si elementos minimales que son únicamente las medidas de probabilidad:

En efecto, si una medida g está contenida en una probabilidad P, $g \subset P$, entonces $P_\sigma = P$ $\forall \sigma \in S_n$ y por tanto $g=P$. Luego P es un elemento minimal.

No se verifica en general la propiedad antisimétrica, aunque más adelante estaremos en condiciones de probarla para la clase de las capacidades de orden dos.

Antes de continuar se plantea el tema de ver si existe relación entre la inclusión y los índices de no aditividad e información para medidas difusas.

Aparentemente, teniendo en cuenta que el índice de no aditividad resulta ser la distancia a la probabilidad promedio de las asociadas a la medida difusa, se podría pensar que si una medida está contenida en otra, la primera debe ser más aditiva (tener un índice $A(\cdot)$ más bajo) que la segunda. Esta intuición es válida en muchos casos, aditividad y precisión van ligados habitualmente, pero falla algunas veces por el siguiente motivo: las probabilidades asociadas a una medida difusa pueden repetirse; para el cálculo de $A(\cdot)$ se tienen en cuenta todas, mientras que para estudiar la inclusión no nos importa cuantas veces puedan repetirse las probabilidades. Veamos un ejemplo que ilustre esta situación:

Ejemplo 4.3

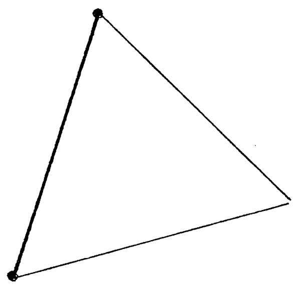
Sean, sobre $X=\{x_1,x_2,x_3\}$ las medidas difusas g y g' definidas por

A	g(A)	g'(A)
$\{x_1\}$	0	0
$\{x_2\}$	0	0
$\{x_3\}$	0	0
$\{x_1,x_2\}$	0	0
$\{x_1,x_3\}$	1	0.9
$\{x_2,x_3\}$	0	0
$\{x_1,x_2,x_3\}$	1	1

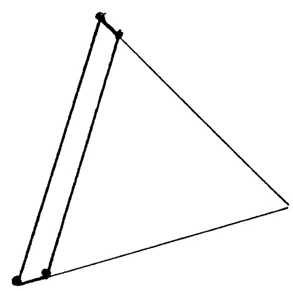
Tanto g como g' son medidas de necesidad. Sus probabilidades asociadas son

σ	$p_{\sigma 1}$	$p_{\sigma 2}$	$p_{\sigma 3}$	$p'_{\sigma 1}$	$p'_{\sigma 2}$	$p'_{\sigma 3}$
(1,2,3)	0	0	1	0	0	1
(1,2,3)	0	0	1	0	0.1	0.9
(2,1,3)	0	0	1	0	0	1
(2,3,1)	1	0	0	1	0	0
(3,1,2)	1	0	0	0.9	0.1	0
(3,2,1)	1	0	0	1	0	0

Gráficamente, g y g' se representan



g



g'

Evidentemente $g \subset g'$. Sin embargo si calculamos el índice de no aditividad obtenemos

$$A(g) = \sqrt{.25} \geq \sqrt{.235} = A(g')$$

y g es menos aditiva que g'.

Tampoco hay relación directa entre los índices de información y la inclusión: una medida poco precisa puede representar una información bastante coherente desde el punto de vista de la determinación de un elemento del referencial, y otra más precisa puede, por el contrario, ser menos informativa en ese sentido (enfocarse menos a un elemento determinado). Geométricamente, una medida es más precisa cuanto menos extensa es su región asociada, y es más informativa cuanto más cercana esté a uno de los puntos correspondientes a medidas de Dirac y más alejada esté de las restantes.

Veamos un ejemplo que muestra esta falta de relación:

Ejemplo 4.4

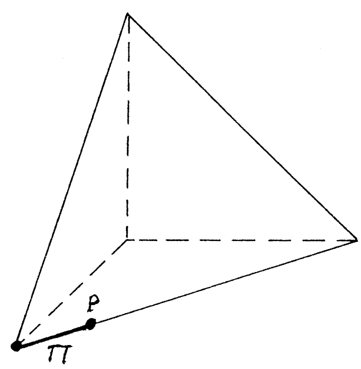
Sean, sobre $X=\{x_1, x_2, x_3\}$ la probabilidad P definida por

$$p(x_1)=0.8, \quad p(x_2)=0.2, \quad p(x_3)=0$$

y la posibilidad Π definida por

$$\pi(x_1)=1, \quad \pi(x_2)=0.2, \quad \pi(x_3)=0$$

La representación gráfica de esas medidas es



Se verifica que $P \subset \Pi$, ya que P es una de las probabilidades asociadas a Π. Sin embargo si calculamos los diversos índices estudiados en el apartado 3.4 obtenemos

$$C(\Pi)=\sqrt{.52}\geq\sqrt{.50\widehat{6}}=C(P)$$

$$I(\Pi)=\sqrt{.02}\leq\sqrt{.04}=I(P)$$

$$Ic(\Pi)=\sqrt{.03\widehat{409}}\leq\sqrt{.078}=Ic(P)$$

y con cualquier índice es preferible Π a P, desde el punto de vista de la incertidumbre.

Volvamos al estudio de las propiedades de la relación de inclusión.

<u>Proposición 4.10</u>

Sea (g,g^*) una pareja de capacidades duales de orden dos, y sea g' una medida difusa cualquiera. Si

$$g'\subset g \text{ (o equivalentemente } g'\subset g^*)$$

entonces

$$g\leq g'\leq g^* \quad \text{y} \quad g\leq g'^*\leq g^*.$$

<u>Demostración:</u>

Por ser g y g^* capacidades de orden dos, $\forall A\subseteq X$

$$g(A)=\min_{\tau\in S_n} P_\tau(A)$$

$$g^*(A)=\max_{\tau\in S_n} P_\tau(A)$$

siendo P_τ $\tau\in S_n$ las probabilidades asociadas a g y g^*.

Dado cualquier subconjunto $A\subseteq X$, existe una permutación $\sigma\in S_n$ tal que $g'(A)=P'_\sigma(A)$ (cualquier permutación que situe los elementos de A en primer lugar). Luego, como $g'\subset g$ ($g'\subset g^*$),

$$g'(A)=P'_\sigma(A)=\sum_{\tau\in S_n}\lambda_{\tau\sigma}P_\tau(A)$$

y

$$g(A)=\min_\tau P_\tau(A)\leq\sum_\tau\lambda_{\tau\sigma}P_\tau(A)\leq\max_\tau P_\tau(A)=g^*(A).$$

Por tanto

$$g(A)\leq g'(A)\leq g^*(A) \quad \forall A\subseteq X$$

y para g'^* se obtiene el mismo resultado.#

Así, que una medida cualquiera esté incluida en una capacidad implica que sus valores están acotados inferiormente por los de la capacidad inferior, y acotados superiormente por los de la capacidad superior.

Inmediatamente nos planteamos si es posible probar el recíproco del resultado anterior. La respuesta es afirmativa, si las dos medidas son capacidades de orden dos. Necesitamos algunos resultados previos, que interpreten geométricamente las capacidades.

<u>Proposición 4.11</u>

Si g es una capacidad de orden dos entonces ninguna de sus probabilidades asociadas puede ser combinación convexa de las restantes (distintas de ella).

<u>Demostración:</u>

Sea $\sigma\in S_n$ una permutación cualquiera, y supongamos que

$$P_\sigma=\sum_{\tau\in S_n}\lambda_\tau P_\tau$$

Por la construcción de P_σ es evidente que

$$P_\sigma(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(i)}\})=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(i)}\})\ \ \forall i\in\{1,..,n\}$$

Por tanto

$$g(\{x_{\sigma(1)},..,x_{\sigma(i)}\})=\sum_{\tau\in S_n}\lambda_\tau P_\tau(\{x_{\sigma(1)},..,x_{\sigma(i)}\}).$$

Pero

$$P_\sigma(\{x_{\sigma(1)},..,x_{\sigma(i)}\})=g(\{x_{\sigma(1)},..,x_{\sigma(i)}\})\le$$

$$\le P_\tau(\{x_{\sigma(1)},..,x_{\sigma(i)}\})\quad \forall\tau\in S_n.$$

Luego, si

$$P_\sigma(\{x_{\sigma(1)},..,x_{\sigma(i)}\})<P_\tau(\{x_{\sigma(1)},..,x_{\sigma(i)}\})$$

debe ser $\lambda_\tau=0$, y por tanto en la combinación convexa solo aparecen probabilidades P_τ tales que

$$P_\tau(\{x_{\sigma(1)},..,x_{\sigma(i)}\})=P_\sigma(\{x_{\sigma(1)},..,x_{\sigma(i)}\}) \quad \forall i\in\{1,..,n\},$$

lo que obliga a que $P_\tau=P_\sigma$, y P_σ solo es combinación convexa de probabilidades iguales a ella.#

Corolario 4.3

Todas las probabilidades asociadas a una capacidad de orden dos son puntos extremos de un poliedro convexo en $\mathbb{R}^n$.

