



Universidad de Oviedo

PROGRAMA DE DOCTORADO: MATEMÁTICAS Y
ESTADÍSTICA

APROXIMACIÓN DE LOS CONJUNTOS
ALEATORIOS BORROSOS DESDE UNA
INTERPRETACIÓN ÓNTICA

TESIS DOCTORAL

LAURA GARRIDO BLANCO

Índice general

Introducción	III
1. Comparación de conjuntos difusos	1
1.1. Introducción y conceptos preliminares	1
1.2. Lista de axiomas	6
1.2.1. Propiedades generales	7
1.2.2. Axiomas para medidas de “igualdad”	8
1.2.3. Axiomas para medidas de “desigualdad” o “diferencia”	10
1.3. Medidas de “igualdad”	11
1.3.1. Definiciones axiomáticas de medida de similaridad	11
1.3.2. Relación formal entre las diferentes medidas	13
1.3.3. Otras definiciones de medida de igualdad	26
1.4. Medidas de “desigualdad” o “diferencia”	31
1.4.1. Definiciones axiomáticas en la literatura	31
1.4.2. Relación formal entre las diferentes medidas	33
1.4.3. El caso particular de los conjuntos nítidos	45
1.5. Dualidad	52
2. El centroide y la d-variación	59
2.1. Motivación y conceptos previos	59
2.2. El centroide.	67
2.2.1. Propiedades del centroide	70
2.3. La d -variación o d -dispersión	86
2.3.1. Propiedades de la d -variación	93
3. Conjuntos Rugosos	99
3.1. Conceptos preliminares	102
3.1.1. Conjuntos aleatorios	102

3.1.2. Variables aleatorias difusas	109
3.1.3. Conjuntos rugosos	110
3.1.4. Conjuntos rugoso difusos	111
3.2. Conjuntos aleatorios rugosos	114
3.3. Probabilidad inducida por un c.a.	119
3.4. Variables aleatorias difusas rugosas	122
3.4.1. Caracterización de la probabilidad inducida por una variable aleatoria difusa	124
3.4.2. Algunos casos especiales	125
3.5. Centroide y dispersión esperada	133
3.6. Rugosidad	136
3.6.1. La diferencia de entropías como grado de “inexactitud”	139
3.6.2. El grado de rugosidad de un conjunto	143
3.6.3. El grado de rugosidad de un conjunto borroso	153
3.6.4. El grado de rugosidad de un conjunto aleatorio y de una variable aleatoria difusa	157
Conclusiones y problemas abiertos	159
Agracecimientos	167

Introducción

El estudio de los conjuntos aleatorios se ha intensificado desde los años 70, con la aparición de los trabajos de Kendall [57] y Matheron [68] en distintas ramas del cálculo de probabilidades y la estadística y aplicaciones derivadas de ellas. Un conjunto aleatorio es una aplicación multi-valuada definida sobre un espacio de probabilidad y que satisface cierta condición de medibilidad que permite trasladar dicha medida de probabilidad, mediante algún mecanismo, al espacio de llegada. Cada uno de los distintos campos de aplicación de los conjuntos aleatorios está conectado con una de las dos visiones de los conjuntos: las visiones “conjuntiva” y “disyuntiva” ([37, 40]). Bajo la perspectiva “conjuntiva”, el conjunto representa información totalmente precisa acerca de un ente complejo, como puede ser, por ejemplo, el conjunto de idiomas que domina una persona. Sin embargo, de acuerdo con la visión “disyuntiva” el conjunto representa información incompleta acerca de cierto elemento del universo: todo lo que conocemos de él es que pertenece al conjunto, y todos los elementos del mismo son posibles candidatos. Por un lado, tanto la Geometría Estocástica [4, 57, 58], como algunas aplicaciones de los conjuntos aleatorios en el campo de la Economía ([12]), están relacionadas con la primera, mientras que la utilización de los conjuntos aleatorios como representaciones de la observación incompleta o imprecisa de las variables aleatorias

([14, 30, 31, 33, 42, 63, 64, 69, 70, 71, 88]) está conectada con la segunda. Dentro de la visión conjuntiva, el conjunto aleatorio es interpretado como un objeto aleatorio cuyos resultados o salidas representan cierta información totalmente precisa acerca de los elementos del espacio inicial, mediante subconjuntos del espacio final. Dentro de este enfoque, el interés del investigador se centra generalmente, en la caracterización de la distribución de probabilidad inducida por el conjunto aleatorio en cierta σ -álgebra formada por familias de subconjuntos del espacio final. De forma alternativa, dentro de la visión disyuntiva, el conjunto aleatorio es interpretado como una familia variables aleatorias (sus llamadas “selecciones medibles”), una de las cuales representa el verdadero comportamiento de la característica estudiada en los elementos del espacio inicial. Dicha característica se modela mediante una variable aleatoria cuyas imágenes se observan con imprecisión, de forma que la información disponible acerca de cada una de ellas es un conjunto más o menos preciso de posibles valores a los que pertenece con seguridad. El interés se centra, en este caso, en caracterizar la información disponible acerca de la distribución de probabilidad (o de algún parámetro determinado asociado a ella) inducida por la variable aleatoria “original” observada de forma incompleta o imprecisa.

El concepto de variable aleatoria difusa ([41]) es más reciente, y generaliza el concepto de conjunto aleatorio. Una variable aleatoria difusa (también llamada “conjunto aleatorio difuso”) es una función definida en un espacio de probabilidad, cuyas imágenes son subconjuntos difusos del espacio final, y que además satisface cierta condición de medibilidad. Los distintos enfoques que se han dado hasta la fecha a las variables aleatorias borro-

sas (véase [19] para una descripción detallada) están conectados con alguna de las dos interpretaciones de los conjuntos borrosos: la interpretación “óptica” y la interpretación “epistémica”. De acuerdo con la interpretación “óptica” el conjunto difuso se entiende como un objeto complejo, en el que cada elemento de cierto universo tiene un mayor o menor grado de participación, cuantificado con un valor en el intervalo unidad. Por ejemplo, una imagen en una escala de grises puede expresarse mediante un conjunto borroso que asocia a cada elemento del universo (píxel) un valor comprendido entre 0 (blanco) y 1 (negro), que indica el grado de intensidad en ese punto. Por otra parte, desde la perspectiva epistémica, el conjunto borroso sirve para representar la información imprecisa disponible acerca de un elemento concreto del universo. Así, por ejemplo, un GPS nos proporciona información incompleta acerca de la localización de un objeto. De acuerdo con las especificaciones habituales de este tipo de herramientas, la información acerca de la localización real viene determinada por una familia de círculos anidados, cada uno de ellos asociado a un nivel de confianza. Cuanto menos preciso es el círculo, mayor es el grado de confianza de que la localización real esté dentro de él. De acuerdo con [28], esa información se puede representar mediante un conjunto borroso que asocia, a cada punto del plano, $\vec{x} \in \mathbb{R}^2$, un número entre 0 y 1. Este número indica el grado de posibilidad de que la localización verdadera del objeto coincida con el punto $\vec{x}$. En la literatura reciente de las variables aleatorias difusas, han aparecido estudios teóricos y aplicados en cada una de las dos vertientes. Así, los estudios iniciados por Puri y Ralescu ([84]) y continuados en por otros autores como Körner ([60, 61]), Lubiano et al. ([66]), Krätschmer ([62]), González-Rodríguez et al. ([50, 49]) o Sinova

et al. [91], se corresponden con la perspectiva óptica, mientras que otros trabajos relacionados con la observación imprecisa de variables aleatorias ([5, 19, 36, 29, 32, 47]), están relacionados con la interpretación epistémica.

En este trabajo, combinaremos la visión óptica de las variables aleatorias borrosas (y, dentro de ella, como caso especial, la visión conjuntiva de los conjuntos aleatorios) con la visión disyuntiva de ciertos “meta”-conjuntos. Nos referimos con este término a conjuntos cuyos elementos son, a su vez, subconjuntos, nítidos o difusos, de un determinado espacio. Nuestro objeto de estudio será una variable aleatoria difusa (o, en particular, un conjunto aleatorio), pero supondremos que no podemos observar de forma directa a los elementos que forman sus imágenes (conjuntos nítidos o difusos), sino que solo podemos percibir el valor (o vector de valores) que les asigna cierto atributo (o cierta colección de atributos). De esta forma, el universo de llegada del conjunto aleatorio o la variable aleatoria borrosa estará implícitamente dividido en una familia de clases de equivalencia, cada una de las cuales contiene a todos los elementos asignados a un valor (o vector de valores) concreto del atributo. Así, no tendremos la posibilidad de diferenciar cada elemento de los demás elementos de su clase de equivalencia, que se considerarán, a todos los efectos, indiscernibles. De esta forma, cuando consideramos un subconjunto arbitrario del universo $A \subseteq U$, no podremos caracterizarlo totalmente, a no ser que sea un conjunto “exacto”, es decir, un conjunto formado por la unión de distintas clases de equivalencia. En el caso de que el conjunto no sea exacto, dado que contiene “trozos” no completos de alguna(s) clase(s) de equivalencia, solo podremos determinar una pareja

de conjuntos exactos que lo acotan, a los que llamaremos sus aproximaciones inferior y superior, de acuerdo con la teoría de los conjuntos rugosos ([80]). La aproximación inferior estará formada por la unión de aquellas clases de equivalencia que están totalmente incluidas en el conjunto. Por su parte, la aproximación superior estará compuesta por la unión de todas las clases de equivalencia que tienen intersección no vacía con el conjunto. De esta forma, todo nuestro conocimiento acerca del conjunto A viene determinado por la familia de subconjuntos del universo que contienen a la aproximación inferior de A y están, a su vez, contenidos en su aproximación superior. Esta última familia de conjuntos o “meta”-conjunto debe ser entendida desde un punto de vista disyuntivo (representa nuestra información incompleta acerca de un conjunto determinado, A), mientras que el conjunto A debe ser interpretado desde un punto de vista conjuntivo.

Sumario

El capítulo 1 se ocupa de la noción de “igualdad” o “similitud” entre conjuntos borrosos, así como del concepto opuesto de “desigualdad” o “disimilitud”. En primer lugar, se proporcionan tres listas separadas de los axiomas que se ajustan a las respectivas nociones de “medida general de comparación”, “medida de igualdad” y “medida de desigualdad”. A continuación, se revisan algunas de las definiciones axiomáticas más importantes de la literatura, estudiadas de forma independiente por distintos grupos de autores, en relación con dichos conceptos, haciendo referencia a los axiomas, dentro de los tres listados mencionados, que satisface cada definición específica. El establecimiento

de este marco común nos permitirá facilitar el estudio formal acerca de las relaciones entre unas y otras definiciones: algunas de ellas, que son aparentemente diferentes, tienen en común aspectos comunes relevantes, tal como vamos a comprobar. El objetivo final del capítulo consiste en presentar un esquema general acerca de estas relaciones, a partir de resultados formales y contraejemplos que reflejan qué definiciones de (di)similitud de la literatura están conectadas por relaciones de implicación y cuáles de ellas, no. El capítulo concluye con un estudio detallado acerca de la dualidad entre las distintas definiciones de “similitud” y “disimilitud” de la literatura. Los resultados obtenidos en este capítulo se encuentran recogidos en la referencia [21].

El capítulo 2 está dedicado al estudio de los conceptos de “medida de tendencia central” y “medida de dispersión”, tanto en el contexto de los conjuntos aleatorios, como en el contexto de las variables aleatorias difusas. Nos servirán como apoyo algunos estudios previos de estos dos conceptos en la literatura. A diferencia de ellos, en nuestro caso, el espacio final no será necesariamente un espacio numérico, ni siquiera un espacio métrico. En este punto, las medidas de “disimilitud” del capítulo 1 representarán un papel fundamental, ya que la dispersión de la variable aleatoria difusa se expresará en función de una medida de este tipo, como la esperanza de las disimilitudes entre las parejas de observaciones de la misma, sin necesidad de que esta medida cumpla las propiedades de las métricas, ni esté definida para parejas de subconjuntos borrosos de $\mathbb{R}^n$, tal como ocurre con otros precedentes de la literatura. Además, antes de abordar el concepto de medida de dispersión, introduciremos un concepto más general, al que llamaremos grado de comparación

esperada entre parejas de imágenes de la variable aleatoria difusa, que, como veremos incluye, como casos particulares, además de algunas varianzas escalares definidas en la literatura reciente ([60, 66]), otros parámetros bien conocidos como la entropía cuadrática o la probabilidad superior de Dempster ([34]). Por otra parte, la idea de medida de tendencia central se materializará en el concepto de “centroide”. Este concepto estará también ligado a las disimilitudes y se determinará, siguiendo un procedimiento análogo al de la construcción de la esperanza de Frèchet, como el subconjunto difuso (único o no) que minimiza la disimilitud esperada con respecto a las imágenes de la variable aleatoria difusa. Comprobaremos que el concepto de centroide engloba algunas definiciones ya existentes en la literatura acerca de variables aleatorias difusas, como la esperanza de Puri y Ralescu ([84]), o la reciente definición de mediana de Sinova et al. ([91]), y también extiende los conceptos de media, mediana y moda de las variables aleatorias clásicas. Parte de los contenidos de este capítulo están recogidos en las referencias [20, 46, 45].

En el capítulo 3, asumiremos la existencia de una relación de equivalencia en un universo U , determinada por una colección de atributos: todos los elementos asociados al mismo vector de valores para todos los atributos serán indistinguibles, y formarán una clase de equivalencia. Consideraremos un conjunto aleatorio, cuyas imágenes son subconjuntos de ese universo. Dicho conjunto aleatorio determinará una distribución de probabilidad en una σ -álgebra contenida en la familia de conjuntos formados, a su vez, por subconjuntos del universo U . Debido a la imprecisión de nuestras observaciones, solo podremos determinar una familia de medidas de probabilidad compatibles

con la información disponible. En primer lugar, trataremos de proporcionar una expresión explícita para el máximo y el mínimo de dicha familia de probabilidades, en aquellos casos en que sea posible. A continuación, abordaremos el estudio de la caracterización del conjunto de posibles “valores” para el centroide y la dispersión del conjunto aleatorio, y las limitaciones prácticas existentes en este punto. Por último, mostraremos una forma adecuada de cuantificar el grado de imprecisión o inexactitud de nuestra información acerca de las imágenes del conjunto aleatorio, generalizando previamente la noción de “rugosidad” introducida por Pawlak en 1991. Para cada uno de estos apartados, comenzaremos con el estudio particular de los conjuntos aleatorios y después lo extenderemos al caso de las variables aleatorias difusas. Parte del material contenido en este capítulo está recogido en las referencias [22, 23, 24, 25].

Capítulo 1

Medidas de comparación entre conjuntos difusos.

1.1. Introducción y conceptos preliminares

Durante las últimas décadas han sido muchos los trabajos realizados sobre la comparación entre conjuntos difusos, debido a su impacto en diferentes áreas, como la minería de datos, la toma de decisiones o el procesamiento de imágenes.

Wang [93] introdujo por primera vez la noción de similaridad de conjuntos difusos, dando una fórmula específica para calcularla. Más tarde, en los años noventa, se propusieron algunas definiciones como “similaridad”, “similitud”, “proximidad” o “semejanza”, además de otras definiciones duales como “disimilaridad”, “disimilitud”, “divergencia” o “distancia”. Todas estas medidas asignan una cantidad (a veces dentro de un intervalo específico, como el intervalo unidad) a cada par de subconjuntos difusos (o de forma más general, a cada par de subconjuntos intuicionistas o intervalo-valorados difusos). Dicha cantidad representará el grado de igualdad o diferencia entre los conjuntos

comparados.

Podemos dividir la literatura existente sobre este tipo de medidas en dos bloques. Por un lado, podemos encontrar algunos trabajos que introducen definiciones axiomáticas (véanse [11, 35, 56, 65, 76, 89, 98, 99, 105, 106], por ejemplo). Por otro, podemos encontrar gran cantidad de trabajos que definen o revisan algunas medidas (o familias paramétricas de medidas) concretas (véanse [2, 8, 9, 15, 52, 53, 79, 96, 97, 94, 95, 100, 108], entre otros), generalmente con el fin de utilizarlas en problemas prácticos específicos.

En este capítulo, nos centraremos en el primer tipo de definiciones, las dadas de forma axiomática. Muchas de estas definiciones han sido introducidas de forma independiente, en diferentes contextos, con diferentes propósitos y usando diferentes nomenclaturas. Nuestro principal objetivo en este capítulo es presentarlas en un marco común y mostrar las relaciones formales entre las distintas definiciones. Demostraremos algunas relaciones de implicación entre los axiomas en algunos casos, y en otros comprobaremos la falta de implicación a través de contraejemplos. Los resultados se resumirán en tablas y gráficos que harán más rápida y sencilla la comprensión de los mismos.

En esta primera sección recordaremos brevemente algunos conceptos básicos sobre conjuntos difusos necesarios para poder realizar el estudio de nuestro interés. En la segunda sección presentaremos la lista de los axiomas que se utilizarán para la definición de las diferentes medidas consideradas en las secciones posteriores. Dividiremos la lista en tres bloques: un primer

bloque de axiomas para medidas de comparación en general, un segundo bloque de axiomas para medidas de “igualdad”, y por último, un bloque de axiomas sobre medidas de “desigualdad” o “diferencia”. En las dos secciones posteriores se analizará la relación formal existente entre las diferentes definiciones de “igualdad” y “desigualdad” entre conjuntos difusos respectivamente. En la quinta sección, se estudiará la “equivalencia” y “dualidad” entre diferentes medidas de comparación.

Y para finalizar, se mostrarán algunas conclusiones, así como problemas abiertos.

A continuación vamos a introducir la notación que utilizaremos a lo largo del capítulo.

Consideremos un conjunto arbitrario U (llamado en ocasiones *universo*) cuyos elementos denotaremos por x . Denotaremos por $\wp(U)$ la familia de los subconjuntos “clásicos” de U , es decir, aquellos cuya función indicador o función de pertenencia es

$$I_A(x) = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{si } x \in U \setminus A \end{cases}$$

Si permitimos que esta función de pertenencia pueda tomar cualquier valor en el intervalo $[0, 1]$, el conjunto se denomina *conjunto difuso* ([104]). Podemos interpretar el valor de esta función como el grado de pertenencia de un elemento x al conjunto $\tilde{A}$ y lo denotaremos por $\mu_{\tilde{A}}(x)$. Un conjunto difuso $\tilde{A} \in \mathcal{F}(U)$ se dice *normal* si existe un elemento $x \in U$ que verifica $\mu_{\tilde{A}}(x) = 1$.

Denotaremos por $\mathcal{F}(U)$ la familia de los subconjuntos difusos de U , también llamada *partes difusas* de U , que podemos identificar con el conjunto de las aplicaciones de U en $[0,1]$ por la

identificación de los subconjuntos difusos con sus funciones de pertenencia.

Observemos que el concepto de conjunto difuso generaliza la noción clásica de conjunto. A partir de ahora, utilizaremos el término *conjunto nítido* para referirnos a un conjunto clásico. Así podemos identificar el conjunto de partes (nítidas) de U , $\wp(U)$, como un subconjunto de las partes difusas, $\mathcal{F}(U)$.

Otro concepto importante en la teoría de conjuntos difusos es el de los α -cortes. Dado un subconjunto difuso $\tilde{A} \in \mathcal{F}(U)$ se define, para cada $\alpha \in [0, 1]$, el α -corte (*débil*) o *conjunto de nivel* α de $\tilde{A}$, como el conjunto nítido $\tilde{A}_\alpha = \{x \in U : \mu_{\tilde{A}}(x) \geq \alpha\}$.

A continuación vemos cómo se extienden algunos conceptos de los conjuntos nítidos al caso de los conjuntos difusos:

Dados $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$, diremos que:

- $\tilde{A}$ está contenido en $\tilde{B}$ (ó $\tilde{B}$ contiene a $\tilde{A}$), denotado por $\tilde{A} \subseteq \tilde{B}$ (ó $\tilde{B} \supseteq \tilde{A}$), si $\mu_{\tilde{A}}(x) \leq \mu_{\tilde{B}}(x) \forall x \in U$. Cuando la desigualdad es estricta para algún $x \in U$, la inclusión se dice estricta y se denota $\tilde{A} \subset \tilde{B}$ ($\tilde{B} \supset \tilde{A}$) ó bien $\tilde{A} \subsetneq \tilde{B}$ ($\tilde{B} \supsetneq \tilde{A}$).
- Son iguales, denotado $\tilde{A} = \tilde{B}$, si $\mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x) \forall x \in U$.

Cuando trabajamos con conjuntos nítidos, $A, B \in \wp(U)$, definimos su unión como $A \cup B = \{x \in U : x \in A \text{ ó } x \in B\}$ y su intersección como $A \cap B = \{x \in U : x \in A \text{ y } x \in B\}$. Estos conceptos pueden extenderse al caso de los conjuntos difusos del

siguiente modo ([104]):

- Dados $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$ se define su unión como el conjunto difuso con función de pertenencia

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = \max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \quad \forall x \in U.$$

- Dados $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$ se define su intersección como el conjunto difuso con función de pertenencia

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \quad \forall x \in U.$$

Dado un conjunto nítido $A \in \wp(U)$ definimos su complementario, A^c , como el conjunto formado por los elementos de U que no pertenecen a A , o dicho de otro modo, el conjunto cuya función indicador es $I_{A^c}(x) = 1 - I_A(x)$. Del mismo modo podemos considerar el complementario de un conjunto difuso ([104]): dado $\tilde{A} \in \mathcal{F}(U)$, definimos su complementario como el conjunto difuso con función de pertenencia

$$\mu_{\tilde{A}^c}(x) = 1 - \mu_{\tilde{A}}(x) \quad \forall x \in U.$$

Observemos que, dado un elemento x , $\mu_{\tilde{A} \cup \tilde{A}^c}(x)$ puede ser distinto de 1 y $\mu_{\tilde{A} \cap \tilde{A}^c}(x)$ puede ser distinto de 0, y por lo tanto $\tilde{A} \cup \tilde{A}^c \neq U$ y $\tilde{A} \cap \tilde{A}^c \neq \emptyset$ en general.

1.2. Lista de axiomas

Como ya hemos indicado en la introducción, daremos una colección de definiciones axiomáticas, procedentes de la literatura, relativas a las nociones de “igualdad” y “desigualdad o diferencia” entre conjuntos difusos, y estudiaremos sus relaciones. Con el objetivo de presentar estas definiciones de forma clara y compacta, enumeraremos en primer lugar los axiomas que aparecerán en ellas, separándolos en tres grupos.

Las propiedades incluidas en el primer grupo son axiomas generales que pueden aparecer en la definición de cualquier tipo de medida de comparación entre conjuntos difusos. Las propiedades de la segunda lista se refieren a la noción de “igualdad” entre conjuntos difusos, y aparecen en las definiciones axiomáticas de similaridad, semejanza y proximidad. Por último, el tercer bloque de axiomas está relacionado con el concepto de “desigualdad o diferencia”. Las propiedades en dicho bloque aparecen en las definiciones de divergencia, distancia y disimilaridad.

La nomenclatura que hemos escogido para denominar todas las propiedades enumeradas a continuación nos permitirá simplificar nuestro análisis sobre las relaciones existentes entre los diferentes axiomas. Por norma general, el asterisco será entendido como “más fuerte que”. Algunas implicaciones son directas, como por ejemplo $G1^* \Rightarrow G1$. Otras relaciones de implicación, como $S2^* \Rightarrow S2$ o $S5^* \Rightarrow S5$ serán probadas en la sección 1.3. Por otro lado, la propiedad $S1^*$ -var es simplemente una variante de la propiedad $S1^*$ (no necesariamente más fuerte, sino relacionada en cierto sentido).

1.2.1. Propiedades generales

- G1.- $0 \leq m(\tilde{A}, \tilde{B}) \leq 1 \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U).$
- G1*.- $0 \leq m(\tilde{A}, \tilde{B}) \leq 1 \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U)$ y existe algún $\tilde{C}, \tilde{D} \in \mathcal{F}(U)$ tal que $m(\tilde{C}, \tilde{D}) = 1.$
- G2.- $m(\tilde{A}, \tilde{B}) = m(\tilde{B}, \tilde{A}) \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U).$
- G3.- Supongamos U finito. Sea $\rho : U \longrightarrow U$ una permutación. Para cada $\tilde{A} \in \mathcal{F}(U)$ denotaremos por $\tilde{A}^\rho$ el conjunto difuso cuya función de pertenencia se define a partir de $\mu_{\tilde{A}}$ como sigue: $\mu_{\tilde{A}^\rho}(x) = \mu_{\tilde{A}}(\rho(x)).$ Entonces

$$m(\tilde{A}, \tilde{B}) = m(\tilde{A}^\rho, \tilde{B}^\rho) \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U).$$

- G3*.- U es finito y, $\forall \tilde{A}, \tilde{B} \in \mathcal{F}(U),$

$$m(\tilde{A}, \tilde{B}) = \sum_{x \in U} h[\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)].$$

- G4.- Existe una aplicación $f : \mathcal{F}(U) \longrightarrow \mathbb{R}$ tal que

$$m(\tilde{A}, \tilde{B}) = f(\tilde{A} \cap \tilde{B}, \tilde{A} \setminus \tilde{B}, \tilde{B} \setminus \tilde{A}) \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U).$$

- G4*.- Existe una función $F_m : \mathbb{R}^3 \longrightarrow \mathbb{R}$ y una medida difusa ¹ $M : \mathcal{F}(U) \longrightarrow \mathbb{R}$ tal que, para cada $\tilde{A}, \tilde{B} \in \mathcal{F}(U),$ $m(\tilde{A}, \tilde{B})$ puede ser expresada del siguiente modo:

$$m(\tilde{A}, \tilde{B}) = F_m(M(\tilde{A} \cap \tilde{B}), M(\tilde{A} \setminus \tilde{B}), M(\tilde{B} \setminus \tilde{A})).$$

¹Pueden encontrarse diferentes definiciones de este concepto. Aquí, función de conjuntos monótona no decreciente satisfaciendo la restricción $M(\emptyset) = 0.$

1.2.2. Axiomas para medidas de “igualdad”

- S1.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$, si $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$ y $\max_{x \in U} \mu_{\tilde{A}}(x) = \max_{x \in U} \mu_{\tilde{B}}(x)$ entonces $s(\tilde{A}, \tilde{C}) \leq s(\tilde{A}, \tilde{B})$.

- S1*.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$,

$$\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C} \Rightarrow s(\tilde{A}, \tilde{C}) \leq s(\tilde{A}, \tilde{B})$$

- S1*-var.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$,

$$\tilde{A} \subsetneq \tilde{B} \subsetneq \tilde{C} \Rightarrow s(\tilde{A}, \tilde{C}) < s(\tilde{A}, \tilde{B})$$

- S2.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$,

$$\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C} \Rightarrow s(\tilde{A}, \tilde{C}) \leq s(\tilde{B}, \tilde{C}).$$

- S2*.- Sean $\tilde{C}, \tilde{A} \in \mathcal{F}(U)$, tales que existe $x_0 \in U$ con $\mu_{\tilde{C}}(x_0) > \mu_{\tilde{A}}(x_0)$. Sea $\tilde{B} \in \mathcal{F}(U)$ tal que:

- $\mu_{\tilde{A}}(x_0) < \mu_{\tilde{B}}(x_0) \leq \mu_{\tilde{C}}(x_0)$, y
- $\mu_{\tilde{B}}(x) = \mu_{\tilde{A}}(x)$, $\forall x \in U, x \neq x_0$

Entonces $s(\tilde{A}, \tilde{C}) < s(\tilde{B}, \tilde{C})$.

- S3.- $\forall D \in \wp(U), s(D, D^c) = 0$.
- S3*.- $s(D, D^c) = 0 \Leftrightarrow D \in \wp(U)$.
- S4.- $s(\tilde{C}, \tilde{C}) = \max_{\tilde{A}, \tilde{B} \in \mathcal{F}(U)} s(\tilde{A}, \tilde{B}) \quad \forall \tilde{C} \in \mathcal{F}(U)$.
- S4*.- $\tilde{C} = \tilde{D} \Leftrightarrow s(\tilde{C}, \tilde{D}) = \max_{\tilde{A}, \tilde{B} \in \mathcal{F}(U)} s(\tilde{A}, \tilde{B})$.
- S5.- Dados dos subconjuntos difusos no vacíos de U , $\tilde{A}$ y $\tilde{B}$, si $s(\tilde{A}, \tilde{B}) = 0$, entonces $\min(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)) = 0$ para todo $x \in U$, es decir, $\tilde{A} \cap \tilde{B} = \emptyset$.
- S5*.- Consideremos $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$ y $x_0 \in U$ arbitrario. Definimos $\tilde{C}$ y $\tilde{D}$ tales que:
 - $\mu_{\tilde{C}}(x) = \mu_{\tilde{A}}(x)$ y $\mu_{\tilde{D}}(x) = \mu_{\tilde{B}}(x), \forall x \neq x_0$
 - $\mu_{\tilde{C}}(x_0) = \mu_{\tilde{A}}(x_0) + \alpha$ y $\mu_{\tilde{D}}(x_0) = \mu_{\tilde{B}}(x_0) + \alpha$, donde $0 \leq \alpha \leq 1 - \max\{\mu_{\tilde{A}}(x_0), \mu_{\tilde{B}}(x_0)\}$

Entonces:

- Si $\max_{x \in U} \mu_{\tilde{C} \cap \tilde{D}}(x) = \max_{x \in U} \mu_{\tilde{A} \cap \tilde{B}}(x)$ entonces

$$s(\tilde{C}, \tilde{D}) = s(\tilde{A}, \tilde{B}).$$
- Si $\max_{x \in U} \mu_{\tilde{C} \cap \tilde{D}}(x) > \max_{x \in U} \mu_{\tilde{A} \cap \tilde{B}}(x)$ entonces

$$s(\tilde{C}, \tilde{D}) > s(\tilde{A}, \tilde{B}).$$
- S6.- Consideremos $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$ y $x_0 \in U$ arbitrario. Definimos $\tilde{C}$ y $\tilde{D}$ tales que:

- $\mu_{\tilde{C}}(x) = \mu_{\tilde{A}}(x)$ y $\mu_{\tilde{D}}(x) = \mu_{\tilde{B}}(x)$, $\forall x \neq x_0$
- $\mu_{\tilde{C}}(x_0) = \mu_{\tilde{A}}(x_0) + \alpha$ (resp. $\mu_{\tilde{C}}(x_0) = \mu_{\tilde{A}}(x_0) - \alpha$) y $\mu_{\tilde{D}}(x_0) = \mu_{\tilde{B}}(x_0) + \alpha$ (resp. $\mu_{\tilde{D}}(x_0) = \mu_{\tilde{B}}(x_0) - \alpha$).

Entonces, $s(\tilde{A}, \tilde{C}) = s(\tilde{B}, \tilde{D})$.

- S7.- Si $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D} \in \mathcal{F}(U)$ satisfacen las siguientes restricciones: $\tilde{A} \cap \tilde{B} \subseteq \tilde{C} \cap \tilde{D}$, $\tilde{A} \setminus \tilde{B} \supseteq \tilde{C} \setminus \tilde{D}$, y $\tilde{B} \setminus \tilde{A} \supseteq \tilde{D} \setminus \tilde{C}$ entonces $s(\tilde{A}, \tilde{B}) \leq s(\tilde{C}, \tilde{D})$.

1.2.3. Axiomas para medidas de “desigualdad” o “diferencia”

- D1.- Si $\tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$ satisfacen las restricciones $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$, entonces:
 - (a) $d(\tilde{A}, \tilde{C}) \geq d(\tilde{A}, \tilde{B})$ y
 - (b) $d(\tilde{A}, \tilde{C}) \geq d(\tilde{B}, \tilde{C})$.
- D2.- $d(D, D^c) = \max_{\tilde{A}, \tilde{B} \in \mathcal{F}(U)} d(\tilde{A}, \tilde{B}) \quad \forall D \in \wp(U)$.
- D3.- $\tilde{A} = \tilde{B} \Rightarrow d(\tilde{A}, \tilde{B}) = 0$.
- D3*.- $\tilde{A} = \tilde{B} \Leftrightarrow d(\tilde{A}, \tilde{B}) = 0$.
- D4.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$, $d(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C}) \leq d(\tilde{A}, \tilde{B})$.

- D5.- $\forall \tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U), d(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}) \leq d(\tilde{A}, \tilde{B})$.
- D6.- Si $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D} \in \mathcal{F}(U)$ satisfacen las siguientes restricciones: $\tilde{A} \cap \tilde{B} \subseteq \tilde{C} \cap \tilde{D}$, $\tilde{A} \setminus \tilde{B} \supseteq \tilde{C} \setminus \tilde{D}$, y $\tilde{B} \setminus \tilde{A} \supseteq \tilde{D} \setminus \tilde{C}$ entonces $d(\tilde{A}, \tilde{B}) \geq d(\tilde{C}, \tilde{D})$.
- D7.- $d(\tilde{A}, \tilde{B}) = d(\tilde{A} \setminus \tilde{B}, \tilde{B} \setminus \tilde{A}) \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U)$.

1.3. Medidas de “igualdad”

1.3.1. Definiciones axiomáticas de medida de similitud

A continuación recordaremos algunas de las definiciones axiomáticas de medida de “igualdad” más relevantes existentes en la literatura. Queremos informar al lector de que hemos modificado la presentación de cada definición tomada de la literatura, con el objetivo de unificar la nomenclatura. Por ejemplo, en la mayor parte de las definiciones las propiedades S1 y S2 se expresan como un solo axioma. Así mismo, muchas de estas definiciones requieren las condiciones G1 y $s(\tilde{A}, \tilde{A}) = 1$ para cualquier $\tilde{A} \in \mathcal{F}(U)$, y las expresaremos aquí de forma equivalente, diciendo que la medida de similitud satisface G1*y S4. Algo similar sucede con las propiedades D4 y D5, que forman un solo axioma en [76], como requerimiento para las medidas de divergencia. Debemos mencionar además que algunas de las definiciones originales fueron propuestas en contextos más generales (subconjuntos intuicionistas o intervalo-valorados difusos). Para llevar a cabo una comparación exhaustiva, restringiremos

su dominio a la familia de los pares de conjuntos difusos.

Definición 1 ([99] L. Xuecheng, 1992) Una función real $s : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}^+$ se denomina medida de similaridad, si verifica las propiedades $G2$, $S1^*$, $S2$, $S3$ y $S4$.

Definición 2 ([35], L. Dengfeng y C. Chuntian, 2002) Diremos que $s(\tilde{A}, \tilde{B})$ es el grado de similaridad entre $\tilde{A} \in \mathcal{F}(U)$ y $\tilde{B} \in \mathcal{F}(U)$ si s satisface las propiedades $G1^*$, $G2$, $S1^*$, $S2$ y $S4$.

Definición 3 ([65] Z. Liang y P. Shi, 2003 y [72] H.B. Mitchell, 2003) Diremos que $s(\tilde{A}, \tilde{B})$ es el grado de similaridad entre $\tilde{A} \in \mathcal{F}(U)$ y $\tilde{B} \in \mathcal{F}(U)$ si s satisface las propiedades $G1^*$, $G2$, $S1^*$, $S2$ y $S4^*$.

Nota 1.1 La definición 3 es simplemente una variante más fuerte de la definición 2, donde la propiedad $S4$ se reemplaza por $S4^*$. (La igualdad entre los conjuntos difusos se requiere en [65, 72], pero no en [35], con el objetivo de alcanzar la similaridad máxima.)

Definición 4 ([105], W. Zeng y H. Li, 2006) Una función real s sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de similaridad cuando satisface las propiedades $G1^*$, $G2$, $S1^*$, $S2$, $S3$ y $S4^*$.

Definición 5 ([98], L. Xudong, C. Zhang, 1999) Una función real s sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina grado de f-cercanía, si

satisface los axiomas $G1^*$, $G2$, $S1^*$, $S2$, $S4^*$ y $S5$.

Nota 1.2 Una función sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ satisfaciendo la definición 5 se denomina medida de similaridad en [74], B. Mondal et al., 2006.

Definición 6 ([106], H. Zhang et al., 2009) Una función real s sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de similaridad, si verifica las propiedades $G1^*$, $G2$, $S1^*$, $S2$, $S3^*$ y $S4^*$.

Definición 7 ([56], I. Jenhani et al., 2010) Una medida de similaridad posibilística debería satisfacer las propiedades $G1^*$, $G2$, $G3$, $S1$, $S3$ y $S4$. Además, las siguientes propiedades serían deseables: $S1^*$ -var, $S2^*$, $S5^*$ y $S6$.

Nota 1.3 El concepto de medida de similaridad posibilística propuesto en [56] estaba originalmente restringido a conjuntos difusos normales, ya que estos pueden ser interpretados como distribuciones de posibilidad. En este capítulo, simplemente estamos considerando la relación formal existente entre las diferentes medidas de similaridad, por lo que no tendremos en cuenta dicha restricción.

1.3.2. Relación formal entre las diferentes medidas de “igualdad”

En la literatura existen algunos estudios sobre las relaciones entre las definiciones axiomáticas consideradas en la sección

anterior. Nuestro objetivo es desarrollar un estudio sistemático sobre ellas. Primero presentaremos las definiciones 1 a 7 de forma concisa, con el propósito de simplificar nuestro estudio. Dichas definiciones pueden resumirse en la siguientes tablas:

Definición	G1*	G2	G3
Def. 1		•	
Def. 2	•	•	
Def. 3	•	•	
Def. 4	•	•	
Def. 5	•	•	
Def. 6	•	•	
Def. 7 (nec.)	•	•	•
Def. 7 (des.)	•	•	•

Definición	S1	S1*	S1*-var	S2	S2*	S3	S3*	S4	S4*	S5	S5*	S7
Def. 1		•		•		•		•				
Def. 2		•		•				•				
Def. 3		•		•					•			
Def. 4		•		•		•			•			
Def. 5		•		•					•	•		
Def. 6		•		•			•		•			
Def. 7 (nec.)	•					•		•				
Def. 7 (des.)	•		•		•	•		•			•	•

A continuación demostraremos algunas implicaciones entre axiomas incluidos en los listados de las secciones 1.2.1 y 1.2.2, que serán utilizadas más tarde para mostrar la relación formal entre las diferentes variantes del concepto “similaridad” propuestas en las definiciones 1 a 7.