Necesitamos además conocer cuales son los bordes del poliedro convexo asociado a una capacidad.

Proposición 4.12

los bordes del poliedro convexo asociado a una capacidad g de orden dos son los 2^n-2 hiperplanos de ecuaciones

$$\sum_{i=1}^{k} X_{\sigma(i)}=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(k)}\}) \quad \forall\sigma\in S_n \quad \forall k\in\{1,..,n\}$$

del hiperplano de $\mathbb{R}^n$ de ecuación $X_1+X_2+..+X_n=1$.

Demostración:

Para ver que los hiperplanos mencionados son los bordes del poliedro asociado a g basta comprobar que existen probabilidades asociadas que los generan y que las restantes probabilidades quedan todas al mismo lado de los hiperplanos.

Supongamos que g es una capacidad inferior de orden dos. Para capacidades superiores el razonamiento es análogo.

$\forall\sigma\in S_n$, $\forall k\in\{1,..,n\}$ es evidente que

$$\sum_{i=1}^{k} p_\tau(x_{\sigma(i)})=P_\tau(\{x_{\sigma(1)},..,x_{\sigma(k)}\})\geq g(\{x_{\sigma(1)},..,x_{\sigma(k)}\}) \quad \forall\tau\in S_n$$

y todas las probabilidades asociadas a g están en el mismo lado del hiperplano

$$\sum_{i=1}^{k} X_{\sigma(i)}=g(\{x_{\sigma(1)},..,x_{\sigma(k)}\}).$$

Eligiendo n-1 probabilidades cualesquiera de entre las $k!(n-k)!$ probabilidades asociadas a g que situan los elementos $x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(k)}$ en los primeros lugares y los restantes elementos después, puede comprobarse que generan el hiperplano anterior.#

Proposición 4.13

Sean (g,g^*) y (g',g'^*) dos parejas de capacidades duales de orden dos. Si

$$g\leq g'\leq g'^*\leq g^*$$

entonces

$$g'\subset g \quad (\text{equivalentemente } g'^*\subset g^*).$$

Demostración:

Probando que todas las probabilidades P'_σ asociadas a g' están al mismo lado de los bordes del poliedro convexo formado por las probabilidades P_σ asociadas a g, entonces es obvio que todas las probabilidades P'_σ estarán en el interior de dicho convexo y por tanto se podrán expresar como combinaciones convexas de sus extremos, y en consecuencia tendremos probado el resultado.

Por la proposición anterior, los bordes del convexo asociado a g son los hiperplanos

$$\sum_{i=1}^{k} X_{\sigma(i)}=g(\{x_{\sigma(1)},x_{\sigma(2)},..,x_{\sigma(k)}\}) \quad \forall\sigma\in S_n \quad \forall k\in\{1,..,n\}.$$

$$\forall\tau\in S_n \quad \sum_{i=1}^{k} P'_\tau(x_{\sigma(i)})=P'_\tau(\{x_{\sigma(1)},..,x_{\sigma(k)}\})\geq$$

$$\geq g'(\{x_{\sigma(1)},..,x_{\sigma(k)}\})\geq g(\{x_{\sigma(1)},..,x_{\sigma(k)}\})$$

y todas las probabilidades asociadas a g' están en el interior del convexo.#

Si las dos medidas no son capacidades no puede asegurarse el resultado, como prueba el siguiente ejemplo.

Ejemplo 4.5

Sean las medidas difusas g y g' definidas sobre $X=\{x_1,x_2,x_3\}$ mediante

A	g(A)	g'(A)
$\{x_1\}$	0.1	0.1
$\{x_2\}$	0.3	0.3
$\{x_3\}$	0.1	0.2
$\{x_1,x_2\}$	0.5	0.5
$\{x_1,x_3\}$	0.2	0.2
$\{x_2,x_3\}$	0.4	0.4
$\{x_1,x_2,x_3\}$	1	1

Es inmediato verificar que

$$g\le g'\le g'^*\le g^*.$$

g es una creencia, y por tanto una capacidad inferior de orden dos, mientras que g' no lo es.

Puede comprobarse sin más que construir las probabilidades asociadas a g y g' que, no solo no se verifica que g' esté incluida en g, sino que es g quien está incluida en g'.

En virtud de las proposiciones 4.10 y 4.13 obtenemos el siguiente corolario:

Corolario 4.4

Si (g,g^*) y (g',g'^*) son dos parejas de capacidades de orden dos, entonces

$$g'\subset g \text{ (ó } g'^*\subset g^*) \quad \text{si y solo si} \quad g\le g'\le g'^*\le g^*.$$

Por tanto, para capacidades de orden dos nuestra definición de inclusión es equivalente a que el "intervalo de imprecisión" de g', $[g'(A),g'^{*}(A)]=[g'_m(A),g'^{m}(A)]$ esté contenido en el de g, $[g(A),g^{*}(A)]=[g_m(A),g^{m}(A)]$, para cualquier subconjunto $A\subseteq X$; luego g' es más precisa que g.

En este contexto estamos ya en condiciones de interpretar la inclusión de medidas de forma coherente con la inclusión geométrica de los poliedros correspondientes, y establecer así un orden entre las capacidades.

<u>Proposición 4.14</u>

La relación de inclusión es una relación de orden parcial en el conjunto de las capacidades de orden dos.

<u>Demostración:</u>

Las propiedades reflexiva y transitiva se verifican en general. Solo resta probar la propiedad antisimétrica.

Sean (g,g^{*}) y (g',g'^{*}) parejas de capacidades duales de orden dos. Si $g\subset g'$ y $g'\subset g$, aplicando la proposición 4.10 obtenemos

$$g'\leq g\leq g^{*}\leq g'^{*}$$
$$g\leq g'\leq g'^{*}\leq g^{*}$$

luego $g=g'$.#

La relación de inclusión por tanto adquiere un sentido pleno en el conjunto de las capacidades de orden dos.

<u>Proposición 4.15</u>

Sea g una capacidad de orden dos y g' una medida difusa cualquiera. Si $g'\subset g$, entonces $g'_m\subset g$ (y $g'^{m}\subset g$).

<u>Demostración:</u>

Si g es una capacidad inferior

$$P'_{\sigma}(A)=\sum_{\tau\in S_n}\lambda_{\tau\sigma}P_{\tau}(A)\geq\sum_{\tau\in S_n}\lambda_{\tau\sigma}\min_{\tau}P_{\tau}(A)=g(A)\quad\forall\sigma\in S_n.$$

Por tanto

$$g'_m(A)= \min_{\sigma\in S_n} P'_\sigma(A)\geq g(A) \quad \forall A\subseteq X$$

y puesto que g'_m y g son capacidades, la expresión anterior equivale a que $g'_m\subset g$.

Si g es un capacidad superior se razona de forma análoga empleando g'^m en lugar de g'_m.#

Corolario 4.5

Sea (g,g^*) una pareja de capacidades duales de orden dos, y g' una medida difusa cualquiera. Si $g'\subset g$, entonces

$$g\leq g'_m\leq g'\leq g'^m\leq g^*.$$

Proposición 4.16

Sea g una medida difusa cualquiera. Se verifica

$$g\subset g^m \quad y \quad g\subset g_m \ .$$

Demostración:

Notemos por P_τ y $P_{m\tau}$ a las probabilidades asociadas a g y g_m respectivamente. Basta probar que todas las probabilidades P_τ están en el interior del poliedro convexo generado por las probabilidades $P_{m\tau}$. Los bordes de esa figura son los hiperplanos

$$\sum_{i=1}^{k} X_{\sigma(i)}=g_m(\{x_{\sigma(1)},..,x_{\sigma(k)}\}).$$

$$\forall k\in\{1,..,n\},\ \forall\sigma,\tau\in S_n \quad \sum_{i=1}^{k} p_\tau(x_{\sigma(i)})=P_\tau(\{x_{\sigma(1)},..,x_{\sigma(k)}\})\geq$$

$$\geq \min_{\gamma\in S_n} P_\gamma(\{x_{\sigma(1)},..,x_{\sigma(k)}\})=g_m(\{x_{\sigma(1)},..,x_{\sigma(k)}\})$$

y todas las probabilidades P_τ están en el interior del convexo.#

Una medida difusa es pues menos dispersa que sus capacidades asociadas, como era de prever.