Con respecto a las propiedades generales (enumeradas en la sección 1.2.1), podemos ver de forma inmediata que G1* implica

G1. Además, por la propiedad conmutativa de la suma, podemos observar también que $G3^*$ implica $G3$. En cuanto a los axiomas referentes a medidas de “igualdad” (enumerados en la sección 1.2.2), se obtiene fácilmente que $S1^*$ implica $S1$, $S3^*$ implica $S3$ y $S4^*$ implica $S4$. A continuación comprobamos otras relaciones de implicación menos evidentes:

Proposición 1 *Si $s : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow \mathbb{R}$ satisface las propiedades $S1^*$ -var y $S4$, entonces satisface $S1^*$.*

Demostración: Sean $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$ tres conjuntos difusos anidados arbitrarios, y supongamos que s satisface $S1^*$ -var y $S4$. Si $\tilde{A} \subsetneq \tilde{B} \subsetneq \tilde{C}$, entonces $s(\tilde{A}, \tilde{C}) \leq s(\tilde{A}, \tilde{B})$, de acuerdo con la propiedad $S1^*$ -var. En el resto de casos:

- Si $\tilde{A} = \tilde{B}$, entonces, de acuerdo con la propiedad $S4$:

$$s(\tilde{A}, \tilde{B}) = \max_{\tilde{D}, \tilde{E} \in \mathcal{F}(U)} s(\tilde{D}, \tilde{E}) \geq s(\tilde{A}, \tilde{C}).$$

- Si $\tilde{A} \neq \tilde{B}$ entonces $\tilde{B}$ debe ser igual a $\tilde{C}$, y por tanto $s(\tilde{A}, \tilde{C}) = s(\tilde{A}, \tilde{B})$.

Así pues, hemos comprobado que $s(\tilde{A}, \tilde{C}) \leq s(\tilde{A}, \tilde{B})$ en cualquier caso, de modo que s satisface $S1^*$. $\square$

Proposición 2 *La propiedad $S2^*$ implica la propiedad $S2$.*

Demostración: Sean $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$ tres conjuntos difusos anidados arbitrarios y supongamos que s satisface la propiedad $S2^*$.

Denotemos por k el número de índices i tales que $\mu_{\tilde{A}}(x_i)$ es estrictamente menor que $\mu_{\tilde{C}}(x_i)$. Así pues, supongamos que

- $\mu_{\tilde{A}}(x_{i(j)}) = \mu_{\tilde{C}}(x_{i(j)}), \forall j = 1, \dots, n - k$ y
- $\mu_{\tilde{A}}(x_{i(j)}) = \mu_{\tilde{C}}(x_{i(j)}), \forall j = n - k - 1, \dots, n$.

Ahora, definamos la siguiente serie finita de conjuntos difusos:

- $\tilde{A}_0 = \tilde{A}$
- $\tilde{A}_1$ es el conjunto difuso cuya función de pertenencia es $\mu_{\tilde{A}_1}(x_{i(j)}) = \mu_{\tilde{A}_0}(x_{i(j)}), \forall j \leq n - 1$ y $\mu_{\tilde{A}_1}(x_{i(j)}) = \mu_{\tilde{B}}(x_{i(j)}),$ y, en general:
- $\mu_{\tilde{A}_{r+1}}(x_{i(j)}) = \mu_{\tilde{A}_r}(x_{i(j)}), \forall j \neq n - r, \mu_{\tilde{A}_{r+1}}(x_{i(n-r)}) = \mu_{\tilde{A}_r}(x_{i(n-r)}), \forall r = 0, \dots, k - 1.$

Podemos comprobar fácilmente que:

- $\tilde{A}_i \subseteq \tilde{A}_{i+1}, \forall i = 0, \dots, k - 1$
- $\tilde{A}_{k-1} = \tilde{B}$
- $\tilde{A}_i, \tilde{A}_{i+1}$ y $\tilde{C}$ satisfacen las propiedades requeridas en S2* para el triplete de conjuntos.

Entonces, de acuerdo con S2*, obtenemos la siguiente cadena finita de desigualdades:

$$s(\tilde{A}, \tilde{C}) = s(\tilde{A}_0, \tilde{C}) \leq s(\tilde{A}_1, \tilde{C}) \leq \dots \leq s(\tilde{A}_{r-1}, \tilde{C}) = s(\tilde{B}, \tilde{C}). \quad \square$$

Proposición 3 *La propiedad S5* implica la propiedad S5.*

Demostración: Consideremos un par de conjuntos difusos $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$ tales que $\tilde{A} \cap \tilde{B} \neq \emptyset$ y s satisfaciendo la propiedad S5*. Para comprobar el resultado dado en esta proposición, es suficiente con encontrar un par diferente de conjuntos difusos $\tilde{C}$ y $\tilde{D}$ tales que $s(\tilde{C}, \tilde{D}) < s(\tilde{A}, \tilde{B})$. Denotemos por α el máximo $\alpha = \max_{x \in U} \mu_{\tilde{A} \cap \tilde{B}}(x)$. De acuerdo con la hipótesis anterior, es estrictamente positivo. Supongamos que este valor de pertenencia es alcanzado en k elementos $x_{i_1}, \dots, x_{i_k}$, con $k \geq 1$. Ahora consideremos un par de series finitas de conjuntos difusos $\tilde{A}_0, \dots, \tilde{A}_{k+1}$ y $\tilde{B}_0, \dots, \tilde{B}_{k+1}$ definidas del siguiente modo:

- $\tilde{A}_0 = \tilde{A}$ y $\tilde{B}_0 = \tilde{B}$
- $\mu_{\tilde{A}_j}(x_{i_j}) = \mu_{\tilde{A}_{j-1}}(x_{i_j}) - \alpha$, $\forall j \in \{1, \dots, k\}$ y $\mu_{\tilde{A}_j}(x_{i_k}) = \mu_{\tilde{A}_{j-1}}(x_{i_k})$, $\forall k \neq j$ y para todo $j \in \{1, \dots, k\}$
- $\mu_{\tilde{B}_j}(x_{i_j}) = \mu_{\tilde{B}_{j-1}}(x_{i_j}) - \alpha$, $\forall j \in \{1, \dots, k\}$ y $\mu_{\tilde{B}_j}(x_{i_k}) = \mu_{\tilde{B}_{j-1}}(x_{i_k})$, $\forall k \neq j$ y para todo $j \in \{1, \dots, k\}$

Entonces, de acuerdo con la propiedad S5*, observamos que:

- $s(\tilde{A}_j, \tilde{B}_j) = s(\tilde{A}_{j-1}, \tilde{B}_{j-1})$, $\forall j = 1, \dots, k$
- $s(\tilde{A}_k, \tilde{B}_k) < s(\tilde{A}_{k-1}, \tilde{B}_{k-1})$

Por tanto, obtenemos la siguiente desigualdad estricta:

$$s(\tilde{A}_k, \tilde{B}_k) < s(\tilde{A}, \tilde{B}). \quad \square$$

Proposición 4 *Si $s : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface las propiedades S3 y S5, entonces satisface S3*.*

Demostración: Supongamos que s satisface las propiedades S3 y S5. Para comprobar que satisface también la propiedad S3*, simplemente necesitamos probar la siguiente implicación: $s(D, D^c) = 0 \Rightarrow D \in \wp(U)$.

De acuerdo con la propiedad S5, $s(D, D^c) = 0$ implica que $\min\{\mu_D(x), \mu_{D^c}(x)\} = \min\{\mu_D(x), 1 - \mu_D(x)\} = 0, \forall x \in U$, y por tanto $\mu_D(x) \in \{0, 1\}, \forall x \in U$. $\square$

Proposición 5 *Si $s : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface las propiedades S2* y S4, entonces satisface S4*.*

Demostración: Supongamos que s satisface las propiedades S2* y S4. Consideremos un par de conjuntos difusos diferentes $\tilde{A}, \tilde{C} \in \mathcal{F}(U)$, $\tilde{A} \neq \tilde{C}$. Para comprobar el resultado dado en esta proposición, será suficiente encontrar un conjunto difuso $\tilde{B}$ tal que $s(\tilde{B}, \tilde{C}) > s(\tilde{A}, \tilde{C})$. Supongamos, sin pérdida de generalidad, que existe algún $x_0 \in U$ tal que $\mu_{\tilde{A}}(x_0) < \mu_{\tilde{C}}(x_0)$, y definamos la función de pertenencia del conjunto difuso $\tilde{B}$ como sigue:

- $\mu_{\tilde{B}}(x) = \mu_{\tilde{A}}(x), \forall x \neq x_0.$
- $\mu_{\tilde{B}}(x_0) = \mu_{\tilde{C}}(x_0)$

Entonces, de acuerdo con la propiedad S2*, $s(\tilde{B}, \tilde{C})$ es estrictamente mayor que $s(\tilde{A}, \tilde{C})$. $\square$

De acuerdo con los anteriores resultados, podemos completar la tabla en la que resumimos las propiedades de las definiciones

dadas, como detallamos a continuación. Si una celda está llena, significa que toda medida de similaridad satisfaciendo la correspondiente definición (fila) también satisface la correspondiente propiedad (columna). Si la celda contiene un punto, significa que la propiedad está incluida en la definición original. Si el contenido es S significa que la propiedad se deduce de forma sencilla, directa. Si el contenido es Px, significa que la propiedad se satisface como consecuencia de la proposición x.

Definición	S1	S1*	S1*-var	S2	S2*	S3	S3*	S4	S4*	S5	S5*	S7
Def. 1	S	•		•		•		•				
Def. 2	S	•		•				•				
Def. 3	S	•		•				S	•			
Def. 4	S	•		•		•		S	•			
Def. 5	S	•		•				S	•	•		
Def. 6	S	•		•		S	•	S	•			
Def. 7 (nec.)	•					•		•				
Def. 7 (des.)	•	P1	•	P2	•	•	P4	•	P5	P3	•	•

De acuerdo con la información mostrada en esta tabla y en la correspondiente a las propiedades generales, podemos comprobar fácilmente varias implicaciones entre las definiciones de similaridad mencionadas en la sección 1.3.1:

Corolario 1 *Las siguientes implicaciones se verifican:*

- *Si s satisface la definición 1, entonces satisface la definición 2, a excepción del axioma $G1^*$ (condiciones de cota). Además, s satisface las propiedades necesarias de la definición 7, a excepción de $G3$.*

- *Si s satisface la definición 3, entonces también satisface la definición 2.*
- *Si s satisface la definición 4, entonces también satisface las definiciones 1 y 3, y por tanto satisface la definición 2 y las propiedades necesarias de la definición 7, a excepción de $G3$.*
- *Si s satisface la definición 5, entonces también satisface la definición 3, y por tanto satisface la definición 2.*
- *Si s satisface la definición 6, entonces también satisface la definición 4, y por tanto también satisface las definiciones 1, 2 y 3, y las propiedades necesarias de la definición 7, a excepción de $G3$.*
- *Si s satisface todas las propiedades deseables enumeradas en la definición 7, entonces satisface todas las definiciones vistas, desde la definición 1 a la definición 6.*

Podemos reflejar estas relaciones en el siguiente diagrama, donde la flecha significa “es al menos tan restrictivo como”.

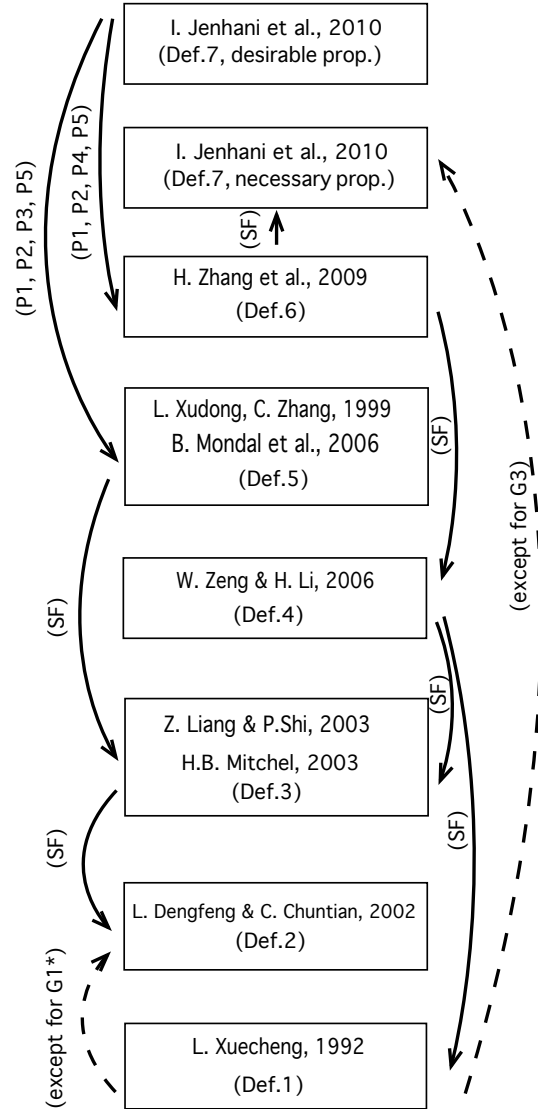


Figura 1.1: Resumen de las relaciones de implicación entre las diferentes variantes del concepto “medida de similaridad”.

Las implicaciones inversas no se satisfacen en general, como puede comprobarse con los siguientes contraejemplos.

Ejemplo 1 *Dada s una medida de similaridad y $f : [0, 1] \rightarrow [0, 1]$ una aplicación no decreciente, $f \circ s$ satisface todas las propiedades que satisfaga s , a excepción de aquellas referentes a su acotación. Entonces, si consideramos s satisfaciendo la definición 2 ó 3 y $f(0) > 0$, $S = f \circ s$ también satisface la definición 2 ó 3. Pero $S(D, D^c) > 0$ para todo D subconjunto nítido de U . Por tanto S no satisface las definiciones 1 y 4. Podemos utilizar también este razonamiento para comprobar que una medida de similaridad satisfaciendo la definición 5 no satisface necesariamente la definición 1.*

Ejemplo 2 *Consideremos un conjunto arbitrario U y una medida s definida para pares de subconjuntos difusos del siguiente modo: $s(\tilde{A}, \tilde{B}) = 0$ si $\tilde{A} \cap \tilde{B} = \emptyset$ y $s(\tilde{A}, \tilde{B}) = 1$ en otro caso. La medida s satisface la definición 1, pero s no satisface la definición 3. En consecuencia, s no satisface las definiciones 4, 5 ó 6, y tampoco satisface las propiedades deseables en la definición 7.*

Ejemplo 3 *Consideremos un conjunto finito $U = \{x_1, x_2, \dots, x_n\}$ y $s(\tilde{A}, \tilde{B}) = 1 - \sup_{x \in U} |\mu_{\tilde{A}}(x) - \mu_{\tilde{B}}(x)|$, $\forall \tilde{A}, \tilde{B} \in \mathcal{F}(U)$. La medida de similaridad s satisface las propiedades $S1^*$, $S2$, $S3$ y $S4^*$, pero s no satisface $S3^*$. Por tanto, s satisface la definición 4 pero no satisface la definición 6.*

Ejemplo 4 Consideremos un universo finito U y una medida s dada por:

- $s(\emptyset, \tilde{A}) = 0 \forall \tilde{A} \in \mathcal{F}(U) \tilde{A} \neq \emptyset.$
- $s(\{x\}, \{y\}) = 0 \forall x, y \in U, x \neq y.$
- $s(\tilde{A}, \tilde{B}) = 0,5 \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U) \text{ tales que } \tilde{A} \neq \tilde{B}, \#\tilde{A} \geq 2 \text{ ó } \#\tilde{B} \geq 2.$
- $s(\tilde{A}, \tilde{A}) = 1 \forall \tilde{A} \in \mathcal{F}(U).$

s satisface la definición 5, pero no satisface el axioma S3. Por tanto, s no satisface las definiciones 4, 6 y 7.

Ejemplo 5 Consideremos una medida s entre pares de conjuntos difusos de un universo U dada por:

- $s(\tilde{A}, \tilde{B}) = 1 \Leftrightarrow \tilde{A} = \tilde{B}$
- $s(\tilde{A}, \tilde{B}) = 0$ si $\tilde{A}(x) \leq 0,1$ y $\tilde{B}(x) = 1 \forall x \in U$
- $s(A, B) = 0$ si $A = B^c$, con A un subconjunto nítido.
- $s(\tilde{A}, \tilde{B}) = 0,5$ en otro caso.

Entonces, s satisface la definición 6 (y, en consecuencia, las definiciones 3 y 4) pero no satisface la definición 5.

Ejemplo 6 Consideremos un conjunto finito $U = \{x_1, x_2, \dots, x_n\}$ y la siguiente medida de similaridad dada en [35]:

$$S_{d\omega}^2(\tilde{A}, \tilde{B}) = 1 - \sqrt{\sum_{i=1}^n \omega_i (\mu_{\tilde{A}}(x_i) - \mu_{\tilde{B}}(x_i))^2}$$

donde $0 \leq \omega_i \leq 1$, $\sum_{i=1}^n \omega_i = 1$. Esta medida satisface las definiciones 1 y 6 pero $S_{d\omega}^2$ no verifica la propiedad G3. En consecuencia, no satisface la definición 7.

La información aportada por estos contraejemplos se encuentra resumida en el siguiente diagrama.

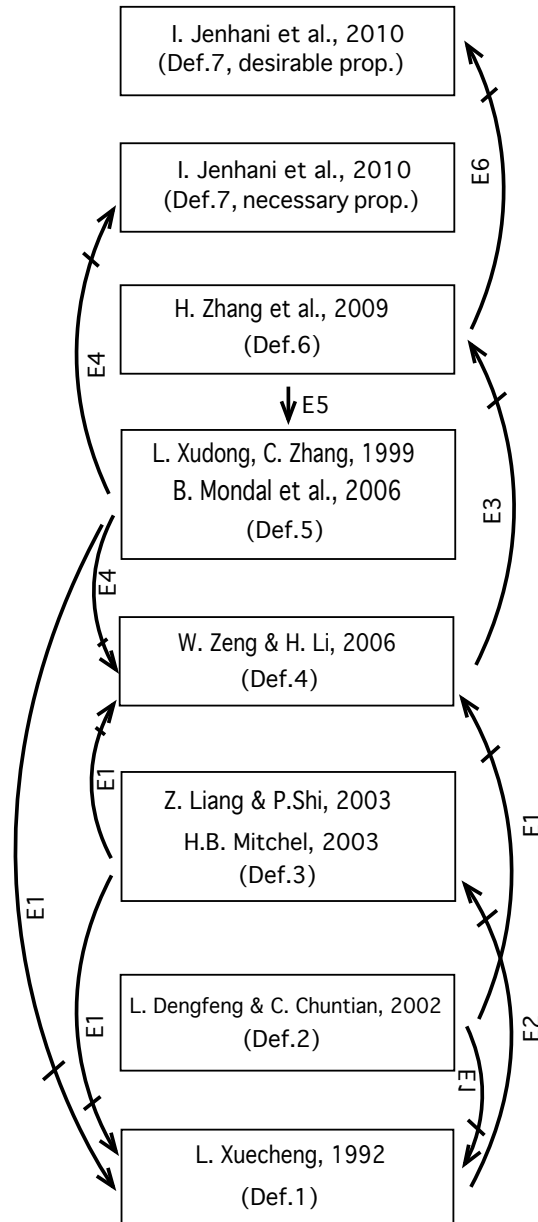


Figura 1.2: Contraejemplos concernientes a las implicaciones entre las diferentes variantes del concepto “medida de similaridad”.

1.3.3. Otras definiciones de medida de igualdad

Como ya apuntábamos en la introducción, la noción de “similaridad” no es el único concepto introducido en la literatura para expresar el grado de “igualdad” entre pares de conjuntos difusos. Veremos a continuación los conceptos de “similitud” y “semejanza” introducidos por Bouchon et. al. en [11].

Definición 8 ([11] Bouchon-Meunier et al., 1996) Una M-medida de comparación sobre U es una función c sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ satisfaciendo $G1$ y $G4^*$.

Definición 9 ([11], Bouchon-Meunier et al., 1996) Una M-medida de similitud sobre U es una M-medida de comparación satisfaciendo $S7$.

Definición 10 ([11], Bouchon-Meunier et al., 1996) Una M-medida de semejanza es una M-medida de similitud satisfaciendo $G1^*$, $G2$ y $S4$.

Al igual que hicimos con las diferentes variantes del concepto “similaridad”, podemos resumir en una tabla las propiedades que verifica cada una de estas definiciones.

Definición	G1	G1*	G2	G3	G4*
Def. 8	•				•
Def.9	•				•
Def. 10	•	•	•		•

Definición	S1	S1*	S1*-var	S2	S2*	S3	S3*	S4	S4*	S5	S5*	S6	S7
Def. 9													•
Def. 10								•					•

A continuación mostraremos las relaciones formales existentes entre las nociones de similitud y semejanza, y las diferentes variantes de la noción similaridad consideradas en las definiciones 1 a 7. Para ello, de nuevo, veremos un resultado que nos dará la relación entre diversos axiomas.

Proposición 6 *Si s satisface las propiedades $G4^*$ y $S7$, entonces satisface las propiedades S_1^* y S_2 .*

Demostración: Consideremos una medida de similitud s , y tomemos tres conjuntos difusos $\tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{F}(U)$ satisfaciendo las restricciones $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$. Podemos expresar $s(\tilde{A}, \tilde{C})$, $s(\tilde{A}, \tilde{B})$ y $s(\tilde{B}, \tilde{C})$ respectivamente como

$$s(\tilde{A}, \tilde{C}) = F_s(M(\tilde{A}), M(\emptyset), M(\tilde{C} \setminus \tilde{A})),$$

$$s(\tilde{A}, \tilde{B}) = F_s(M(\tilde{A}), M(\emptyset), M(\tilde{B} \setminus \tilde{A})),$$

y

$$s(\tilde{B}, \tilde{C}) = F_s(M(\tilde{B}), M(\emptyset), M(\tilde{C} \setminus \tilde{B})).$$

Por tanto, teniendo en cuenta la monotonicidad de M podemos comprobar fácilmente que s satisface las propiedades S_1^* y S_2 . $\square$

De acuerdo con este resultado, podemos completar la tabla anterior del siguiente modo:

Definición	G1	G1*	G2	G3	G4*
Def. 8	•				•
Def. 9	•				•
Def. 10	•	•	•		•

Definición	S1	S1*	S1*-var	S2	S2*	S3	S3*	S4	S4*	S5	S5*	S6	S7
Def. 9	S	P6		P6									•
Def. 10	S	P6		P6				•					•

Teniendo en cuenta la información de esta tabla, podemos deducir fácilmente que el concepto de semejanza es más restrictivo que el concepto de similaridad considerado en la definición 2, como mostramos en el siguiente corolario.

Corolario 2 *Cualquier medida de similitud satisface las propiedades S_1^* y S_2 . Por tanto, toda medida de semejanza satisface la definición 2.*

De acuerdo con el corolario 2, toda medida de semejanza satisface la definición 2. Podemos preguntarnos además si satisface el resto de definiciones de similaridad enumeradas en la sección 1.3.1 ó no. El siguiente ejemplo demuestra que, en general, las medidas de semejanza no cumplen el resto de definiciones.

Ejemplo 7 *Sea U un conjunto finito con cardinal n , y sea la M -medida de semejanza definida para pares de subconjuntos nítidos*

de U dada por: $s(A, B) = F(M(A \cap B), M(B \setminus A), M(A \setminus B))$ con $M(A) = \frac{\#A}{n}$, donde $\#A$ denota el cardinal del conjunto A , y $F : [0, 1]^3 \rightarrow [0, 1]$ dada por $F(u, v, w) = \frac{u+(1-v)+(1-w)}{3}$ si $v, w \neq 0$ y $F(u, v, w) = 1$ en otro caso.

Esta medida no satisface los axiomas $S3$ y $S4^$. Por tanto, no satisface las definiciones 1, 3, 4, 5, 6 y 7.*

Finalmente, podemos resumir los resultados obtenidos en esta sección concernientes a las relaciones formales entre las diferentes variantes de la noción de “igualdad” de conjuntos difusos dadas en las definiciones 1 a 10. La parte izquierda de la figura 1.3 completa la información dada en la figura 1.1, y su parte derecha completa la información dada en la figura 1.2

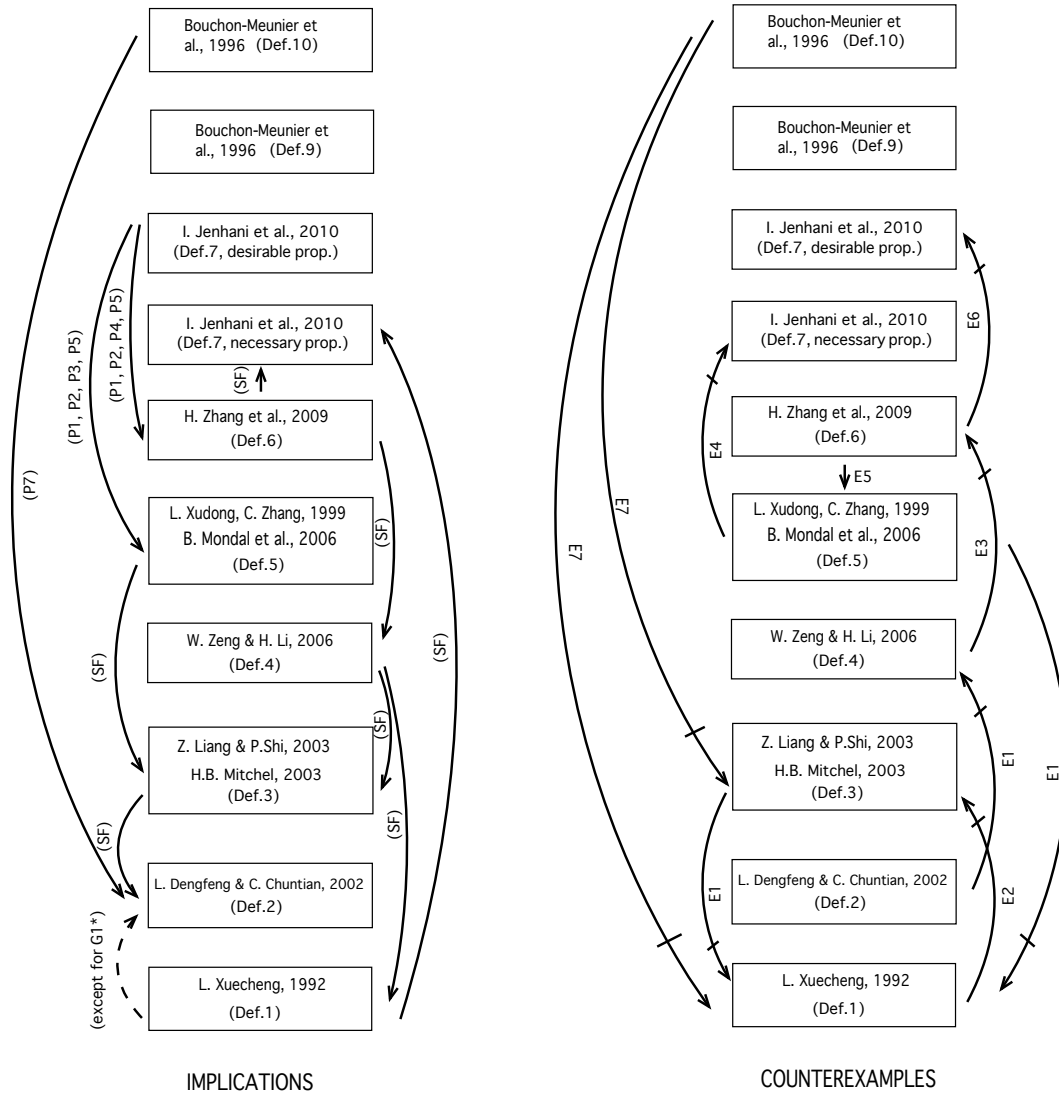


Figura 1.3: Relaciones de implicación y contraejemplos concernientes a las definiciones 1 a 10.

1.4. Medidas de “desigualdad” o “diferencia”

1.4.1. Definiciones axiomáticas en la literatura

A continuación veremos algunas definiciones de la literatura que están relacionadas con la idea de proporcionar el grado de “desigualdad” o “diferencia” entre pares de conjuntos difusos.

Definición 11 ([11], Bouchon-Meunier et al., 1996) *d es una M-medida de disimilaridad si satisface los axiomas $G1$, $G4^*$, $D3$ y la función asociada $F_d : \mathbb{R}^3 \rightarrow \mathbb{R}$ no depende del primer argumento y es creciente en el segundo y tercer argumento.*

Nota 1.4 *En la definición original, se considera una definición axiomática para el concepto de “diferencia” entre conjuntos difusos, diferente de otras definiciones propuestas por Kaufman, Roberts ó Dubois y Prade. Cuando, además, suponemos que se satisface la siguiente propiedad:*

$$(\tilde{A} \setminus \tilde{B}) \setminus (\tilde{B} \setminus \tilde{A}) = \tilde{A} \setminus \tilde{B}, \quad \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U), \quad (1.1)$$

la definición anterior de disimilaridad es equivalente a la siguiente:

d es una función de disimilaridad si satisface los axiomas $G1$, $G4^$, $D3$, $D6$ y $D7$.*

De hecho, puede comprobarse fácilmente que toda medida de comparación d satisface los axiomas $D6$ y $D7$ si y solo si F_d satisface las propiedades requeridas en la definición 11. A partir de

este punto, supondremos que todas las operaciones de diferencia consideradas en el resto del capítulo satisfacen la Ecuación 1.1.

Definición 12 ([99] L. Xuecheng, 1992) *Una función con valores reales d sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de distancia si satisface las propiedades $G2$, $D1$, $D2$ y $D3$.*

Definición 13 ([106], H. Zhang et al., 2009) *Una función con valores reales d sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de distancia si satisface las propiedades $G1^*$, $G2$, $D1$, $D2$ y $D3^*$.*

Nota 1.5 *El término medida de distancia aparece en [99, 106], pero no se refiere al concepto matemático de métrica.*

Definición 14 ([76], S. Montes et al., 2002) *Una función con valores reales d sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de divergencia si satisface las propiedades $G2$, $D3$, $D4$ y $D5$.*

Definición 15 ([76], S. Montes et al., 2002) *Una función con valores reales d sobre $\mathcal{F}(U) \times \mathcal{F}(U)$ se denomina medida de divergencia local si satisface las propiedades de medida de divergencia ($G2$, $D3$, $D4$ y $D5$) y, además, la propiedad $G3^*$.*

1.4.2. Relación formal entre las diferentes medidas de “desigualdad”

En esta sección, estudiaremos las relaciones formales entre las definiciones axiomáticas mencionadas en la anterior sección (de la definición 11 a la definición 14). Para proporcionar la información de una forma organizada y clara, y para simplificar su posterior análisis, resumiremos dicha información a través de la siguiente tabla.

Definición	G1	G1*	G2	G3*	G4*
Def. 11	•				•
Def. 12			•		
Def. 13		•	•		
Def. 14			•		
Def. 15			•	•	

Definición	D1	D2	D3	D3*	D4	D5	D6	D7
Def. 11			•				•	•
Def. 12	•	•	•					
Def. 13	•	•		•				
Def. 14			•		•	•		
Def. 15			•		•	•		

A continuación revisaremos las conexiones existentes entre los diferentes axiomas considerados en estas definiciones. La literatura sobre ello es escasa y, de hecho, se reduce a algunos resultados dados en [76]. Resumiremos esos resultados en la siguiente proposición.

Proposición 7 *Sea d una aplicación definida sobre $\mathcal{F}(U) \times \mathcal{F}(U)$:*

- *Si d satisface $D4$ y $D5$ entonces satisface $D1$. ([76], proposición 2.4)*
- *Si, además, d satisface $G3^*$ entonces satisface $D2$. ([76], proposición 3.11)*

Nota 1.6 *Hemos mencionado en la proposición 7 que toda medida satisfaciendo los axiomas $D4$ y $D5$ también satisface el axioma $D1$. Concretamente, puede comprobarse que el axioma $D4$ implica $D1a$ y el axioma $D5$ implica $D1b$. Por otro lado, ninguno de ellos por separado implica el axioma $D1$, como fue probado en [75].*

De acuerdo con los últimos resultados, podemos completar la tabla anterior. Utilizaremos la misma nomenclatura que en la sección anterior: si una celda está llena, significará que la correspondiente definición (fila) satisface la correspondiente propiedad (columna). Si está rellena con un punto, significa que la propiedad está incluida en la definición original. Si el contenido es S, significa que la propiedad se deriva de forma sencilla, directa. Si el contenido es Px, significa que la propiedad se satisface según lo demostrado en la proposición x.

Definición	G1	G1*	G2	G3*	G4*
Def. 11	•				•
Def. 12			•		
Def. 13	S	•	•		
Def. 14			•		
Def. 15			•	•	

Definición	D1	D2	D3	D3*	D4	D5	D6	D7
Def. 11			•				•	•
Def. 12	•	•	•					
Def. 13	•	•	S	•				
Def. 14	P7		•		•	•		
Def. 15	P7	P7	•		•	•		

A partir de la información mostrada en la tabla anterior, podemos establecer el siguiente resultado.

Corolario 3 *Se verifican las siguientes implicaciones:*

- *Si d satisface la definición 13, entonces satisface la definición 12.*
- *Si d satisface la definición 15, entonces satisface la definición 14.*
- *Si d satisface la definición 15 entonces satisface la definición 12.*

Podemos ilustrar este resultado a través del siguiente diagrama donde la flecha significa “es al menos tan restrictivo como”.

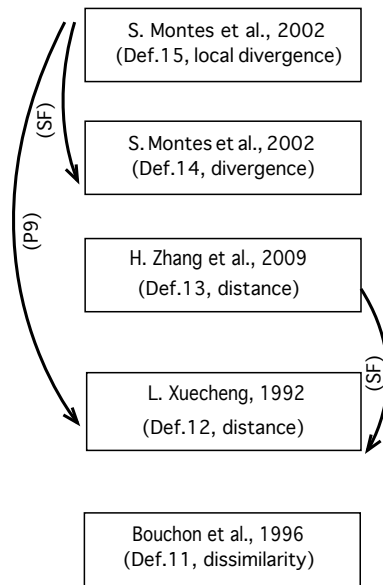


Figura 1.4: Resumen de las relaciones de implicación entre las diferentes definiciones axiomáticas para la noción de “desigualdad”.

La figura 1.4 muestra que existen algunas relaciones de implicación entre algunas de las definiciones mencionadas en la sección 1.4.1. Para completar el estudio sobre las relaciones formales entre las medidas de “desigualdad” entre conjuntos difusos, mostraremos a continuación que el resto de implicaciones posibles no son ciertas en general a través de los siguientes contraejemplos.

Ejemplo 8 *Consideremos un universo U de cardinal 1 y una aplicación $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida como sigue:*

$$d(\emptyset, U) = d(U, \emptyset) = 5 \text{ y } d(\tilde{A}, \tilde{B}) = 0 \text{ en otro caso.}$$

Puede comprobarse fácilmente que d satisface los axiomas $G2$, $D1$, $D2$ y $D3$. Por otro lado, no verifica los axiomas $G1^$ y $D3^*$. Por tanto, d es una “medida de distancia ” según la definición 12, pero no según la definición 13.*

Ejemplo 9 *Consideremos un universo arbitrario U y la aplicación $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida del siguiente modo:*

$$d(\tilde{A}, \tilde{B}) = \begin{cases} 1 & \text{si } \tilde{A} \subsetneq \tilde{B} \text{ o } \tilde{B} \subsetneq \tilde{A} \text{ o } \tilde{A} = \tilde{B}^c, \\ 0 & \text{si } \tilde{A} = \tilde{B} \\ 0,5 & \text{en otro caso.} \end{cases}$$

Esta aplicación satisface todas las propiedades incluidas en la definición 13. De hecho, toma valores dentro del intervalo $[0, 1]$, y alcanza el valor 1 al menos para el par de conjuntos $(\emptyset, U)$. Además, satisface los axiomas $G2$, $D2$ y $D3^$ por definición. Es sencillo comprobar que satisface el axioma $D1$. De hecho, si*

consideramos tres conjuntos difusos anidados $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$, de acuerdo con la definición de la aplicación d , observamos que

$$d(\tilde{A}, \tilde{C}) = \begin{cases} 1 & \text{si } \tilde{A} \neq \tilde{C}, \\ 0 & \text{si } \tilde{A} = \tilde{C}. \end{cases}$$

En ambas situaciones, es fácil comprobar que

$$\max\{d(\tilde{A}, \tilde{B}), d(\tilde{B}, \tilde{C})\} \leq d(\tilde{A}, \tilde{C}).$$

Por otro lado, d no satisface la definición 14. Para demostrarlo, tomemos dos subconjuntos difusos diferentes pero no disjuntos, $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$, tales que ninguno de ellos está incluido en el otro. Sea $\tilde{C} = \tilde{A}$. Bajo estas condiciones

$$d(\tilde{A} \cup \tilde{C}, \tilde{B} \cup \tilde{C}) = d(\tilde{A}, \tilde{B} \cup \tilde{A}) = 1$$

y

$$d(\tilde{A} \cap \tilde{C}, \tilde{B} \cap \tilde{C}) = d(\tilde{A}, \tilde{A} \cap \tilde{B}) = 1.$$

Sin embargo, $d(\tilde{A}, \tilde{B})$ es igual a 0,5. De acuerdo con esto, tenemos que d no satisface la propiedad D4 ni la propiedad D5, y por tanto no es una medida de divergencia.

En conclusión, este ejemplo nos muestra que la definición 13 no implica la definición 14 en general.

Ejemplo 10 Consideremos un universo arbitrario U y una aplicación $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida del siguiente modo:

$$d(\tilde{A}, \tilde{B}) = \frac{\#\{x : \mu_{\tilde{A}}(x) \neq \mu_{\tilde{B}}(x)\}^{\frac{1}{\#(\tilde{A} \cap \tilde{B}) + 1}}}{\#U},$$

donde $\#\tilde{A}$ denota el “cardinal” del conjunto difuso $\tilde{A}$:

$$\#\tilde{A} = \sum_{x \in U} \mu_{\tilde{A}}(x).$$

La aplicación d satisface todas las propiedades incluidas en la definición 13: es positiva y satisface el axioma de simetría (G2) por construcción, es creciente con respecto a $\#\{x : \mu_{\tilde{A}}(x) \neq \mu_{\tilde{B}}(x)\}$ y decreciente con respecto a $\#(\tilde{A} \cap \tilde{B})$. Por tanto, comprobamos fácilmente que satisface el axioma D1. Además, d asigna su máximo valor al par $A = \emptyset$ y $B = U$, y observamos que $d(\emptyset, U) = 1$. Así pues, d satisface G1*. También es sencillo ver que $d(D, D^c) = 1$, para cualquier conjunto nítido D , y por tanto se satisface la propiedad D2. Finalmente, $d(\tilde{A}, \tilde{B})$ es nulo si y solo si $\tilde{A} = \tilde{B}$, por construcción.

Por otro lado, d no satisface la definición 11. Para demostrarlo, consideremos el universo $U = \{1, 2, 3, 4\}$ y los conjuntos nítidos $A = \{1, 2, 3\}$ y $B = \{2, 3, 4\}$. Entonces, $d(A, B) = \frac{\sqrt[3]{2}}{4}$ es menor que $d(A \setminus B, B \setminus A) = 1$.

En conclusión, este ejemplo nos muestra que la definición 13 no implica la definición 11 en general.

Ejemplo 11 Consideremos un universo de cardinal 2, y denotémoslo por $U = \{x_1, x_2\}$. Sea $d : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow \mathbb{R}$ una aplicación definida del siguiente modo:

- La restricción de la aplicación d a los pares de subconjuntos nítidos de U queda representada por la siguiente tabla de doble entrada:

	$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_1, x_2\}$
$\emptyset$	0	1	1	3
$\{x_1\}$	1	0	2	1
$\{x_2\}$	1	2	0	1
$\{x_1, x_2\}$	3	1	1	0

- En la familia de los pares de subconjuntos difusos, se define a partir de la definición dada para los conjuntos nítidos a través de la siguiente igualdad:

$$d(\tilde{A}, \tilde{B}) = d(C_{\tilde{A}}, C_{\tilde{B}}),$$

donde $C_{\tilde{A}}$ representa el “conjunto nítido más cercano a $\tilde{A}$ ”, es decir, $C_{\tilde{A}} = \{x \in U : \mu_{\tilde{A}}(x) > 0,5\}$.

Teniendo en cuenta que $C_{\tilde{A} \cap \tilde{B}} = C_{\tilde{A}} \cap C_{\tilde{B}}$ y $C_{\tilde{A} \cup \tilde{B}} = C_{\tilde{A}} \cup C_{\tilde{B}}$, podemos comprobar fácilmente que d satisface todas las propiedades incluidas en la definición 14.

Pero observamos que no satisface la propiedad D2. De hecho, si consideramos el conjunto $A = \{x_1\}$, observamos que la divergencia entre él y su complementario, $A^c = \{x_2\}$, no alcanza el máximo valor de d .

De acuerdo con todo esto, tenemos que no toda medida de divergencia (no local) satisface la definición 12 y la definición 13.

Ejemplo 12 Consideremos un universo unipuntual $U = \{x\}$ y la medida de comparación $d : \mathcal{F}(U) \times \mathcal{F}(U)$ determinada por la medida difusa $M : \mathcal{F}(U) \rightarrow \mathbb{R}$ dada por $M(\tilde{A}) = \mu_{\tilde{A}}(x)$ y la

función $F : \mathbb{R}^3 \longrightarrow \mathbb{R}$ definida como $F(u, v, w) = \frac{v+w}{2}$. F no depende de su primer argumento, u , y es creciente en v y w . Por tanto d es una medida de disimilaridad. Ahora, probaremos que no satisface el axioma D1. Consideremos los conjuntos difusos anidados $\tilde{A} \subseteq \tilde{B} \subseteq \tilde{C}$ definidos como sigue: $\mu_{\tilde{A}}(x) = 0,2$, $\mu_{\tilde{B}}(x) = 0,8$, $\mu_{\tilde{C}}(x) = 0,9$. Observemos que $\mu_{\tilde{A} \setminus \tilde{B}}(x) = 0,2$, $\mu_{\tilde{A} \setminus \tilde{C}}(x) = 0,1$, $\mu_{\tilde{B} \setminus \tilde{A}}(x) = 0,8$, $\mu_{\tilde{C} \setminus \tilde{A}}(x) = 0,8$. En consecuencia, $d(\tilde{A}, \tilde{B}) = 0,5 > d(\tilde{A}, \tilde{C}) = 0,45$.

De acuerdo con este ejemplo, la definición 11 no implica el axioma D1 y por tanto no implica las definiciones 12, 13, 14 y 15.

Ejemplo 13 Consideremos un universo finito arbitrario U y la aplicación $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida como

$$d(\tilde{A}, \tilde{B}) = \sum_{x \in U} h(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)),$$

donde

$$h(a, b) = \begin{cases} |a - b| & \text{si } a < 0,5 \text{ ó } b < 0,5, \\ 0 & \text{en otro caso.} \end{cases}$$

Podemos comprobar que d es una medida de divergencia local, ya que h satisface la siguiente desigualdad:

$$\max\{h(\max\{a, c\}, \max\{b, c\}), h(\min\{a, c\}, \min\{b, c\})\} \leq h(a, b),$$

$$\forall a, b, c \in [0, 1].$$

Por otro lado, d no es una medida de disimilaridad. No existe ninguna medida difusa $M : \mathcal{F}(U) \longrightarrow \mathbb{R}$ tal que $d(\tilde{A}, \tilde{B})$ pueda

expresarse en función de $M(\tilde{A} \setminus \tilde{B})$ y $M(\tilde{B} \setminus \tilde{A})$. Para probarlo, consideremos los conjuntos difusos $\tilde{A}, \tilde{B}, \tilde{A}'$ y $\tilde{B}'$ cuyas funciones de pertenencia están definidas del siguiente modo para cierto $x^* \in U$:

$$\mu_{\tilde{A}}(x^*) = 0,1, \mu_{\tilde{B}}(x^*) = 0,4, \mu_{\tilde{A}'}(x^*) = 0,6, \mu_{\tilde{B}'}(x^*) = 0,9$$

$$\mu_{\tilde{A}}(x) = \mu_{\tilde{B}}(x) = \mu_{\tilde{A}'}(x) = \mu_{\tilde{B}'}(x) = 0, \quad \forall x \neq x^*.$$

Entonces $d(\tilde{A}, \tilde{B}) = 0,3$ no coincide con $d(\tilde{A}', \tilde{B}') = 0$. Por otro lado, se tienen las siguientes igualdades:

$$\min\{\mu_{\tilde{A}}(x), 1 - \mu_{\tilde{B}}(x)\} = \min\{\mu_{\tilde{A}'}(x), 1 - \mu_{\tilde{B}'}(x)\}$$

y

$$\min\{\mu_{\tilde{B}}(x), 1 - \mu_{\tilde{A}}(x)\} = \min\{\mu_{\tilde{B}'}(x), 1 - \mu_{\tilde{A}'}(x)\}, \quad \forall x \in U.$$

Por tanto, $\tilde{A} \setminus \tilde{B} = \tilde{A}' \setminus \tilde{B}'$ y $\tilde{B} \setminus \tilde{A} = \tilde{B}' \setminus \tilde{A}'$. Es decir, no se verifica el axioma D7.

De acuerdo con este ejemplo, las definiciones 15 y 14 no implican la definición 11.

Ejemplo 14 Consideremos un universo finito arbitrario U y la aplicación $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida como

$$d(\tilde{A}, \tilde{B}) = \sum_{x \in U} h(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)),$$

donde

$$h(a, b) = \begin{cases} 1 & \text{si } a = 0 \text{ ó } a = 1 \text{ y } b = 1 - a, \\ 0 & \text{en otro caso.} \end{cases}$$

Puede comprobarse fácilmente que d es una medida de divergencia local, dado que h satisface la desigualdad:

$$\begin{aligned} \max\{h(\max\{a, c\}, \max\{b, c\}), h(\min\{a, c\}, \min\{b, c\})\} &\leq h(a, b), \\ \forall a, b, c &\in [0, 1]. \end{aligned}$$

Por otro lado, d no satisface la propiedad $D3^$. Según esto, la definición 15 no implica la definición 13.*

Para resumir la información que acabamos de proporcionar, podemos colocar todos nuestros contraejemplos en el siguiente diagrama.

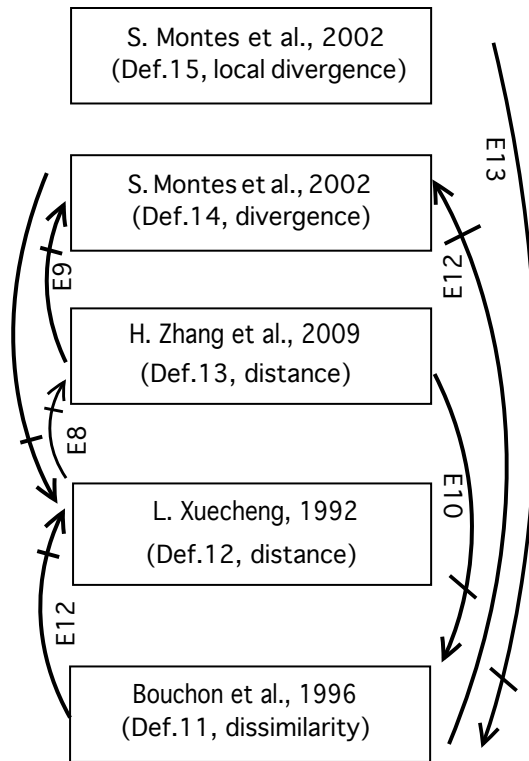


Figura 1.5: Contraejemplos relativos a las relaciones de implicación entre las diferentes variantes de los conceptos de “disimilaridad”, “distancia” y “divergencia”.

1.4.3. El caso particular de los conjuntos nítidos

Según los contraejemplos vistos, no hay más relaciones de implicación entre las definiciones 11 a 15, aparte de aquellas implicaciones representadas en la figura 1.4. Sin embargo, podemos encontrar más relaciones si nos restringimos a la familia de los conjuntos nítidos. Lo comprobamos a continuación.

Proposición 8 *Si d satisface el axioma D6 entonces satisface el axioma D1 (cuando nos restringimos a la familia de pares de conjuntos nítidos).*

Demostración:

Considerar tres subconjuntos nítidos anidados del universo, $A \subseteq B \subseteq C$.

- a) $A \cap B = A \cap C = A$, $A \setminus B = \emptyset = A \setminus C$ y $B \setminus A \subseteq C \setminus A$, y por tanto, de acuerdo con D6, $d(A, B) \leq d(A, C)$.
- b) $A \cap C \subseteq B \cap C$, $A \setminus C = B \setminus C = \emptyset$ y $C \setminus B \subseteq C \setminus A$. Por tanto, de acuerdo con D6, $d(B, C) \leq d(A, C)$. $\square$

Proposición 9 *Si d satisface los axiomas D6 y D7, entonces d verifica los axiomas D4 y D5 (cuando nos restringimos a la familia de pares de conjuntos nítidos).*

Demostración: Supongamos que d satisface D6 y D7 y consideremos tres subconjuntos nítidos arbitrarios A, B y $C \in \wp(U)$.

- Para comprobar que se verifica D4, compararemos los valores de divergencia $d(A, B)$ y $d(A \cup C, B \cup C)$. De acuerdo con D7,

$$d(A, B) = d(A \setminus B, B \setminus A)$$

y

$$d(A \cup C, B \cup C) = d((A \cup C) \setminus (B \cup C), (B \cup C) \setminus (A \cup C)).$$

Además, observamos que

$$(A \cup C) \setminus (B \cup C) = A \cap \overline{B} \cap \overline{C}$$

y

$$(B \cup C) \setminus (A \cup C) = B \cap \overline{A} \cap \overline{C}.$$

Entonces, de acuerdo con D6, obtenemos que

$$d(A \setminus B, B \setminus A) \geq d(A \cap \overline{B} \cap \overline{C}, B \cap \overline{A} \cap \overline{C}),$$

y entonces

$$d(A, B) \geq d(A \cup C, B \cup C).$$

- Para comprobar que se verifica D5, compararemos los valores de divergencia $d(A, B)$ y $d(A \cap C, B \cap C)$. De acuerdo con D7,

$$d(A, B) = d(A \setminus B, B \setminus A)$$

y

$$d(A \cap C, B \cap C) = d((A \cap C) \setminus (B \cap C), (B \cap C) \setminus (A \cap C)).$$

Además, podemos comprobar fácilmente las igualdades

$$(A \cap C) \setminus (B \cap C) = A \cap \overline{B} \cap C$$

y

$$(B \cap C) \setminus (A \cap C) = B \cap \bar{A} \cap C.$$

De acuerdo con D6, obtenemos que

$$d(A \setminus B, B \setminus A) \geq d(A \cap \bar{B} \cap C, B \cap \bar{A} \cap C).$$

Por tanto

$$d(A, B) \geq d(A \cap C, B \cap C). \quad \square$$

Lema 1 *Consideremos el operador diferencia asociado a la t -norma del mínimo y la t -conorma del máximo. Si m satisface $G3^*$, entonces satisface $G4$.*

Demostración: Supongamos que m satisface $G3^*$, es decir, que existe una función $h : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ tal que

$$m(A, B) = \sum_{x \in U} h(\mu_A(x), \mu_B(x)).$$

Entonces, existe una aplicación $f_h : \mathbb{R}^3 \rightarrow \mathbb{R}$ tal que

$$m(A, B) = \sum_{x \in U} f_h(\mu_{A \cap B}(x), \mu_{A \setminus B}(x), \mu_{B \setminus A}(x)).$$

Tal aplicación puede expresarse del siguiente modo:

$$f_h(\lambda_1, \lambda_2, \lambda_3) = \begin{cases} h(1 - \lambda_3, \lambda_1) & \text{si } \lambda_1 < \lambda_2, \text{ y } \lambda_1 = 1 - \lambda_2, \\ h(\lambda_2, \lambda_1) & \text{si } \lambda_1 > \lambda_2, \text{ y } \lambda_1 = 1 - \lambda_3, \\ h(1 - \lambda_3, 1 - \lambda_2) & \text{si } \lambda_1 > \lambda_2, \\ h(1 - \lambda_3, 0,5) & \text{si } \lambda_1 = \lambda_2 = 0,5, \\ h(\lambda_2, \lambda_3) & \text{si } \lambda_1 = \lambda_2 < 0,5. \end{cases} \quad \square$$

Proposición 10 *Consideremos la función $d : \wp(U) \times \wp(U) \longrightarrow \mathbb{R}$.*

(a) *Si d satisface el axioma D5, entonces*

$$d(A, B) \leq d(A \setminus B, B \setminus A), \quad \forall A, B \in \wp(U).$$

(b) *Si d satisface el axioma D4, entonces*

$$d(A, B) \geq d(A \setminus B, B \setminus A), \quad \forall A, B \in \wp(U).$$

(c) *Si d satisface el axioma D4, entonces*

$$d(A, B) \leq d(C, B),$$

para todo $A, B, C \in \wp(U)$ tal que $C \cap B = \emptyset$ y $A \subseteq C$.

(d) *Si d satisface el axioma D4 y D5, entonces*

$$d(A, B) \leq d(B^c, B), \quad \forall A, B \in \wp(U).$$

Demostración:

(a) Si aplicamos el axioma D5 a los conjuntos $A' = (A \setminus B)$, $B' = (B \setminus A)$ y $C = A \cap B$ demostramos la primera implicación.

(b) Si aplicamos el axioma D4 a los conjuntos A , B y $C = A^c \cup B^c$ demostramos la segunda implicación.

(c) Si aplicamos el axioma D4 a los conjuntos C , B y $B \cup A$ demostramos la tercera implicación.

(d) De acuerdo con el axioma D5 tenemos que

$$d(A, B) = d((A \setminus B) \cup (B \cap A), B \cup (B \cap A)) \leq d(A \setminus B, B).$$

Además, teniendo en cuenta que $A \setminus B \subseteq B^c$ y $B^c \cap B = \emptyset$ y, de acuerdo con el apartado (c), obtenemos que

$$d(A \setminus B, B) \leq d(B^c, B). \quad \square$$

Nota 1.7 *Teniendo en cuenta la proposición anterior, toda aplicación d que verifique $D4$ y $D5$ satisface $D6$ y $D7$, cuando nos restringimos a la familia de los conjuntos nítidos. Además, d satisface la siguiente propiedad, la cual es menos restrictiva que $D2$:*

$$d(A, B) \leq d(B^c, B), \quad \forall A, B \in \wp(U).$$

A partir de este punto, denotaremos dicha propiedad por $D2^-$.

Notar que no toda medida de divergencia (no local) satisface el axioma $D2$, porque no necesariamente satisface las igualdades

$$d(A^c, A) = d(B^c, B), \quad \forall A, B \in \wp(U).$$

Este hecho queda demostrado en el Ejemplo 11, donde consideramos una medida de divergencia no local que no satisface el axioma $D2$, incluso aunque nos restrinjamos a la familia de los conjuntos nítidos.

De acuerdo con los últimos resultados, si nos restringimos a la familia de los subconjuntos nítidos del universo, podemos completar la tabla resumen de las propiedades verificadas por cada definición. Utilizaremos el mismo tipo de nomenclatura que en secciones anteriores. Si una celda está llena, la correspondiente

definición axiomática (fila) satisface la correspondiente propiedad (columna). Si está rellena con un punto, significa que la propiedad está incluida en la definición original. Si el contenido es S, significa que la propiedad se comprueba de forma sencilla, directa. Si el contenido es Px, significa que la propiedad se satisface por lo demostrado en la proposición x.

Definición	G1	G1*	G2	G3*	G4*
Def. 11	•				•
Def. 12			•		
Def. 13	S	•	•		
Def. 14			•		
Def. 15			•	•	

Definición	D1	D2 ⁻	D2	D3	D3*	D4	D5	D6	D7
Def. 11	P8	P9+P10		•		P9	P9	•	•
Def. 12	•	S	•	•					
Def. 13	•	S	•	S	•				
Def. 14	P7	P10		•		•	•	P10	P10
Def. 15	P7	S	P7	•		•	•	P10	P10

De acuerdo con la información contenida en la tabla anterior, podemos probar de forma directa varias implicaciones entre las definiciones de disimilaridad, divergencia (local/no local) y distancia mencionadas en la sección 1.4.1, cuando nos restringimos a la familia de los subconjuntos nítidos:

Corolario 4 *Consideremos un universo arbitrario U .*

- *Toda medida de disimilaridad simétrica es una medida de divergencia.*
- *Toda medida de disimilaridad simétrica es una medida de distancia de acuerdo con la definición 12, excepto por la propiedad $d(A, A^c) = d(B, B^c)$, $\forall A, B \in \wp(U)$.*

- Toda medida de divergencia que pueda expresarse como una medida de comparación es una medida de disimilaridad.
- Toda medida de divergencia es una medida de distancia según la definición 12, excepto por la propiedad $d(A, A^c) = d(B, B^c)$, $\forall A, B \in \wp(U)$.

Podemos resumir toda esta información a través del siguiente diagrama.

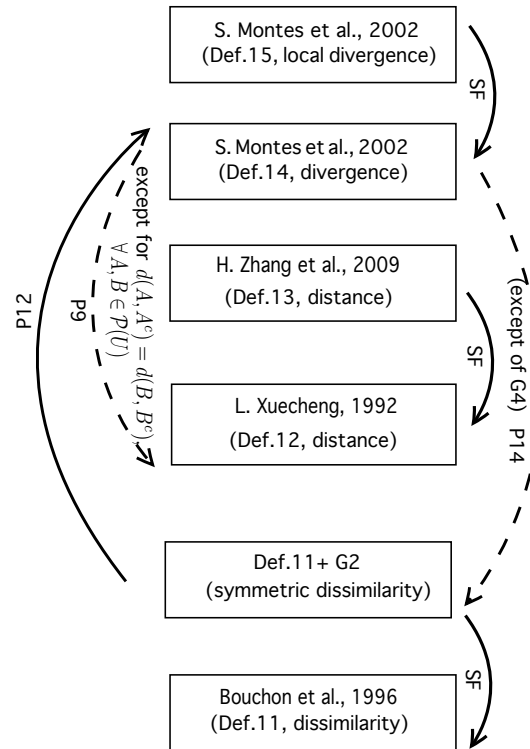


Figura 1.6: Resumen de las relaciones de implicación entre las diferentes variantes de las nociones de ‘disimilaridad’, ‘distancia’ y ‘divergencia’, cuando nos restringimos a la familia de los conjuntos nítidos.

De acuerdo con esto, las nociones de “divergencia” y “disimilaridad simétrica” son muy similares. Toda disimilaridad simétrica, cuando se restringe a los conjuntos nítidos, satisface los axiomas de medida de divergencia. Inversamente, toda medida de divergencia que puede expresarse como una medida de comparación es una medida de disimilaridad. Además, si tenemos en cuenta el resultado dado en el lema 1, toda medida de divergencia local, puede expresar la divergencia entre A y B como una función de $A \cap B$, $A \setminus B$ y $B \setminus A$. Por tanto, la restricción a la familia de los conjuntos nítidos de toda medida de divergencia local satisface todas las propiedades de una medida de disimilaridad simétrica, siempre que exista una medida difusa M que permita expresar la divergencia entre A y B como una función de $M(A \cap B)$, $M(A \setminus B)$ y $M(B \setminus A)$.

1.5. Dualidad entre las nociones de igualdad y desigualdad

Algunos de los trabajos mencionados en las secciones 1.3 y 1.4 proponen pares de conceptos duales de medidas de “igualdad” y “desigualdad”, mientras que otros no lo hacen. Por ejemplo, las definiciones de “medidas de similaridad” y “medidas de distancia” dadas por Zhang et al. en [106] son duales, en el sentido de que, dada una medida de similaridad, s , la función $d = 1 - s$ es una medida de distancia, y viceversa. Una relación similar existe entre los conceptos dados en [99]. En este caso, observamos que cualquier función no creciente de una medida de similaridad satisface las propiedades de una medida de distancia y viceversa. Por otro lado, Montes et al. introducen el concepto de divergen-

cia en [76], pero no introducen un concepto dual, y por ello no se ha dado en la sección 1.2.2 ningún axioma dual de los axiomas D4 y D5. En cuanto a Bouchon et al. introducen las definiciones de similitud y semejanza, correspondientes al bloque de medidas de “igualdad”, y el concepto de disimilaridad, correspondiente al bloque de medidas de “desigualdad”.

La siguiente proposición compila algunos resultados respecto a la relación formal existente entre algunos de los axiomas de “igualdad” enumerados en la sección 1.2.2 y sus axiomas duales enumerados en la sección 1.2.3. La demostración es inmediata.

Proposición 11 *Consideremos un universo arbitrario U . Denotemos por $\mathcal{F}(U)$ la familia de los subconjuntos difusos de U .*

1. $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ *satisface el apartado a) del axioma D1 si y solo si $s = g \circ d$ satisface $S1^*$ para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$. Además:*
 - *Si s satisface el axioma $S1^*$ y h es una función decreciente (no necesariamente estrictamente) arbitraria entonces $h \circ s$ satisface el apartado a) del axioma D1.*
 - *Si d satisface el apartado a) del axioma D1 y g es una función decreciente (no necesariamente estrictamente) arbitraria entonces $g \circ d$ satisface el axioma $S1^*$.*
2. $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ *satisface el apartado b) del axioma D1 si y solo si $s = g \circ d$ satisface $S2$ para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$. Además:*

- Si s satisface el axioma $S2$ y h es una función decreciente (no necesariamente estrictamente) arbitraria entonces $h \circ s$ satisface el apartado b) del axioma $D1$.
 - Si d satisface el apartado b) del axioma $D1$ y g es una función decreciente (no necesariamente estrictamente) arbitraria entonces $g \circ d$ satisface el axioma $S2$.
3. $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface el axioma $D2$ si y solo si $s = g \circ d$ satisface el axioma $S3$ y toma valores no negativos para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$ tal que $g(\max \text{Im}(d)) = 0$. Además:
- Si d satisface el axioma $D2$ y $g : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $g(\max \text{Im}(d)) = 0$, entonces $g \circ d$ satisface el axioma $S3$.
 - Si s satisface el axioma $S3$ y $h : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $h(0) = \max \text{Im}(h \circ s)$, entonces $h \circ s$ satisface el axioma $D2$.
4. $s : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface el axioma $S4$ si y solo si $d = g \circ s$ satisface el axioma $D3$ y toma valores no negativos para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$ tal que $g(\max \text{Im}(s)) = 0$. Además:
- Si s satisface el axioma $S4$ y $g : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $g(\max \text{Im}(s)) = 0$, entonces $g \circ d$ satisface el axioma $D3$.
 - Si d satisface el axioma $D3$ y $h : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $h(0) = \max \text{Im}(h \circ d)$, entonces $h \circ d$ satis-

face el axioma $S4$.

5. $s : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface el axioma $S4^*$ si y solo si $d = g \circ s$ satisface el axioma $D3^*$ y toma valores no negativos para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$ tal que $g(\max \text{Im}(s)) = 0$. Además:

- Si s satisface el axioma $S4^*$ y $g : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $g(x) = 0 \Leftrightarrow x = \max \text{Im}(s)$ entonces $g \circ d$ satisface el axioma $D3$.
- Si d satisface el axioma $D3^*$ y $h : \mathbb{R} \longrightarrow \mathbb{R}$ satisface la restricción $h(y) = \max \text{Im}(h \circ d) \Leftrightarrow y = 0$, entonces $h \circ d$ satisface el axioma $S4^*$.

6. $d : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ satisface el axioma $D6$ si y solo si $s = g \circ d$ satisface el axioma $S7$ para toda función estrictamente decreciente $g : \mathbb{R} \longrightarrow \mathbb{R}$. Además:

- Si s satisface el axioma $S7$ y h es una función decreciente (no necesariamente estrictamente) arbitraria entonces $h \circ s$ satisface el axioma $D6$.
- Si d satisface el axioma $D6$ y g es una función decreciente (no necesariamente estrictamente) arbitraria entonces $g \circ d$ satisface el axioma $S7$.

La siguiente tabla resume la información dada en la proposición 11.

Igualdad	Desigualdad
S1*	D1(a)
S2	D1(b)
S3	D2
S4	D3
S4*	D3*
S7	D6

Hemos dividido las definiciones axiomáticas dadas en este capítulo en dos bloques: medidas de “igualdad” y medidas de “desigualdad” entre conjuntos difusos. Si componemos cualquiera de esas medidas con una función estrictamente creciente, la medida resultante permanecerá en el mismo bloque o familia de medidas. Si, por el contrario, la componemos con una función estrictamente decreciente entonces la medida resultante pertenecerá al otro bloque. Este hecho está relacionado con la noción de “*equivalencia en orden*” entre dos medidas de similaridad considerada por Omhover et al. en [87] y por Bouchon-Meunier et al. en [10]. Dos medidas de similaridad se dicen equivalentes [87] cuando son co-monótonas, es decir, cuando

$$s(\tilde{A}, \tilde{B}) \leq s(\tilde{C}, \mathcal{D}) \Leftrightarrow s'(\tilde{A}, \tilde{B}) \leq s'(\tilde{C}, \mathcal{D}), \forall \tilde{A}, \tilde{B}, \tilde{C}, \mathcal{D} \in \mathcal{F}(U).$$

En otras palabras, ambas establecen la misma ordenación cuando comparan cualquier subconjunto difuso con otro fijado. Es sencillo comprobar que dos medidas de similaridad son equivalentes en orden si y solo si existe una función estrictamente creciente $f : \mathbb{R} \longrightarrow \mathbb{R}$ relacionándolas.

En nuestro contexto, toda función creciente de una medida de similaridad fijada va a satisfacer la misma definición axiomática

que la medida original, excepto aquellas condiciones de acotación exigidas en algunas definiciones ($s(\tilde{A}, \tilde{A}) = 1$, etc.). Si queremos garantizar que también se cumplen dichas acotaciones, necesitaremos exigir algunas condiciones adicionales a f .

De modo similar, podemos introducir el concepto de “dualidad”. A partir de ahora diremos que una medida $d : \mathcal{F}(U) \times \mathcal{F}(U)$ es *dual* a cierta medida s si y solo si existe una función estrictamente decreciente g tal que $d = g \circ s$. Esto es, obviamente, equivalente a decir que ambas medidas son “antítonas”, es decir,

$$s(\tilde{A}, \tilde{B}) \leq s(\tilde{C}, \mathcal{D}) \Leftrightarrow d(\tilde{A}, \tilde{B}) \geq d(\tilde{C}, \mathcal{D}), \quad \forall \tilde{A}, \tilde{B}, \tilde{C}, \mathcal{D} \in \mathcal{F}(U).$$

Si una medida satisface ciertas propiedades de la columna izquierda de la tabla anterior, entonces su medida dual satisface las correspondientes propiedades de la columna derecha.

Capítulo 2

El centroide y la d -variación de conjuntos aleatorios y variables aleatorias difusas.

2.1. Motivación y conceptos previos

En determinadas situaciones, un experimento aleatorio no puede ser descrito a través de una variable aleatoria ordinaria. Para solventar este problema existen otras herramientas matemáticas, como son los conjuntos aleatorios y las variables aleatorias difusas.

De acuerdo con Dubois y Prade ([40]), la definición formal de conjunto admite dos interpretaciones llamadas “conjuntiva” y “disyuntiva”. Si, desde la perspectiva “conjuntiva”, el conjunto representa una colección de objetos que forman un ente “compuesto”, proporcionando información precisa acerca de él, un conjunto “disyuntivo” representa información incompleta acerca de cierto elemento del universo, que denotaremos por x ; cualquier elemento del conjunto es un posible candidato para x , mientras que cualquier elemento fuera del conjunto es conside-

rado como imposible.

Del mismo modo, las aplicaciones multivaluadas, es decir, aquellas cuyos valores no son elementos del universo en estudio sino subconjuntos del mismo, admiten ambas interpretaciones. En este trabajo consideraremos la interpretación basada en el enfoque conjuntista.

La dualidad de las visiones conjuntiva y diyuntiva se puede generalizar al caso de los conjuntos difusos. Los autores Dubois y Prade ([39]) distinguen, en este caso más general, entre las perspectivas “óptica” y “epistémica”. De acuerdo con Dubois ([37]), los conjuntos epistémicos sirven para representar información imprecisa acerca de cierto elemento del universo, incluyendo grados de posibilidad. Así, el grado de pertenencia de un elemento “e” al conjunto difuso representará el grado de posibilidad de que “e” sea realmente el elemento desconocido (véase [27]). Por el contrario, los conjuntos difusos ópticos representan objetos contruidos originalmente como conjuntos, pero para los cuales la representación difusa es más expresiva, debido a sus fronteras graduales. Los grados de pertenencia en estos casos representan el grado en que cada componente participa de la entidad global.

De acuerdo con estas interpretaciones, las variables aleatorias difusas, es decir, las aplicaciones cuyo espacio de llegada es el de los subconjuntos difusos de cierto universo, también pueden ser vistas desde dos perspectivas. A lo largo de este capítulo, se tendrá en cuenta la perspectiva óptica. Además de este hecho, debemos recordar que existen distintas definiciones del concepto de variable aleatoria difusa, basadas la naturaleza del espacio de

llegada considerada y la condición de medibilidad exigida.

En este capítulo, profundizaremos en el estudio de los conceptos de medidas de tendencia central y dispersión dentro de este contexto. Generalizaremos, así, algunos estudios precedentes acerca de los conceptos de esperanza (Aumann [1], Puri y Ralescu [84]), mediana (Sinova et al. [91]) o varianza (Körner [60], Lubiano et al. [66]).

Para generalizar estos conceptos, haremos uso de las medidas de comparación y diferencia entre conjuntos difusos estudiadas en el capítulo 1. Nuestras generalizaciones tendrán la importante ventaja de que los conjuntos aleatorios o, en general, las variables aleatorias difusas, no necesitarán tener imágenes convexas en espacios numéricos, tal como ocurre con otros precedentes de la literatura.

A lo largo del resto del documento, por comodidad, utilizaremos la misma nomenclatura para los conjuntos difusos y sus funciones de pertenencia, al contrario de lo que hemos venido haciendo en el capítulo 1. Es decir, denotaremos por $\tilde{A}(x)$ a la función de pertenencia del conjunto difuso $\tilde{A}$ para el elemento x , en lugar de $\mu_{\tilde{A}}(x)$.

Para comenzar, resulta de interés el estudio de los conjuntos aleatorios, acerca de los cuales existe una literatura muy extensa. Un conjunto aleatorio es una aplicación multi-valuada definida sobre un espacio probabilístico que satisface cierta condición de medibilidad, la cual garantiza la posibilidad de definir correctamente la probabilidad inducida por el conjunto aleatorio.

Definición 16 ([78], Nguyen, 1978) Sean $(\Omega, \mathcal{A})$ y $(U, \mathcal{A}')$ dos espacios medibles. Diremos que una aplicación multivaluada, $\Gamma : \Omega \longrightarrow \mathcal{P}(U)$, es fuertemente medible si y solo si:

$$\forall B \in \mathcal{A}' \quad \Gamma^*(B) := \{\omega \in \Omega : \Gamma(\omega) \cap B \neq \emptyset\} \in \mathcal{A}$$

En este trabajo llamaremos *conjunto aleatorio* a toda aplicación multi-valuada fuertemente medible.

En esta memoria combinaremos las dos visiones de los conjuntos anteriormente mencionadas. Por una parte, consideraremos la existencia de un conjunto aleatorio, entendido desde un punto de vista conjuntivo, cuya distribución de probabilidad deseamos caracterizar. Por otra parte, asumiremos que observamos sus imágenes (conjunto-valoradas) de forma incompleta. Este aspecto será estudiado en el siguiente capítulo.

Diversos autores relacionan el estudio de los conjuntos aleatorios con la Teoría de las Capacidades de Choquet ([16]). De forma independiente a la mencionada teoría, A.P.Dempster ([34]) define las probabilidades superior e inferior inducidas por una aplicación multivaluada. La relación de estas probabilidades con las capacidades de Choquet puede encontrarse en algunos trabajos de D. G. Kendall ([57]) y G. Matheron ([68]).

Definición 17 ([34], Dempster, 1967) Sean $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea $(U, \mathcal{A}')$ un espacio medible. Dado el conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$, la probabilidad superior inducida por él es la función $P^* : \wp(U) \longrightarrow [0, 1]$ dada por

$$P^*(A) = \frac{P(\Gamma^*(A))}{P(\Gamma^*(U))}.$$

Recordemos que, para cada conjunto $B \subseteq U$, $\Gamma^*(B)$ es el subconjunto de Ω definido por

$$\Gamma^*(B) = \{\omega \in \Omega : \Gamma(\omega) \cap B \neq \emptyset\}.$$

La probabilidad inferior inducida por Γ es la función $P_* : \wp(U) \longrightarrow [0, 1]$ dada por

$$P_*(A) = \frac{P(\Gamma_*(A))}{P(\Gamma^*(U))}$$

donde, para cada conjunto $B \subseteq U$, $\Gamma_*(B)$ es el subconjunto de Ω definido por

$$\Gamma_*(B) = \{\omega \in \Omega : \Gamma(\omega) \subseteq B, \Gamma(\omega) \neq \emptyset\}.$$

Si, además, las imágenes del conjunto aleatorio Γ son no vacías, $\Gamma^*(U)$ es el conjunto total Ω , y por tanto su probabilidad es 1, simplificándose las expresiones anteriores.

El término variable aleatoria difusa ha sido empleado por diversos autores en diversos contextos (como L.A. Zadeh [104] o H.T. Nguyen [77]). En nuestro caso, una variable aleatoria difusa se define como una aplicación que toma, como valores, subconjuntos difusos de cierto universo, y que cumple determinada condición de medibilidad. Esta forma de definir la variable aleatoria difusa ha sido considerada, entre otros, por R. Féron [41] o M.L. Puri y D.A. Ralescu ([84]). Como ya señalamos al comienzo del capítulo, seguiremos una perspectiva óptica.

Definición 18 Sean $(\Omega, \mathcal{A})$ y $(U, \mathcal{A}')$ dos espacios medibles. Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ una aplicación entre ambos. Diremos que $\tilde{\Gamma}$ es una variable aleatoria difusa si, para cada $\alpha \in [0, 1]$, la aplicación multivaluada $\tilde{\Gamma}_\alpha$ definida como:

$$\tilde{\Gamma}_\alpha(\omega) := \left[\tilde{\Gamma}(\omega) \right]_\alpha \quad \forall \omega \in \Omega$$

es una aplicación fuertemente medible, donde $\tilde{A}_\alpha$ denota el α -corte (débil) del conjunto difuso $\tilde{A}$.

Las aplicaciones multivaluadas $\tilde{\Gamma}_\alpha, \alpha \in [0, 1]$ se denominan α -cortes de la variable aleatoria difusa $\tilde{\Gamma}$, y determinan completamente a la variable, es decir, $\tilde{\Gamma}_1 = \tilde{\Gamma}_2 \Leftrightarrow (\tilde{\Gamma}_1)_\alpha = (\tilde{\Gamma}_2)_\alpha \quad \forall \alpha \in [0, 1]$.