Corolario 4.6

Con el orden de la relación de inclusión, las capacidades g_m y g^m asociadas a una medida difusa g son las menores capacidades de orden dos que contienen a g.

Demostración:

Se deduce inmediatamente de las proposiciones 4.15 y 4.16.#

Proposición 4.17

Sean g y g' dos medidas difusas cualesquiera. Si $g' \subset g$ entonces

$$g_m \leq g'_m \leq g'^m \leq g^m,$$

o equivalentemente

$$g'_m \subset g_m \quad (y \ g'^m \subset g^m).$$

Demostración:

La equivalencia entre las dos afirmaciones se deduce del corolario 4.4, puesto que g'_m , g_m , g'^m y g^m son capacidades de orden dos.

si $g' \subset g$, por la proposición 4.16 se tiene $g \subset g_m$, y por la transitividad de la inclusión, $g' \subset g_m$. Aplicando ahora la proposición 4.15, $g'_m \subset g_m$.#

Por tanto las aplicaciones que llevan una medida difusa en sus capacidades asociadas inferior y superior preservan la relación de inclusión (son aplicaciones monótonas).

Para medidas difusas generales ocurre igual que para capacidades: si $g' \subset g$, el intervalo de imprecisión de g', $[g'_m(A), g'^m(A)]$, está contenido en el de g, $[g_m(A), g^m(A)]$, para cualquier subconjunto $A \subseteq X$, y por tanto g' es una medida más precisa que g.

De todos modos, la relación de inclusión se comporta muy bien para capacidades de orden dos, pero no tanto en medidas difusas más generales. Vamos a dar otra definición de inclusión parecida a la anterior, y válida para la clase de medidas representables, que será de utilidad en el apartado siguiente de este capítulo. Para ello, previamente debemos poner de manifiesto algunas propiedades de este tipo de medidas.

Proposición 4.18

Sea (g,g^*) una pareja de medidas representables. Entonces podemos expresar tales medidas mediante

$$g(A)=\min_{P\in\mathcal{E}_g} P(A) \quad y \quad g^*(A)=\max_{P\in\mathcal{E}_g} P(A),$$

donde $\mathcal{E}_g=\{P\in PR \ / \ g(A)\le p(A)\le g^*(A),\ \forall A\subseteq X\}$.

Demostración:

Como g y g^* son representables, existe un subconjunto $\mathcal{H}\subseteq PR$ tal que

$$g(A)=\inf_{P\in\mathcal{H}} P(A) \quad y \quad g^*(A)=\sup_{P\in\mathcal{H}} P(A)$$

Evidentemente se verifica $g(A)\le P(A)\le g^*(A)\ \forall A\subseteq X,\ \forall P\in\mathcal{H}$, por lo que $\mathcal{H}\subseteq\mathcal{E}_g$. Por tanto

$$g^*(A)=\sup_{P\in\mathcal{H}} P(A)\le\sup_{P\in\mathcal{E}_g} P(A)$$

$$g(A)=\inf_{P\in\mathcal{H}} P(A)\ge\inf_{P\in\mathcal{E}_g} P(A)$$

Pero como $g(A)\le P(A)\le g^*(A)\ \forall P\in\mathcal{E}_g$, entonces

$$g(A)\le\inf_{P\in\mathcal{E}_g} P(A), \qquad g^*(A)\ge\sup_{P\in\mathcal{E}_g} P(A),$$

por lo que

$$g^*(A)=\sup_{P\in\mathcal{E}_g} P(A) \quad y \quad g(A)=\inf_{P\in\mathcal{E}_g} P(A).$$

Además el supremo y el ínfimo se alcanzan en $\mathcal{E}_g$ puesto que éste es un conjunto cerrado. #

Por tanto $\mathcal{E}_g$ es el mayor conjunto de probabilidades que representa a (g,g^*), y es evidente que es un poliedro convexo cuyos bordes son los hiperplanos de ecuaciones

$$\sum_{i\in I} X_i = g(A),\quad A=\{x_i\in X \ / \ i\in I\},\quad \forall I\subseteq\{1,2,..,n\},$$

es decir, $\mathcal{E}_g$ es el convexo determinado por las desigualdades

$$\sum_{i\in I} X_i \geq g(A) \ \ (\text{o} \ \sum_{i\in I} X_i \leq g^*(A)\ ),\quad \forall I\subseteq\{1,2,..,n\}.$$

Definición 4.4

Sean $(g,g^*),(g',g'^*)$ medidas representables, y sean $\mathcal{E}_g$ y $\mathcal{E}'_g$ sus poliedros asociados. Diremos que (g',g'^*) está incluida en (g,g^*), y lo notaremos por $g'\leq g$, si $\mathcal{E}'_g\subseteq\mathcal{E}_g$.

Esta relación obviamente verifica las propiedades reflexiva y transitiva, tiene elemento máximo, que es la pareja de medidas duales de ignorancia total (bel_o,Pl_o), y aunque no hay mínimo si existen elementos minimales que son las medidas de probabilidad. También es antisimétrica, porque distintas medidas representables tienen asociados poliedros necesariamente distintos, con lo que la relación es un orden parcial en MR.

Esta nueva definición de inclusión en general no coincide con la anterior (restringida a la clase MR), debido a que las probabilidades asociadas a una medida representable no tienen que ser los puntos extremos de su poliedro asociado, ni siquiera tienen que pertenecer a él.

No obstante, ambas definiciones si coinciden para la clase de las capacidades de orden dos, como podemos deducir de la siguiente caracterización de la inclusión:

Proposición 4.19

Sean $(g,g^*),(g',g'^*)$ medidas representables. Entonces

$$g'\leq g \iff g(A)\leq g'(A)\leq g'^*(A)\leq g^*(A)\ \ \forall A\subseteq X.$$

Demostración:

Condición necesaria:

Si $g'\leq g$ entonces $\mathcal{E}'_g\subseteq\mathcal{E}_g$. Por tanto

$$g(A)=\min_{P\in\mathcal{E}_g} P(A)\leq\min_{P\in\mathcal{E}'_g} P(A)=g'(A)$$

$$g^*(A)=\max_{P\in\mathcal{E}_g} P(A)\geq\max_{P\in\mathcal{E}'_g} P(A)=g'^*(A)$$

Condición suficiente:

Si $g(A)\leq g'(A)\leq g'^*(A)\leq g^*(A)$ $\forall A\subseteq X$, y suponemos que g' no está incluida en g, entonces $\mathcal{E}'_g$ no está contenido en $\mathcal{E}_g$. Por tanto existe $P\in\mathcal{E}'_g$ y $P\notin\mathcal{E}_g$. Teniendo en cuenta la definición de $\mathcal{E}_g$ y $\mathcal{E}'_g$,

$$P\notin\mathcal{E}_g \Rightarrow \exists B\subseteq X \text{ tal que } P(B)=\sum_{x_i\in B} p(x_i)<g(B).$$

$$P\in\mathcal{E}'_g \Rightarrow g'(A)\leq P(A)\ \forall A\subseteq X.$$

Asi pues,

$$\exists B\subseteq X \ / \ g'(B)<g(B),$$

en contradicción con la hipótesis. Por tanto $g'\leq g$. #

Corolario 4.7

Sean $(g,g^*),(g',g'^*)$ capacidades de orden dos. Entonces

$$g'\subset g \Leftrightarrow g'\leq g$$

Demostración:

Se deduce inmediatamente de la proposición anterior y el corolario 4.4. #

Asi pues, para capacidades de orden dos, nuestras dos definiciones de inclusión son equivalentes.

Finalmente analizamos la relación entre nuestras definiciones y las que se conocen para el caso particular de medidas de evidencia.

Yager [61] y Moral [38], independientemente, han definido una misma relación de inclusión para evidencias, fundamentada en la idea de que una información adicional compatible con la existente debe producir una atomización de la masa de evidencia.

Definición 4.5

Sean m_1 y m_2 dos asignaciones básicas de probabilidad. La evidencia representada por m_1 está incluida en la evidencia representada por m_2 ($m_1 \subseteq m_2$) si

$$\forall A \subseteq X \quad \exists \; m_A : \mathcal{P}(A) \longrightarrow [0,1] \text{ tal que}$$

$$\sum_{B \subseteq A} m_A(B) = m_1(A) \quad \text{y} \quad \sum_{B \subseteq A \subseteq X} m_A(B) = m_2(B).$$

Esta inclusión es también una relación de orden parcial con elemento mínimo la asignación m_0 que representa la ignorancia total, y elementos maximales que son las asignaciones asociadas a las medidas de probabilidad.

También se verifica que si $m_1 \subseteq m_2$ entonces

$$bel_1(A) \le bel_2(A) \le Pl_2(A) \le Pl_1(A) \quad \forall A \subseteq X,$$

pero el recíproco no es cierto en general. En Delgado y Moral [13] se prueba el recíproco para medidas de posibilidad, por lo que esta relación de inclusión generaliza la inclusión habitual de medidas de posibilidad (o conjuntos difusos).

Dubois y Prade [22] distinguen entre una inclusión fuerte (la de Yager o Moral) y una débil definida por la relación

$$m_1 \subseteq m_2 \iff bel_1(A) \le bel_2(A) \le Pl_2(A) \le Pl_1(A) \quad \forall A \subseteq X.$$

Salvo una inversión en el orden que establece, esta última definición coincide con las nuestras para el caso particular de evidencias, en virtud de las caracterizaciones que proporcionan el corolario 4.4 y la proposición 4.19. Las relaciones existentes entre nuestras definiciones y la inclusión fuerte son las siguientes:

a) Para medidas de posibilidad todas las definiciones son equivalentes.

b) Para evidencias cualesquiera la inclusión fuerte implica las inclusiones definidas en esta memoria.

c) Para otro tipo de medidas difusas, las nuestras son las únicas definiciones que tienen sentido.

Pensamos que nuestra formulación es preferible, puesto que es más sencilla (sobre todo para medidas representables y en especial para capacidades), menos restrictiva, mucho más general, y posee una sólida fundamentación basada en la caracterización de las medidas difusas por probabilidades, a la vez que un valor intuitivo.

4.3. COMBINACION DE MEDIDAS DIFUSAS.

En este apartado estudiamos formas de combinar informaciones representadas mediante medidas difusas. Estos métodos son específicos para medidas representables, obteniéndose como resultado de la combinación otra medida de este tipo. No obstante, tales métodos podrían aplicarse a medidas difusas generales, aproximando éstas mediante sus capacidades asociadas (que son medidas representables), en caso de no disponer de otros mecanismos de combinación más generales.

La idea en la que se basan los métodos de combinación que expondremos es sencilla: una medida representable genera en $\mathbb{R}^n$ un poliedro convexo de a lo más n! caras paralelas a ciertos hiperplanos. La conjunción de dos medidas representables g y g' es entonces la medida representable asociada al poliedro intersección de los asociados a g y g' (si esta intersección existe), que también tiene sus caras paralelas a los hiperplanos citados (ver figura 1). La disyunción de dos medidas representables g y g' será la medida representable asociada al menor poliedro de lados paralelos a los hiperplanos que contenga a los poliedros asociados a g y g' (ver figura 2). Veamos la definición más formalmente:

<u>Definición 4.6</u>

Sean $(g,g^*),(g',g'^*)$ medidas representables, y sean $\mathcal{E}_g$ y $\mathcal{E}'_g$ sus poliedros asociados.

La conjunción de estas medidas es el par dual $(g\wedge g',g^*\wedge g'^*)$ (notado abreviadamente $(g\wedge g')$), definido por

$$(g\wedge g')(A)=\inf_{P\in\mathcal{E}_g\cap\mathcal{E}'_g} P(A), \qquad (g^*\wedge g'^*)(A)=\sup_{P\in\mathcal{E}_g\cap\mathcal{E}'_g} P(A),$$

siempre que $\mathcal{E}_g\cap\mathcal{E}'_g\neq\emptyset$.

La disyunción de estas medidas es el par dual $(g\vee g',g^*\vee g'^*)$

(notado abreviadamente por ($g\vee g'$)), definido por

$$(g\vee g')(A)=\inf_{P\in\mathcal{C}_g\cup\mathcal{C}'_g} P(A), \qquad (g^{*}\vee g'^{*})(A)=\sup_{P\in\mathcal{C}_g\cup\mathcal{C}'_g} P(A),$$

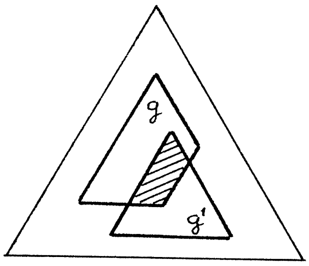


Figura 1.

Conjunción de capacidades

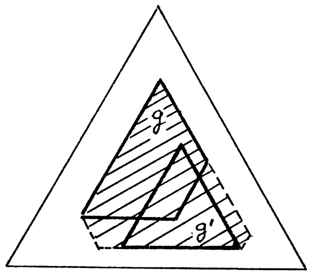


Figura 2.

disyunción de capacidades

<u>Nota:</u> Nuestra intención inicial era obtener mecanismos de combinación para capacidades de orden dos, pero tales métodos no proporcionan otra medida de este tipo. Para obtener una clase cerrada frente a la conjunción y disyunción hemos tenido que considerar la familia de medidas difusas inmediatamente superior en el esquema que vimos en el capítulo I, la clase de las medidas representables.

Existe una estrecha relación entre estas operaciones y la inclusión de medidas representables definida en el apartado anterior, como muestran las siguientes proposiciones.

<u>Proposición 4.20</u>

Sean (g,g^{*}) y (g',g'^{*}) medidas representables. Si existe $(g\wedge g')$, se verifica:

a) $(g\wedge g')$ es el ínfimo de g y g' respecto del orden parcial generado en MR por la relación de inclusión $\subseteq$, es decir

$(g\wedge g')\subseteq g$; $(g\wedge g')\subseteq g'$.

$\forall g''\in MR$ / $g''\subseteq g$ y $g''\subseteq g' \Rightarrow g''\subseteq(g\wedge g')$.

b) $\max(g,g') \leq (g \wedge g') \leq (g^{*} \wedge g'^{*}) \leq \min(g^{*},g'^{*})$.

Demostración:

a) Es evidente, teniendo en cuenta la definición de inclusión y conjunción de medidas representables.

b) se deduce inmediatamente de a) y de la caracterización de la inclusión (dada en la proposición 4.19).#

Obsérvese que si $(\max(g,g'),\min(g^{*},g'^{*}))$ fuese un par representable, entonces coincidiría con la conjunción de (g,g^{*}) y (g',g'^{*}). Como en general no es así, para calcular cada valor $(g \wedge g')(A)$ debemos resolver el problema de programación lineal

$$\min \sum_{x_i \in A} X_i$$

$$\text{sujeto a} \quad \sum_{x_j \in B} X_j \geq \max(g(B), g'(B)) \quad \forall B \subseteq X.$$

Proposición 4.21

Sean (g,g^{*}) y (g',g'^{*}) medidas representables. Se verifica:

a) $(g \vee g')$ es el supremo de g y g' respecto del orden parcial generado en MR por la relación de inclusión $\subseteq$, es decir

$g \subseteq (g \vee g')$; $g' \subseteq (g \vee g')$.

$\forall g'' \in MR$ / $g \subseteq g''$ y $g' \subseteq g''$ $\Rightarrow$ $(g \vee g') \subseteq g''$.

b) $(g \vee g') = \min(g,g') \leq \max(g^{*},g'^{*}) = (g^{*} \vee g'^{*})$.

Demostración:

a) Es evidente, por la definición de inclusión y disyunción.

b) Se deduce de la proposición 4.19 y del hecho de que $(\min(g,g'),\max(g^{*},g'^{*}))$ es una pareja de medidas representable.#

Así pues, el cálculo de $(g \vee g')$ es mucho más sencillo que el de $(g \wedge g')$, basta con tomar el mínimo de las medidas inferiores o el máximo de las superiores.

Nota: Obsérvese que el conjunto MR, con las operaciones de conjunción y disyunción no tiene estructura de retículo, porque no siempre existe la conjunción (o el ínfimo) de dos medidas representables, debido a que no tiene sentido la "medida vacia", hipotéticamente opuesta a la ignorancia.

La conjunción y la disyunción son ricas en propiedades algebraicas, como muestra la siguiente proposición.