Del mismo modo que los conceptos de conjunto aleatorio y variable aleatoria difusa extienden el concepto de variable aleatoria, diversos autores buscaron dar una extensión del concepto de esperanza. En primer lugar, recordaremos el concepto de esperanza de Fréchet, propuesto en contextos aún más generales.

Definición 19 ([43], M. Fréchet, 1948) Sea X un elemento aleatorio de un espacio métrico. Llamamos esperanzas de Fréchet de X a aquellos puntos que minimizan la distancia cuadrática esperada a X . Ese mínimo, si es finito, se llama varianza. En general, la esperanza de Fréchet no tiene por qué existir o ser única.

Posteriormente, J. Aumann ([1]) definió la esperanza de una aplicación multivaluada de un modo diferente, aunque puede comprobarse fácilmente que esta esperanza constituye un caso

particular de esperanza de Fréchet, considerando la distancia adecuada.

Definición 20 ([1], J. Aumann, 1965) Sea $(\Omega, \mathcal{A}, P)$ espacio probabilístico. Sea E^n un espacio euclideo de dimensión n . Dado un conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(E^n)$ se define su esperanza (de Aumann) como el subconjunto de E^n formado por las esperanzas de todas las selecciones integrables de Γ , es decir:

$$(A) \int_{\Omega} \Gamma dP = \left\{ \int_{\Omega} X dP : X \in S(\Gamma) \right\}$$

donde $S(\Gamma) = \{X : \Gamma \longrightarrow E^n \text{ integrable, } X(\omega) \in \Gamma(\omega) \text{ c.s.}(P)\}$

La extensión del concepto de esperanza de Aumann para una variable aleatoria difusa fue realizado por Puri y Ralescu, quienes la definían de la siguiente forma:

Definición 21 ([84], M.L. Puri, D. Ralescu, 1986) Sea $(\Omega, \mathcal{A}, P)$ espacio probabilístico. Sea U un espacio de Banach. Dada una variable aleatoria difusa $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$, su esperanza es el conjunto difuso cuyos α -cortes son los conjuntos:

$$[E(\tilde{\Gamma})]_{\alpha} = (A) \int_{\Omega} \tilde{\Gamma}_{\alpha} dP, \alpha \in [0, 1]$$

Es decir, la esperanza de la variable aleatoria difusa es el conjunto difuso que tiene por α -cortes la esperanza de Aumann de los α -cortes de la variable. Teniendo en cuenta esto, la función de pertenencia de la esperanza de una variable aleatoria difusa viene dada por:

$$E(\tilde{\Gamma})(x) = \sup\{\alpha \in [0, 1] : x \in E(\tilde{\Gamma}_{\alpha})\}$$

En artículos existentes ya en la literatura (véase [60]) ha sido probado que las esperanzas de Aumann y de Puri y Ralescu coinciden con la esperanza de Fréchet calculada a partir de la distancia entre conjuntos difusos de $\mathbb{R}^n$ dada por

$$d_2(\tilde{A}, \tilde{B}) = \left(n \int_0^1 \int_{S^{n-1}} |s_{\tilde{A}}(\alpha, u) - s_{\tilde{B}}(\alpha, u)|^2 \mu(du) d\alpha \right)^{1/2}$$

donde $s_{\tilde{A}}$ representa la función soporte del conjunto difuso $\tilde{A}$ ¹.

Igualmente, es fácil comprobar que dichas esperanzas coinciden con la de Fréchet calculada a partir de la distancia de Bertoluzza, que recordamos a continuación.

Definición 22 ([7], Bertoluzza et al., 1995) Denotaremos por $\mathcal{F}_c(\mathbb{R})$ el conjunto de los subconjuntos difusos de $\mathbb{R}$ con α -cortes convexos, compactos y no vacíos. Dada una pareja de difusos pertenecientes a dicho conjunto, $\tilde{A}, \tilde{B} \in \mathcal{F}_c(\mathbb{R})$, la distancia de Bertoluzza entre ellos viene dada por:

$$D_{\vec{\lambda}}(\tilde{A}, \tilde{B}) = \sqrt{\int_{(0,1]} [d_{\vec{\lambda}}(\tilde{A}_\alpha, \tilde{B}_\alpha)]^2 d\alpha}$$

donde

$$\begin{aligned} [d_{\vec{\lambda}}(\tilde{A}_\alpha, \tilde{B}_\alpha)]^2 &= \lambda_1(\sup \tilde{A}_\alpha - \sup \tilde{B}_\alpha)^2 + \lambda_2(\text{mid} \tilde{A}_\alpha - \text{mid} \tilde{B}_\alpha)^2 \\ &\quad + \lambda_3(\inf \tilde{A}_\alpha - \inf \tilde{B}_\alpha)^2, \end{aligned}$$

con $\lambda_i \in [0, 1]$, $i = 1, 2, 3$ y $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

¹ $s_{\tilde{A}}(\alpha, u) = \sup\{\langle u, a \rangle : a \in \tilde{A}_\alpha\}$, $u \in S^{n-1}$, $\alpha \in [0, 1]$

Además de los diversos conceptos de esperanza, existen también varias extensiones al caso de las variables aleatorias difusas del concepto de mediana, algunos de ellos basados en la esperanza de Fréchet. Como ejemplo tenemos el trabajo de Sinova et al. ([91]), donde se define la mediana de un número aleatorio difuso como la esperanza de Fréchet del mismo con respecto a la distancia entre números difusos dada por $\rho_1(\tilde{A}, \tilde{B}) = \frac{1}{2} \int_0^1 |\inf \tilde{A}_\alpha - \inf \tilde{B}_\alpha| + |\sup \tilde{A}_\alpha - \sup \tilde{B}_\alpha| d\alpha$.

2.2. El centroide.

El objetivo de esta sección es proporcionar una generalización del concepto de medida de tendencia central para conjuntos aleatorios y variables aleatorias difusas. Esta generalización engloba algunas propuestas anteriormente ([84, 91], y aporta algunas ventajas adicionales en algunas situaciones concretas.

Definición 23 Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}$ una función definida sobre un espacio probabilístico $(\Omega, \mathcal{A}, P)$, con $\mathcal{F}$ pudiendo ser el universo U , sus partes nítidas $\wp(U)$ o bien sus partes difusas $\mathcal{F}(U)$. Sea d una medida de “diferencia” entre pares de elementos de $\mathcal{F}$. Se denomina *centroide* de $\tilde{\Gamma}$ con respecto a d , si existe, al elemento de $\mathcal{F}$ definido como

$$C_d(\tilde{\Gamma}) = \arg \left(\min_{\tilde{A} \in \mathcal{F}} E[d(\tilde{\Gamma}, \tilde{A})] \right)$$

El centroide está bien definido siempre que la aplicación $g(\omega) = d(\tilde{\Gamma}(\omega), \tilde{A})$ sea $\mathcal{A} \otimes \mathcal{A} - \beta_{\mathbb{R}}$ medible $\forall \tilde{A} \in \mathcal{F}$, donde $\beta_{\mathbb{R}}$ es la σ -álgebra de Borel en $\mathbb{R}$. En caso de existir, puede no ser único.

A la hora de buscar el centroide de una variable exigiremos, por convenio, que éste sea de la misma naturaleza que las imágenes de la variable. Es decir, si las imágenes son elementos de U , el centroide también será un elemento de U . Si estamos ante un conjunto aleatorio, de modo que sus imágenes son subconjuntos nítidos del universo, también ha de serlo el centroide, no permitiéndole que sea un conjunto difuso. Más adelante estaremos en condiciones de ilustrar con un ejemplo cierta situación que podría darse si no consideráramos este convenio.

El centroide de una variable aleatoria difusa (y, como caso particular, de un conjunto aleatorio) será, en cierto modo, análogo a la esperanza de Fréchet de esa variable pero con respecto a una medida de diferencia, que no ha de ser necesariamente una distancia. Bastará con que cumpla alguno de los axiomas generales y de diferencia vistos en la secciones 1.2.1 y 1.2.3 del capítulo 1. Cabe destacar que, para obtener el centroide de una variable, no será necesario que el universo de definición de la misma sea numérico, lo cual supondrá una gran ventaja en determinados experimentos.

Tal como hemos mencionado, este nuevo concepto supone una generalización de otros ya existentes. Por ejemplo, escogiendo la medida adecuada, el centroide extiende los conceptos de media, mediana y moda de una variable aleatoria clásica X , como podemos ver a continuación:

- Si X es discreta y consideramos $d_1(x, y) = \begin{cases} 1, & \text{si } x \neq y; \\ 0, & \text{si } x = y, \end{cases}$ el centroide asociado coincide con la moda de la variable X .
- Si la variable toma sus valores en $\mathbb{R}$ y consideramos $d_2(x, y) =$

$|x - y|^2$, el centroide coincide con la esperanza de la variable.

- Si la variable toma sus valores en $\mathbb{R}$ y consideramos $d_3(x, y) = |x - y|$, el centroide coincide con la mediana de la variable.

Esbozcamos brevemente la demostración de estas afirmaciones. El centroide de la variable discreta X con respecto a d_1 es el elemento y que hace mínima la suma $\sum_{i=1}^n d_1(x_i, y)P(X = x_i)$, donde $x_1, \dots, x_n$ son los distintos valores que toma la variable X . Dado que $d_1(x_i, y)$ valdrá 1 en aquellos casos en los que y sea distinto de x_i , y cero en caso contrario, la suma que deseamos minimizar es la suma de las probabilidades de aquellas imágenes de X que son distintas de y . Se hace mínima por tanto cuando y coincide con el valor que X toma con mayor probabilidad (de modo que “eliminamos” de la suma esa probabilidad). Es decir, el centroide de X es su moda.

En cuanto al centroide de una variable X con respecto a la medida d_2 , será por definición $\arg \left(\min_y E(|X - y|^2) \right)$; es conocido que el elemento que verifica tal condición es la esperanza de la variable. De forma similar, el centroide de una variable X con respecto a d_3 es $\arg \left(\min_y E(|X - y|) \right)$, elemento que se corresponde con la mediana de la variable.

Además, es obvio que el centroide generaliza la esperanza de Puri y Ralescu ([84]), así como el concepto de mediana para variables aleatorias difusas estudiado por Sinova et. al ([91]).

2.2.1. Propiedades del centroide

Algunas propiedades del centroide podrán ser estudiadas en función de los axiomas verificados por la medida de diferencia considerada. Otras, sin embargo, estarán relacionadas con la forma concreta de la medida, por lo que deberán ser estudiadas para cada caso concreto. En este trabajo se han analizado algunas propiedades generales (o la ausencia de ellas) y, a modo de ejemplo, las propiedades verificadas por el centroide con respecto a la distancia de Hamming. La elección de esta distancia se debe a que verifica la mayor parte de los axiomas generales y de desigualdad analizados en el capítulo 1 (secciones 1.2.1 y 1.2.3), así como al hecho de que esta distancia puede utilizarse en conjuntos no numéricos.

I.S. Molchanov realiza en uno de sus artículos ([73]) una clasificación de las propiedades de la esperanza de un conjunto aleatorio que considera “razonables”, dividiéndolas en 3 grupos. El primero de ellos es el de las propiedades relacionadas con las relaciones de inclusión, que enumeramos a continuación. Aunque el autor considera en su trabajo únicamente los conjuntos aleatorios, adaptamos estas propiedades a nuestro contexto, considerando una variable aleatoria difusa.

A1 Si $\tilde{\Gamma}$ es determinista, es decir, $\tilde{\Gamma}(\omega) = \tilde{A}$ para todo $\omega \in \Omega$, entonces $E[\tilde{\Gamma}] = \tilde{A}$.

A2 Si $\tilde{A} \subseteq \tilde{\Gamma}$ c.s., donde $\tilde{A} \in \mathcal{F}(U)$, entonces $\tilde{A} \subseteq E[\tilde{\Gamma}]$.

A3 Si $\tilde{\Gamma} \subseteq \tilde{A}$ c.s. donde $\tilde{A} \in \mathcal{F}(U)$, entonces $E[\tilde{\Gamma}] \subseteq \tilde{A}$.

A4 Si $\tilde{\Gamma} \subseteq \tilde{\Gamma}'$ c.s., entonces $E[\tilde{\Gamma}] \subseteq E[\tilde{\Gamma}']$.

A continuación estudiaremos qué propiedades de entre las mencionadas verifica el centroide.

Proposición 12 *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea d una medida de diferencia sobre un conjunto U que toma valores positivos. Entonces:*

1. *Si d verifica el axioma $D3$, el centroide verifica la propiedad $A1$. Es decir, el conjunto imagen de la variable aleatoria difusa es uno de sus centroides, pero puede no ser único. Si d verifica el axioma $D3^*$, se da además la unicidad del centroide.*
2. *Si d verifica el axioma $D5$, al menos uno de los centroides verifica la propiedad $A2$ (en caso de que no haya unicidad del centroide, no todos los centroides verifican la propiedad $A2$ necesariamente).*
3. *Si d verifica el axioma $D4$, al menos uno de los centroides verifica la propiedad $A3$ (en caso de que no haya unicidad del centroide, no todos los centroides verifican la propiedad $A3$ necesariamente).*

Demostración:

1. Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ una variable aleatoria difusa constante, es decir, $\tilde{\Gamma}(\omega) = \tilde{A}$ para todo $\omega \in \Omega$. El centroide de $\tilde{\Gamma}$ es, por definición,

$$C_d(\tilde{\Gamma}) = \arg \left(\min_{\tilde{B} \in \mathcal{F}(U)} E[d(\tilde{\Gamma}, \tilde{B})] \right).$$

Dado que $\tilde{\Gamma}$ es constante, tenemos que

$$C_d(\tilde{\Gamma}) = \arg \left(\min_{\tilde{B} \in \mathcal{F}(U)} E[d(\tilde{A}, \tilde{B})] \right).$$

Si la medida de diferencia d verifica el axioma D3, el mínimo se alcanza en $\tilde{A}$, dado que la diferencia sería 0. Por tanto, $\tilde{A}$ sería un centroide, pero no necesariamente el único, ya que pueden existir otros conjuntos de modo que su diferencia respecto a $\tilde{A}$ sea 0. Sin embargo, si d verifica el axioma $D3^*$, esto no sería posible, siendo $\tilde{A}$ el único centroide de $\tilde{\Gamma}$.

2. Dado que $\tilde{A} \subseteq \tilde{\Gamma}$ c.s., y debido al axioma D5, se verifica que

$$d(\tilde{\Gamma}(\omega), C_d(\tilde{\Gamma}) \cup \tilde{A}) \leq d(\tilde{\Gamma}(\omega), C_d(\tilde{\Gamma})) \text{ c.s.}$$

Por tanto, $E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}) \cup \tilde{A})] \leq E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}))]$. Pero por definición del centroide, $E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}))]$ debe ser menor o igual que $E[d(\tilde{\Gamma}, \tilde{B})]$ para cualquier subconjunto $\tilde{B} \in \mathcal{F}(U)$. En total, ambas esperanzas son iguales, lo que nos indica que $C_d(\tilde{\Gamma}) \cup \tilde{A}$ también es un centroide de la variable $\tilde{\Gamma}$. Por tanto, el conjunto $\tilde{A}$ está contenido en al menos uno de los centroides de $\tilde{\Gamma}$.

3. En este caso $\tilde{\Gamma} \subseteq \tilde{A}$ c.s. unido al axioma D4 nos conduce, de forma análoga a la propiedad anterior, a la desigualdad

$$E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}) \cap \tilde{A})] \leq E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}))] \text{ c.s.}$$

Por el mismo razonamiento, vemos que se verifica A3 para al menos uno de los centroides. $\square$

Veamos ahora que, en caso de que el centroide no sea único, no todos los centroides de un conjunto aleatorio verifican las propiedad A2 y A3, mediante los siguientes contraejemplos.

Ejemplo 15 Consideremos un espacio probabilístico arbitrario $(\Omega, \mathcal{A}, P)$, y un universo finito $U = \{x_1, x_2, x_3, x_4\}$. Consideremos la medida de diferencia d entre pares de subconjuntos nítidos de U dada por $d(B, C) = 0$ si $B \cap C \neq \emptyset$ y $d(B, C) = 1$ en otro caso. Esta medida verifica el axioma D5. Sea $\Gamma : \Omega \longrightarrow \wp(U)$ un conjunto aleatorio constante con imagen $\Gamma(\omega) = \{x_1, x_2, x_3\} \forall \omega \in \Omega$. Sea $\tilde{A} = \{x_1, x_2\}$. Entonces $\tilde{A} \subseteq \Gamma(\omega) \forall \omega \in \Omega$. No existe unicidad del centroide de Γ : todo aquel conjunto que minimice su diferencia con el conjunto $\{x_1, x_2, x_3\}$ es centroide, y esto sucede con todos aquellos subconjuntos que tienen intersección no vacía con él (ya que la diferencia es entonces 0). Así pues, el conjunto $\{x_3\}$ es un centroide de Γ y no contiene al conjunto $\tilde{A}$.

Ejemplo 16 Consideremos el mismo espacio probabilístico y el mismo conjunto U del ejemplo anterior. Consideremos la medida de diferencia d entre pares de subconjuntos nítidos de U dada por $d(\tilde{B}, \tilde{C}) = 1$ si $\tilde{B}^c = \tilde{C}$ y $d(\tilde{B}, \tilde{C}) = 0$ en otro caso. Esta medida verifica el axioma D4. Consideremos el conjunto aleatorio constante $\Gamma : \Omega \longrightarrow \wp(U)$ con imagen $\Gamma(\omega) = \{x_1, x_2\} \forall \omega \in \Omega$. Sea $\tilde{A} = \{x_1, x_2, x_3\}$. Entonces $\tilde{A} \supseteq \Gamma(\omega) \forall \omega \in \Omega$. No existe unicidad del centroide de Γ : todo aquel conjunto que minimice su diferencia con el conjunto $\{x_1, x_2\}$ es centroide, y esto sucede con todos los subconjuntos de U a excepción de su complementario (el conjunto $\{x_3, x_4\}$). Por tanto, el conjunto $\{x_4\}$ es un centroide de Γ y sin embargo no está contenido en $\tilde{A}$.

En cuanto a la relación de contenido existente entre dos variables, no ha de ser necesariamente “heredada” por sus centroides. Es decir, el centroide no verifica, en general, la propiedad A4 considerada por Molchanov ([73]). Ilustremos este hecho mediante el siguiente contraejemplo.

Ejemplo 17 *Consideremos los conjuntos finitos $\Omega = \{\omega_1, \omega_2\}$ y $U = \{x_1, x_2, x_3\}$, la σ -álgebra $\mathcal{A} = \wp(\Omega)$ y una probabilidad P arbitraria sobre ella. Sea d la medida de diferencia para subconjuntos nítidos de U dada por: $d(A, B) = 1$ si $A \cap B = \emptyset$ y $d(A, B) = 0$ en otro caso. Esta medida verifica todos los axiomas generales y de diferencia estudiados en las secciones 1.2.1 y 1.2.3, a excepción de los axiomas $D3^*$, $D4$ y $D7$. Por último, consideremos dos conjuntos aleatorios $\Gamma_1, \Gamma_2 : \Omega \rightarrow \wp(U)$ dados por: $\Gamma_1(\omega_1) = \{x_1\}$, $\Gamma_1(\omega_2) = \{x_2\}$ y $\Gamma_2(\omega_1) = \{x_1, x_3\}$, $\Gamma_2(\omega_2) = \{x_2, x_3\}$. Podrán ser considerados centroides de un conjunto aleatorio con respecto a d todos aquellos subconjuntos de U con intersección no vacía con todas y cada una de las imágenes de dicho conjunto aleatorio. Elijamos entre ellos, por convenio, el de menor cardinalidad. Así, $C_d(\Gamma_1) = \{x_1, x_2\}$ y $C_d(\Gamma_2) = \{x_3\}$. Por tanto, $\Gamma_1 \subseteq \Gamma_2$ y sin embargo la intersección de sus centroides es el conjunto vacío.*

Veamos un ejemplo más, en este caso considerando una métrica como medida de desigualdad para el cálculo de los centroides. De nuevo, este ejemplo deja patente el hecho de que la relación de contenido entre dos conjuntos aleatorios no es heredada por sus centroides, ya que incluso pueden tener intersección vacía.

Ejemplo 18 Consideremos la distancia euclídea entre intervalos cerrados no vacíos de $[0, 1]$,

$$d_e([a_1, b_1], [a_2, b_2]) = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2}.$$

A partir de esta distancia, definimos una medida d para pares de intervalos cerrados de $[0, 1]$ y el vacío. Esta medida coincidirá con la distancia euclídea en los pares de intervalos cerrados no vacíos. Por otro lado, la distancia de un intervalo respecto al vacío será $d(\emptyset, [a, b]) = d([a, b], \emptyset) = \sqrt{2}$, $\forall a, b \in [0, 1]$, $a < b$ y $d(\emptyset, \emptyset) = 0$. Puede comprobarse fácilmente que d también es una métrica.

Consideremos el conjunto $\Omega = \{\omega_1, \omega_2\}$, la sigma-álgebra $\wp(\Omega)$ y la probabilidad sobre ella dada por $P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}$. Consideremos dos conjuntos aleatorios definidos sobre Ω , Γ_1 y Γ_2 , con espacio de llegada el de los intervalos cerrados de $[0, 1]$ y el conjunto vacío. El primero de ellos vendrá dado por $\Gamma_1(\omega_1) = [0, 1, 0, 2]$ y $\Gamma_1(\omega_2) = \emptyset$. El segundo estará definido como $\Gamma_2(\omega_1) = [0, 0, 5]$ y $\Gamma_2(\omega_2) = [0, 7, 0, 8]$. De este modo, $\Gamma_1(\omega) \subseteq \Gamma_2(\omega)$ para todo $\omega \in \Omega$.

Sin embargo, esta relación de contenido no es verificada por sus centroides. Para Γ_1 , el centroide es el conjunto γ que minimiza $d([0, 1, 0, 2], \gamma) + d(\emptyset, \gamma)$ (notar que ambos elementos de Ω tienen la misma probabilidad). Dado que la segunda distancia es 1 para cualquier conjunto γ , debemos minimizar $d([0, 1, 0, 2], \gamma)$. Es obvio entonces que el centroide de Γ_1 con respecto a d es $C_d(\Gamma_1) = [0, 1, 0, 2]$. En cuanto a Γ_2 , deberemos minimizar la esperanza $E[d(\Gamma_2, \gamma)]$. Dado que la distancia de las imágenes de

Γ_2 al vacío es la mayor posible, $\sqrt{2}$, sabemos que el centroide no puede ser el vacío. Por tanto, calculamos el mínimo entre los intervalos no vacíos de $[0, 1]$, de modo que las distancias de las imágenes del conjunto aleatorio al conjunto γ calculadas en la esperanza, coinciden con las distancias euclídeas respectivas. Es decir, el centroide de Γ_2 con respecto a d coincide con el centroide con respecto a la distancia euclídea, d_e . Y este último coincide con la esperanza de Aumann de Γ_2 , siendo por tanto $C_d(\Gamma_2) = [0,35, 0,65]$.

En total, tenemos que $\Gamma_1(\omega) \subseteq \Gamma_2(\omega)$ para todo $\omega \in \Omega$, pero sus centroides ($C_d(\Gamma_1) = [0,1, 0,2]$ y $C_d(\Gamma_2) = [0,35, 0,65]$) tienen intersección vacía.

Otra de las propiedades consideradas por Molchanov ([73]) es la de la homogeneidad. En el caso de una variable X que toma valores en la recta real, su esperanza verifica que $E[\lambda \cdot X] = \lambda \cdot E[X]$, con $\lambda \in \mathbb{R}$. Molchanov considera que es razonable que la esperanza de un conjunto aleatorio verifique esta propiedad. Sin embargo, este hecho no es necesariamente verificado por el centroide de una variable aleatoria difusa; únicamente podemos afirmar que será así en el caso de que la medida de diferencia considerada verifique $d(\lambda \cdot \tilde{A}, \lambda \cdot \tilde{B}) = \lambda \cdot d(\tilde{A}, \tilde{B})$. Como contraejemplo, podemos tomar la medida de diferencia d considerada en el ejemplo 8 del capítulo 1. Para dicha medida, $E[\lambda \cdot X] = E[X]$. Por otro lado, el producto por un escalar puede no tener sentido en algunos espacios, mientras que precisamente una de las grandes ventajas del centroide es el hecho de que puede ser utilizado en universos de cualquier carácter, incluso en aquellos en los que no está definido el producto de sus elementos por escalares.

Como hemos mencionado con anterioridad, puede ser complicado analizar en general las propiedades del centroide respecto a una medida de diferencia, ya que muchas de ellas se basan en la forma concreta de dicha medida más que en los axiomas que verifica. Además, algunas de las propiedades habituales de la esperanza de las variables aleatorias pueden no satisfacerse o no tener sentido según el universo de interés y la medida de diferencia empleados. Para medidas concretas sí es posible realizar el estudio. A modo de ejemplo, analizamos a continuación algunas propiedades interesantes que son verificadas por el centroide de una variable aleatoria difusa con respecto a la distancia de Hamming.

Caso concreto: la distancia de Hamming

Para comenzar, es posible proporcionar una fórmula general para la obtención del centroide de una variable aleatoria difusa con respecto a la distancia de Hamming, como queda reflejado en el siguiente resultado. En él se ha tenido en cuenta la exigencia de que el centroide tenga la misma naturaleza que las imágenes de la variable, a la cual hicimos mención en su definición.

Proposición 13 *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea U el universo en estudio. Sea d_{Ham} la distancia de Hamming entre subconjuntos difusos de U .*

1. *Si $X : \Omega \longrightarrow U$ es una variable aleatoria discreta, el conjunto de sus centroides con respecto a la distancia de Ham-*

ming coincide con el conjunto de modas de la variable. El centroide de X es único si y solo si X es unimodal.

2. Si $\Gamma : \Omega \longrightarrow \wp(U)$ es un conjunto aleatorio con un número finito de imágenes, su centroide es el conjunto

$$C_{d_{Ham}}(\Gamma) = \{x \in U : P_{\Gamma}^*(\{x\}) > 0,5\} \in \wp(U)$$

donde P_{Γ}^* es la probabilidad superior de Dempster del conjunto aleatorio.

3. Si $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ es una variable aleatoria difusa que toma un número finito de imágenes, su centroide es el conjunto difuso $C_{d_{Ham}}(\tilde{\Gamma}) \in \mathcal{F}(U)$ cuya función de pertenencia viene dada por

$$C_{d_{Ham}}(\tilde{\Gamma})(x) = Me(\tilde{\gamma}_1(x), \tilde{\gamma}_2(x), \dots, \tilde{\gamma}_n(x))$$

donde Me representa la mediana y los subconjuntos difusos de U $\tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_n \in \mathcal{F}(U)$ son las n imágenes de la variable aleatoria difusa.

Demostración:

1. Partimos de una variable aleatoria $X : \Omega \longrightarrow U$ tomando un número finito de valores $x_1, x_2, \dots, x_n$. Representaremos por $p_1, p_2, \dots, p_n$ las respectivas probabilidades de cada uno de dichos valores. Debemos tener en cuenta que la distancia de Hamming entre elementos del conjunto U viene dada por: $d_{Ham}(x, y) = \begin{cases} 0, & \text{si } x = y; \\ 2, & \text{si } x \neq y. \end{cases}$

Por definición, el centroide será el siguiente elemento de U :

$$C_{d_{Ham}}(X) = \arg \left(\min_{x \in U} E[d_{Ham}(X, x)] \right).$$

Por otro lado, $E[d_{Ham}(X, x)] = \sum_{i=1}^n d_{Ham}(x_i, x) \cdot p_i = 2 \sum_{i, x_i \neq x} p_i = 2(1 - P(X = x))$. Por tanto, se alcanzará el mínimo cuando $P(X = x)$ sea máximo, es decir, cuando x sea la moda de la variable.

2. Consideremos ahora un conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$ que toma los valores $\gamma_1, \gamma_2, \dots, \gamma_n \in \wp(U)$ con probabilidad $p_1, p_2, \dots, p_n$ respectivamente. Debemos recordar que $d_{Ham}(A, B) = \sharp A \triangle B \ \forall A, B \in \wp(U)$.

El centroide será el conjunto

$$C_{d_{Ham}} = \arg \left(\min_{\gamma \in \wp(U)} E[d_{Ham}(\Gamma, \gamma)] \right) =$$

$$\arg \left(\min_{\gamma \in \wp(U)} \sum_{i=1}^n d_{Ham}(\gamma_i, \gamma) \cdot p_i \right)$$

Podemos expresar la distancia de Hamming del siguiente modo: $d_{Ham}(\gamma_i, \gamma) = \sum_{j=1}^k d_j(\gamma_i, \gamma)$, donde

$$d_j(\gamma_i, \gamma) = \begin{cases} 1, & \text{si } x_j \in \gamma_i \triangle \gamma; \\ 0, & \text{en otro caso.} \end{cases}$$

De este modo,

$$E[d_{Ham}(\Gamma, \gamma)] = \sum_{i=1}^n \left(\sum_{j=1}^k d_j(\gamma_i, \gamma) \right) \cdot p_i =$$

$$\sum_{j=1}^k \left(\sum_{i=1}^n d_j(\gamma_i, \gamma) \cdot p_i \right).$$

Buscaremos, para cada $j \in 1, 2, \dots, k$, el valor de $\gamma(x_j)$ que minimiza la suma en el índice i :

$$\sum_{i=1}^n d_j(\gamma_i, \gamma) \cdot p_i = \begin{cases} \sum_{i: x_j \in \gamma_i} p_i, & \text{si } x_j \notin \gamma \\ \sum_{i: x_j \notin \gamma_i} p_i, & \text{si } x_j \in \gamma \end{cases} =$$

$$\begin{cases} P_{\Gamma}^*(\{x_j\}), & \text{si } x_j \notin \gamma \\ 1 - P_{\Gamma}^*(\{x_j\}), & \text{si } x_j \in \gamma \end{cases}$$

Donde P_{Γ}^* denota la probabilidad superior de Dempster para Γ , es decir, $P_{\Gamma}^*(A) = P(\{\omega \in \Omega : \Gamma(\omega) \cap A \neq \emptyset\})$

Para cada $j \in 1, 2, \dots, k$, si $P_{\Gamma}^*(\{x_j\}) \leq 0,5$ es obvio que $P_{\Gamma}^*(\{x_j\}) \leq 1 - P_{\Gamma}^*(\{x_j\})$. Pero si $P_{\Gamma}^*(\{x_j\}) > 0,5$ el menor de los dos es $1 - P_{\Gamma}^*(\{x_j\})$.

De todo esto se deduce que

$$C_{d_{Ham}}(\Gamma) = \{x \in U : P_{\Gamma}^*(x) > 0,5\}.$$

3. Consideremos ahora una variable aleatoria difusa $\tilde{\Gamma}$ que toma los valores $\tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_n \in \mathcal{F}(U)$ con probabilidad $p_1, p_2, \dots, p_n$ respectivamente. Debemos recordar que la distancia de Hamming entre conjuntos difusos es $d_{Ham}(\tilde{A}, \tilde{B}) = \sum_{j=1}^k |\tilde{A}(x_j) - \tilde{B}(x_j)| \forall \tilde{A}, \tilde{B} \in \mathcal{F}(U)$.

El centroide de $\tilde{\Gamma}$ será por tanto:

$$\begin{aligned}
 C_{d_{Ham}}(\tilde{\Gamma}) &= \arg \left(\min_{\tilde{\gamma} \in \mathcal{F}(U)} E[d_{Ham}(\tilde{\Gamma}, \tilde{\gamma})] \right) \\
 &= \arg \left(\min_{\tilde{\gamma} \in \mathcal{F}(U)} \sum_{i=1}^n d_{Ham}(\tilde{\gamma}_i, \tilde{\gamma}) \cdot p_i \right) \\
 &= \arg \left(\min_{\tilde{\gamma} \in \mathcal{F}(U)} \sum_{i=1}^n \left(\sum_{j=1}^k |\tilde{\gamma}_i(x_j) - \tilde{\gamma}(x_j)| \right) \cdot p_i \right) = \\
 &\arg \left(\min_{\tilde{\gamma} \in \mathcal{F}(U)} \sum_{j=1}^k \left(\sum_{i=1}^n |\tilde{\gamma}_i(x_j) - \tilde{\gamma}(x_j)| \cdot p_i \right) \right).
 \end{aligned}$$

Para cada $j \in 1, 2, \dots, k$, el valor de $\tilde{\gamma}(x_j)$ que hace mínima la suma $\sum_{i=1}^n |\tilde{\gamma}_i(x_j) - \tilde{\gamma}(x_j)| \cdot p_i$ es la mediana de los valores $\tilde{\gamma}_i(x_j)$, $i = 1, 2, \dots, n$. Por tanto el centroide de $\tilde{\Gamma}$ con respecto a la distancia de Hamming es el conjunto difuso $C_{d_{Ham}}(\tilde{\Gamma})$ cuya función de pertenencia viene dada, para $j = 1, 2, \dots, k$, por

$$C_{d_{Ham}}(\tilde{\Gamma})(x_j) = Me(\tilde{\gamma}_1(x_j), \tilde{\gamma}_2(x_j), \dots, \tilde{\gamma}_n(x_j)). \quad \square$$

Ahora que conocemos la expresión del centroide para la distancia de Hamming, la utilizaremos para ver un ejemplo de la situación que puede darse en el caso de que permitiésemos al centroide de una variable aleatoria ser un conjunto en lugar de un elemento. Para ello, consideraremos dos medidas de diferencia entre subconjuntos de un universo U . La primera de ellas,

como hemos indicado, será la distancia de Hamming. La segunda, d , estará definida como $d(A, A) = 0$ y $d(A, B) = 1$ si $A \neq B$, $\forall A, B \subseteq U$. Si nos restringimos a conjuntos unipuntuales, es decir, vemos d y d_{Ham} como distancias entre elementos de U , tenemos que $d_{Ham}(x, y) = 2d(x, y)$. Consideremos una variable aleatoria $X : \Omega \rightarrow U$. Dado que las imágenes de X son elementos de U , y que la distancia de Hamming entre ellas será 2 veces la diferencia si consideramos d , parece deseable que los centroides de la variable con respecto a ambas medidas sean iguales. Si imponemos que ambos centroides sean elementos de U , coinciden y son la moda de la variable (en el caso de que la variable no fuese unimodal, la variable tendría varios centroides, pero los mismos para las dos distancias). Sin embargo, si permitimos que estos centroides sean conjuntos, obtenemos lo siguiente:

- $C_{d_{Ham}}(X) = \arg \left(\min_{A \subseteq U} E[d_{Ham}(X, A)] \right) = \{x \in U : P(X = x) > 0,5\}$, es decir, el conjunto de las imágenes de X con probabilidad superior a 0,5, o bien el vacío si no existiese ninguna.
- $C_d(X) = \{x \in U : P(X = x) = \max_{y \in U} P(X = y)\}$, es decir, el conjunto formado por todas las modas de la variable X , o bien el vacío si no existiese ninguna.

En este segundo caso, es obvio que puede no haber ningún tipo de relación entre los centroides. Si, por ejemplo, la moda de la variable fuese única, y apareciese con probabilidad 0,4, el centroide respecto a d sería el conjunto formado por ese elemento, mientras que el centroide con respecto a la distancia de Hamming sería el conjunto vacío.

Pasemos a analizar algunas propiedades interesantes verificadas por el centroide con respecto a la distancia de Hamming en la siguiente proposición. En cuanto a la última propiedad, debemos recordar que tendrá sentido únicamente en los casos en los que U sea un conjunto cerrado para una multiplicación escalar; denotaremos por $\mathbb{K}$ el conjunto de dichos escalares. En cuanto al producto de un escalar por un conjunto difuso podemos considerar, en primer lugar, la definición clásica, basada en el Principio de Extensión de Zadeh: dados $\lambda \in \mathbb{K}$ y $\tilde{A} \in \mathcal{F}(U)$, el conjunto difuso $\lambda\tilde{A}$ viene dado por la función de pertenencia

$$(\lambda\tilde{A})(x) = \begin{cases} \tilde{A}(\frac{1}{\lambda}x) & \text{si } \lambda \neq 0; \\ 0 & \text{si } \lambda = 0. \end{cases}$$

Por otro lado, puede resultar interesante otra interpretación del producto, en la que consideraremos $\mathbb{K} \subseteq [0, \infty)$. Veámoslo mediante un ejemplo. Consideremos como universo el siguiente conjunto de idiomas $U = \{\text{Chino (c), Español (e), Holandés (h), Francés (f), Italiano (i), Inglés (g)}\}$. Sea $\tilde{A}$ un subconjunto difuso de U de modo que $\tilde{A}(x)$ representa el grado de dominio del idioma $x \in U$ que tiene cierta persona de una población. Esta persona está interesada en mejorar su conocimiento de los idiomas de los que tiene ya ciertas nociones (es decir, aquellos para los que la función de pertenencia es mayor que 0). Realiza una serie de cursos, hasta llegar a dominar cada idioma el doble de lo que lo hacía al principio. El conjunto que representa su grado de conocimiento de los idiomas de U al finalizar su formación es $2\tilde{A}$, donde $(2\tilde{A})(x) = \min\{1, 2 \cdot \tilde{A}(x)\}$.

En la siguiente proposición se han tenido en cuenta las dos interpretaciones posibles del producto de un difuso por un escalar, verificándose la propiedad en ambos casos.

Proposición 14 *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea U el universo en estudio. Sean $\tilde{\Gamma}, \tilde{\Gamma}' : \Omega \longrightarrow \mathcal{F}(U)$ dos variables aleatorias difusas. Sea d_{Ham} la distancia de Hamming entre subconjuntos difusos de U . Se verifican las siguientes propiedades:*

- a) *Si $\tilde{\Gamma}(\omega) = \tilde{A} \ \forall \omega \in \Omega$, entonces $C_{d_{Ham}}(\tilde{\Gamma}) = \tilde{A}$.*
- b) *Si $\tilde{\Gamma} \subseteq \tilde{\Gamma}'$ entonces $C_{d_{Ham}}(\tilde{\Gamma}) \subseteq C_{d_{Ham}}(\tilde{\Gamma}')$.*
- c) *$C_{d_{Ham}}(\lambda \tilde{\Gamma}) = \lambda \cdot C_{d_{Ham}}(\tilde{\Gamma})$ para todo $\lambda \in \mathbb{K}$.*

Demostración: Las dos primeras propiedades son obvias por la propia expresión del centroide.

Probemos ahora la veracidad de la propiedad c). Dado que el producto de un escalar por un conjunto difuso admite (al menos) dos definiciones, consideraremos por separado ambos casos:

1. Consideremos un conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$. Tenemos que

$$\lambda \cdot \Gamma(\omega) = \{\lambda \cdot x : x \in \Gamma(\omega)\}$$

y

$$P_{\lambda, \Gamma}^*(\{x\}) = P\left(\{\omega \in \Omega : x \in (\lambda \cdot \Gamma)(\omega) = \lambda \cdot (\Gamma(\omega))\}\right) =$$

$$P\left(\{\omega \in \Omega : \frac{1}{\lambda}x \in \Gamma(\omega)\}\right) = P_{\Gamma}^*\left(\left\{\frac{1}{\lambda}x\right\}\right).$$

Teniendo esto en cuenta, así como la definición del centroide para un conjunto aleatorio, podemos ver que

$$\begin{aligned} x \in C_{d_{Ham}}(\lambda \cdot \Gamma) &\Leftrightarrow P_{\Gamma}^*(\{\frac{1}{\lambda}x\}) > 0,5 \Leftrightarrow \frac{1}{\lambda}x \in C_{d_{Ham}}(\Gamma) \\ &\Leftrightarrow x \in \lambda \cdot C_{d_{Ham}}(\Gamma). \end{aligned}$$

2. Consideremos una variable aleatoria difusa $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$. Para cada elemento $\omega \in \Omega$, $[\lambda \cdot \tilde{\Gamma}](\omega) = \lambda \cdot [\tilde{\Gamma}(\omega)]$. Podemos interpretar ese producto de dos formas, como ha sido indicado anteriormente:

- Si $(\lambda \cdot \tilde{A})(x) = \tilde{A}(\frac{1}{\lambda} \cdot x)$, entonces:

$$\begin{aligned} &[C_{d_{Ham}}(\lambda \cdot \tilde{\Gamma})](x) = \\ &Me\left((\lambda \cdot \tilde{\Gamma})(\omega_1)(x), (\lambda \cdot \tilde{\Gamma})(\omega_2)(x), \dots, (\lambda \cdot \tilde{\Gamma})(\omega_n)(x)\right) = \\ &Me\left(\tilde{\Gamma}(\omega_1)(\frac{1}{\lambda} \cdot x), \tilde{\Gamma}(\omega_2)(\frac{1}{\lambda} \cdot x), \dots, \tilde{\Gamma}(\omega_n)(\frac{1}{\lambda} \cdot x)\right) = \\ &C_{d_{Ham}}(\tilde{\Gamma})(\frac{1}{\lambda} \cdot x) = [\lambda \cdot C_{d_{Ham}}(\tilde{\Gamma})](x) \quad \forall x \in U \end{aligned}$$

- Si $(\lambda \cdot \tilde{A})(x) = \min\{1, \lambda \cdot (\tilde{A}(x))\}$. Entonces:

$$\begin{aligned} &[C_{d_{Ham}}(\lambda \cdot \tilde{\Gamma})](x) = \\ &Me\left((\lambda \cdot \tilde{\Gamma})(\omega_1)(x), (\lambda \cdot \tilde{\Gamma})(\omega_2)(x), \dots, (\lambda \cdot \tilde{\Gamma})(\omega_n)(x)\right) = \\ &Me\left(\lambda \cdot (\tilde{\Gamma}(\omega_1)(x)), \lambda \cdot (\tilde{\Gamma}(\omega_2)(x)), \dots, \lambda \cdot (\tilde{\Gamma}(\omega_n)(x))\right) = \\ &\lambda \cdot Me\left(\tilde{\Gamma}(\omega_1)(x), \tilde{\Gamma}(\omega_2)(x), \dots, \tilde{\Gamma}(\omega_n)(x)\right) = \\ &\lambda \cdot \left(C_{d_{Ham}}(\tilde{\Gamma})(x)\right) = \left(\lambda \cdot C_{d_{Ham}}(\tilde{\Gamma})\right)(x) \quad \square \end{aligned}$$

2.3. La d -variación o d -dispersión de una variable aleatoria difusa

Una vez generalizado el concepto de medida de tendencia central, es natural buscar una generalización de la idea de medida de dispersión al caso de las variables aleatorias difusas. La varianza de una variable con imágenes reales X puede obtenerse como la esperanza de la variable $(X - E(X))^2$, siendo necesario para este cálculo la existencia de la esperanza de X . Por otro lado, la varianza de X también puede ser obtenida como la mitad de la esperanza de la variable $(X_1 - X_2)^2$, donde X_1 y X_2 son dos copias idénticamente distribuidas de X . De este modo no es necesaria la existencia y el cálculo de la esperanza de la variable X . Sustituyendo la diferencia entre variables por una medida de desigualdad arbitraria, obtendremos dos generalizaciones de la varianza, correspondientes a las dos formulaciones anteriores. Además, no será necesario para su obtención que el universo de discurso sea numérico.

Antes de introducir la primera de las mencionadas generalizaciones, comenzaremos por un concepto más general aún, basado en medidas de comparación genéricas.

Definición 24 Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea $\tilde{\Gamma} : \Omega \rightarrow \mathcal{F}(U)$ una variable aleatoria difusa definida sobre él. Sean $\tilde{\Gamma}_1$ y $\tilde{\Gamma}_2$ dos copias (idénticamente distribuidas) de $\tilde{\Gamma}$. Sea m una medida de comparación entre subconjuntos de $\mathcal{F}(U)$. La m -comparación esperada de $\tilde{\Gamma}$, denotada por $V_m(\tilde{\Gamma})$, se define como:

$$V_m(\tilde{\Gamma}) = E[m(\tilde{\Gamma}_1, \tilde{\Gamma}_2)]$$

donde la esperanza anterior se calcula con respecto a la distribución de probabilidad producto $P \otimes P$ definida en $\mathcal{A} \otimes \mathcal{A}$. Esta cantidad está bien definida siempre que la aplicación $g(\omega, \omega') = m(\tilde{\Gamma}_1(\omega), \tilde{\Gamma}_2(\omega'))$, $\forall (\omega, \omega') \in \Omega \times \Omega$ sea $\mathcal{A} \otimes \mathcal{A} - \beta_{\mathbb{R}}$ medible.

La definición anterior generaliza varios conceptos conocidos, como ilustramos a través de los siguientes ejemplos.

Ejemplo 19 Consideremos una población finita Ω y el conjunto de idiomas $U = \{\text{Chino } (c), \text{Español } (e), \text{Holandés } (h), \text{Francés } (f), \text{Italiano } (i), \text{Inglés } (g)\}$. Consideremos la aplicación multivaluada $\Gamma : \Omega \longrightarrow \wp(U)$ que asigna a cada persona $\omega \in \Omega$ el subconjunto de los idiomas de U que puede hablar con fluidez. Supongamos que todos los individuos de la población pueden hablar al menos uno de los idiomas de U , de modo que las todas las imágenes de Γ son no vacías. Utilizando diversas M -medidas de comparación (véase la definición 8 en la sección 1.3.3) obtenemos información interesante sobre Γ :

- Fijemos un subconjunto arbitrario $D \subseteq U$. Consideremos la medida de comparación s_1 tal que G_{s_1} se define como $G_{s_1}(A, B, C) = \max\{M(A \cap B), M(A \cap B^c)\}$, donde

$$M(E) = \begin{cases} 1 & \text{si } E \cap D \neq \emptyset, \\ 0 & \text{en otro caso.} \end{cases}$$

La medida de comparación esperada de Γ , $V_{s_1}(\Gamma)$, coincide con la probabilidad superior de Dempster de D y representa la proporción de personas de la población que hablan alguno de los idiomas incluidos en el conjunto D . Si, por ejemplo,

D es igual a $\{f\}$, entonces $V_{s_1}(\Gamma)$ representa la proporción de personas que pueden hablar, al menos, francés.

- Fijemos de nuevo un subconjunto arbitrario $D \subseteq U$. Consideremos la medida de comparación s_2 tal que G_{s_2} se define como $G_{s_2}(A, B, C) = \min\{M(A \cap B), M(A \cap B^c)\}$, donde

$$M(E) = \begin{cases} 1 & \text{si } E \subseteq D, \\ 0 & \text{en otro caso.} \end{cases}$$

La medida de comparación esperada de Γ , $V_{s_2}(\Gamma)$, coincide con la probabilidad inferior de Dempster de D y representa la proporción de personas de la población que no hablan ningún idioma que no esté incluido en D .

- Consideremos ahora la medida de comparación $s_3(A, B) = \#(A \cap B)$. La medida de comparación esperada $V_{s_3}(\Gamma)$ cuantifica (en media) la capacidad de comunicación entre pares de personas de la población.
- Consideremos ahora la distancia de Hamming, que es una medida de comparación, $d_{Ham}(A, B) = \#(A \triangle B)$. La medida de comparación esperada de Γ , $V_{d_{Ham}}(\Gamma)$, representa un grado de divergencia sobre las habilidades comunicativas de la gente de la población en estudio.

Ejemplo 20 El ejemplo anterior puede ser modificado si disponemos de una información más refinada acerca de las habilidades comunicativas de la población. Podríamos considerar una

variable aleatoria difusa $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ para representar esas habilidades. El valor de la función de pertenencia $\tilde{\Gamma}(\omega)(x)$ representará el grado de preferencia ([39]) en una escala del 0 al 1 para el idioma $x \in U$. Entonces $\tilde{\Gamma}(\omega)(x) > \tilde{\Gamma}(\omega)(x')$ indicará que la persona ω prefiere hablar el idioma x frente a x' , porque está más familiarizada con él. Dichos grados de preferencia pueden determinarse como una función de los niveles MCERL, por ejemplo.

Para cierta medida de comparación específica, en la que no entraremos aquí dado que la medida concreta no tiene interés, la comparación esperada de $\tilde{\Gamma}$ representa un grado esperado de diferencia respecto a las habilidades comunicativas entre un par de personas de la población.

Ejemplo 21 Consideremos un conjunto de días, Ω , y la aplicación multivaluada $\Gamma : \Omega \longrightarrow \wp(\mathbb{R})$, donde $\Gamma(\omega) = [L(\omega), U(\omega)]$ representa el intervalo de mínima y máxima temperatura alcanzadas en Gijón en la fecha $\omega \in \Omega$. Diversas medidas de comparación esperadas (en las que no entraremos en detalle aquí) proporcionan diferentes medidas informativas de interés: la varianza de las temperaturas mínimas, la varianza de las temperaturas máximas, una mezcla (una combinación lineal) de ambas varianzas, la varianza de las amplitudes de los intervalos, la proporción de días en los que la temperatura mínima sobrepasa cierto límite, la varianza de los puntos medios de los intervalos, etc.

Cabe destacar el caso concreto en el que la medida de comparación es una media de diferencia. Éste extiende algunas no-

ciones clave existentes en la literatura, como veremos a continuación.

Definición 25 Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea $\tilde{\Gamma} : \Omega \rightarrow \mathcal{F}(U)$ una variable aleatoria difusa definida sobre él. Sean $\tilde{\Gamma}_1$ y $\tilde{\Gamma}_2$ dos copias (idénticamente distribuidas) de $\tilde{\Gamma}$. Sea d una medida de “diferencia” entre subconjuntos de $\mathcal{F}(U)$. La d -variación o d -dispersión de $\tilde{\Gamma}$, denotada por $V_d(\tilde{\Gamma})$, se define como:

$$V_d(\tilde{\Gamma}) = E[d(\tilde{\Gamma}_1, \tilde{\Gamma}_2)]$$

donde la esperanza anterior se calcula con respecto a la distribución de probabilidad producto $P \otimes P$ definida en $\mathcal{A} \otimes \mathcal{A}$. Esta cantidad está bien definida siempre que la aplicación $g(\omega, \omega') = d(\tilde{\Gamma}_1(\omega), \tilde{\Gamma}_2(\omega'))$, $\forall (\omega, \omega') \in \Omega \times \Omega$ sea $\mathcal{A} \otimes \mathcal{A} - \beta_{\mathbb{R}}$ medible.

Por un lado, la d -variación extiende el concepto de entropía de una variable aleatoria. Sea $\tilde{\Gamma}$ una variable aleatoria difusa cuyas imágenes son conjuntos nítidos unipuntuales y que, por tanto, se puede identificar con una variable aleatoria clásica X . Supongamos que el universo de llegada es finito, $U = \{x_1, \dots, x_n\}$. Si consideramos como medida de diferencia la distancia de Hamming, $d_{Ham}(A, B) = \#A \triangle B$, entonces la d -variación de $\tilde{\Gamma}$ es 2 veces la entropía de X :

$$E_{d_{Ham}}(\tilde{\Gamma}) = \sum_{i=1}^n \sum_{j=1}^n d_H(\{x_i\}, \{x_j\}) p_i \cdot p_j = 2(1 - \sum_{i=1}^n p_i^2), \quad (2.1)$$

donde p_i denota la probabilidad $P(X = x_i)$, $i = 1, \dots, n$.

Por otro lado, la d -variación extiende algunos conceptos existentes de varianza escalar de una variable aleatoria difusa ([19]). Podemos encontrar en la literatura reciente diversas métricas y pseudo-métricas definidas en la familia de los conjuntos difusos, como la ya ampliamente comentada distancia de Hamming, la pseudo-métrica de Puri-Ralescu²([84]), así como otras familias de métricas propuestas por C. Bertoluzza ([7]), R. Körner ([60]) o W. Trutschnig ([92]), definidas para familias específicas de conjuntos difusos convexos.

Si construimos la siguiente medida $D = \frac{d^2}{2}$ donde d es cualquiera de las (pseudo-)métricas mencionadas, se extienden los conceptos existentes de varianza escalar (como los de R. Körner [60] o A. Lubiano et al. [66]) mediante la D -variación correspondiente.

Una de las ventajas de la d -variación es el hecho de que puede ser aplicada incluso cuando las imágenes de la variable aleatoria difusa no son convexas o no son subconjuntos de $\mathbb{R}^n$, al contrario que las varianzas escalares anteriormente mencionadas, y además no es necesaria la existencia de la esperanza de la variable para obtener la d -variación.

Aunque para el cálculo de la d -variación no sea necesario considerar un conjunto difuso “central” con respecto al cual calcular las diferencias de las imágenes de la variable, podríamos tener en cuenta el centroide de la misma. De este modo, puede considerarse una definición alternativa para cuantificar el grado de

²Puri y Ralescu introducen una métrica en la clase de los subconjuntos difusos de $\mathbb{R}^n$ cuyos α -cortes son compactos y no vacíos. Esta medida puede ser extendida fácilmente a familias más generales de subconjuntos difusos como una pseudo-métrica.

dispersión, análoga a la utilizada en las definiciones de varianza escalar mencionadas.

Definición 26 Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ una variable aleatoria difusa definida sobre él. Sea d una medida de “diferencia” entre subconjuntos de $\mathcal{F}(U)$. La d -variación o d -dispersión de $\tilde{\Gamma}$ respecto a su centroide, en el caso de que este exista, denotada por $V_{C_d}(\tilde{\Gamma})$, se define como:

$$V_{C_d}(\tilde{\Gamma}) = E[d(\tilde{\Gamma}, C_d(\tilde{\Gamma}))]$$

Esta cantidad estará bien definida siempre que la aplicación $g(\omega) = d(\tilde{\Gamma}(\omega), C_d(\tilde{\Gamma}))$, $\forall \omega \in \Omega$ sea $\mathcal{A} - \beta_{\mathbb{R}}$ medible.

La d -variación y la d -variación respecto al centroide de una variable aleatoria difusa no coinciden necesariamente. Ilustremos este hecho mediante un contraejemplo.

Ejemplo 22 Consideremos el espacio inicial $\Omega = \{\omega_1, \omega_2, \omega_3\}$ con la sigma-álgebra $\mathcal{A} = \wp(\Omega)$ y la distribución de probabilidad dada por $P(\{\omega_1\}) = 0,1$, $P(\{\omega_2\}) = 0,6$ y $P(\{\omega_3\}) = 0,3$. Consideremos así mismo un conjunto $U = \{x_1, x_2, x_3, x_4\}$ y un conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$ definido como: $\Gamma(\omega_1) = \{x_1, x_2\}$, $\Gamma(\omega_2) = \{x_2, x_3, x_4\}$ y $\Gamma(\omega_3) = \{x_1, x_3\}$. El centroide de Γ con respecto a la distancia de Hamming es el conjunto $\{x_2, x_3, x_4\}$. La d_{Ham} -variación de Γ es 1,56, y la d_{Ham} -variación con respecto al centroide es 1,2.

Además de no coincidir numéricamente, ambas medidas pueden tener diferentes interpretaciones. Veámoslo mediante un nuevo ejemplo. Consideremos un conjunto aleatorio Γ que asigna a

cada individuo de una población el conjunto de los idiomas que habla. El centroide de esta aplicación, con respecto a la distancia de Hamming, es el conjunto de los idiomas que son hablados por más de la mitad de la población. Considerando la población de España, el centroide sería el conjunto {español, inglés}. La d_{Ham} -variación de Γ representa el número medio de idiomas que, al considerar las distintas parejas de individuos, son dominados por uno solo de ellos y no por el otro. Por otro lado, la d_{Ham} -variación respecto al centroide, representa el número medio de idiomas que habla un individuo y no habla más de la mitad de la población, y viceversa. En este caso, la d_{Ham} -variación nos proporciona mejor información sobre la capacidad de comunicación entre dos individuos cualesquiera de la población. Por tanto, si ese es nuestro objetivo, escogeremos dicha medida de dispersión. Sin embargo, en otros casos puede ser más conveniente la utilización de la variación respecto al centroide. Por ello no podemos justificar que, en general, una medida sea mejor que la otra. En cada caso, el investigador escogerá aquella que se adapte mejor a los objetivos de su trabajo.

En el resto de nuestro estudio, nos centraremos en la d -variación, debido a la ventaja que presenta por no necesitar el cálculo del centroide.

2.3.1. Propiedades de la d -variación

Al igual que sucedía con el centroide, algunas propiedades de las d -variaciones podrán ser estudiadas en función de los axiomas verificados por la medida de diferencia considerada. Otras, de nuevo, estarán relacionadas con la forma concreta de la me-

dida, por lo que deberán ser estudiadas para cada caso. Comenzamos viendo algunas propiedades generales.

Proposición 15 *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ una variable aleatoria difusa definida sobre él. Sea d una medida de “diferencia” entre subconjuntos de $\mathcal{F}(U)$. Entonces:*

1. *Si d verifica el axioma $D3$ y $\tilde{\Gamma}$ es constante, $V_d(\tilde{\Gamma}) = 0$.*
2. *Si d verifica el axioma $D5$, $V_d(X \cup A) \leq V_d(X) \forall \tilde{A} \in \mathcal{F}(U)$.*
3. *Si d verifica el axioma $D4$, $V_d(X \cap A) \leq V_d(X) \forall \tilde{A} \in \mathcal{F}(U)$.*
4. *Si d verifica el axioma $D7$, $V_d(X) = V_d(X \setminus \cap X_i)$ donde $\{X_i\}_i$ es la familia de las imágenes de la variable.*

Demostración: La demostración de estas propiedades es obvia.
□

Ilustremos el significado de estas propiedades utilizando de nuevo el conjunto de idiomas de ejemplos anteriores. La primera de las propiedades significa que la d -variación es nula cuando toda la población habla el mismo conjunto de idiomas. Para la segunda propiedad, supongamos que todas las personas de la población que no hablan español realizan un curso para aprender dicho idioma. Entonces, la d -variación debe decrecer. La interpretación de la tercera propiedad es similar; en este caso, todas las personas de la población deciden utilizar únicamente aquellos idiomas que conocen y que se encuentran dentro de cierto conjunto. Al tomar esa decisión, la dispersión en las habilidades comunicativas de la población debe disminuir. En cuanto a

la cuarta propiedad, supongamos que todos los miembros de la población hablan un determinado conjunto de idiomas y se les prohíbe utilizarlos. La dispersión en las destrezas de comunicación seguiría siendo la misma.

Caso concreto: la distancia de Hamming

A modo de ejemplo, analizamos a continuación algunas propiedades interesantes que son verificadas por las d-variaciones de una variable aleatoria difusa con respecto a la distancia de Hamming.

Proposición 16 *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico. Sea U el universo en estudio. Sea $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$ una variable aleatoria difusa. Sea d_{Ham} la distancia de Hamming entre subconjuntos difusos de U . La d_{Ham} -variación de la variable aleatoria difusa $\tilde{\Gamma}$, denotada por $V_{d_{Ham}}(X)$, verifica las siguientes propiedades:*

1. $\tilde{\Gamma}$ es constante si y solo si $V_{d_{Ham}}(\tilde{\Gamma}) = 0$.
2. $V_{d_{Ham}}(a \cdot \tilde{\Gamma}) = V_{d_{Ham}}(\tilde{\Gamma})$, $\forall a, b \in \mathbb{K}$. La operación considerada es la resultante del Principio de Extensión de Zadeh, es decir, $(\lambda \cdot \tilde{A})(x) = \tilde{A}(\frac{1}{\lambda} \cdot x)$.
3. $V_{d_{Ham}}(a \cdot \tilde{\Gamma}) = a \cdot V_{d_{Ham}}(\tilde{\Gamma})$ si interpretamos el producto de un escalar por un difuso como $(\lambda \cdot \tilde{A})(x) = \min\{1, \lambda \cdot (\tilde{A}(x))\}$.
4. Sea $g : U \longrightarrow U$ una aplicación biyectiva, a partir de la cual contruimos una función $G : \mathcal{F}(U) \longrightarrow \mathcal{F}(U)$ como $G(A)(x) = A(g^{-1}(x))$. Entonces $V_{d_{Ham}}(G(\tilde{\Gamma})) = V_{d_{Ham}}(\tilde{\Gamma})$.

Demostración:

1. Es obvio que si $\tilde{\Gamma}$ es constante, $V_{d_{Ham}}(\tilde{\Gamma}) = 0$. Veamos el recíproco:

Estamos suponiendo que

$$V_{d_{Ham}}(\tilde{\Gamma}) = \int_{\Omega \times \Omega} d_{Ham}(\tilde{\Gamma}_1(\omega), \tilde{\Gamma}_2(\omega')) d(P \otimes P)(\omega, \omega') = 0.$$

Como la función que integramos es no negativa (es una distancia), que la integral tenga valor 0 implica que la función es constante e igual a 0. Es decir, para cada par $(\omega, \omega') \in \Omega^2$, $d_{Ham}(\tilde{\Gamma}_1(\omega), \tilde{\Gamma}_2(\omega')) = 0$, lo que implica que $\tilde{\Gamma}$ es constante.

2. El caso de la variable aleatoria clásica es obvio, ya que en tal caso la distancia de Hamming esperada solo depende del conjunto de probabilidades que toma la variable y no de los valores concretos: $V_{d_{Ham}}(\tilde{\Gamma}) = 2\left(1 - \sum_{i=1}^n P(\tilde{\Gamma} = x_i)\right)$ donde $\{x_1, \dots, x_n\} = U$.

El caso del conjunto aleatorio y la v.a.d. (considerando el producto por escalar resultante del principio de extensión de Zadeh) es también inmediato, teniendo en cuenta que la distancia de Hamming entre dos conjuntos es la misma distancia existente entre sus imágenes por la función $f(x) = \frac{1}{a}x$.

3. En este tercer caso, con esa interpretación del producto por a , es obvio que

$$d_{Ham}(a\tilde{\Gamma}_1(\omega), a\tilde{\Gamma}_2(\omega')) = |a|d_{Ham}(\tilde{\Gamma}_1(\omega), \tilde{\Gamma}_2(\omega')),$$

de donde se deduce la tesis del resultado.

4. La demostración de esta propiedad es análoga a la de la propiedad 2. $\square$

Capítulo 3

Conjuntos aleatorios rugosos y variables aleatorias rugosas difusas

El rápido desarrollo de la teoría de conjuntos aleatorios durante las últimas cuatro décadas ha sido debido principalmente a las siguientes aplicaciones: la geometría estocástica y la representación de mediciones aleatorias imprecisas. La geometría estocástica se ocupa del estudio de estructuras aleatorias como simples puntos, segmentos o conjuntos cerrados arbitrarios. Por otra parte, los conjuntos aleatorios resultan ser la representación natural a la hora de representar cantidades aleatorias observadas de forma imprecisa, cuando lo único que podemos determinar es un conjunto de valores que contiene a cada resultado. Cada una de estas ramas de la teoría de los conjuntos aleatorios está conectada a cada una de las dos interpretaciones diferentes de los conjuntos mencionadas en el capítulo anterior (las interpretaciones *conjuntiva* y *disyuntiva* ([37, 40])). Tal como señalábamos entonces, bajo la primera interpretación, un conjunto es considerado como un objeto complejo, y proporciona información precisa acerca de una entidad objetiva, mientras que bajo la vi-

sión disyuntiva, el conjunto representa información incompleta acerca de un elemento x del universo. El conjunto “disyuntivo” E representa el conjunto de posibles candidatos para x , mientras que todos los elementos que están fuera de E son considerados como imposibles.

El concepto de variable aleatoria difusa, de más reciente aparición, también admite varias interpretaciones diferentes (ver [19] para una descripción detallada de las distintas interpretaciones). Cada una de ellas está relacionada con cada uno de los dos enfoques generales de los conjuntos difusos, el óntico y el epistémico, ya mencionados en el capítulo 2. Tal como recordábamos entonces, bajo la interpretación “epistémica” el conjunto difuso representa cierta información imprecisa o incompleta acerca de un elemento del universo, $x \in U$. El grado de pertenencia de cada elemento del universo $y \in U$ refleja el grado de posibilidad de que éste coincida con x . En otras palabras, basándonos en algunos resultados de [27], y teniendo en cuenta las propiedades de las medidas de posibilidad, nuestra información acerca del elemento x se puede expresar en términos de los α -cortes del conjunto difuso de la forma siguiente: la probabilidad de que el corte de nivel α contenga a x es mayor o igual que $1 - \alpha$, para cada $\alpha \in (0, 1]$. De forma alternativa, bajo la interpretación “óntica”, el conjunto borroso es considerado un objeto complejo cuyas fronteras están “difuminadas”. El grado en que cada elemento del universo contribuye a ese objeto complejo viene dado por su valor de pertenencia.

En este capítulo, combinaremos ambos enfoques, el enfoque conjuntivo de los conjuntos nítidos y el óntico de los conjuntos

(borrosos). Por una parte, consideraremos un conjunto aleatorio conjuntivo (o, en general, una variable aleatoria difusa óptica), cuya distribución de probabilidad queremos caracterizar. Por otra, asumiremos que disponemos de información incompleta acerca de sus imágenes: partiremos de una relación de equivalencia determinada por una colección de atributos de los elementos del universo. Si no podemos observar directamente dichos elementos, sino que solo podemos determinar el vector de valores que la colección de atributos asigna a cada uno de ellos, todos los elementos que pertenezcan a la misma clase de equivalencia (cada conjunto de elementos asociados al mismo vector de valores de los atributos) serán considerados indistinguibles o indiscernibles a todos los efectos. De esta forma, nuestra información acerca de cada conjunto (difuso) del universo vendrá determinada por una pareja de aproximaciones superior e inferior. La aproximación superior de un conjunto nítido es la familia de elementos que posiblemente pertenecen a él (es decir, la unión de todas las clases de equivalencia que tienen intersección no vacía con él). Por su parte, la aproximación inferior de un conjunto estará constituido por la colección de elementos que pertenecen a él con seguridad (la unión de todas las clases de equivalencia incluidas en el conjunto). Diremos que un subconjunto del universo es “exacto” cuando sus aproximaciones inferior y superior coinciden entre ellas, y por tanto, coinciden con él. En caso contrario, diremos que el conjunto es “rugoso”. Los conceptos de aproximación superior e inferior también se pueden extender al caso en que el conjunto que se desea aproximar es un conjunto difuso, de acuerdo con el concepto de conjunto difuso rugoso, tal como veremos en secciones posteriores.

En este contexto, nuestra información acerca del conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$ estará determinada por la pareja de conjuntos aleatorios, $(\underline{\Gamma}, \overline{\Gamma})$, que asigna a cada elemento $\omega \in \Omega$ la pareja de aproximaciones inferior e superior de $\Gamma(\omega)$. De forma más general, nuestra información acerca de la variable aleatoria difusa estudiada vendrá representada por una pareja de variables aleatorias difusas, $(\tilde{\underline{\Gamma}}, \tilde{\overline{\Gamma}})$ a las que denominaremos también aproximaciones inferior y superior de $\tilde{\Gamma}$. Los objetivos principales de este capítulo son tres: primero, caracterizaremos la información disponible acerca de probabilidad inducida por Γ (o $\tilde{\Gamma}$). Esta información vendrá determinada, en ciertas situaciones, por medio de la distribución conjunta del par formado por las aproximaciones inferior y superior. En segundo lugar, mostraremos la forma general de caracterizar la información acerca de cualquier parámetro de la distribución inducida por $\tilde{\Gamma}$, como su centroide o su dispersión esperada, por ejemplo. En tercer lugar, trataremos de cuantificar el grado de “falta de exactitud” o “rugosidad” de un conjunto aleatorio y, en general, de una variable aleatoria difusa, cuando la información acerca de ella viene “granularizada” de la forma que acabamos de explicar.

3.1. Conceptos preliminares

3.1.1. Conjuntos aleatorios

Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y un espacio medible $(U, \mathcal{A}')$. Tal como indicábamos en el capítulo anterior, se denomina *conjunto aleatorio* a una aplicación multivaluada $\Gamma : \Omega \rightarrow \wp(U)$ definida entre ellos, cuando ésta satisface cierta condición de medibilidad. En el capítulo anterior hemos

elegido la condición de medibilidad fuerte, tal como haremos también en este. Sin embargo, en la literatura se han considerado otras condiciones de medibilidad diferentes (consúltense, por ejemplo, las referencias [13, 54, 59] para una información más detallada).

En este capítulo vamos a considerar, además, el concepto de grafo-medibilidad (véase [54], por ejemplo):

Definición 27 *Consideremos un espacio probabilístico, $(\Omega, \mathcal{A}, P)$, y un espacio medible, $(U, \mathcal{A}')$. La aplicación multi-valuada $\Gamma : \Omega \rightarrow \wp(U)$ es grafo-medible si y solo si su grafo, $Gr(\Gamma) = \{(\omega, x) : \omega \in \Omega \text{ y } x \in \Gamma(\omega)\}$, pertenece a la σ -álgebra producto $\mathcal{A} \otimes \mathcal{A}'$.*

El lector interesado puede consultar un estudio detallado acerca de las relaciones de implicación entre diferentes condiciones de medibilidad en [54]. Aquí, vamos a recordar algunas relaciones formales existentes entre las condiciones de grafo-medibilidad y medibilidad fuerte, en relación con la estructura topológica del espacio final y las propiedades del espacio inicial. Recordemos que un espacio probabilístico $(\Omega, \mathcal{A}, P)$ es *completo* si todo conjunto nulo es medible. Por otra parte, se dice que un espacio topológico (U, τ) es:

- *polaco*, si es metrizable con alguna métrica d , de forma que (U, d) es completo y separable.
- *Souslin* si es la imagen biyectiva de un espacio polaco.

Teorema 1 ([54]) *Sean $(\Omega, \mathcal{A}, P)$ un espacio probabilístico, (U, τ) un espacio topológico y $\Gamma : \Omega \rightarrow \wp(U)$ una aplicación multivaluada. Se verifican las siguientes implicaciones:*

- (a) Si Γ es fuertemente medible y sus imágenes son cerradas, entonces es grafo-medible.
- (b) Si $(\Omega, \mathcal{A}, P)$ es completo, (U, τ) es Souslin y Γ es grafo-medible, entonces es fuertemente medible.

Por otra parte, debemos hacer notar que la condición de medibilidad fuerte puede ser expresada en términos de la medibilidad de la función $\Gamma : \Omega \rightarrow \wp(U)$ considerada como una función unipuntual que toma valores en la familia de subconjuntos del espacio final, $\wp(U)$. Más específicamente, consideremos, para cada $B \in \mathcal{A}'$, la familia de conjuntos:

$$\wp_B = \{C \subseteq U : C \cap B \neq \emptyset\},$$

y denotemos por $\sigma_\wp$ a la σ -álgebra generada por la clase $\{\wp_B : B \in \mathcal{A}'\}$. Podemos observar fácilmente que Γ es fuertemente medible si y solo si es $\mathcal{A} - \sigma_\wp$ medible (considerada como una aplicación unipuntual con valores en $\wp(U)$).

Tal como hemos mencionado en el capítulo anterior, las probabilidades superior e inferior de Dempster asociadas a un conjunto medible $B \in \mathcal{A}'$ vienen dadas por las expresiones siguientes:

- La probabilidad superior de B es $P_*(B) = \frac{P(\Gamma_*(B))}{P(\Gamma^*(U))}$.
- La probabilidad inferior de B es $P^*(B) = \frac{P(\Gamma^*(B))}{P(\Gamma^*(U))}$.

Si, además, las imágenes de Γ son subconjuntos no vacíos de U , entonces la inversa superior de U es el conjunto total Ω ,

de manera que las expresiones anteriores pueden escribirse de forma más simple como sigue:

- $P^*(B) = P(\Gamma^*(B)) = P(\{\omega \in \Omega : \Gamma(\omega) \cap B \neq \emptyset\})$.
- $P_*(B) = P(\Gamma_*(B)) = P(\{\omega \in \Omega : \Gamma(\omega) \subseteq B\})$.

La pareja de probabilidades superior e inferior tiene un claro significado dentro de la interpretación disyuntiva de los conjuntos aleatorios. Supongamos que el conjunto aleatorio con imágenes no vacías Γ representa nuestro conocimiento incompleto acerca de una variable aleatoria $X_0 : \Omega \rightarrow U$, es decir, supongamos que no podemos determinar el valor, $X_0(\omega)$, asociado a cada resultado $\omega \in \Omega$, sino que solo podemos afirmar que pertenece al conjunto de elementos $\Gamma(\omega)$. En otras palabras, supongamos que todo lo que sabemos acerca de X_0 es que es una “selección medible” de Γ . Bajo estas condiciones, podemos comprobar fácilmente que las imágenes superior e inferior de un conjunto medible arbitrario B satisfacen las siguientes restricciones:

$$\begin{aligned} \Gamma_*(B) &= \{\omega \in \Omega : \Gamma(\omega) \subseteq B\} \subseteq \{\omega \in \Omega : X_0(\omega) \in B\} \subseteq \\ &\quad \{\omega \in \Omega : \Gamma(\omega) \cap B \neq \emptyset\} = \Gamma^*(B). \end{aligned}$$

Así, la pareja de probabilidades inferior y superior de B acotan a su “verdadera” probabilidad, $P_{X_0}(B)$, es decir;

$$P_*(B) \leq P_{X_0}(B) \leq P^*(B), \quad \forall B \in \mathcal{A}'. \quad (3.1)$$

Denotemos por $S(\Gamma)$ al conjunto de selecciones medibles de Γ :

$$S(\Gamma) = \{Y : \Omega \longrightarrow U \text{ medible con } Y(\omega) \in \Gamma(\omega) \ \forall \omega \in \Omega\},$$

y sea $\mathcal{P}(\Gamma) = \{P_Y : Y \in S(\Gamma)\}$ la familia de medidas de probabilidad inducidas por ellas. De acuerdo con la ecuación 3.1 el conjunto convexo de probabilidades $M(P^*) = \{P \text{ probabilidad} : P_* \leq P \leq P^*\}$ incluye a la familia $\mathcal{P}(\Gamma)$. El recíproco no es cierto en general, tal como se muestra, por ejemplo, en [29]. Podemos encontrar relaciones formales adicionales entre estas dos clases de medidas de probabilidad en [69, 70].

En el apartado siguiente, tendremos en cuenta este resultado que indica cierta relación entre las dos familias en el caso particular de que el conjunto aleatorio sea simple:

Teorema 2 ([28]) *Sea $(\Omega, \mathcal{A}, P)$ un espacio probabilístico, y sea $(U, \mathcal{A}')$ un espacio medible. Sea $\Gamma : \Omega \rightarrow \wp(U)$ una aplicación multi-valuada definida entre ellos. Si $\Gamma : \Omega \rightarrow \wp(U)$ es simple (es decir, su número de posibles resultados es finito), entonces:*

$$P^*(B) = \max_{P \in \mathcal{P}(\Gamma)} P(B) \text{ y } P_*(B) = \min_{P \in \mathcal{P}(\Gamma)} P(B).$$

Las probabilidades superior e inferior son duales, en el sentido de que $P^*(B^c) = 1 - P_*(B)$, $\forall B \in \mathcal{A}'$ y, por lo tanto, cada una de ellas determina unívocamente a la otra. Podemos considerar, de forma alternativa, la medida de probabilidad inducida por el conjunto aleatorio, $P \circ \Gamma^{-1} : \sigma_{\wp} \longrightarrow [0, 1]$. Dicha probabilidad

inducida surge de forma natural de la interpretación conjuntiva (representa la medida de probabilidad asociada a un objeto aleatorio cuyos resultados son subconjuntos del espacio final), de la misma forma que las probabilidades superior e inferior tienen un significado válido dentro de la interpretación disyuntiva, tal como hemos aclarado anteriormente. Pero, independientemente de la interpretación elegida, existe una conexión formal entre la probabilidad inducida y la pareja de probabilidades superior e inferior: podemos comprobar fácilmente que la medida de probabilidad inducida, $P \circ \Gamma^{-1}$ determina a la probabilidad superior, P^* , dado que se satisfacen las siguientes igualdades:

$$P \circ \Gamma^{-1}(\wp_B) = P(\Gamma^{-1}(\wp_B)) = P(\Gamma^*(B)) = P^*(B).$$

Recíprocamente, la probabilidad superior determina unívocamente a la probabilidad inducida, $P \circ \Gamma^{-1}$, dado que la clase $\{\wp_B^c : B \in \mathcal{A}'\}$ es cerrada para intersecciones.