Proposición 4.22

Sean g, g', g'' y g''' medidas representables. Se verifican las siguientes propiedades:

a) Idempotencia: $(g \wedge g) = g$; $(g \vee g) = g$.

b) Asociatividad: $((g \wedge g') \wedge g'') = (g \wedge (g' \wedge g''))$

$$((g \vee g') \vee g'') = (g \vee (g' \vee g'')).$$

c) Conmutatividad: $(g \wedge g') = (g' \wedge g)$; $(g \vee g') = (g' \vee g)$.

d) Si $g \leq g'$ entonces $(g \wedge g') = g$ y $(g \vee g') = g'$.

e) Si g_o es la medida de ignorancia total, entonces

$$(g_o \wedge g) = g \quad \text{y} \quad (g_o \vee g) = g_o \ ,$$

con lo que g_o actua como neutro para la conjunción.

f) Si P_o es una probabilidad degenerada en el punto $x_o \in X$, y (g, g^*) es tal que $g(A) = 0$ si $x_o \notin A$ (ó $g^*(A) = 1$ si $x_o \in A$), entonces $(P_o \wedge g) = P_o$.

g) Monotonía: Si $g \leq g'$ y $g'' \leq g'''$ entonces

$$(g \wedge g'') \leq (g' \wedge g''') \quad \text{y} \quad (g \vee g'') \leq (g' \vee g''').$$

h) $((g \wedge g') \vee g'') = ((g \vee g'') \wedge (g' \vee g''))$.

i) $((g \wedge g'') \vee (g' \wedge g'')) = ((g \vee g') \wedge g'')$.

Demostración:

Las propiedades a), b), c) y d) son evidentes.

e) se deduce de d), ya que $g \leq g_o$.

f) En esas condiciones, $g \leq P_o \leq g^*$, y por tanto $P_o \leq g$.

Nuevamente el resultado se deduce de d).

g) $(g\wedge g'')\leq g\leq g' \Rightarrow (g\wedge g'')\leq g'$; $(g\wedge g'')\leq g''$. Luego

$$(g\wedge g'')\leq(g'\wedge g'').$$

$(g'\wedge g'')\leq g'$; $(g'\wedge g'')\leq g''\leq g''' \Rightarrow (g'\wedge g'')\leq g'''$. Por tanto

$$(g'\wedge g'')\leq(g'\wedge g'''),$$

y así $(g\wedge g'')\leq(g'\wedge g'')\leq(g'\wedge g''')$.

Con la disyunción se procede analogamente.

h) Puesto que $(\mathcal{C}_g\cap\mathcal{C}'_g)\cup\mathcal{C}''_g=(\mathcal{C}_g\cup\mathcal{C}''_g)\cap(\mathcal{C}'_g\cup\mathcal{C}''_g)$, el resultado es evidente.

i) Como $(\mathcal{C}_g\cap\mathcal{C}''_g)\cup(\mathcal{C}'_g\cap\mathcal{C}''_g)=(\mathcal{C}_g\cup\mathcal{C}'_g)\cap\mathcal{C}''_g$, se verifica la igualdad propuesta. #

Entre las propiedades que no se verifican destaca la inexistencia de neutro para la disyunción.

Una capacidad de orden dos siempre puede considerarse como la disyunción de sus probabilidades asociadas:

<u>Proposición 4.23</u>

Sea g una capacidad de orden dos y sean P_σ $\sigma\in S_n$ sus probabilidades asociadas. Entonces

$$g=(\bigvee_{\sigma\in S_n} P_\sigma).$$

<u>Demostración:</u>

Es evidente, ya que $g=\min_\sigma P_\sigma$ y $g^*=\max_\sigma P_\sigma$. #

Como hemos visto, una cuestión importante es la de la existencia o no de la conjunción. Pretendemos determinar en qué casos existe y en cuales no.

<u>Definición 4.7</u>

Dadas las medidas representables (g,g^*) y (g',g'^*), diremos que son compatibles cuando exista $(g\wedge g')$, y que son incompatibles si no existe $(g\wedge g')$.

Es evidente que g y g' serán compatibles cuando sus poliedros asociados tengan intersección no vacía, e incompatibles en caso contrario. A continuación probamos dos caracterizaciones de la compatibilidad.

<u>Proposición 4.24</u>

Sean (g,g^*) y (g',g'^*) medidas representables. Entonces g y g' son compatibles si y solamente si $(\max(g,g'),\min(g^*,g'^*))$ es una pareja de medidas duales que acotan una probabilidad.

<u>Demostración:</u>

Condición necesaria:

Si existe $(g\wedge g')$ entonces $\exists\ P_o \in PR\ /\ P_o \in \mathscr{C}_g \cap \mathscr{C}'_g$. Por tanto

$$g = \min_{P\in\mathscr{C}_g} P \leq P_o \leq \max_{P\in\mathscr{C}_g} P = g^* \quad \text{y} \quad g' = \min_{P\in\mathscr{C}'_g} P \leq P \leq \max_{P\in\mathscr{C}'_g} P = g'^*.$$

Entonces

$$\max(g,g') \leq P_o \leq \min(g^*,g'^*)$$

es decir

$$(\max(g,g'),\min(g^*,g'^*)) \in PA.$$

Condición suficiente:

Si $(\max(g,g'),\min(g^*,g'^*)) \in PA$, entonces $\exists P_o \in PR$ tal que

$$g \leq P_o \leq g^* \quad \text{y} \quad g' \leq P_o \leq g'^*,$$

luego $P_o \in \mathscr{C}_g \cap \mathscr{C}'_g$, y por tanto $\mathscr{C}_g \cap \mathscr{C}'_g \neq \emptyset$. #

Podemos dar una caracterización más sencilla basada en un teorema de separación de convexos:

<u>Proposición 4.25</u>

Sean (g,g^*) y (g',g'^*) medidas representables. Entonces son compatibles si y solamente si

$$g'(A) \leq g^*(A) \quad \forall A \subseteq X$$

o equivalentemente

$$g(A) \leq g'^*(A) \quad \forall A \subseteq X.$$

Demostración:

g y g' son incompatibles si y solo si las figuras asociadas son disjuntas. Como tales figuras son convexos cerrados, serán disjuntos si y solo si son separables, es decir, si

$$\exists\alpha\in\mathbb{R}\quad \exists(a_1,a_2,\ldots,a_n)\in\mathbb{R}^n \;/\; \sum_{i=1}^{n} a_i p(x_i)<\alpha \quad \forall P\in\mathcal{C}_g$$
$$\text{y} \quad \sum_{i=1}^{n} a_i p'(x_i)>\alpha \quad \forall P'\in\mathcal{C}'_g \,.$$

Como esos convexos tienen lados paralelos a $\sum_{i\in I} X_i=c_I$, $I\subseteq\{1,\ldots,n\}$, entonces como hiperplano separador podemos escoger uno paralelo a alguno de los anteriores. Así, g y g' son incompatibles $\Leftrightarrow \exists\alpha\in\mathbb{R},\ \exists I\subseteq\{1,\ldots,n\} \;/\; \sum_{i\in I} p(x_i)<\alpha \quad \forall P\in\mathcal{C}_g$ y $\sum_{i\in I} p'(x_i)>\alpha \quad \forall P'\in\mathcal{C}'_g$ $\Leftrightarrow$

$\Leftrightarrow \exists\alpha\in\mathbb{R}\,,\ \exists A\subseteq X \;/\; P(A)<\alpha \ \forall P\in\mathcal{C}_g$ y $P'(A)>\alpha \ \forall P'\in\mathcal{C}'_g \Leftrightarrow \exists\alpha\in\mathbb{R}\,,\ \exists A\subseteq X \;/\; \max_{P\in\mathcal{C}_g} P(A)<\alpha<\min_{P'\in\mathcal{C}'_g} P'(A) \Leftrightarrow \exists\alpha\in\mathbb{R}\,,\ \exists A\subseteq X \;/\; g^*(A)<\alpha<g'(A) \Leftrightarrow \exists A\subseteq X \;/\; g^*(A)<g'(A)$.

Por tanto g y g' son compatibles si y solo si

$$\forall A\subseteq X \quad g'(A)\leq g^*(A).$$

La otra equivalencia se prueba de igual forma.#

La caracterización dada por esta última proposición es a la vez potente y razonable. En efecto, por un lado proporciona una manera muy simple de comprobar si dos medidas representables son compatibles o no, y por otro responde a la idea intuitiva de que cuando dos medidas representables son incompatibles es porque no existe ninguna medida de probabilidad (y por consiguiente ninguna medida representable) compatible simultáneamente con las dos implicadas. Si, como es la norma que hemos adoptado, ambas informaciones se consideran ciertas, no cabe sino concluir en tal caso que las dos fuentes son contradictorias.

Resulta ilustrativo analizar los casos particulares de medidas de tipo crisp, de posibilidades y de evidencias.