Cuando, en particular, las imágenes de la aplicación multivaluada son subconjuntos no vacíos de un universo finito U , existe una relación formal entre las nociones anteriores y los conceptos de medidas de plausibilidad y creencia, así como el concepto de asignación básica de probabilidad descritos en la Teoría de la Evidencia ([90]). Cualquier aplicación $\mathcal{A} - \wp(U)$ -fuertemente medible $\Gamma : \Omega \longrightarrow \wp(U)$ es, de hecho, una aplicación $\mathcal{A} - \wp(\wp(U))$ -medible, vista como una función medible clásica, cuyas imágenes son elementos de $\wp(U)$. La función de masa inducida por esta aplicación medible satisface las propiedades de una asignación básica de probabilidad. Las probabilidades superior e inferior, por su parte, satisfacen las propiedades de

medida de plausibilidad y creencia respectivamente. Denotemos estas tres funciones por $m : \wp(U) \longrightarrow [0, 1]$, $Pl : \wp(U) \longrightarrow [0, 1]$ y $Bel : \wp(U) \longrightarrow [0, 1]$ y comprobemos este hecho.

Las imágenes de dichas funciones quedan determinadas por las siguientes expresiones:

- $m(C) = P(\Gamma = C) = P(\{\omega \in \Omega : \Gamma(\omega) = C\}), \forall C \in \wp(U).$
- $Pl(B) = P(\{\omega \in \Omega : \Gamma(\omega) \cap B \neq \emptyset\}).$
- $Bel(B) = P(\{\omega \in \Omega : \Gamma(\omega) \subseteq B\}).$

Resulta sencillo probar las siguientes igualdades:

- $m(\emptyset) = 0, \sum_{C \in \wp(U)} m(C) = 1,$
- $Pl(B) = \sum_{C \cap B \neq \emptyset} m(C)$ y
- $Bel(B) = \sum_{C \subseteq B} m(C), \forall B \subseteq U.$

Por tanto, m , Bel y Pl satisfacen respectivamente las propiedades formales de asignación básica de probabilidad, medida de creencia y medida de plausibilidad.

A continuación centraremos nuestra atención al caso particular en el que las imágenes de la aplicación multivaluada pueden expresarse a través de productos cartesianos. Es decir, consideraremos una aplicación $\Gamma = \Gamma_1 \times \Gamma_2$, donde $\Gamma_1 : \Omega \longrightarrow \wp(U_1)$ y $\Gamma_2 : \Omega \longrightarrow \wp(U_2)$ quedan determinados por las proyecciones de las imágenes de Γ sobre U_1 y U_2 respectivamente. Estas proyecciones son estocásticamente independientes si su distribución de

probabilidad conjunta puede expresarse como el producto de sus respectivas probabilidades. En tal caso, la probabilidad inducida por Γ puede obtenerse como una función de las probabilidades marginales $P \circ \Gamma_1^{-1}$ y $P \circ \Gamma_2^{-1}$.

Cuando, en particular, los dos universos son finitos, podemos expresar las respectivas funciones de masa asociadas a Γ_1 y Γ_2 por medio de sus respectivas asignaciones básicas de probabilidad, $m_1 : \wp(U_1) \rightarrow [0, 1]$ y $m_2 : \wp(U_2) \rightarrow [0, 1]$.

Además, el conjunto aleatorio Γ induce otra asignación básica de probabilidad m sobre $\wp(U_1 \times U_2)$, cuyos conjuntos focales son productos cartesianos. Cuando ambos conjuntos aleatorios, Γ_1 y Γ_2 , son estocásticamente independientes, la asignación m puede descomponerse como el producto de m_1 y m_2 como sigue:

$$m(A \times B) = m_1(A) \cdot m_2(B), \quad \forall A \subseteq U_1, B \subseteq U_2.$$

3.1.2. Variables aleatorias difusas

En el capítulo 2, hemos introducido el concepto de variable aleatoria difusa como el de una aplicación con valores borrosos, cuyos α -cortes satisfacen la condición de medibilidad fuerte. Esta condición de medibilidad es equivalente a la medibilidad (en el sentido clásico) con respecto a cierta σ -álgebra, como veremos a continuación y, por tanto, tendrá sentido hablar de la medida de probabilidad inducida en dicha σ -álgebra por la variable aleatoria difusa. Consideremos, para cada $A \in \mathcal{A}'$, la siguiente familia de conjuntos:

$$\mathcal{F}_A^\alpha = \{F \in \mathcal{F}(U) : F_\alpha \cap A \neq \emptyset\}.$$

Denotemos por $\sigma_{\mathcal{F}}$ la σ -álgebra generada por la clase:

$$\{\mathcal{F}_A^\alpha : A \in \mathcal{A}', \alpha \in [0, 1]\} \subseteq \wp(\mathcal{F}(U)).$$

Resulta sencillo comprobar que una aplicación cuyos valores son subconjuntos difusos del universo U es una variable aleatoria difusa si y solo si es $\mathcal{A} - \sigma_{\mathcal{F}}$ medible. De acuerdo con la notación establecida al comienzo de esta sección, denotaremos por $P \circ \tilde{\Gamma}^{-1}$ a la probabilidad inducida por $\tilde{\Gamma}$ sobre $\sigma_{\mathcal{F}}$.

3.1.3. Conjuntos rugosos

Sea $\Pi = \{A_1, \dots, A_k\}$ una partición finita del universo arbitrario U . Denotemos por R a la relación de equivalencia asociada a Π . Para un elemento arbitrario $x \in U$, $[x]_R$ denotará su clase de equivalencia, es decir, el único conjunto de la partición Π al cual pertenece. Sea C un subconjunto arbitrario de U .

- La *aproximación inferior* de C es el conjunto formado por todos aquellos elementos que pueden ser clasificados con certeza como C con respecto a R (son verdaderamente C con respecto a R), $\underline{C} = \{x \in U : [x]_R \subseteq C\}$.
- La *aproximación superior* de C es el conjunto formado por todos aquellos elementos que pueden ser clasificados posiblemente como C con respecto a R (son posiblemente C desde el punto de vista de R), $\overline{C} = \{x \in U : [x]_R \cap C \neq \emptyset\}$.
- La región frontera de un conjunto C con respecto a R es el conjunto formado por todos aquellos elementos que no pueden clasificarse como C ni como no- C con respecto a R , $\overline{C} \setminus \underline{C}$.

Un conjunto formado por elementos indiscernibles (similares) se denomina *conjunto elemental*, y forma un gránulo (átomo) básico de conocimiento sobre el universo. Cualquier unión de conjuntos elementales se denomina conjunto *preciso* o *exacto*, y en otro caso el conjunto se denomina *rugoso* o *inexacto*.

3.1.4. Conjuntos rugoso difusos

El concepto de conjunto rugoso difuso (*fuzzy rough set*) fue introducido por primera vez por Dubois y Prade en 1990 ([38]). La formulación actual, que se refiere a un concepto más general, se expresa de la forma que se indica en la definición 31, en la que se introducen las nociones de aproximación inferior y superior de un conjunto borroso. Para poder introducir estos conceptos, necesitamos apoyarnos en dos definiciones auxiliares, que presentamos a continuación.

Definición 28 *Llamamos relación de similaridad difusa a toda función $R : U \times U \longrightarrow [0, 1]$ que satisface las siguientes condiciones:*

- $R(x, x) = 1, \forall x \in U.$
- $R(x, y) = R(y, x), \forall (x, y) \in U \times U$
- $R(x, z) \leq \min\{R(x, y), R(y, z)\}, \forall x, y, z \in U$

Es bien sabido que el concepto de relación de similaridad generaliza el concepto de equivalencia utilizado en el apartado anterior, en el sentido de que una relación de equivalencia no es más que una relación de similaridad que toma valores en el conjunto binario $\{0, 1\}$. Por esta razón, utilizaremos la misma nomenclatura, R , en ambos casos.

Definición 29 Llamamos implicación difusa a cualquier función $I : [0, 1] \times [0, 1] \longrightarrow [0, 1]$ que cumple las propiedades siguientes:

- Si $x \leq z$, entonces $I(x, y) \geq I(z, y), \forall y \in [0, 1]$
- Si $y \leq t$, entonces $I(x, y) \leq I(x, t), \forall y \in [0, 1]$
- $I(0, y) = 1$
- $I(x, 1) = 1$
- $I(1, 0) = 0$

Definición 30 Llamamos t -norma a cualquier función $T : [0, 1] \times [0, 1] \longrightarrow [0, 1]$ que satisface las propiedades siguientes:

- $T(x, y) = T(y, x), \forall (x, y) \in [0, 1] \times [0, 1]$
- $T(x, T(y, z)) = T(T(x, y), z), \forall x, y, z \in [0, 1]$
- $T(x, 1) = x$
- Si $x \leq y$, entonces $T(x, z) \leq T(y, z), \forall z \in [0, 1]$

Ahora ya podemos pasar a definir el concepto mencionado anteriormente.

Definición 31 Consideremos una medida de similaridad R , una implicación difusa, I y una t -norma, T , definidas sobre un universo U . Consideremos un subconjunto borroso de U , $\tilde{C} \in \mathcal{F}(U)$. Llamamos aproximación inferior de $\tilde{C}$ al conjunto difuso cuya función de pertenencia viene dada por la expresión siguiente:

$$\underline{\tilde{C}}(y) = \inf_{x \in U} I(R(x, y), A(x)).$$

Llamamos aproximación superior de $\tilde{C}$ al conjunto difuso cuya función de pertenencia viene dada por la siguiente expresión:

$$\overline{\tilde{C}}(y) = \sup_{x \in U} T((R(x, y), A(x)))$$

En la definición inicial de Dubois y Prade de 1990 ([38]), se consideró solamente el caso particular en que T era la t-norma del mínimo, y la implicación difusa venía dada por la expresión $I(x, y) = \max\{1 - x, y\}$, $\forall (x, y) \in [0, 1] \times [0, 1]$. Otros autores, como por ejemplo Radzikowska y Kerre ([85]) consideraron posteriormente familias más amplias de implicaciones difusas y t-normas. En este trabajo nos restringiremos a la definición original, y dentro de ella, consideraremos el caso particular en que R toma valores en el conjunto binario $\{0, 1\}$, es decir, el caso en que R representa una relación de equivalencia, determinada por una colección de atributos observados sobre los elementos del universo U , y que además determina, al igual que en la sección anterior, una partición finita del universo. En esta situación especial, se puede comprobar fácilmente que las expresiones de las aproximaciones inferior y superior se reducen a las siguientes:

$$\underline{\tilde{C}}(y) = \inf_{x \in [y]_R} C(x) \text{ y } \overline{\tilde{C}}(y) = \sup_{x \in [y]_R} C(x). \quad (3.2)$$

Observe el lector que, en el caso particular de que $\tilde{C}$ sea un conjunto nítido, las expresiones anteriores se reducen a las expresiones de las funciones indicador de las respectivas aproximaciones inferior y superior del conjunto, estudiadas en la sección anterior.

De forma análoga al caso particular de los conjuntos nítidos, diremos que el conjunto difuso $\tilde{C}$ es *exacto* cuando sus aproximaciones inferior y superior coinciden. En otro caso, diremos que es *rugoso*.

3.2. Conjuntos aleatorios rugosos

Consideremos un espacio probabilístico $(\Omega, \mathcal{A}, P)$, un espacio medible arbitrario $(U, \mathcal{A}')$ y una aplicación fuertemente medible $\Gamma : \Omega \rightarrow \wp(U)$. Supongamos además que no podemos observar directamente los elementos de U , sino que únicamente podemos observar el valor de cierta colección de atributos, $\mathbf{a} : U \rightarrow V = \{v_1, \dots, v_k\}$, asignados a ellos. La función atributo induce una partición finita sobre U , $\Pi = \{A_1, \dots, A_k\}$. Sólomente somos capaces de determinar el índice $i \in \{1, \dots, k\}$ tal que $x \in A_i$, para cada elemento arbitrario $x \in U$. De acuerdo con este hecho, únicamente podremos proporcionar las aproximaciones superior e inferior de los subconjuntos arbitrarios de U . Para cada $\omega \in \Omega$, consideremos las aproximaciones superior e inferior de $\Gamma(\omega)$,

$$\begin{aligned}\underline{\Gamma}(\omega) &= \{x \in U : [x]_R \subseteq \Gamma(\omega)\}, \\ \overline{\Gamma}(\omega) &= \{x \in U : [x]_R \cap \Gamma(\omega) \neq \emptyset\}.\end{aligned}$$

Denotemos respectivamente por $\underline{\Gamma}$ y $\overline{\Gamma}$ las aplicaciones multi-valuadas que asignan, a cada $\omega \in \Omega$, las aproximaciones inferior y superior de $\Gamma(\omega)$,

$$\underline{\Gamma}(\omega) := \underline{\Gamma}(\omega), \overline{\Gamma}(\omega) := \overline{\Gamma}(\omega), \forall \omega \in \Omega.$$

$$\underline{\Gamma}(\omega) \subseteq \Gamma(\omega) \subseteq \overline{\Gamma}(\omega), \forall \omega \in \Omega.$$

A partir de este punto, $\underline{\Gamma}$ y $\bar{\Gamma}$ serán denominadas *aproximación inferior* y *superior* de Γ respectivamente. Denotaremos por $\Lambda = (\underline{\Gamma}, \bar{\Gamma})$ la aplicación con imágenes rugosas que asigna, a cada $\omega \in \Omega$, el conjunto rugoso $(\underline{\Gamma}(\omega), \bar{\Gamma}(\omega))$. La siguiente definición será fundamental durante el resto del capítulo.

Definición 32 *Consideremos un espacio probabilístico $(\Omega, \mathcal{A}, P)$ y un espacio medible $(U, \mathcal{A}')$. Consideremos además una partición finita Π sobre U . Un conjunto aleatorio rugoso es un par ordenado $(\underline{\Gamma}, \bar{\Gamma}) : \Omega \longrightarrow \wp(U) \times \wp(U)$ satisfaciendo:*

- *Existe una aplicación multivaluada $\Gamma : \Omega \longrightarrow \wp(U)$ tal que $\underline{\Gamma}(\omega)$ y $\bar{\Gamma}(\omega)$ determinan las aproximaciones inferior y superior de $\Gamma(\omega)$ con respecto a Π , para cada $\omega \in \Omega$.*
- *$\underline{\Gamma}$ y $\bar{\Gamma}$ son $\mathcal{A} - \mathcal{A}'$ fuertemente medibles.*

El propósito principal de este trabajo es caracterizar la información que el conjunto aleatorio rugoso proporciona sobre la probabilidad inducida por el conjunto aleatorio. A fin de conseguirlo, contemplaremos el conjunto aleatorio Γ como una aplicación medible con valores en $\wp(U)$, e identificaremos el conjunto aleatorio rugoso con una aplicación multi-valuada que lo contiene. Nuestro conjunto aleatorio Γ será visto entonces como una selección medible de tal aplicación multi-valuada. De acuerdo con algunos resultados existentes en la literatura reciente acerca de los conjuntos aleatorios, enumeraremos algunas condiciones suficientes para que la probabilidad superior inducida por tal aplicación multi-valuada caracterice la información que el conjunto aleatorio rugoso proporciona acerca de la probabilidad inducida por Γ .

Definición 33 Sean $(\Omega, \mathcal{A}, P)$ un espacio probabilístico y $(U, \mathcal{A}')$ un espacio medible. Consideremos además una partición finita Π sobre U y un conjunto aleatorio rugoso $\Lambda = (\underline{\Gamma}, \bar{\Gamma})$. La aplicación $G_\Lambda : \Omega \longrightarrow \wp(\wp(U))$ definida como sigue:

$$G_\Lambda(\omega) = \{C \subseteq U : \underline{\Gamma}(\omega) \subseteq C \subseteq \bar{\Gamma}(\omega)\}.$$

se llamará meta-conjunto aleatorio asociado a Λ .

El resto de esta sección se encargará de la caracterización de la medibilidad fuerte de G_Λ en relación con la medibilidad fuerte de $\bar{\Gamma}$, $\underline{\Gamma}$ y Γ .

Lema 2 Sean $(\Omega, \mathcal{A}, P)$ y $(U, \mathcal{A}')$ dos espacios medibles. Sea $H : \Omega \longrightarrow \wp(U)$ una aplicación multi-valuada simple. H es fuertemente medible si y solo si

$$H^{-1}(\{C\}) = \{\omega \in \Omega : H(\omega) = C\} \in \mathcal{A}, \quad \forall C \in \wp(U).$$

Demostración: Denotemos por $C_1, \dots, C_k$ las diferentes imágenes de H , y consideremos un conjunto arbitrario C . Podemos observar fácilmente que:

$$H^*(C) = \cup_{i: C_i \cap C \neq \emptyset} H^{-1}(\{C_i\}),$$

y además este conjunto puede ser expresado como una unión finita de anti-imágenes de salidas de H . Inversamente, observamos que:

$$H^{-1}(\{C_i\}) = \cap_{j: C_j \subseteq C_i} H^*(C_j) \cap [H^*(C_i^c)]^c, \quad \forall i \in \{1, \dots, k\}.$$

Por tanto, todos los conjuntos de la forma $H^*(C)$ son medibles si y solo si todos los conjuntos de la forma $H^{-1}(\{C_i\})$ son medibles, con lo que queda demostrado el lema. $\square$

Corolario 5 Sean $(\Omega, \mathcal{A})$ y $(U, \mathcal{A}')$ dos espacios medibles. Consideremos una partición finita, Π , de U , y un conjunto aleatorio rugoso $\Lambda = (\underline{\Gamma}, \overline{\Gamma}) : \Omega \longrightarrow \wp(U) \times \wp(U)$. El meta-conjunto aleatorio G_Λ es fuertemente medible si y solo si

$$G_\Lambda^{-1}(\{\mathcal{C}\}) = \{\omega \in \Omega : G_\Lambda(\omega) = \mathcal{C}\} \in \mathcal{A}, \quad \forall \mathcal{C} \in \wp(\wp(U)).$$

Demostración: En primer lugar, tengamos en cuenta que el número de posibles imágenes diferentes de G_Λ es finito: de hecho, debido a que la partición Π es finita, el número de diferentes imágenes de $\underline{\Gamma}$ y $\overline{\Gamma}$ es finita, y por tanto también lo es el cardinal de la imagen de G_Λ . Así pues, podemos deducir fácilmente el resultado a partir del lema 2. $\square$

Lema 3 Sean $(\Omega, \mathcal{A})$ y $(U, \mathcal{A}')$ dos espacios medibles arbitrarios, y sea $\Gamma : \Omega \longrightarrow \wp(U)$ una aplicación multivaluada definida entre ambos. Sea Π una partición finita de U . Denotemos por $\underline{\Gamma}$ y $\overline{\Gamma}$ las aproximaciones inferior y superior de Γ . Denotemos por G_Λ el meta-conjunto aleatorio asociado a $\Lambda = (\underline{\Gamma}, \overline{\Gamma})$. Entonces:

- Si Γ es fuertemente medible entonces $\overline{\Gamma}$ es fuertemente medible.
- Si Γ^c es fuertemente medible entonces $\underline{\Gamma}$ es fuertemente medible.
- Si $\underline{\Gamma}$ y $\overline{\Gamma}$ son fuertemente medibles entonces G_Λ es fuertemente medible.

Demostración: Las tres afirmaciones pueden deducirse fácilmente si tenemos en cuenta el lema 2 y las siguientes igualdades, donde $\cup_\Pi = \{A_i\}_i$ es la familia de todas las uniones finitas de conjuntos de la partición Π :

- $\bar{\Gamma}^{-1}(\{A\}) = [\cap_{i: A_i \subseteq A} \Gamma^*(A_i)] \cap [\cap_{i: A_i \cap A = \emptyset} [\Gamma^*(A_i)]^c]$.
- $\underline{\Gamma}^{-1}(\{A\}) = [(\Gamma^c)^*(A)]^c \cap [\cap_{i: A_i \cap A = \emptyset} (\Gamma^c)^*(A_i)]$
- $G_{\Lambda}^{-1}(\{\mathcal{C}_{(A,B)}\}) = \underline{\Gamma}^{-1}(\{A\}) \cap \bar{\Gamma}^{-1}(\{B\}), \forall A, B \in \cup_{\Pi}$, donde $\mathcal{C}_{(A,B)} = \{C : A \subseteq C \subseteq B\}$. $\square$

Teorema 3 *Consideremos un espacio probabilístico completo $(\Omega, \mathcal{A}, P)$ y un espacio Souslin (U, τ) . Sea $\Gamma : \Omega \rightarrow \wp(U)$ una aplicación multi-valuada con imágenes cerradas. Sea Π una partición finita sobre U y sea G_{Λ} el meta-conjunto aleatorio asociado a $\Lambda = (\underline{\Gamma}, \bar{\Gamma})$. Si Γ es fuertemente medible entonces $\underline{\Gamma}$, $\bar{\Gamma}$ y G_{Λ} son fuertemente medibles.*

Demostración: Si Γ es fuertemente medible entonces, de acuerdo con el teorema 1, es grafo-medible. Entonces, teniendo en cuenta que el grafo de Γ^c coincide con el complementario del grafo de Γ , deducimos que Γ^c es así mismo grafo-medible. Por tanto, de acuerdo con el teorema 1, es también fuertemente medible. Finalmente, de acuerdo con el lema 3, $\underline{\Gamma}$ y $\bar{\Gamma}$ y, por tanto, G_{Λ} , son fuertemente medibles. $\square$

Corolario 6 *Consideremos un espacio probabilístico completo $(\Omega, \mathcal{A}, P)$ y un espacio Souslin (U, τ) . Sea Π una partición finita sobre U .*

$(\underline{\Gamma}, \bar{\Gamma})$ es un conjunto aleatorio rugoso si y solo si existe una aplicación conjunto-valorada fuertemente medible $\Gamma : \Omega \rightarrow \wp(U)$ tal que $\underline{\Gamma}(\omega) = \underline{\Gamma}(\omega)$ y $\bar{\Gamma}(\omega) = \overline{\Gamma(\omega)}$, $\forall \omega \in \Omega$.

3.3. Caracterización de la probabilidad inducida por un conjunto aleatorio observado de forma incompleta

Teorema 4 *Consideremos un espacio probabilístico completo $(\Omega, \mathcal{A}, P)$ y un espacio Souslin (U, τ) . Sea $\Gamma : \Omega \longrightarrow \wp(U)$ una aplicación multi-valuada con imágenes cerradas. Sea Π una partición finita sobre U y sea G_Λ el meta-conjunto aleatorio asociado a $\Lambda = (\underline{\Gamma}, \bar{\Gamma})$. Denotemos por $\mathcal{P}(\Lambda)$ el conjunto de las posibles distribuciones de probabilidad inducidas por Γ , de acuerdo con la información proporcionada por Λ (o, equivalentemente, por G_Λ), es decir, la familia de medidas de probabilidad siguiente:*

$$\mathcal{P}(\Lambda) = \mathcal{P}(G_\Lambda) = \{P \circ H^{-1} : \underline{\Gamma}(\omega) \subseteq H(\omega) \subseteq \bar{\Gamma}(\omega), \forall \omega \in \Omega\}.$$

Entonces, el máximo y el mínimo de $\mathcal{P}(G_\Lambda)$ están caracterizados por la probabilidad conjunta de $(\underline{\Gamma}, \bar{\Gamma})$, de acuerdo con las siguientes igualdades:

$$\min \mathcal{P}(G_\Lambda)(\mathcal{C}) = P \left(\bigcup_{(A,B): \mathcal{C}_{(A,B)} \subseteq \mathcal{C}} \{\omega \in \Omega : \underline{\Gamma}(\omega) = A, \bar{\Gamma}(\omega) = B\} \right) \quad (3.3)$$

y

$$\max \mathcal{P}(G_\Lambda)(\mathcal{C}) = P \left(\bigcup_{(A,B): \mathcal{C} \subseteq \mathcal{C}_{(A,B)}} \{\omega \in \Omega : \underline{\Gamma}(\omega) = A, \bar{\Gamma}(\omega) = B\} \right). \quad (3.4)$$

Demostración: Por el teorema 3, G_Λ es fuertemente medible. Además, de acuerdo con el teorema 2 (recordemos que G_Λ es simple debido a que la partición Π es finita), $P_{G_\Lambda}^*(\mathcal{C})$ y $P_{* G_\Lambda}(\mathcal{C})$ coinciden respectivamente con el máximo y el mínimo de $\mathcal{P}(G_\Lambda)(\mathcal{C})$ para todo $\mathcal{C}$. Además, podemos comprobar fácilmente que dichas probabilidades inferior y superior coinciden

respectivamente con los términos de la parte derecha de las ecuaciones 3.3 y 3.4. $\square$

Nota 3.1 *En las ecuaciones 3.3 y 3.4 las probabilidades de las respectivas uniones no se pueden descomponer como las sumas de las probabilidades de los sucesos que se unen, dado que éstos no son, en general, disjuntos. En efecto, dados dos pares de subconjuntos exactos de U , $A \subseteq B$ y $C \subseteq D$, la intersección $\mathcal{C}_{(A,B)} \cap \mathcal{C}_{(C,D)}$ no es, en general, vacía, sino que coincide con $\mathcal{C}_{(A \cup C, B \cap D)}$.*

Ejemplo 23 *Consideremos la distribución uniforme discreta definida en $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, el universo finito $U = \{1, 2, 3, 4, 5, 6\}$ y la partición $\Pi = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$. Consideremos los conjuntos aleatorios $\Gamma_1, \Gamma_2 : \Omega \rightarrow \wp(U)$ definidos respectivamente del siguiente modo:*

$$\Gamma_1(\omega_1) = \{1, 2, 3\}, \quad \Gamma_1(\omega_2) = \{1, 2, 3, 5\}, \quad \Gamma_1(\omega_3) = \{2, 3, 4\},$$

$$\Gamma_1(\omega_4) = \{2, 3, 4, 5\}.$$

$$\Gamma_2(\omega_1) = \Gamma_2(\omega_2) = \{1, 2, 3\}, \quad \Gamma_2(\omega_3) = \Gamma_2(\omega_4) = \{2, 3, 4, 5\}.$$

Probaremos que las distribuciones de probabilidad marginales de las aproximaciones inferior y superior de ambos conjuntos aleatorios coinciden. De hecho, de acuerdo con la partición anterior, tenemos que:

$$P(\underline{\Gamma}_1 = \{1, 2\}) = P(\underline{\Gamma}_1 = \{3, 4\}) = 0,5,$$

$$\text{y } P(\bar{\Gamma}_1 = \{1, 2, 3, 4\}) = P(\bar{\Gamma}_1 = \{1, 2, 3, 4, 5, 6\}) = 0,5.$$

Además:

$$\begin{aligned}\underline{\Gamma}_1(\omega_1) = \underline{\Gamma}_1(\omega_2), \underline{\Gamma}_1(\omega_3) = \underline{\Gamma}_1(\omega_4), \bar{\Gamma}_1(\omega_1) = \bar{\Gamma}_1(\omega_3), \text{ y} \\ \bar{\Gamma}_1(\omega_2) = \bar{\Gamma}_1(\omega_4).\end{aligned}$$

Por lo tanto, $\underline{\Gamma}_1$ y $\bar{\Gamma}_1$ son estocásticamente independientes.

Por otro lado, las distribuciones de probabilidad marginales de $\underline{\Gamma}_2$ y $\bar{\Gamma}_2$ coinciden respectivamente con las de $\underline{\Gamma}_1$ y $\bar{\Gamma}_1$. Sin embargo, son funcionalmente dependientes, dado que $\underline{\Gamma}_2(\omega_1) = \underline{\Gamma}_2(\omega_2)$, $\bar{\Gamma}_2(\omega_1) = \bar{\Gamma}_2(\omega_2)$, $\underline{\Gamma}_2(\omega_3) = \underline{\Gamma}_2(\omega_4)$, $\bar{\Gamma}_2(\omega_3) = \bar{\Gamma}_2(\omega_4)$. Estas diferencias entre las distribuciones conjuntas de $(\underline{\Gamma}_1, \bar{\Gamma}_1)$ y $(\underline{\Gamma}_2, \bar{\Gamma}_2)$ tienen repercusión en el cálculo de las cotas más ajustadas para las probabilidades de la forma $P(\Gamma_i = B)$. Por ejemplo, si consideramos el conjunto $A = \{1, 2, 3\}$ obtenemos dos cotas superiores diferentes. De hecho:

$$P_{G_{\Lambda_1}}^*(\{\{1, 2, 3\}\}) = P(\underline{\Gamma}_1 = \{1, 2\}, \bar{\Gamma}_1 = \{1, 2, 3, 4\}) = 0,25, \text{ pero}$$

$$P_{G_{\Lambda_2}}^*(\{\{1, 2, 3\}\}) = P(\underline{\Gamma}_2 = \{1, 2\}, \bar{\Gamma}_2 = \{1, 2, 3, 4\}) = 0,5.$$

De acuerdo con la información disponible, el mínimo de la probabilidad de $\{1, 2, 3\}$ coincide en ambos casos, y es el valor 0. Sin embargo, si consideramos la “cadena” de conjuntos determinada por el par $(\{1, 2\}, \{1, 2, 3, 4\})$, es decir, la familia de conjuntos:

$$\mathcal{C}_{(\{1,2\},\{1,2,3,4\})} = \{\{1, 2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\}$$

entonces, también el mínimo de los posibles valores de la probabilidad $P(\Gamma_1 \in \mathcal{C}_{(\{1,2\},\{1,2,3,4\})})$ es diferente del mínimo asociado

a $P(\Gamma_2 \in \mathcal{C}_{(\{1,2\},\{1,2,3,4\})})$. En efecto, podemos comprobar fácilmente que

$$P_{*,G_{\Lambda_1}}(\mathcal{C}_{(\{1,2\},\{1,2,3,4\})}) = P(\underline{\Gamma}_1 = \{1, 2\}, \bar{\Gamma}_1 = \{1, 2, 3, 4\}) = 0,25,$$

pero $P_{*,G_{\Lambda_2}}(\mathcal{C}_{(\{1,2\},\{1,2,3,4\})}) = P(\underline{\Gamma}_2 = \{1, 2\}, \bar{\Gamma}_2 = \{1, 2, 3, 4\}) = 0,5$.

3.4. Variables aleatorias difuso rugosas como representación de la percepción imprecisa de variables aleatorias difusas

En primer lugar, vamos a extender el concepto de conjunto aleatorio rugoso, e introduciremos el concepto más general de *variable aleatoria difuso rugosa*. Siguiendo un esquema similar al de la sección 3.2, consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, un espacio medible $(U, \mathcal{A}')$ y una variable aleatoria difusa $\tilde{\Gamma} : \Omega \longrightarrow \mathcal{F}(U)$. De la misma forma que en aquella sección, supongamos además que no podemos observar directamente los elementos de U , sino que únicamente podemos observar el valor de una colección de atributos, $\mathbf{a} : U \longrightarrow V = \{v_1, \dots, v_k\}$, asignados a ellos. La función atributo induce una partición finita sobre U , $\Pi = \{A_1, \dots, A_k\}$. Sólomente somos capaces de determinar el índice $i \in \{1, \dots, k\}$ tal que $x \in A_i$, para cada elemento arbitrario $x \in U$. De acuerdo con este supuesto, únicamente podremos proporcionar las aproximaciones superior e inferior de un subconjunto difuso arbitrario de U . Para cada $\omega \in \Omega$, consideremos las aproximaciones inferior y superior de $\tilde{\Gamma}(\omega)$, $\underline{\tilde{\Gamma}}(\omega)$ y $\overline{\tilde{\Gamma}}(\omega)$, definidas, de acuerdo con la sección 3.1.4, de la forma siguiente:

$$\underline{\tilde{\Gamma}}(\omega)(x) = \inf_{y \in [x]_R} \tilde{\Gamma}(\omega)(y) \text{ y } \overline{\tilde{\Gamma}}(\omega)(x) = \sup_{y \in [x]_R} \tilde{\Gamma}(\omega)(y), \quad \forall \omega \in \Omega.$$

Denotemos ahora por $\underline{\tilde{\Gamma}}$ y $\overline{\tilde{\Gamma}}$, respectivamente, a las funciones borroso valoradas que asignan, a cada $\omega \in \Omega$, las aproximaciones inferior y superior de $\tilde{\Gamma}(\omega)$, $\underline{\tilde{\Gamma}}(\omega) := \underline{\tilde{\Gamma}(\omega)}$, $\overline{\tilde{\Gamma}}(\omega) := \overline{\tilde{\Gamma}(\omega)}$, $\forall \omega \in \Omega$. Podemos observar fácilmente que, de acuerdo con la inclusión de conjuntos borrosos, tenemos que:

$$\underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{\Gamma}(\omega) \subseteq \overline{\tilde{\Gamma}}(\omega), \quad \forall \omega \in \Omega.$$

A partir de ahora, llamaremos *aproximaciones inferior y superior* de $\tilde{\Gamma}$ a las aplicaciones $\underline{\tilde{\Gamma}}$ y $\overline{\tilde{\Gamma}}$. Denotaremos por $\tilde{\Lambda} = (\underline{\tilde{\Gamma}}, \overline{\tilde{\Gamma}})$ a función que asigna, a cada $\omega \in \Omega$, el conjunto rugoso difuso $(\underline{\tilde{\Gamma}}(\omega), \overline{\tilde{\Gamma}}(\omega))$.

Definición 34 *Consideremos un espacio probabilístico $(\Omega, \mathcal{A}, P)$ y un espacio medible $(U, \mathcal{A}')$. Sea Π una partición finita sobre el conjunto U . Una variable aleatoria difuso rugosa es un par ordenado $(\underline{\tilde{\Gamma}}, \overline{\tilde{\Gamma}}) : \Omega \longrightarrow \mathcal{F}(U) \times \mathcal{F}(U)$ que satisface:*

- *Existe una aplicación borroso-valuada $\tilde{\Gamma} : \Omega \longrightarrow U$ tal que $\underline{\tilde{\Gamma}}(\omega)$ y $\overline{\tilde{\Gamma}}(\omega)$ determinan las aproximaciones inferior y superior de $\tilde{\Gamma}(\omega)$ con respecto a Π , para cada $\omega \in \Omega$.*
- *Los α -cortes de $\underline{\tilde{\Gamma}}$ y $\overline{\tilde{\Gamma}}$ son $\mathcal{A} - \mathcal{A}'$ fuertemente medibles.*

De acuerdo con el teorema 3, si el espacio probabilístico inicial, $(\Omega, \mathcal{A}, P)$, es completo, el espacio final, (U, τ) , es Souslin y los α -cortes de las imágenes de $\tilde{\Gamma}$ son subconjuntos cerrados de U , entonces la medibilidad fuerte de las aproximaciones inferior

y superior de los α -cortes de $\tilde{\Gamma}$ se deducen de la medibilidad fuerte de los respectivos α -cortes de $\tilde{\Gamma}$. Además, los α -cortes de las aproximaciones coinciden, trivialmente, con las aproximaciones de los α -cortes. De esta forma deducimos que, si $\tilde{\Gamma}$ es una variable aleatoria difusa bajo las condiciones mencionadas (espacio probabilístico inicial completo, espacio final Souslin y α -cortes cerrados), entonces también lo son las funciones difuso-valoradas $\underline{\tilde{\Gamma}}$ y $\overline{\tilde{\Gamma}}$. De esta forma, tiene sentido considerar la pareja de distribuciones de probabilidad inducidas, así como la distribución de probabilidad conjunta. Al igual que en el caso de los conjuntos aleatorios difusos, hablaremos del meta conjunto asociado a la variable aleatoria difuso rugosa:

Definición 35 Sean $(\Omega, \mathcal{A}, P)$ un espacio probabilístico y $(U, \mathcal{A}')$ un espacio medible. Consideremos además una partición finita Π sobre U y una variable aleatoria difuso rugosa $\tilde{\Lambda} = (\underline{\tilde{\Gamma}}, \overline{\tilde{\Gamma}})$. La aplicación $G_{\tilde{\Lambda}} : \Omega \longrightarrow \wp(\mathcal{F}(U))$ definida como sigue:

$$G_{\tilde{\Lambda}}(\omega) = \{\tilde{C} \in \mathcal{F}(U) : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{C} \subseteq \overline{\tilde{\Gamma}}(\omega)\}.$$

se llamará meta-conjunto aleatorio asociado a $\tilde{\Lambda}$.

3.4.1. Caracterización de la probabilidad inducida por una variable aleatoria difusa

Teniendo en cuenta que la variable aleatoria difusa $\tilde{\Gamma}$ se puede ver como una selección $\mathcal{A} - \sigma_{\mathcal{F}}$ medible de $G_{\tilde{\Lambda}}$, y teniendo en cuenta que las probabilidades superior e inferior de Dempster de un conjunto aleatorio acotan a las medidas de probabilidad de sus selecciones medibles, podemos deducir que, la probabilidad

asignada por $\tilde{\Gamma}$ a un subconjunto arbitrario $\mathcal{C} \subseteq \Gamma(U)$ está acotada por los valores siguientes, siempre que la función $G_{\tilde{\Lambda}}$ sea $\mathcal{A} - \sigma_{\mathcal{F}}$ fuertemente medible:

$$P_{*G_{\tilde{\Lambda}}}(\mathcal{C}) = P\left(\bigcup_{(\tilde{A}, \tilde{B}) : \mathcal{C}_{(\tilde{A}, \tilde{B})} \subseteq \mathcal{C}} \{\omega \in \Omega : \underline{\tilde{\Gamma}}(\omega) = \tilde{A}, \bar{\tilde{\Gamma}}(\omega) = \tilde{B}\}\right) \quad (3.5)$$

y

$$P_{G_{\tilde{\Lambda}}}^*(\mathcal{C}) = P\left(\bigcup_{(\tilde{A}, \tilde{B}) : \mathcal{C} \subseteq \mathcal{C}_{(\tilde{A}, \tilde{B})}} \{\omega \in \Omega : \underline{\tilde{\Gamma}}(\omega) = \tilde{A}, \bar{\tilde{\Gamma}}(\omega) = \tilde{B}\}\right). \quad (3.6)$$

De acuerdo con los resultados de la sección anterior, dichas cotas coinciden con el máximo y el mínimo en algunos casos particulares, como los que veremos a continuación.

3.4.2. Algunos casos especiales

La variable aleatoria difusa es simple

Si la variable aleatoria difusa $\tilde{\Gamma}$ es simple, entonces también lo son sus aproximaciones inferior y superior, $\underline{\tilde{\Gamma}}$ y $\bar{\tilde{\Gamma}}$, y por tanto, también lo es la aplicación multivaluada $G_{\tilde{\Lambda}}$ que, a cada $\omega \in \Omega$, le asigna la “cadena”:

$$\mathcal{C}_{(\underline{\tilde{\Gamma}}(\omega), \bar{\tilde{\Gamma}}(\omega))} = \{\tilde{C} \in \mathcal{F}(U) : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{C} \subseteq \bar{\tilde{\Gamma}}(\omega)\}.$$

(Observe el lector que las imágenes de $G_{\tilde{\Lambda}}$ son familias de subconjuntos borrosos de U y que se entiende, entonces, como aplicación multivaluada entre Ω y $\mathcal{F}(U)$). Así, de acuerdo con el teorema 2, las probabilidades superior e inferior de $G_{\tilde{\Lambda}}$ coinciden con el máximo y el mínimo, respectivamente, del conjunto

$\mathcal{P}(G_{\tilde{\Lambda}})$, para cualquier familia de subconjuntos difusos de U . La familia de distribuciones de probabilidad sobre $\sigma_{\mathcal{F}}$, $\mathcal{P}(G_{\tilde{\Lambda}})$, determina, por otra parte, el conocimiento disponible acerca de la distribución de probabilidad inducida por la variable aleatoria borrosa $\tilde{\Gamma}$.

El número de α -cortes distintos es finito

En este caso, a pesar de que $\tilde{\Gamma}$ no sea, necesariamente, una función simple (es decir, con un número finito de imágenes distintas), sí lo son, sin embargo, sus aproximaciones inferior y superior. Esto se puede deducir del hecho de que los α -cortes de dichas aproximaciones lo son, tal como hemos visto en la sección 3.2, y del hecho de que el número de α -cortes distintos es finito. De esta forma, deducimos que la aplicación multivaluada $G_{\tilde{\Lambda}}$ es, también, simple, y entonces sus probabilidades inferior y superior determinan el máximo y el mínimo del conjunto buscado.

El espacio final es finito

En este apartado veremos que las probabilidades superior e inferior de $G_{\tilde{\Lambda}}$ también determinan el máximo y el mínimo de los posibles valores para la distribución de probabilidad inducida por la variable aleatoria difusa. Para ello, tendremos en cuenta la siguiente construcción. Supongamos que el universo tiene n elementos, que denotaremos por $U = \{x_1, \dots, x_n\}$. Consideremos la función biyectiva $f : \wp(U) \longrightarrow \{0, 1\}^n$ definida de la forma siguiente $f(A) := (I_A(x_1), \dots, I_A(x_n))$, donde I_A denota la función indicador asociada al conjunto A . A continuación, estudiaremos la composición de la función f con algunas de las

aplicaciones multivaluadas consideradas en las secciones precedentes. De ahora en adelante, se utilizará el símbolo “ $\circ$ ” para denotar la composición entre funciones.

Lema 4 *Si $\Gamma : \Omega \longrightarrow \wp(U)$ es fuertemente medible, entonces $f \circ \Gamma : \Omega \longrightarrow \{0, 1\}^n$ es $\mathcal{A}$ - $\wp(\{0, 1\}^n)$ medible. Además, la distribución de probabilidad de $f \circ \Gamma$ determina unívocamente a la distribución de probabilidad inducida por Γ .*

Demostración: Puesto que el espacio final de la función $f \circ \Gamma$ es finito, es suficiente comprobar que la anti-imagen de cualquier elemento de $\{0, 1\}^n$ es medible. Elijamos un vector arbitrario, $\vec{v} \in \{0, 1\}^n$. Resulta fácil comprobar que $(f \circ \Gamma)^{-1}(\{\vec{v}\})$ coincide con $\Gamma^{-1}(A_{\vec{v}})$, donde $A_{\vec{v}}$ es el conjunto formado por aquellos elementos x_i tales que la correspondiente componente v_i es igual a 1. La tesis del lema se deduce fácilmente, si tenemos en cuenta el resultado del teorema 2, y el hecho de que la aplicación multivaluada Γ es simple (dado que el espacio de llegada es finito, también lo es la familia de sus subconjuntos). $\square$

De acuerdo con el lema anterior, y teniendo en cuenta que U es un conjunto finito, si $(\Omega, \mathcal{A}, P)$ es completo, tenemos que las funciones multi-valuadas $\underline{\Gamma}$ y $\bar{\Gamma}$ son fuertemente medibles y, por tanto, $f \circ \underline{\Gamma}$ y $f \circ \bar{\Gamma}$ son también $\mathcal{A}$ - $\mathcal{A}'$ medibles.

Ejemplo 24 *Consideremos el conjunto inicial $\Omega = \{1, 2, 3\}$ y la distribución de Laplace definida sobre la σ -álgebra discreta, $\wp(\Omega)$ (la distribución de probabilidad uniforme discreta), a la cual denotaremos por P . Consideremos también el conjunto*

$U = \{1, 2, 3, 4, 5, 6\}$ y la partición $\Pi = \{\{1, 2, 3\}, \{4\}, \{5, 6\}\}$. Consideremos el conjunto aleatorio $\Gamma : \Omega \longrightarrow \wp(U)$ definido de la forma siguiente:

$$\Gamma(\omega_1) = \{2, 3, 4\}, \Gamma(\omega_2) = \{4, 5, 6\}, \Gamma(\omega_3) = \{1, 3, 4, 5\}.$$

Las aproximaciones inferior y superior de Γ vienen dadas por las siguientes expresiones:

$$\underline{\Gamma}(\omega_1) = \{4\}, \underline{\Gamma}(\omega_2) = \{4, 5, 6\}, \underline{\Gamma}(\omega_3) = \{4\} \text{ y}$$

$$\overline{\Gamma}(\omega_1) = \{1, 2, 3, 4\}, \overline{\Gamma}(\omega_2) = \{4, 5, 6\}, \overline{\Gamma}(\omega_3) = \{1, 2, 3, 4, 5, 6\}.$$

Por otra parte las respectivas composiciones de f con las aplicaciones multi-valuadas anteriores son las siguientes:

$$f \circ \Gamma(\omega_1) = (0, 1, 1, 1, 0, 0), f \circ \Gamma(\omega_2) = (0, 0, 0, 1, 1, 1),$$

$$f \circ \Gamma(\omega_3) = (1, 0, 1, 1, 1, 0),$$

$$f \circ \underline{\Gamma}(\omega_1) = (0, 0, 0, 1, 0, 0), f \circ \underline{\Gamma}(\omega_2) = (0, 0, 0, 1, 1, 1),$$

$$f \circ \underline{\Gamma}(\omega_3) = (0, 0, 0, 1, 0, 0),$$

$$f \circ \overline{\Gamma}(\omega_1) = (1, 1, 1, 1, 0, 0), f \circ \overline{\Gamma}(\omega_2) = (0, 0, 0, 1, 1, 1),$$

$$f \circ \overline{\Gamma}(\omega_3) = (1, 1, 1, 1, 1, 1).$$

De acuerdo con la relación de orden parcial $\leq$ definida en $\{0, 1\}^n$ de la forma $\vec{v} \leq \vec{w}$ si y solo si $v_i \leq w_i, \forall i = 1, \dots, n$, podemos observar que, en efecto,

$$f \circ \underline{\Gamma}(\omega_i) \leq f \circ \Gamma(\omega_i) \leq f \circ \overline{\Gamma}(\omega_i), \quad i = 1, 2, 3.$$

Sea G es la aplicación multivaluada que, a cada ω_i , le asigna el siguiente conjunto de vectores de $\{0, 1\}^4$:

$$G(\omega_i) = \{\vec{v} : f \circ \underline{\Gamma}(\omega_i) \leq \vec{v} \leq f \circ \bar{\Gamma}(\omega_i)\}.$$

Es sencillo comprobar que $G = f \circ G_\Lambda$, donde $\Lambda = (\underline{\Gamma}, \bar{\Gamma})$.

Extendamos ahora los conceptos y resultados anteriores desde el contexto de los conjuntos aleatorios al contexto de las variables aleatorias borrosas, reemplazando la función indicador de un conjunto nítido por la función de pertenencia de un conjunto difuso. Así, identificaremos cada conjunto difuso, $\tilde{C} \in \mathcal{F}(U)$ por un vector $f(\tilde{C}) \in [0, 1]^n$, donde $f : \mathcal{F}(U) \rightarrow [0, 1]^n$ asocia, a cada $\tilde{C} \in \mathcal{F}(U)$, el vector $(\tilde{C}(x_1), \dots, \tilde{C}(x_n))$. Así, cada variable aleatoria borrosa $\tilde{\Gamma}$ será identificada con un vector aleatorio, cuyas imágenes pertenecen al conjunto $[0, 1]^n$ en lugar de al conjunto $\{0, 1\}^n$.

Extendemos a continuación el resultado del lema 4 a este caso más general en el que consideramos variables aleatorias borrosas, en lugar de conjuntos aleatorios.

Lema 5 *Si los α -cortes de $\tilde{\Gamma} : \Omega \rightarrow \mathcal{F}(U)$ son fuertemente medibles, entonces $f \circ \tilde{\Gamma} : \Omega \rightarrow [0, 1]^n$ es $\mathcal{A}$ - $\beta_{[0,1]^n}$ medible, donde $\beta_{[0,1]^n}$ denota la σ -álgebra de Borel habitual restringida al rectángulo unidad. Además, la distribución de probabilidad de $f \circ \tilde{\Gamma}$ determina unívocamente a la distribución de probabilidad inducida por $\tilde{\Gamma}$.*

Demostración: Para comprobar la medibilidad de la función $f \circ \tilde{\Gamma}$, es suficiente comprobar que la anti-imagen de cualquier rectángulo de la forma $[\vec{0}, \vec{v}) = \{\vec{y} \in [0, 1]^n : y_i < v_i, i =$

$1, \dots, n\}$ es medible. Por otra parte, para un vector arbitrario $\vec{v} \in [0, 1]^n$, tenemos que $[\vec{0}, \vec{v}) = \cap_{i=1}^n [\vec{0}, \vec{v}_i)$, donde $\vec{v}_i$ denota el vector cuyas componentes son unos, excepto para la componente i -ésima, la cual coincide con la componente i -ésima de v . Teniendo en cuenta lo anterior, resulta suficiente comprobar que la anti-imagen de cualquier rectángulo de la forma $[\vec{0}, \vec{v}_i)$ es medible. Podemos observar que

$$f \circ \tilde{\Gamma}^{-1}([\vec{0}, \vec{v}_i)) = \{\omega \in \Omega : \tilde{\Gamma}(\omega)(x_i) < v_i\} = [\tilde{\Gamma}_{v_i}^*]^c(\{x_i\}).$$

Teniendo en cuenta que, por hipótesis, todas las aplicaciones multi-valuadas de la forma $\tilde{\Gamma}_{v_i}$ son fuertemente medibles, la tesis del resultado queda probada. $\square$

Consideremos ahora la aplicación multivaluada $G = f \circ G_{\tilde{\Lambda}} : \Omega \longrightarrow \wp([0, 1]^n)$ que, a cada ω , le asigna el conjunto de vectores $G(\omega) = f(G_{\tilde{\Lambda}}(\omega)) = \{f(\tilde{E}) : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{E} \subseteq \overline{\tilde{\Gamma}}(\omega)\}$.

Lema 6 *Denotemos por $\leq$ el orden parcial definido en $[0, 1]^n$ de la forma: $\vec{v} \leq \vec{w}$ si y solo si $v_i \leq w_i, \forall i = 1, \dots, n$. Entonces,*

$$G(\omega) = \{\vec{v} : f(\underline{\tilde{\Gamma}}(\omega)) \leq \vec{v} \leq f(\overline{\tilde{\Gamma}}(\omega))\}, \forall \omega \in \Omega.$$

Demostración: Se puede comprobar fácilmente que, para dos subconjuntos borrosos arbitrarios del universo, $\tilde{A}, \tilde{B} \in \mathcal{F}(U)$, se cumple la relación $\tilde{A} \subseteq \tilde{B}$ si y solo si $f(\tilde{A}) \leq f(\tilde{B})$, donde “ $\leq$ ” denota la relación de orden parcial señalada en el enunciado. La tesis del lema se deduce de forma inmediata de este hecho. $\square$

Lema 7 *Si los α -cortes de $\tilde{\Gamma}$ son fuertemente medibles, entonces $G_{\tilde{\Lambda}}$ es fuertemente medible.*

Demostración: En estas condiciones, dado que el espacio final es finito, se puede comprobar fácilmente, haciendo uso del teorema 3 (y teniendo en cuenta que se verifican las condiciones del mismo: espacio probabilístico completo y espacio final Souslin, Γ con imágenes cerradas y Π , partición finita), que los α -cortes de $\underline{\tilde{\Gamma}}$ y de $\overline{\tilde{\Gamma}}$ también son fuertemente medibles. Así, de acuerdo con el lema 4, sus composiciones con f son vectores aleatorios (medibles). Es un resultado conocido que cualquier rectángulo aleatorio cerrado en $\mathbb{R}^n$ es fuertemente medible si sus extremos son vectores aleatorios medibles. Teniendo en cuenta esta circunstancia, por el lema anterior y por ser f biyectiva, deducimos que $G_{\tilde{\Lambda}}$ es fuertemente medible. $\square$

Lema 8 *Consideremos un espacio de probabilidad $(\Omega, \mathcal{A}, P)$ y un espacio topológico Polaco (U, τ) . Sea $H : \Omega \longrightarrow \wp(U)$ una aplicación multi-valuada fuertemente medible definida entre ellos. Si sus imágenes son cerradas, entonces:*

$$P_H^*(A) = \max \mathcal{P}(H)(A) \text{ y } P_{*H}(A) = \min \mathcal{P}(H)(A), \quad \forall A \in \beta_\tau.$$

De todo lo anterior, deducimos que las probabilidades superior e inferior de Dempster generadas por la aplicación multi-valuada $G_{\tilde{\Lambda}}$ determinan el máximo y el mínimo para la distribución de probabilidad inducida por $\tilde{\Gamma}$ en $\sigma_{\mathcal{F}}$, de acuerdo con la información rugosa disponible.

Teorema 5 *Para cualquier subconjunto arbitrario $\mathcal{C} \in \sigma_{\mathcal{F}}$, las probabilidades superior e inferior inducidas por $G_{\tilde{\Lambda}}$ determinan el máximo y el mínimo del conjunto de valores que el conjunto de medidas de probabilidad:*

$$\{P \circ \tilde{H}^{-1} : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{H}(\omega) \subseteq \overline{\tilde{\Gamma}}(\omega), \forall \omega \in \Omega\} \quad (3.7)$$

asigna al conjunto $\mathcal{C}$.

Demostración: De acuerdo con los resultados anteriores, sabemos que $G_{\tilde{\Lambda}}$ es fuertemente medible. Además, sus imágenes son subconjuntos cerrados de la topología usual en $[0, 1]^n$, que determina un espacio polaco. Así, de acuerdo con el lema 8, sus probabilidades superior e inferior determinan el máximo y el mínimo del conjunto de probabilidades $\mathcal{P}(G_{\tilde{\Lambda}})$, que, a su vez, coincide con el conjunto determinado en la ecuación 3.7. $\square$

Las imágenes de la variable aleatoria difusa tienen funciones de pertenencia continuas en un intervalo cerrado

Denotemos por $[a, b]$ al soporte de las funciones de pertenencia. En este caso, podemos construir una función biyectiva que asigne, a cada subconjunto difuso, su función de pertenencia, que no es más que una función continua de $[a, b]$, en $[0, 1]$. Utilizando esta identificación, podemos construir un conjunto aleatorio, $G_{\tilde{\Lambda}}$, cuyas imágenes son subconjuntos cerrados de funciones, y que determina nuestro conocimiento acerca de los valores de $\tilde{\Gamma}$. Se puede comprobar que esta función es fuertemente medible y que, por tanto, de acuerdo con el resultado del lema 8, determina los valores máximo y mínimo de los posibles valores de probabilidad asociados a $P \circ \tilde{\Gamma}^{-1}$.

3.5. Caracterización del centroide y la dispersión esperada de conjuntos aleatorios y variables aleatorias difusas

En las secciones anteriores, hemos tratado de caracterizar la información disponible acerca de la distribución de probabilidad inducida por un conjunto aleatorio o, en un contexto más general, de una variable aleatoria difusa a partir de la información proporcionada por sus respectivas aproximaciones inferior y superior. Hemos visto que, bajo ciertas condiciones, el máximo y el mínimo de los posibles valores para la probabilidad de una familia de subconjuntos (difusos) $\mathcal{C} \subset \mathcal{F}(U)$ vienen determinados por la distribución de probabilidad conjunta del par $(\tilde{\Gamma}, \bar{\Gamma})$, aunque no por medio de las distribuciones de probabilidad marginales. En esta sección nos proponemos caracterizar la información disponible acerca del centroide o la desigualdad esperada de la variable aleatoria difusa, si disponemos de información granularizada, como en las secciones precedentes. Parece obvio que, de acuerdo con la información imprecisa disponible, nuestro conocimiento acerca del centroide de $\tilde{\Gamma}$ viene dado por el siguiente conjunto (interpretado desde un punto de vista disyuntivo):

$$\{C_d(\tilde{H}) : \tilde{\Gamma}(\omega) \subseteq \tilde{H}(\omega) \subseteq \bar{\Gamma}(\omega), \forall \omega \in \Omega\}. \quad (3.8)$$

En el capítulo anterior veíamos que, para determinadas medidas de desigualdad, d , el centroide no mantenía la propiedad de la inclusión, en el sentido de que $\tilde{H}(\omega) \subseteq \tilde{H}'(\omega)$, $\forall \omega \in \Omega$ no implicaba que el centroide de H estuviese incluido en el centroide de H' . Teniendo en cuenta este hecho, el conjunto mostrado en la ecuación 3.8 no está acotado, en general, por los respectivos

centroides la pareja de aproximaciones de $\tilde{\Gamma}$.

Veamos a continuación un ejemplo:

Ejemplo 25 *Consideremos la distancia d y el espacio probabilístico $(\Omega, \wp(\Omega), P)$ utilizados anteriormente en el ejemplo 18. Sea R una relación de equivalencia que divide al intervalo $[0, 1]$ en 10 clases: $[0, 0,1)$, $[0,1, 0,2]$, $(0,2, 0,3]$, $\dots$, $(0,8, 0,9)$, $[0,9, 1]$. Sea Γ el conjunto aleatorio dado por $\Gamma(\omega_1) = [0,05, 0,35]$ y $\Gamma(\omega_2) = [0,92, 0,96]$. El centroide de este conjunto aleatorio con respecto a d es $C_d(\Gamma) = [0,485, 0,655]$.*

Por otro lado, la aproximación inferior de Γ viene dada por $\underline{\Gamma}(\omega_1) = [0,1, 0,3]$ y $\underline{\Gamma}(\omega_2) = \emptyset$. En este caso, el centroide respecto a d es $C_d(\underline{\Gamma}) = [0,1, 0,3]$. Así pues, los centroides del conjunto aleatorio y de su aproximación inferior tienen intersección vacía.

En el caso de la d -dispersión, nos encontramos con una situación similar. Todo lo que conocemos acerca de la d -dispersión de la variable aleatoria borrosa $\tilde{\Gamma}$ es que pertenece al siguiente conjunto de valores:

$$\{V_d(\tilde{H}) : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{H}(\omega) \subseteq \overline{\tilde{\Gamma}}(\omega), \forall \omega \in \Omega\}. \quad (3.9)$$

En el ejemplo siguiente, veremos que las d -dispersiones de las aproximaciones inferior y superior de $\tilde{U}$ no coinciden con el mínimo o el máximo del conjunto de valores mostrado en la ecuación 3.9.

Ejemplo 26 *Consideremos de nuevo el conjunto aleatorio del ejemplo 25 y calculemos las d -dispersiones del conjunto y su*

aproximación inferior. Tenemos que

$$V_d(\Gamma) = 2d(\Gamma(\omega_1), \Gamma(\omega_2))P(\{\omega_1\})P(\{\omega_2\}) = \\ \frac{1}{2}d([0,05, 0,35], [0,92, 0,96]) = \frac{1}{2} \cdot \sqrt{1,129}.$$

Por otro lado, la d-dispersión de su aproximación inferior es

$$V_d(\underline{\Gamma}) = 2d(\underline{\Gamma}(\omega_1), \underline{\Gamma}(\omega_2))P(\{\omega_1\})P(\{\omega_2\}) = \\ \frac{1}{2}d([0,1, 0,3], \emptyset) = \frac{1}{2} \cdot \sqrt{2}.$$

En cuanto a la aproximación superior, viene dada por $\bar{\Gamma}(\omega_1) = [0, 0,4]$ y $\bar{\Gamma}(\omega_2) = [0,9, 1]$; su d-variación es

$$V_d(\bar{\Gamma}) = 2d(\bar{\Gamma}(\omega_1), \bar{\Gamma}(\omega_2))P(\{\omega_1\})P(\{\omega_2\}) = \\ \frac{1}{2}d([0, 0,4], [0,9, 1]) = \frac{1}{2} \cdot \sqrt{1,17}.$$

Por tanto, la d-dispersión del conjunto aleatorio es estrictamente menor que la d-dispersión de cualquiera sus aproximaciones inferior.

Veamos otro ejemplo. Si consideramos el conjunto aleatorio Γ' dado por $\Gamma'(\omega_1) = [0,09, 0,48]$ y $\Gamma'(\omega_2) = [0,05, 0,42]$, su aproximación inferior es un conjunto aleatorio constante con imagen $[0,2, 0,4]$. Es obvio que en tal caso la d-dispersión de la aproximación inferior es 0, y que la del conjunto aleatorio es estrictamente mayor que ella. De hecho, en este caso, también la aproximación superior es constante, con valor $[0, 0,5]$, y por tanto también la d-variación de esta aproximación es 0, estrictamente menor que la del conjunto aleatorio Γ' .

En resumen, no existe una relación de orden entre las d -dispersiones de un conjunto aleatorio y sus aproximaciones.

En general, nuestra información acerca de cualquier parámetro, $\theta_{\tilde{\Gamma}}$, que dependa de la distribución de probabilidad inducida por $\tilde{\Gamma}$ (el valor de probabilidad asignado a una familia de difusos concreta, la esperanza, la varianza, la entropía, etc.) se puede expresar en términos del conjunto:

$$\{\theta_H : \underline{\tilde{\Gamma}}(\omega) \subseteq \tilde{H}(\omega) \subseteq \overline{\tilde{\Gamma}}(\omega), \forall \omega \in \Omega\}.$$

Sin embargo, el método para caracterizar explícitamente dicho conjunto en términos de un parámetro análogo que dependa de las distribuciones marginales de $\underline{\tilde{\Gamma}}$ y $\overline{\tilde{\Gamma}}$ o de su distribución conjunta puede resultar más o menos complejo.

3.6. Grado de rugosidad de los conjuntos aleatorios y las variables aleatorias difusas

En las secciones precedentes hemos mostrado la forma de caracterizar la información disponible acerca de la distribución de probabilidad inducida (o acerca de algún parámetro asociado a la misma) por un conjunto aleatorio o, de forma más general, por una variable aleatoria difusa, cuando no tenemos la posibilidad de observar directamente sus imágenes, dado que la información acerca de los elementos del universo está “granularizada”. Es evidente que, si conseguimos “refinar” la información acerca de los elementos del universo U , utilizando una partición Π' más fina que la partición original (debido al hecho de poder observar una colección más amplia de atributos, por ejemplo), nuestra

información acerca de la distribución de probabilidad inducida por $\tilde{\Gamma}$ o acerca de cualquiera de sus parámetros se hará más precisa.

En esta sección nos proponemos encontrar una herramienta que nos permita cuantificar el grado de “imprecisión” o “inexactitud” de la información proporcionada por un conjunto aleatorio rugoso o por una variable aleatoria difuso rugosa. Una primera opción consiste en comparar la entropía del conjunto aleatorio original con la de cualquiera de sus aproximaciones (inferior o superior). Es evidente que varias imágenes distintas del conjunto aleatorio pueden dar lugar a la misma aproximación inferior o superior. Sin embargo, el recíproco no es cierto: el hecho de que las aproximaciones de dos imágenes coincidan, no implica que éstas lo hagan. De esta forma, resulta evidente que la entropía de cualquiera de las dos aproximaciones del conjunto aleatorio es siempre inferior o igual a la entropía original. En primer lugar, investigaremos esa vía a la hora de cuantificar el grado de “inexactitud” de un conjunto aleatorio. Mostraremos algunos inconvenientes importantes que nos harán decantarnos por otra vía alternativa, basada en un concepto introducido por Pawlak en su libro de 1991. En un contexto finito, Pawlak ([82]) introdujo el concepto de *exactitud* de un conjunto X con respecto a la relación de equivalencia R con el objetivo de cuantificar el grado de precisión de nuestro conocimiento (determinado por R) acerca de X . El autor lo denotó por $\alpha_R(X)$ y lo definió como el cociente entre los cardinales de sus aproximaciones inferior y superior. Obviamente, $0 \leq \alpha_R(X) \leq 1$ y $\alpha_R(X) = 1$ indica que X es un conjunto exacto o preciso. En otro caso, la región frontera de X es no vacía y X es un conjunto rugoso.

Pawlak introdujo así mismo la noción opuesta de *R-rugosidad* o simplemente *rugosidad* de un conjunto X , con el propósito de cuantificar el grado de imprecisión o incompletitud del conocimiento acerca de X proporcionado por R . El autor lo define como $\rho_R(X) = 1 - \alpha_R(X)$. En [102], Yao menciona que dicha medida de rugosidad de Pawlak puede expresarse como la MZ-distancia entre la aproximación inferior y la superior de X con respecto a la relación R . De acuerdo con esto, y teniendo en cuenta nuestros estudios previos acerca de las nociones de medida de “distancia”, “disimilaridad” y “divergencia” en la teoría de conjuntos (difusos), proponemos una generalización del concepto de Pawlak, y estudiamos las propiedades de esta definición general, no solo para el caso de los conjuntos nítidos definidos en universos finitos, sino para el caso general de conjuntos borrosos definidos en espacios arbitrarios.

Dentro de esta sección, utilizaremos una nomenclatura diferente a la utilizada hasta ahora para referirnos a las aproximaciones superior e inferior de los conjuntos, de manera que se muestre explícitamente la relación de equivalencia R a la que se refieren estas aproximaciones. Esto no era necesario en las secciones anteriores, dado que siempre se consideraba la misma relación de equivalencia o partición. Sin embargo, en esta sección, hablaremos de refinamientos de particiones, de forma que convivirán en una misma expresión matemática las referencias a las aproximaciones (superiores o inferiores) con respecto a diferentes relaciones de equivalencia definidas sobre el mismo universo U .

Sea $R : U \times U \longrightarrow \{0, 1\}$ una relación de equivalencia (tam-

bién llamada *relación de indiscernibilidad*) definida sobre un universo U . Sea C un subconjunto arbitrario de U .

- La aproximación inferior de C con respecto a R será denotada de la forma $\underline{R}(C) = \{x \in U : [x]_R \subseteq C\}$.
- La aproximación superior de C con respecto a R será denotada de la forma $\overline{R}(C) = \{x \in U : [x]_R \cap C \neq \emptyset\}$.

3.6.1. La diferencia de entropías como grado de “inexactitud”

Como ya hemos comentado, una primera opción para medir el grado de “inexactitud” de la información proporcionada por una aproximación consistiría en comparar la entropía del conjunto aleatorio original con el de cualquiera de sus aproximaciones. Denotemos por $H(\Gamma)$ la entropía de la distribución de probabilidad asociada al conjunto aleatorio Γ . Dado que $\overline{\Gamma}$ y $\underline{\Gamma}$ son funciones de Γ , se verifica que $H(\Gamma) \geq \max\{H(\overline{\Gamma}), H(\underline{\Gamma})\}$. En particular, si Γ es simple y F_Γ es el conjunto de sus focales, entonces $\#F_{\underline{\Gamma}} \leq \#F_\Gamma$ y $\#F_{\overline{\Gamma}} \leq \#F_\Gamma$. Así, podemos definir las diferencias $d_1 = H(\Gamma) - H(\overline{\Gamma})$ y $d_2 = H(\Gamma) - H(\underline{\Gamma})$. Estas medidas aportan información acerca de la falta de exactitud de las imágenes del conjunto aleatorio: d_1 tiene en cuenta el número de veces que coinciden las aproximaciones inferiores de las imágenes de Γ , y d_2 por su parte tiene en cuenta el número de veces que coinciden las aproximaciones superiores. Sin embargo, no tienen en cuenta cuán distintas son dos imágenes de Γ si sus aproximaciones inferiores o superiores coinciden.

Por otro lado, no existe una relación de desigualdad concreta entre la entropía de la aproximación superior y la entropía de

la inferior. Podemos encontrar sencillos ejemplos en los que la primera es mayor que la segunda y viceversa, como vemos a continuación.

Ejemplo 27 *Consideremos el espacio probabilístico $(\Omega, \wp(\Omega), P)$, donde $\Omega = \{\omega_1, \omega_2\}$ y P es la probabilidad dada por $P(\{\omega_i\}) = \frac{1}{2}$, $i = 1, 2$. Sea U el universo en estudio, el cual se encuentra dividido en clases de equivalencia por una relación de indiscernibilidad. Sea $\Pi = \{C_1, C_2, \dots, C_n\}$ la partición correspondiente. Sea $\Gamma : \Omega \rightarrow \wp(U)$ el conjunto aleatorio dado por: $\Gamma(\omega_1) = C_1$, $\Gamma(\omega_2) = C_1 \cup B$, donde B es un subconjunto de U contenido estrictamente en C_2 . De este modo, $\underline{\Gamma}$ toma el valor constante C_1 ; $\bar{\Gamma}$ viene dado por $\bar{\Gamma}(\omega_1) = C_1$ y $\bar{\Gamma}(\omega_2) = C_1 \cup C_2$. Las entropías de cada uno de estos conjuntos aleatorios son $H(\underline{\Gamma}) = 0$ bit y $H(\Gamma) = H(\bar{\Gamma}) = 1$ bit. Por tanto, en este caso, $H(\underline{\Gamma}) < H(\bar{\Gamma})$. Sin embargo, para el conjunto aleatorio $\Gamma'(\omega_1) = C_1 \cup B$ y $\Gamma'(\omega_2) = C_1 \cup C_2$ tenemos la desigualdad contraria entre las entropías de sus aproximaciones.*

Así pues, si nos planteamos una medida del grado de inexactitud basada en las entropías de las dos aproximaciones del conjunto, dicha medida debería ser la diferencia en valor absoluto. Sin embargo, esta no es una opción adecuada. Dicha diferencia es obviamente 0 en el caso en el que Γ toma valores exactos, pero también puede serlo en el caso en el que no sea así, y por tanto esta diferencia de entropías no mide correctamente el grado de inexactitud. Basta tomar como ejemplo el caso en el que la partición del universo esté formada únicamente por el conjunto total U . En tal caso, las aproximaciones inferior y superior de cualquier conjunto aleatorio son constantes, tomando como valores el conjunto vacío y U respectivamente, de modo que am-

bas tienen entropía 0 y por tanto su diferencia también es 0. Esta medida nos indicaría por tanto que el grado de inexactitud de cualquier conjunto aleatorio en esas circunstancias es 0, lo cual lógicamente no es correcto. Sin llegar al extremo de tener una partición formada por un único elemento, tenemos también ejemplos que demuestran la no idoneidad de esta medida. Consideremos, por ejemplo, el conjunto $[0, 1]$, la partición y el conjunto aleatorio Γ' expuestos en el ejemplo 26. De nuevo, las aproximaciones inferior y superior de Γ' son constantes, de modo que ambas entropías (y por tanto su diferencia) tienen valor 0, mientras que las imágenes de Γ' no son conjuntos exactos de la partición.

Retomemos el análisis de las diferencias d_1 y d_2 . Éstas verifican dos propiedades importantes como candidatos a representar el grado de rugosidad. Dichas propiedades se enumeran en la siguiente proposición.

Proposición 17 *Las medidas d_1 y d_2 verifican las siguientes propiedades:*

1. *Son nulas si y sólo si el conjunto aleatorio en estudio tiene imágenes exactas, es decir, si son unión de clases de equivalencia. De hecho, las siguientes afirmaciones son equivalentes:*

- a) $H(\Gamma) = H(\underline{\Gamma})$.
- b) $\Gamma(\omega)$ es exacto c.s.(P).
- c) *Consideradas como función medible, la distribución de Γ , la de $\underline{\Gamma}$ y la de $\bar{\Gamma}$ son idénticas.*
- d) $H(\Gamma) = H(\bar{\Gamma})$.

2. *Disminuyen a medida que la partición se hace más fina.*

Demostración:

1. La equivalencia de estas cuatro afirmaciones se deduce fácilmente del hecho de que ambas aproximaciones son funciones de Γ y por tanto sus entropías sólo coinciden con las del conjunto aleatorio Γ cuando éste también es función de ellas, lo cual sucede únicamente cuando las imágenes de Γ son exactas c.s. (P).
2. Sea Γ un conjunto aleatorio, $\underline{R}(\Gamma)$ su aproximación inferior respecto a la partición determinada por la relación R y $\underline{R}'(\Gamma)$ su aproximación respecto a una partición más fina determinada por la relación R' . Es obvio que $\underline{R}(\Gamma)$ es además la aproximación inferior del conjunto aleatorio $\underline{R}'(\Gamma)$ con respecto a la primer partición. Por tanto, $H(\underline{R}(\Gamma)) \leq H(\underline{R}'(\Gamma))$, de modo que $H(\Gamma) - H(\underline{R}(\Gamma)) \geq H(\Gamma) - H(\underline{R}'(\Gamma))$. Es decir, la diferencia d_2 disminuye a medida que la partición se hace más fina. De forma análoga, se tiene el mismo resultado para d_1 . $\square$

Tras este estudio de la diferencia de entropías, descartamos esta opción por las razones que detallamos a continuación. En primer lugar, tenemos el hecho de que no sería de utilidad una medida expresada como función de las entropías de las aproximaciones: como hemos visto, la diferencia de entropías de las dos aproximaciones no informa acerca del grado de inexactitud o imprecisión, en general. Aunque las medidas d_1 y d_2 tengan buenas propiedades, exigen para su cálculo el conocimiento de la

entropía del conjunto aleatorio original, cuando la base de este estudio es el hecho de que desconocemos los valores de Γ , teniendo únicamente la información proporcionada por sus aproximaciones. Además, la diferencia de entropías no tiene en cuenta el grado de diferencia entre las imágenes de las aproximaciones, sino solo si hay diferencias o no, con las imágenes del conjunto aleatorio original.

3.6.2. El grado de rugosidad de un conjunto

Pawlak introdujo en [82] el concepto de *exactitud* de un conjunto, con el objetivo de cuantificar su grado de exactitud (o completitud de nuestro conocimiento acerca de él), con respecto a una relación de equivalencia particular. Definió la exactitud de un subconjunto $A \subseteq U$ con respecto a la relación de equivalencia R como el cociente $\alpha_R(A) = \frac{\#R(A)}{\#R(A)}$. También definió la noción opuesta de *rugosidad* de un conjunto A como la cantidad $\rho_R(A) = 1 - \alpha_R(A)$. Yao ([102]) mostró que ésta puede ser vista como la ampliamente conocida métrica de Marczewski-Steinhaus (métrica MZ, para abreviar) aplicada al par formado por las aproximaciones inferior y superior de un conjunto. La métrica MZ ([67]) cuantifica la distancia entre dos conjuntos arbitrarios X e Y del siguiente modo:

$$d(X, Y) = \frac{\#(X \Delta Y)}{\#(X \cup Y)} = 1 - \frac{\#(X \cap Y)}{\#(X \cup Y)}.$$

De acuerdo con esta interpretación de Yao, proponemos generalizar el concepto de rugosidad dado por Pawlak como sigue:

Definición 36 Sea U un universo arbitrario (no necesariamente finito). Definimos la *d-rugosidad* de un conjunto $A \subseteq U$ como

la cantidad $\rho_R^d(A) = d(\underline{R}(A), \overline{R}(A))$, donde d satisface alguno de los axiomas de medida de desigualdad enumerados en la sección 1.2.3. Hablaremos de rugosidad con respecto a una disimilaridad, a una distancia o a una divergencia, cuando d satisfaga los correspondientes axiomas (recordemos que las definiciones de estas medidas se encuentran en la sección 1.4.1).