Corolario 4.8

Sean g_B y $g_{B'}$ dos medidas difusas de tipo crisp, focalizadas en los subconjuntos B y B' respectivamente.

g_B y $g_{B'}$ son compatibles si y solamente si $B\cap B'\neq\emptyset$.

Demostración:

Las medidas g_B y $g^*_{B'}$ están definidas como

$$g_B(A)=\begin{cases}1 & \text{si } B\subseteq A\\ 0 & \text{en otro caso}\end{cases} \qquad g^*_{B'}(A)=\begin{cases}1 & \text{si } B'\cap A\neq\emptyset\\ 0 & \text{en otro caso}\end{cases}$$

Aplicando la proposición anterior, g_B y $g_{B'}$ son compatibles si y solo si $g_B(A)\leq g^*_{B'}(A)\ \forall A\subseteq X \iff$

$$\iff \Big[\text{ si } B\subseteq A \Rightarrow B'\cap A\neq\emptyset, \text{ y si } B'\cap A=\emptyset \Rightarrow B \text{ no contenido en } A\Big] \iff$$

$\iff B\cap B'\neq\emptyset$. #

Las combinaciones de medidas de tipo crisp vuelven a ser medidas del mismo tipo:

Proposición 4.26

Sean g_B y $g_{B'}$ medidas difusas de tipo crisp focalizadas en los subconjuntos B y B' respectivamente. Entonces

a) $(g_B\vee g_{B'})=g_{B\cup B'}$, medida crisp focalizada en $B\cup B'$.

b) Si $B\cap B'\neq\emptyset$, $(g_B\wedge g_{B'})=g_{B\cap B'}$, medida crisp focalizada en $B\cap B'$.

Demostración:

a) Se deduce inmediatamente por la proposición 4.21.b, ya que $\min(g_B,g_{B'})=g_{B\cup B'}$

b) Se deduce de que $\max(g_B,g_{B'})$ es una medida representable, y $\max(g_B,g_{B'})=g_{B\cap B'}$. #

Es importante resaltar la coherencia de la disyunción y la conjunción que hemos definido entre medidas representables con las operaciones clásicas de unión e intersección de conjuntos, puesta de manifiesto a través de las medidas de tipo crisp. Nuestras operaciones parecen extensiones naturales en la línea de interpretar las medidas de tipo crisp a partir de la función de pertenencia de un conjunto crisp del referencial.

Si esta idea se extiende a subconjuntos difusos, esto es, a medidas de posibilidad, la coherencia se mantiene para la disyunción en relación con la unión de subconjuntos difusos:

<u>Proposición 4.27</u>

Si Π_A y Π_B son medidas de posibilidad, correspondientes a los subconjuntos difusos normalizados $\underset{\sim}{A}$ y $\underset{\sim}{B}$, entonces

$$(\Pi_A \vee \Pi_B) = \Pi_{A \cup B} \text{ ,}$$

medida de posibilidad correspondiente al subconjunto difuso normalizado $\underset{\sim}{A} \cup \underset{\sim}{B}$.

<u>Demostración:</u>

Por la proposición 4.21, el resultado es evidente, ya que el máximo de dos medidas de posibilidad (o el mínimo de dos medidas de necesidad) es otra medida de posibilidad (otra medida de necesidad, respectivamente). #

En cuanto se refiere a la intersección de subconjuntos difusos, la correspondencia se rompe al no tener garantizada la normalización de aquella, y por tanto su equivalencia a una medida de posibilidad.

Con respecto a las medidas de evidencia, analizaremos en particular el comportamiento de la conjunción. La combinación de informaciones a través de las medidas de evidencia que las expresan ha sido y es un tema destacado de investigación, con amplia trascendencia en la elaboración de modelos de la realidad. Partiendo de la idea de extraer la información común a dos evidencias supuestas verdaderas, Dempster [16] (ver también Shafer [46] y Ruspini [45]) dió su conocida regla, que puede resumirse de la forma siguiente:

Supongamos que dos evidencias representadas por sus a.b.p. m_1 y m_2 tienen elementos focales $A_1,..,A_k$ y $B_1,..,B_l$ respectivamente. Si no existe contradicción total entre ellas, es decir si

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

entonces la combinación de m_1 y m_2 según la regla de Dempster es otra evidencia con a.b.p.

$$m_1 \oplus m_2(A) = \frac{\displaystyle\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \displaystyle\sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

donde $m_1 \oplus m_2(A)=0$ si $A_i \cap B_j \neq A$ $\forall i=1,..,k,\ j=1,..,l$.

La regla de Dempster verifica las propiedades b), c), e) y f) de la proposición 4.22 correspondientes a la conjunción, no así las propiedades a), d) y g). Esta regla funciona correctamente en muchas situaciones, pero no se pueden combinar evidencias sin tener en cuenta la forma efectiva en que estas interaccionan. Si por ejemplo disponemos de dos a.b.p. idénticas $m_1 \equiv m_2 \equiv m$, no es razonable combinarlas con la regla de Dempster ya que, debido a la falta de idempotencia, en general $m \oplus m \neq m$.

Como ya se ha mencionado, nuestra conjunción permite una interpretación semántica en la misma línea, esto es, la búsqueda de una información común a dos, supuestas verdaderas. Funciona válidamente como mecanismo de combinación, extensible a un campo mucho más general que el de las evidencias, y, al menos en términos algebraicos, presenta mejores propiedades que la regla de Dempster.

Algunos ejemplos serán de utilidad para analizar las diferencias conceptuales:

Ejemplo 4.6

Sean, sobre $X=\{x_1, x_2, x_3\}$ las evidencias (g_1, g_1^*) y (g_2, g_2^*) definidas por

A	$g_1(A)$	$g_1^*(A)$	$g_2(A)$	$g_2^*(A)$	$m_1(A)$	$m_2(A)$
$\{x_1\}$	0.4	0.6	0.42	0.72	0.4	0.42
$\{x_2\}$	0.4	0.6	0.28	0.58	0.4	0.28
$\{x_3\}$	0	0.2	0	0.3	0	0
$\{x_1, x_2\}$	0.8	1	0.7	1	0	0
$\{x_1, x_3\}$	0.4	0.6	0.42	0.72	0	0
$\{x_2, x_3\}$	0.4	0.6	0.28	0.58	0	0
$\{x_1, x_2, x_3\}$	1	1	1	1	0.2	0.3

Las probabilidades asociadas a (g_1, g_1^*) son

(0.4, 0.4, 0.2), (0.4, 0.6, 0) y (0.6, 0.4, 0)

y las asociadas a (g_2, g_2^*) son

(0.42, 0.28, 0.3), (0.72, 0.28, 0) y (0.42, 0.58, 0).

Las combinaciones $(g_1 \wedge g_2)$ y $g_1 \oplus g_2$ son

A	$(g_1 \wedge g_2)$	$(g_1 \wedge g_2)^*$	$m_\wedge$	$g_1 \oplus g_2$	$(g_1 \oplus g_2)^*$	$m_\oplus$
$\{x_1\}$	0.42	0.6	0.42	0.516	0.6	0.516
$\{x_2\}$	0.4	0.58	0.4	0.4	0.483	0.4
$\{x_3\}$	0	0.18	0	0	0.083	0
$\{x_1, x_2\}$	0.82	1	0	0.916	1	0
$\{x_1, x_3\}$	0.42	0.6	0	0.516	0.6	0
$\{x_2, x_3\}$	0.4	0.58	0	0.4	0.483	0
$\{x_1, x_2, x_3\}$	1	1	0.18	1	1	0.083

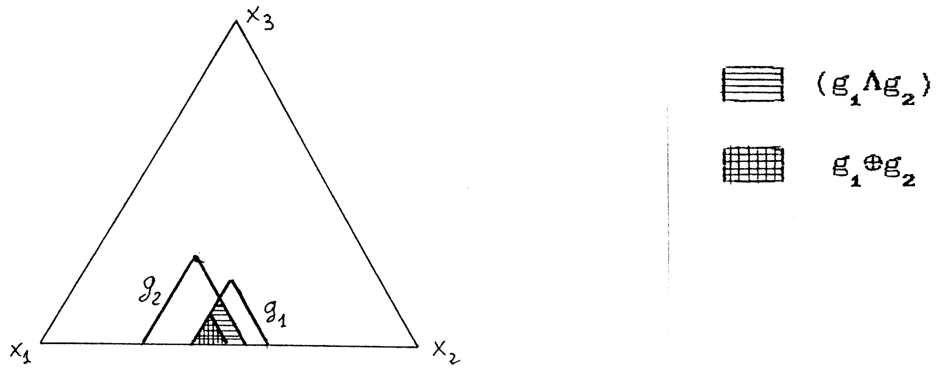
con probabilidades asociadas a $(g_1 \wedge g_2)$

(0.42, 0.58, 0), (0.6, 0.4, 0) y (0.42, 0.4, 0,18)

y las asociadas a $g_1 \oplus g_2$

(0.516, 0.4, 0.083), (0.516, 0.483, 0) y (0.6, 0.4, 0).

En este caso $g_1 \oplus g_2 \subseteq (g_1 \wedge g_2)$



No siempre se verifica esa relación de inclusión; veamos otro ejemplo:

<u>Ejemplo 4.7</u>

Sean, sobre $X=\{x_1, x_2, x_3\}$ las evidencias (g_1, g_1^*) y (g_2, g_2^*) definidas por

A	$g_1(A)$	$m_1(A)$	$g_2(A)$	$m_2(A)$
$\{x_1\}$	0.5	0.5	0	0
$\{x_2\}$	0	0	0.4	0.4
$\{x_3\}$	0	0	0	0
$\{x_1, x_2\}$	0.5	0	0.4	0
$\{x_1, x_3\}$	0.5	0	0	0
$\{x_2, x_3\}$	0	0	0.4	0
$\{x_1, x_2, x_3\}$	1	0.5	1	0.6

con probabilidades asociadas a g_1

$$(0.5,\ 0,\ 0.5),\ (0.5,\ 0.5,\ 0)\ y\ (1,\ 0,\ 0)$$

y asociadas a g_2

$$(0,\ 0.4,\ 0.6),\ (0.6,\ 0.4,\ 0)\ y\ (0,\ 1,\ 0).$$

En este caso, las combinaciones $(g_1 \wedge g_2)$ y $g_1 \oplus g_2$ son

A	$(g_1 \wedge g_2)$	$m_\wedge$	$g_1 \oplus g_2$	$m_\oplus$
$\{x_1\}$	0.5	0.5	0.375	0.375
$\{x_2\}$	0.4	0.4	0.25	0.25
$\{x_3\}$	0	0	0	0
$\{x_1, x_2\}$	0.9	0	0.625	0
$\{x_1, x_3\}$	0.5	0	0.375	0
$\{x_2, x_3\}$	0.4	0	0.25	0
$\{x_1, x_2, x_3\}$	1	0.1	1	0.375

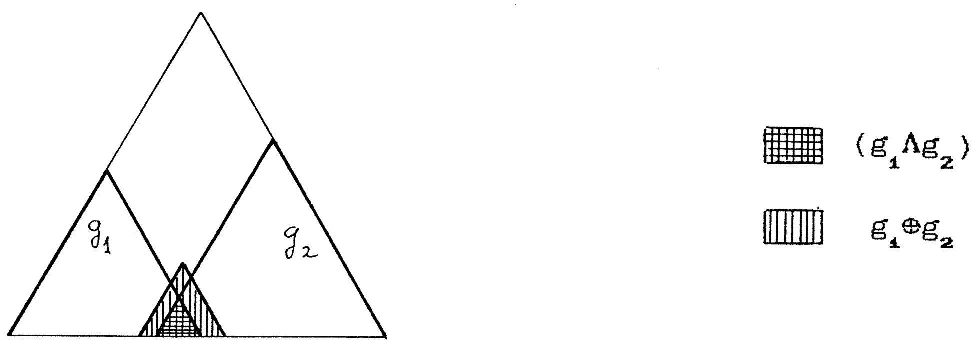
con probabilidades asociadas a $(g_1 \wedge g_2)$

$$(0.6,\ 0.4,\ 0),\ (0.5,\ 0.5,\ 0)\ y\ (0.5,\ 0.4,\ 0.1)$$

y asociadas a $g_1 \oplus g_2$

$$(0.375,\ 0.25,\ 0.375),\ (0.375,\ 0.625,\ 0)\ y\ (0.75,\ 0.25,\ 0),$$

y ahora se verifica $(g_1 \wedge g_2) \subseteq g_1 \oplus g_2$.



A la vista de estos ejemplos queda claro que ninguno de los métodos de combinación es más preciso que el otro: dependiendo del caso un método u otro proporcionará una medida de combinación más ajustada.

Otro tema importante es cuándo se pueden combinar dos medidas con ambos métodos. En este aspecto, la regla de Dempster es mucho menos restrictiva que nuestra conjunción.

Proposición 4.28

Consideremos dos evidencias con a.b.p. m_1 y m_2 . Si no se pueden combinar mediante la regla de Dempster, entonces son incompatibles, y en consecuencia no se pueden combinar mediante la conjunción.

Demostración:

Si m_1 y m_2 no son combinables con la regla de Dempster, entonces

$$\sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)=1$$

donde A_i y B_j son los elementos focales de m_1 y m_2 respectivamente. En este caso $A_i \cap B_j = \emptyset \ \forall i,j$.

Tomemos un elemento focal A_k de m_1. Dada la condición anterior

$$Pl_2(A_k)=\sum_{B_j \cap A_k \neq \emptyset} m_2(B_j)=0<m_1(A_k)\leq bel_1(A_k),$$

y por la proposición 4.25 estas dos evidencias son incompatibles.#

En nuestra opinión, la regla de Dempster es demasiado permisiva (y nuestra conjunción demasiado restrictiva). Puede combinar medidas con muy alto grado de contradicción, tratando de compatibilizarlas a toda costa, con lo que el resultado no es adecuado.

Por ejemplo, si tenemos una evidencia crisp focalizada en un subconjunto B, y una probabilidad P, de modo que $P(B)>0$, si $P(B)<1$ nuestro método de combinación considera incompatibles esas medidas. La combinación con la regla de Dempster proporciona la probabilidad P condicionada al subconjunto B. Así, si consideramos en $X=\{x_1,x_2,x_3\}$ la medida crisp g focalizada en $B=\{x_2,x_3\}$, y $p(x_1)=0.9$, $p(x_2)=p(x_3)=0.05$, $g\oplus P=P(./B)$, es decir

$$p(x_1/B)=0,\ p(x_2/B)=p(x_3/B)=0.5,$$

que no parece una combinación razonable de g y P.

Nuestro método de combinación es más prudente, y no da lugar a este tipo de situaciones, pero es demasiado restrictivo y declara incompatibles medidas que no son apenas contradictorias.

Si consideramos ahora la medida crisp anterior g, y la probabilidad P definida por $p(x_1)=0.01$, $p(x_2)=0.49$ y $p(x_3)=0.5$, g y P son incompatibles y no se pueden combinar, pero no son informaciones contradictorias.

Así pues, sería deseable encontrar mecanismos de combinación para el caso de incompatibilidad, que seguramente deberían estar relacionados con alguna medida del grado de contradicción entre dos informaciones. Tal vez la consideración de medidas difusas no normalizadas podría ser un mecanismo adecuado para abordar este problema, que nos proponemos estudiar en el futuro.

BIBLIOGRAFIA.

[1] Banon, G. (1981)
Distinction between several subsets of fuzzy measures. Fuzzy Sets and Systems 5, pp. 325-339.

[2] Batle, N. y Trillas, E. (1979)
Entropy and fuzzy integral. Journal of Math. An. and Appl. 69, pp. 469-474.

[3] Beran, R. J. (1970)
Upper and lower risk and minimax procedures. Proc. 6th. Berkeley Symp. on Math. Statist. and Prob. vol.I, pp. 1-16.

[4] Bolaños, M. J. (1984)
Caracterización de medidas y valoraciones difusas a partir de medidas ordinarias. Tesis doctoral. Universidad de Granada.

[5] Bolaños, M. J., Lamata, M. T. y Moral, S. (1985)
La esperanza monótona: una generalización de la esperanza probabilística. XV Reunión Nacional de Estadística, I.O. e Informática. Gijón.

[6] Bolaños, M. J., Lamata, M. T. y Moral, S. (1985)
Alternativas para la extensión de medidas difusas. XV Reunión Nacional de Estadística, I.O. e Informática. Gijón.

[7] Bolaños, M. J., Lamata, M. T. y Moral, S. (1985)
A decision model under general information. Busefal 24, pp.56-63.

[8] Bolaños, M. J. y de Campos, L. M. (1986)
Medidas e integración difusa: aplicación a la valoración de tumores malignos del S.N.C.. I Conferencia Española de Biometría. Granada.

[9] Burbea, J. y Rao, C. R. (1982)
Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. Journal of Multivariate Analysis 12, pp. 575-596.

[10] Butnariu, D. (1983)
Additive fuzzy measures and integrals I. Journal of Math. An. and Appl. 93, pp. 436-452.

[11] Choquet, G. (1953)
Theory of capacities. Ann. Inst. Fourier 5, pp. 131-295.

[12] Delgado, M. y Verdegay, J. L. (1984)
On the concept of fuzzy number. Proc. FISAL-84, pp. 79-90.

[13] Delgado, M. y Moral, S. (1985)
A definition of inclusion for evidences. Fuzzy Mathematics. Por aparecer.

[14] De Luca, A. y Termini, S. (1972)
A definition of a non probabilistic entropy in the setting of fuzzy sets Theory. Information and Control 20, pp. 301-312.

[15] De Luca, A. y Termini, S. (1979)
Entropy and energy measures of a fuzzy set. Advances in Fuzzy Sets Theory and Applications, M.M. Gupta, R.K. Ragade y R.R. Yager (eds). North-Holland, pp. 321-338.

[16] Dempster, A.P. (1967)
Upper and lower probabilities induced by a multivaluated mapping. Ann. Math. Stat. 38, pp. 325-339.

[17] Dubois, D. y Prade, H. (1980)
Fuzzy sets and systems. Theory and applications. Academic Press.

[18] Dubois, D. y Prade, H. (1982)
On several representations of an uncertain body of evidence. Fuzzy Information and Decision Processes, M.M. Gupta y E. Sanchez (eds). North-Holland, pp. 167-182.

[19] Dubois, D. y Prade, H. (1983)
Unfair coins and necessity measures: towards a possibilistic interpretation of histograms. Fuzzy Sets and Systems 10, pp. 15-20.

[20] Dubois, D. y Prade, H. (1984)
A note on measures of specificity for fuzzy sets. Busefal 19, pp. 83-89.

[21] Dubois, D. y Prade, H. (1985)
Théorie des possibilités. Applications à la représentation des connaissances en informatique. Masson.

[22] Dubois, D. y Prade, H. (1986)
A set-theoretic view of belief functions-logical operations and approximations by fuzzy sets. Int. J. of General Systems. Por aparecer.

[23] Giles, R. (1982)
Foundations for a theory of possibility. Fuzzy Information and Decision Processes, M.M. Gupta y E. Sanchez (eds). North-Holland, pp. 183-195.

[24] Gupta, M.M., Nikiforuk, P.N., Tsukamoto, Y. y Martin-Clouaire, R. (1982)
An application of fuzzy integral to medical diagnosis. IFAC Theory and Applications of Digital Control, pp. 449-454. New Delhi, India.

[25] Higashi, M. y Klir, G. J. (1983)
Measures of uncertainty and information based on possibility distributions. Int. J. of General Systems 9, pp. 43-58.

[26] Höhle, U. (1981)

Fuzzy measures as extensions of stochastic measures. Proc. of the Second International Seminar on Fuzzy Sets Theory. Johannes Kepler Universität, Linz, Austria.

[27] Höhle, U. (1982)

A mathematical Theory of uncertainty. Fuzzy Sets and Possibility Theory. Recent Developments, R.R. Yager (ed). Pergamon Press, pp. 344-355.

[28] Huber, P. J. (1973)

The use of Choquet capacities in statistics. Bull. Inter. Statist. Institute, vol XLV, pp. 181-188.

[29] Huber, P. J. (1981)

Robust Statistics. Wiley.

[30] Kandel, A. (1979)

On fuzzy statistics. Advances in Fuzzy Sets Theory and Applications, M.M. Gupta, R.K. Ragade y R.R. Yager (eds). North-Holland, pp. 181-198.

[31] Kandel, A. (1982)

Fuzzy Techniques in Pattern Recognition. Wiley Interscience.

[32] Klement, E. P. (1980)

Fuzzy σ-algebras and fuzzy measurable functions. Fuzzy Sets and Systems 4, pp. 83-93.

[33] Klement, E. P., Lowen, R. y Schwyhla, W. (1981)
Fuzzy probability measures. Fuzzy Stes and Systems 5, pp. 21-30.

[34] Klement, E. P. y Ralescu, D. (1983)
Non linearity of the fuzzy integral. Fuzzy Sets and Systems 11, pp. 309-316.

[35] Kruse, R. (1982)
A note on λ-additive fuzzy measures. Fuzzy Sets and Systems 8, pp. 219-222.

[36] Kruse, R. (1983)
Fuzzy integrals and conditional fuzzy measures. Fuzzy Sets and Systems 10, pp. 309-313.

[37] Lamata, M. T. (1985)
Modelos de decisión con información general. Tesis doctoral. Universidad de Granada.

[38] Moral, S. (1985)
Información difusa, relaciones entre probabilidad y posibilidad. Tesis doctoral. Universidad de Granada.

[39] Natvig, B. (1983)
Possibility versus probability. Fuzzy Sets and Systems 10, pp. 299-310.

[40] Nguyen, H. T. (1979)
Toward a calculus of the mathematical notion of possibility. Advances in Fuzzy Sets Theory and Applications, M.M. Gupta, R.K. Ragade y R.R. Yager (eds). North-Holland, pp. 235-246.

[41] Nguyen, H. T. (1982)
Possibility measures and related topics. Fuzzy Information and Decision Processes, M.M. Gupta y E. Sanchez (eds). North-Holland, pp. 197-201.

[42] Puri, L. y Ralescu, D. (1982)
A possibility measure is not a fuzzy measure. Fuzzy Sets and Systems 7, pp. 311-313.

[43] Ralescu, D. y Adams, G. (1980)
The fuzzy integral. Journal of Math. An. and Appl. 75, pp. 562-570.

[44] Ralescu, D. (1982)
Toward a general theory of fuzzy variables. Journal of Math. An. and Appl. 86, pp. 176-193.

[45] Ruspini, E. H. (1986)
The logical foundations of evidential reasoning. Technical note 408, SRI International.

[46] Shafer, G. (1976)
A mathematical Theory of evidence. Princeton Univ. Press.

[47] Shafer, G. (1982)
Belief functions and parametric models. J. R. Statistical Soc. 44, pp. 322-352.

[48] Schweizer, B. y Sklar, A. (1963)
Associative functions and abstract semi-groups. Publ. Math. Debrecen 10, pp. 69-81.

[49] Schweizer, B. y Sklar, A. (1983)
Probabilistic metric spaces. North-Holland.

[50] Smets, P. (1981)
The degree of belief in a fuzzy event. Information Sci. 25, pp. 1-19.

[51] Suárez, F. (1983)
Familias de integrales difusas y medidas de entropía relacionadas. Tesis doctoral. Universidad de Oviedo.

[52] Sugeno, M. (1974)
Theory of fuzzy integrals and its applications. Ph.D. Tesis. Tokio Inst. of Technology.

[53] Vila, M. A. y Delgado, M. (1983)
On medical diagnosis using possibility measures. Fuzzy Sets and Systems 10, pp. 211-222.

[54] Wang, Z. (1984)
The autocontinuity of set functions and the fuzzy integral. Journal of Math. An. and Appl. 99, pp. 195-218.

[55] Weber, S. (1984)
⊥-Decomposable measures and integrals for archimedean t-conorms ⊥. Journal of Math. An. and Appl. 101, pp. 114-138.

[56] Weber, S. (1986)
Two integrals and some modified versions. Critical remarks. Fuzzy Sets and Systems 20, pp. 97-105.

[57] Wolf, G. (1977)
Obere und untere Wahrscheinlichkeiten. Ph.D. Thesis, Eidgen Technische Hochschule, Zurich.

[58] Yager, R. R. (1982)
Measuring tranquility and anxiety in decision making; and application of fuzzy sets. Int. J. of General Systems 8, pp. 139-146.

[59] Yager, R. R. (1983)
Entropy and specificity in a mathematical Theory of evidence. Int. J. of General Systems 9, pp. 249-260.

[60] Yager, R. R. (1983)
An introduction to applications of possibility theory. Human Systems Management 3, pp. 246-253.

[61] Yager, R. R. (1985)
The entailment principle for Dempster-Shafer granule. Tech. Report MII-512. Iona College, New Rochelle.

[62] Zadeh, L. A. (1978)

Fuzzy sets as a basis for a Theory of possibility. Fuzzy Sets and Systems 1, pp. 3-28.

[63] Zi-Xiao, W. (1984)

Fuzzy measures and measures of fuzziness. Journal of Math. An. and Appl. 104, pp. 589-601.