En el primer capítulo de este documento, hemos revisado las relaciones formales entre los diferentes axiomas de medidas de diferencia (propiedades generales y axiomas específicos), cuando restringimos su dominio a la familia de los pares de subconjuntos nítidos. En particular, hemos visto que se cumplen los resultados recogidos en el siguiente teorema. El lector puede volver a consultar la sección 1.4.3 para más detalles.

Teorema 6 *Consideremos la restricción de una función $d : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow \mathbb{R}$ a la familia de los pares de subconjuntos nítidos, $\wp(U)$.*

- *Si d satisface $D5$, entonces $d(A, B) \leq d(A \setminus B, B \setminus A)$, $\forall A, B \in \wp(U)$.*
- *Si d satisface $D4$, entonces $d(A, B) \geq d(A \setminus B, B \setminus A)$, $\forall A, B \in \wp(U)$.*
- *Si d satisface $D4$, entonces $d(A, B) \leq d(C, B)$, para todo $A, B, C \in \wp(U)$ tales que $C \cap B = \emptyset$ y $A \subseteq C$.*
- *Si d satisface $D4$ y $D5$ entonces $d(A, B) \leq d(B^c, B)$, $\forall A, B \in \wp(U)$.*
- *Si d satisface $G2$, $D3$, $D4$, $D5$ y $G5$, entonces satisface $G4^*$ en $\wp(U)$ y por tanto puede ser expresada como una composición $F_d \circ M$. Además, la aplicación F_d no depende de*

su primer argumento, y es creciente en su segundo y tercer argumento.

- *Si d satisface $D6$ entonces satisface $D1$.*
- *Si d satisface $D6$ y $D7$, entonces satisface $D4$ y $D5$.*

Como consecuencia de lo anterior, hemos demostrado que:

Teorema 7

- *Cualquier medida de disimilaridad simétrica es una medida de divergencia.*
- *Cualquier medida de disimilaridad simétrica es una medida de distancia, excepto por la propiedad $d(A, A^c) = d(B, B^c)$, $\forall A, B \in \wp(U)$.*
- *Cualquier medida de divergencia satisfaciendo $G5$ es una medida de disimilaridad simétrica, excepto por el axioma $G1$ (condiciones de acotación).*
- *Cualquier medida de divergencia local d es una medida de disimilaridad simétrica, excepto por el axioma $G1$.*
- *Cualquier medida de divergencia es una medida de distancia, excepto por la propiedad $d(A, A^c) = d(B, B^c)$, $\forall A, B \in \mathcal{P}(U)$.*

Si nos restringimos a la familia de los pares de conjuntos nítidos anidados, $\{(A, B) \in \wp(U) \times \wp(U) : A \subseteq B\}$, podemos comprobar, además, que cualquier medida satisfaciendo los axiomas $D4$ y $D5$ satisface $G5$ y $G4^*$. Por otro lado, la evaluación de las

igualdades $d(A, A^c) = d(B, B^c)$, $\forall A, B \in \wp(U)$ no es aplicable en este caso. Como consecuencia, podemos completar el teorema 7 y obtener el siguiente resultado:

Teorema 8

- *Cualquier medida de disimilaridad simétrica es una medida de divergencia.*
- *Cualquier medida de divergencia es una medida de disimilaridad simétrica, excepto por el axioma G1 (condiciones de acotación).*
- *Cualquier medida de disimilaridad simétrica es una medida de distancia.*
- *Cualquier medida de divergencia es una medida de distancia.*

Por tanto, en este contexto, los conceptos de “disimilaridad simétrica” y “divergencia” son equivalentes entre sí, excepto por las condiciones de acotación. Del mismo modo, ambos satisfacen las propiedades de medida de distancia. Note el lector además que, incluso en este marco particular (pares de conjuntos anidados) las medidas de distancia no satisfacen necesariamente la implicación inversa, es decir, no son necesariamente medidas de divergencia o de disimilaridad. La distancia MZ es un ejemplo de este hecho, como puede comprobarse a continuación.

Ejemplo 28 *Consideremos el universo finito $U = \{1, 2, 3, 4\}$ y la distancia MZ,*

$$d_{MZ}(X, Y) := \frac{\#(Y \triangle X)}{\#(Y \cup X)}, \quad \forall X, Y \in \wp(U).$$

Podemos comprobar fácilmente que satisface los axiomas G2, D1, D2 y D3, y por tanto es una medida de distancia. Sin embargo, no verifica las propiedades de medida de disimilaridad y divergencia, incluso cuando nos restringimos a la familia de los pares de conjuntos anidados, $\{(X, Y) \in \wp(U) \times \wp(U) : X \subseteq Y\}$. De hecho, si consideramos, por ejemplo los conjuntos $A = \{1\}$, $B = \{1, 3\}$, $C = \{1, 2\}$ y $D = \{1, 2, 3\}$, observamos que $A \setminus B = C \setminus D = \emptyset$ y $B \setminus A = D \setminus C = \{3\}$. Es fácilmente comprobable que $d_{MZ}(A, B) = 1 \neq 0,5 = d_{MZ}(C, D)$. Por tanto, d_{MZ} no es una disimilaridad, y tampoco una medida de divergencia, incluso aunque nos restrinjamos el dominio a la clase de los pares de conjuntos anidados.

De acuerdo con el teorema 8 y el ejemplo 28, si desviamos nuestra atención de las condiciones de acotación hacia propiedades específicas de las medidas de “diferencia”, podemos dividir los conceptos recordados en la sección 1.4.1 en dos grandes grupos: el grupo de las medidas de “disimilaridad” y “divergencia”, y el grupo constituido por el resto de medidas de “distancia”. Las medidas pertenecientes al primer grupo centran su atención en las diferencias entre los conjuntos comparados, pero no tienen en cuenta aquello que tienen en común, dado que satisfacen la propiedad:

$$d(A, B) = d(\emptyset, B \setminus A), \quad \forall A, B \in \wp(U), A \subseteq B. \quad (3.10)$$

Contrariamente, las medidas pertenecientes al segundo grupo tienen en cuenta lo que tienen en común. En particular, hemos comprobado en el ejemplo 28 que la distancia MZ no satisface la ecuación 3.10, y por tanto la rugosidad dada por Pawlak constituye un caso particular de medida de rugosidad con respecto a una medida de distancia que no satisface las propiedades

de divergencia o disimilaridad. Además, la distancia MZ entre el conjunto vacío y cualquier otro subconjunto del universo es igual a 1, no importando como de “grande” o “pequeño” sea, es decir:

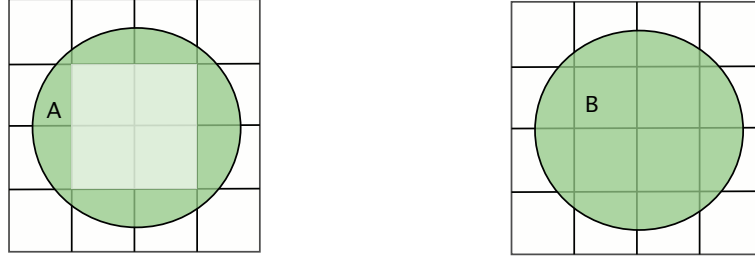
$$d_{MZ}(\emptyset, C) = 1, \quad \forall C \in \wp(U) \setminus \{\emptyset\}. \quad (3.11)$$

Podemos considerar esto como una desventaja de la medida de rugosidad de Pawlak. De acuerdo con la ecuación 3.11, cualquier conjunto rugoso con aproximación inferior vacía alcanza el máximo grado de rugosidad, no importando como de “grande” o “pequeña” sea la aproximación superior. Por el contrario, cualquier medida de divergencia/disimilaridad no trivial asignará diferentes grados de rugosidad, dependiendo del “tamaño” de la aproximación superior.

El siguiente ejemplo compara la rugosidad de Pawlak y la medida de rugosidad derivada de la distancia de Hamming, que es una divergencia/disimilaridad.

Ejemplo 29 *Consideremos un universo finito arbitrario U y denotemos por $d_{Ham} : \wp(U) \times \wp(U) \rightarrow \mathbb{R}$ la distancia de Hamming, que cuantifica el cardinal de la diferencia simétrica entre dos conjuntos, $d_{Ham}(X, Y) = \#(X \Delta Y)$. Podemos comprobar fácilmente que es una divergencia y una medida de disimilaridad simétrica. Comparemos la medida de rugosidad derivada de ella con la medida de rugosidad dada por Pawlak. Para este fin, consideremos los conjuntos A y B que pueden verse en el dibujo. La aproximación inferior de A es el conjunto vacío, mientras que la aproximación inferior de B es un rectángulo no vacío. Además, es obvio que $\overline{R}(A) \setminus \underline{R}(A) = \overline{R}(B) \setminus \underline{R}(B)$. Por tanto, la d_{Ham} -rugosidad de A y B coinciden. Por el contra-*

rio, si consideramos la medida de Pawlak, la rugosidad de A es estrictamente mayor que la rugosidad de B .



Algunas propiedades de la medida de rugosidad de Pawlak así como de la medida de precisión dual han sido estudiadas por diferentes autores como Gediga y Duntsch ([48]), Yao ([103]), y Zhu [107].

A continuación, relacionaremos esas propiedades con la colección de axiomas enumerados en el capítulo 1.

Teorema 9

- P1.- Si d satisface el axioma D3, entonces $\rho_R^d(A) = 0$ para cada conjunto exacto A .*
- P2.- Si d satisface el axioma D3*, entonces $\rho_R^d(A) = 0$ si y solo si A es un conjunto exacto.*
- P3.- Sea d una medida satisfaciendo el axioma D1 o el axioma D6. Si $R' \leq R$ (o equivalentemente, si R' induce un refinamiento de la partición inducida por R) entonces $\rho_{R'}^d(A) \leq \rho_R^d(A)$.*
- P4.- Si d satisface el axioma D1 (b) y $\overline{R}(A) = \overline{R}(B)$ entonces $\rho_R^d(A \cup B) \leq \min\{\rho_R^d(A), \rho_R^d(B)\}$.*

P5.- Si d satisface el axioma D1 (a) y $\underline{R}(A) = \underline{R}(B)$ entonces $\rho_R^d(A \cap B) \leq \min\{\rho_R^d(A), \rho_R^d(B)\}$.

Demostración:

- Las propiedades P1 y P2 se derivan fácilmente del hecho de que un conjunto es exacto si y solo si sus aproximaciones inferior y superior coinciden.
- Vamos a demostrar ahora la propiedad P3. Si R' es un refinamiento de R entonces, la aproximaciones superiores e inferiores de cualquier conjunto con respecto a ellas satisfacen las siguientes relaciones:

$$\underline{R}(A) \subseteq \underline{R}'(A) \subseteq \overline{R}'(A) \subseteq \overline{R}(A), \quad \forall A \in \wp(U). \quad (3.12)$$

- Supongamos, en primer lugar, que d satisface el axioma D1. Entonces, de acuerdo con la parte (a) de dicho axioma, podemos asegurar que $d(\underline{R}(A), \overline{R}(A)) \geq d(\underline{R}(A), \overline{R}'(A))$ y, además, por la parte (b) del mismo, podemos afirmar que $d(\underline{R}(A), \overline{R}'(A)) \geq d(\underline{R}'(A), \overline{R}'(A))$. Teniendo en cuenta ambas desigualdades, llegamos a la conclusión de que $\rho_R^d(A) \geq \rho_{R'}^d(A)$, que es lo que queríamos demostrar.
- Supongamos ahora que d satisface el axioma D6. A partir de la ecuación 3.12, deducimos que $\underline{R}(A) \cap \overline{R}(A) = \underline{R}(A)$ está incluida en $\underline{R}'(A) \cap \overline{R}'(A) = \underline{R}'(A)$ y que la diferencia $\overline{R}(A) \setminus \underline{R}(A)$ (la frontera de A con respecto a R) incluye a la diferencia $\overline{R}'(A) \setminus \underline{R}'(A)$. De esta forma, aplicando el axioma D6, derivamos fácilmente la desigualdad $\rho_{R'}^d(A) \leq \rho_R^d(A)$.

- Vamos a demostrar ahora la propiedad P4. Debemos tener en cuenta que la aproximación superior de la unión de dos conjuntos cuyas aproximaciones superiores coinciden, coincide con ambas aproximaciones superiores. Por otra parte, su aproximación inferior incluye a ambas aproximaciones inferiores. De esta forma, de acuerdo con el axioma D1 (a), el grado de rugosidad de la unión es inferior o igual al grado de rugosidad de cualquiera de los dos conjuntos.
- La demostración de la propiedad P5 es análoga a la de la propiedad P4. $\square$

Las siguientes propiedades no han sido enumeradas por los autores anteriormente mencionados. Además, la Propiedad P7 no es verificada en general por la medida de rugosidad de Pawlak. Sin embargo, ambas parecen propiedades razonables a tener en cuenta en el contexto de las medidas de rugosidad.

Teorema 10

P6.- Sea d una medida satisfaciendo D5 y consideremos un conjunto exacto C . Entonces $\rho_R^d(A \cup C) \leq \rho_R^d(A)$, $\forall A \in \wp(U)$.

P7.- Sea d una medida satisfaciendo D4 y consideremos un conjunto exacto C . Entonces $\rho_R^d(A \cap C) \leq \rho_R^d(A)$, $\forall A \in \wp(U)$.

Demostración:

P6.- Se puede comprobar fácilmente que, tanto la aproximación inferior, como la aproximación superior de la unión entre un conjunto A y otro conjunto exacto C coincide con la unión de la respectiva aproximación (inferior o superior) de

A y el conjunto C . Así, si la función d satisface el axioma D5, entonces cumple la propiedad enunciada.

P7.- La demostración de esta propiedad es análoga a la de la anterior. Basta sustituir la unión por la intersección donde corresponda. $\square$

La rugosidad de la unión de dos conjuntos, A y B está acotada superiormente por la rugosidad de A , si suponemos que B es un conjunto exacto o si sus aproximaciones superiores coinciden. No obstante, la d -rugosidad de la unión de dos conjuntos no está necesariamente acotada por la rugosidad de cualquiera de ellos, en general. Veamos un ejemplo al respecto.

Ejemplo 30 Consideremos el universo finito $U = \{1, 2, 3, 4\}$ y la relación de equivalencia R asociada a la partición $\Pi = \{\{1, 2\}, \{3, 4\}\}$. Consideremos ahora la distancia de Hamming normalizada, $d_{Ham}(X, Y) = \frac{\#X \Delta Y}{\#U}$, $\forall X, Y \in \wp(U)$, que satisface todos los axiomas de medida divergencia, disimilaridad y distancia enumerados en la sección 1.4.1. Denotemos por $\rho_R^{d_{Ham}}$ la medida de rugosidad inducida por dicha métrica:

$$\rho_R^{d_{Ham}}(A) = \frac{\#\bar{R}(A) - \#R(A)}{4}, \quad \forall A \in \wp(U).$$

Consideremos los conjuntos $A = \{1\}$ y $B = \{3\}$. Sus rugosidades son iguales a $\frac{1}{2}$ y la rugosidad de su unión es igual a 1, siendo por tanto estrictamente mayor que sus respectivas rugosidades.

3.6.3. El grado de rugosidad de un conjunto borroso

En la literatura se han propuesto distintas generalizaciones del concepto de rugosidad introducido por Pawlak. Banerjee y Pal ([3]) definen una familia parametrizada de medidas de rugosidad de la forma siguiente:

Definición 37 ([3], Banerjee y Pal, 1996) *Se define la medida de rugosidad del conjunto difuso $\tilde{A} \in \mathcal{F}(U)$ con respecto a los parámetros (α, β) , donde $0 < \beta \leq \alpha \leq 1$ al valor:*

$$\rho_R^{\alpha, \beta}(A) = 1 - \frac{\# \underline{R}(A_\alpha)}{\# \overline{R}(A_\beta)}.$$

Más adelante, Huynh y Nakamori ([55]) definieron una medida de rugosidad para conjuntos difusos “libre” de parámetros, utilizando la noción de asignación básica, de la forma siguiente:

Definición 38 ([55], Huynh y Nakamori, 2005) *Consideremos un universo finito arbitrario $U = \{x_1, \dots, x_n\}$ y un subconjunto difuso $\tilde{A} \in \mathcal{F}(U)$ cuyo rango es $Rg(\tilde{A}) = \{\alpha_1, \dots, \alpha_n\}$, donde $\alpha_i > \alpha_{i+1} > 0$, para $i = 1, \dots, n-1$ y $\alpha_1 = 1$. Consideremos ahora la asignación básica, m , determinada por $\tilde{A}$ de la forma siguiente:*

$$m(\tilde{A}_{\alpha_i}) = \alpha_i - \alpha_{i+1}, i = 1, \dots, n-1, m(\tilde{A}_{\alpha_n}) = \alpha_n.$$

Se define la medida de rugosidad del conjunto difuso $\tilde{A}$ de la forma siguiente:

$$\rho_R(\tilde{A}) = \sum_{i=1}^n m(\tilde{A}_{\alpha_i}) \left(1 - \frac{\underline{R}(\tilde{A}_{\alpha_i})}{\overline{R}(\tilde{A}_{\alpha_i})} \right).$$

De acuerdo con estos precedentes, y teniendo en cuenta la generalización de la definición de Pawlak considerada en la sección anterior, podríamos generalizar el concepto anterior, sustituyendo los valores de la rugosidad de Pawlak de los α -cortes de $\tilde{A}$ por la d -rugosidad de los mismos, con respecto a una medida de “desigualdad” d . Ahora bien, la suma anterior, en la que los pesos de las rugosidades de los distintos α -cortes son los valores que les asocia la asignación básica m , se podría haber expresado, de forma alternativa, como la esperanza, con respecto a la distribución uniforme en el intervalo $[0, 1]$ de las rugosidades de dichos α -cortes:

$$\rho_R(A) = \int_0^1 \left(1 - \frac{R(\tilde{A}_\alpha)}{\bar{R}(\tilde{A}_\alpha)} \right) d\alpha.$$

Teniendo en cuenta todo lo anterior, podemos generalizar la definición de Huynh y Nakamori ([55]) de la forma siguiente:

Definición 39 *Consideremos un universo arbitrario (no necesariamente finito). Se define la d -rugosidad de un conjunto difuso $\tilde{A} \in \mathcal{F}(U)$ como la cantidad:*

$$\rho_R^d(\tilde{A}) = \int_0^1 \rho_R^d(\tilde{A}_\alpha) d\alpha.$$

Dado que todo conjunto nítido se puede ver como un conjunto difuso cuyos α -cortes coinciden con él mismo en todo el intervalo $(0, 1]$, vemos que la definición anterior generaliza la definición 36, y por ello hemos utilizado la misma nomenclatura ρ_R^d , para denotar a la d -rugosidad definida sobre la familia de todos los subconjuntos difusos de un universo arbitrario U .

Aún más, si, dada la restricción de una medida de desigualdad d a la familia de parejas de conjuntos nítidos, consideramos la función $\tilde{d} : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$ definida de la forma: $\tilde{d}(\tilde{A}, \tilde{B}) = \int_0^1 d(\tilde{A}_\alpha, \tilde{B}_\alpha) d\alpha$, entonces, la d -rugosidad de un conjunto difuso $\tilde{A}$ se puede expresar, alternativamente, de la forma:

$$\rho_R^d(\tilde{A}) = \tilde{d}(\underline{R}(\tilde{A}), \overline{R}(\tilde{A})), \quad (3.13)$$

donde las expresiones anteriores se refieren a las aproximaciones inferior y superior del difuso $\tilde{A}$ con respecto a la relación de equivalencia R , de acuerdo con la fórmula de la ecuación 3.2. Si, además de las consideraciones anteriores, tenemos en cuenta que los α -cortes de las aproximaciones inferior y superior de un difuso coinciden con las aproximaciones inferior y superior de sus α -cortes, observaremos fácilmente que la igualdad mostrada en la ecuación 3.13 es cierta.

La argumentación anterior nos da pie para introducir una definición de rugosidad aún más general:

Definición 40 *Consideremos un universo arbitrario (no necesariamente finito). Consideremos una medida de “desigualdad” $\tilde{d} : \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow \mathbb{R}$. Se define la $\tilde{d}$ -rugosidad de un conjunto difuso $\tilde{A} \in \mathcal{F}(U)$ como la cantidad:*

$$\rho_R^{\tilde{d}}(\tilde{A}) = \tilde{d}(\underline{R}(\tilde{A}), \overline{R}(\tilde{A})).$$

A continuación, vamos a enumerar algunas propiedades de la $\tilde{d}$ -rugosidad, cuyas demostraciones se derivan fácilmente de las propiedades análogas probadas en la sección anterior para el caso particular de los conjuntos nítidos.

Teorema 11

- P1.- Si $\tilde{d}$ cumple el axioma D3, entonces $\rho_R^{\tilde{d}}(\tilde{A}) = 0$ para todo conjunto difuso exacto $\tilde{A}$.*
- P2.- Si $\tilde{d}$ satisface el axioma D3*, entonces $\rho_R^{\tilde{d}}(A) = 0$ si y solo si A es un conjunto exacto.*
- P3.- Sea $\tilde{d}$ una medida que satisface el axioma D1 o el axioma D6. Si $R' \leq R$ (o equivalentemente, si R' induce un refinamiento de la partición inducida por R) entonces $\rho_{R'}^{\tilde{d}}(A) \leq \rho_R^{\tilde{d}}(A)$.*
- P4.- Si $\tilde{d}$ satisface el axioma D1 (b) y $\overline{R}(\tilde{A}) = \overline{R}(\tilde{B})$ entonces $\rho_R^{\tilde{d}}(\tilde{A} \cup \tilde{B}) \leq \min\{\rho_R^{\tilde{d}}(A), \rho_R^{\tilde{d}}(B)\}$.*
- P5.- Si $\tilde{d}$ satisface axioma D1 (a) y $\underline{R}(\tilde{A}) = \underline{R}(\tilde{B})$ entonces $\rho_R^{\tilde{d}}(\tilde{A} \cap \tilde{B}) \leq \min\{\rho_R^{\tilde{d}}(\tilde{A}), \rho_R^{\tilde{d}}(\tilde{B})\}$.*
- P6.- Sea $\tilde{d}$ una medida que satisface D5 y consideremos un conjunto difuso exacto $\tilde{C}$. Entonces $\rho_R^{\tilde{d}}(A \cup C) \leq \rho_R^{\tilde{d}}(A)$, $\forall A \in \wp(U)$.*
- P7.- Sea $\tilde{d}$ una medida que satisface D4 y consideremos un conjunto difuso exacto $\tilde{C}$. Entonces $\rho_R^{\tilde{d}}(A \cap C) \leq \rho_R^{\tilde{d}}(A)$, $\forall A \in \wp(U)$.*

En la sección siguiente, utilizaremos este concepto como herramienta para definir el grado de rugosidad de un conjunto aleatorio.

3.6.4. El grado de rugosidad de un conjunto aleatorio y de una variable aleatoria difusa

Definiremos el grado de rugosidad de un conjunto aleatorio o, de forma más general, el grado de rugosidad de una variable aleatoria difusa, como la esperanza de las rugosidades de sus imágenes.

Definición 41 *Consideremos un espacio probabilístico $(\Omega, \mathcal{A}, P)$ y un espacio medible $(U, \mathcal{A}')$. Sea $\tilde{\Gamma}$ una variable aleatoria difusa definida entre ellos. Sea R una relación de equivalencia definida sobre U y sea $\tilde{d} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow \mathbb{R}$ una medida de desigualdad entre conjuntos difusos. Llamamos grado de rugosidad de $\tilde{\Gamma}$ a la cantidad: $\text{Rug}_R^{\tilde{d}}(\tilde{\Gamma}) = E(\rho_R^{\tilde{d}}(\tilde{\Gamma}))$, siempre y cuando la función $\rho_R^{\tilde{d}} \circ \tilde{\Gamma} : \Omega \rightarrow \mathbb{R}$ sea $\mathcal{A} - \beta_{\mathbb{R}}$ medible.*

El grado de rugosidad de las variables aleatorias difusas satisface las siguientes propiedades, que se derivan fácilmente de las propiedades enunciadas en la sección anterior.

Teorema 12

- P1.- Si $\tilde{d}$ cumple el axioma D3, y las imágenes de la variable aleatoria difusa son conjuntos difusos exactos con respecto a R , entonces su grado de rugosidad es nulo.*
- P2.- Si $\tilde{d}$ satisface el axioma D3*, entonces el grado de rugosidad de la variable aleatoria difusa es nulo si y solo si sus imágenes son conjuntos exactos [c.s.(P)]*
- P3.- Supongamos que $\tilde{d}$ cumple el axioma D1 o el axioma D6. Si $R' \leq R$ (o equivalentemente, si R' induce un refinamiento de la partición inducida por R), entonces el grado de rugosidad de la variable aleatoria difusa con respecto a R' es menor o igual que su grado de rugosidad con respecto a R .*

- P4.- Si $\tilde{d}$ cumple la parte (b) del axioma D1, y las aproximaciones superiores de dos variables aleatorias difusas coinciden, entonces la rugosidad de la variable aleatoria difusa que viene determinada por la unión de sus imágenes es inferior o igual a la de cualquiera de ellas.*
- P5.- Si $\tilde{d}$ cumple la parte (a) del axioma D1, y las aproximaciones inferiores de dos variables aleatorias difusas coinciden, entonces la rugosidad de la variable aleatoria difusa que viene determinada por la intersección de sus imágenes es inferior o igual a la de cualquiera de ellas.*
- P6.- Supongamos que $\tilde{d}$ cumple el axioma D5 y consideremos un conjunto difuso exacto $\tilde{C}$. Consideremos una variable aleatoria difusa $\tilde{\Gamma}$ y la variable aleatoria difusa $\tilde{\Gamma}'$ cuyas imágenes son el resultado de unir las imágenes de $\tilde{\Gamma}$ con el conjunto difuso $\tilde{C}$. Entonces el grado de rugosidad de $\tilde{\Gamma}'$ es inferior o igual al grado de rugosidad de $\tilde{\Gamma}$.*
- P7.- Supongamos que $\tilde{d}$ cumple el axioma D4 y consideremos un conjunto difuso exacto $\tilde{C}$. Consideremos una variable aleatoria difusa $\tilde{\Gamma}$ y la variable aleatoria difusa $\tilde{\Gamma}'$ cuyas imágenes son el resultado de intersecar las imágenes de $\tilde{\Gamma}$ con el conjunto difuso $\tilde{C}$. Entonces el grado de rugosidad de $\tilde{\Gamma}'$ es inferior o igual al grado de rugosidad de $\tilde{\Gamma}$.*

Conclusiones y problemas abiertos

A lo largo del capítulo 1 hemos revisado y organizado algunas definiciones destacables de medidas de (des)igualdad entre conjuntos difusos, incluyendo todas ellas en un marco común. Posteriormente hemos analizado y mostrado sus relaciones formales. El lector tendrá, gracias al análisis formal proporcionado en dicho capítulo, una ventaja a la hora de escoger la definición axiomática que mejor se adapte a su problema concreto. Hemos mostrado la existencia de algunas relaciones muy cercanas entre definiciones axiomáticas aparentemente diferentes. Por ejemplo, las definiciones de disimilaridad [11] y divergencia [76] son muy próximas cuando nos restringimos a la familia de los subconjuntos nítidos del universo. De hecho, ambas se centran en las diferencias entre los conjuntos comparados, pero no en sus coincidencias. Para ilustrarlo, consideremos como universo, por ejemplo, el siguiente conjunto de idiomas $U = \{\text{Chino (c), Inglés (e), Holandés (h), Francés (f), Italiano (i), Ruso (r)}\}$. Sean $E = \{h, e, f, i, c\}$, $G = \{h, e, f, r, c\}$, $A = \{i\}$ y $V = \{r\}$ los conjuntos nítidos que representan las habilidades comunicativas de cuatro personas llamadas Erik, Gerard, Angelo y Vladimir. Erik y Gerard comparten muchos más idiomas que Angelo y Vladimir, pero esas coincidencias no pueden detectarse a través

de medidas de divergencia y disimilaridades. Si solamente queremos centrarnos en las diferencias, dichas medidas pueden ser útiles. Pero, si también queremos tener en cuenta sus habilidades comunes, debemos utilizar unas medidas de comparación diferentes. Hemos revisado otras definiciones axiomáticas de medidas de desigualdad que tienen en cuenta las coincidencias entre los conjuntos comparados (véanse las definiciones 12 y 13, por ejemplo).

Existen, por otra parte, varias cuestiones que merecen un estudio futuro. Por un lado, algunos artículos recientes establecen clases de equivalencia de medidas de similaridad y proponen diferentes modos para clasificar de acuerdo con su poder discriminante (véanse [10, 86], por ejemplo). Teniendo en cuenta los resultados obtenidos en este primer capítulo, podríamos tratar de extender estos estudios a clases más generales de medidas de comparación.

Por otro lado, en la literatura reciente se han introducido muchas y diferentes definiciones no axiomáticas de similaridad y disimilaridad/distancia entre conjuntos difusos (intervalo valorados) para ser utilizadas en aplicaciones específicas. En el futuro próximo, nos proponemos desarrollar un estudio bibliográfico detallado de dichos trabajos, relacionándolos con este trabajo más teórico. En nuestra opinión, esta conexión nos permitirá detectar qué axiomas satisface cada familia paramétrica de medidas, y además, cuáles de esas medidas son co-monótonas, es decir, equivalentes.

Finalmente, debemos destacar que nos hemos restringido a la comparación de conjuntos difusos estándar. Podemos extender nuestro estudio formal a un marco más general, como los conjuntos difusos intervalo-valorados, etc.

En el segundo capítulo, hemos proporcionado una generalización del concepto de “medida de tendencia central”, al que hemos denominado centroide. Su definición se obtiene siguiendo un procedimiento análogo al de la construcción de la esperanza de Frèchet, minimizando la “diferencia” o “desigualdad” esperada con respecto a las imágenes de la variable aleatoria difusa. Hemos comprobado que este concepto generaliza algunas definiciones ya existentes en la literatura acerca de variables aleatorias difusas, como la esperanza de Puri y Ralescu ([84]), la definición de mediana dada por Sinova et al. ([91]), y también extiende los conceptos de media, mediana y moda de las variables aleatorias clásicas. Además, este concepto aporta la gran ventaja de que el espacio final de la variable no ha de ser necesariamente un espacio numérico, ni siquiera un espacio métrico.

En dicho capítulo hemos comprobado así mismo que los conceptos de medida de comparación esperada y d -variación generalizan varios parámetros conocidos asociados a variables aleatorias, conjuntos aleatorios y variables aleatorias difusas. En particular, generaliza el concepto de entropía de una variable aleatoria y algunas varianzas escalares para variables aleatorias difusas. Las definiciones de varianza escalar que podemos encontrar en la literatura ([60, 66]) se restringen a aquellas situaciones en las que los valores de la variable aleatoria difusa son subconjuntos convexos de $\mathbb{R}^n$. El nuevo concepto de d -variación, sin

embargo, puede aplicarse en varias situaciones, incluso aunque los valores de la variable no sean reales. En trabajos futuros, estudiaremos nuevas propiedades de la medida de comparación esperada para caso concretos, tratando de establecer su conexión con las nociones de entropía y dispersión.

En el tercer capítulo de esta memoria hemos introducido el concepto de conjunto aleatorio rugoso, con el fin de representar formalmente la percepción incompleta de un conjunto aleatorio. Hemos comprobado que, bajo condiciones bastante generales, la pareja de aproximaciones superior e inferior de un conjunto aleatorio borroso son medibles, con respecto a cierta σ -álgebra definida sobre la familia de subconjuntos del espacio final y que, además, su distribución conjunta determina el mínimo y el máximo del conjunto de distribuciones de probabilidad compatibles con la información disponible. A continuación, hemos estudiado el caso general de las variables aleatorias difusas, y hemos encontrado algunas condiciones suficientes bajo las cuales la distribución conjunta de la pareja de aproximaciones proporciona cotas adecuadas para el conjunto de posibles valores para la probabilidad de una familia de subconjuntos difusos. Para demostrar todos estos resultados, nos hemos apoyado en resultados previos de la literatura que relacionan la familia de las medidas de probabilidad asociadas a las selecciones medibles de un conjunto aleatorio con la familia de distribuciones de probabilidad dominadas por su probabilidad superior de Dempster. Tal como mencionábamos al principio de la sección, esta segunda familia contiene, en general, a la primera, y no necesariamente coincide con ella. En este trabajo hemos encontrado una forma equivalente de expresar la información proporcionada por la variable aleatoria difusa rugosa $\tilde{\Lambda} = (\tilde{\Gamma}, \tilde{\bar{\Gamma}})$, en términos de una aplicación

multi-valuada $G_{\tilde{\Lambda}} : \Omega \longrightarrow \wp(\wp(U))$, cuyos valores son familias de subconjuntos difusos del universo U : todo lo que sabemos acerca de la verdadera variable aleatoria difusa es que es una selección medible de dicha aplicación multi-valuada. Apoyándonos en los resultados antes mencionados de la literatura reciente, en algunas situaciones no muy restrictivas, podemos afirmar que la familia de medidas de probabilidad acotadas por las probabilidades inferior y superior de Dempster de $G_{\tilde{\Lambda}}$ proporciona cotas adecuadas para la familia de medidas de probabilidad asociadas a sus selecciones $\mathcal{A} - \sigma_{\mathcal{F}}$ medibles. Esto nos permite afirmar que, en esas situaciones, la pareja de probabilidades superior e inferior de Dempster determinan el máximo y el mínimo del conjunto de medidas de probabilidad sobre $\sigma_{\mathcal{F}}$ compatible con la información disponible acerca de $\tilde{\mathcal{F}}$. Pero, aún en esos casos, esta familia de probabilidades puede estar estrictamente contenida en el conjunto de medidas de probabilidad acotadas por la pareja de probabilidades inferior y superior, debido a su falta de convexidad, por ejemplo. Como ampliación de este trabajo, nos proponemos estudiar la relación entre los dos conjuntos de medidas de probabilidad desde un punto de vista más amplio. Teniendo en cuenta algunos ejemplos proporcionados en trabajos previos (véase [29]), conjeturamos que las diferencias entre ellas, incluso en aquellos casos en que sus máximos y mínimos coinciden, pueden tener impacto en el estudio de la información acerca de probabilidades condicionadas, de las distribuciones conjuntas de parejas de variables aleatorias difusas, o de la d -dispersión.

Por otra parte, hemos propuesto una generalización de la medida de rugosidad introducida por Pawlak en 1982, con el propósito de utilizarla a la hora de cuantificar el grado de inexac-

titud de los conjuntos aleatorios y las variables aleatorias difusas. La rugosidad de un conjunto nítido se define como el grado de “desigualdad” (formalmente entendido como una “distancia”, “divergencia” o “disimilaridad”) entre sus aproximaciones inferior y superior. Hemos comprobado que, en el contexto actual (pares de subconjuntos anidados del universo) cualquier divergencia y cualquier disimilaridad simétrica satisface las propiedades de medida de “distancia” de [99], pero el recíproco no es cierto. Además, las nociones de divergencia y disimilaridad son equivalentes en esencia. Por consiguiente, podemos diferenciar dos clases de medidas de “diferencia”: aquellas que satisfacen las propiedades de disimilaridad y divergencia y aquellas que no las satisfacen. Por lo tanto, podemos dividir el conjunto de las medidas de d -rugosidad en dos clases disjuntas. Las medidas en la primera clase cuantifican el “tamaño” de la diferencia entre las aproximaciones, de modo que cada una de ellas puede expresarse en términos de una función de conjuntos monótona M del siguiente modo:

$$\rho_R^d(A) = M(\overline{R}(A) \setminus \underline{R}(A)), \quad \forall A \in \wp(U). \quad (3.14)$$

La segunda clase no sólo tiene en cuenta el “tamaño” de la diferencia entre la aproximación superior y la inferior, sino que también tiene en cuenta el “tamaño” de la primera. En el caso de que ambas estén calculadas de forma separada, la correspondiente medida de rugosidad puede expresarse en términos de un par de funciones de conjunto monótonas, M_1 y M_2 , y una función $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ como sigue:

$$\rho_R^d(A) = f(M_1(\underline{R}(A)), M_2(\overline{R}(A) \setminus \underline{R}(A))), \quad \forall A \in \wp(U), \quad (3.15)$$

donde f es decreciente (no-creciente) con respecto a la primer componente y creciente con respecto a la segunda.

De acuerdo con el teorema 10, la rugosidad de la unión o la intersección con un conjunto exacto está acotada por la rugosidad inicial, siempre que la función f no dependa de la primera componente y, por tanto, la medida de rugosidad asociada pueda ser formulada en términos de la ecuación 3.14.

Además, de acuerdo con el teorema 9, cualquier medida de rugosidad satisfaciendo la fórmula general expuesta en la ecuación 3.15 satisface las propiedades más relevantes de la medida de rugosidad de Pawlak reseñadas en la literatura.

Podemos probar fácilmente que la medida de Pawlak es un caso particular de la segunda clase de medidas de rugosidad (aquellas consideradas en la ecuación 3.15), donde $M_1 = M_2$ es el cardinal del conjunto, y $f(x, y) = \frac{y}{x+y}$. Observamos que $f(0, y) = 1$ sin importar el valor de $y \in \mathbb{N}$, y esto significa que la medida de rugosidad de Pawlak asocia el máximo valor de rugosidad a todo conjunto rugoso con una aproximación inferior vacía. Esto puede verse como una desventaja de la rugosidad de Pawlak que puede ser evitada con una función f diferente.

Tal como hemos mencionado al principio de la sección 3.6, la medida de precisión de Pawlak, α_R , puede expresarse como la composición de la función $g(x) = 1 - x$ con la medida de rugosidad ρ_R . En este trabajo nos hemos centrado únicamente en la noción de rugosidad. Podemos obtener de forma sencilla una medida de d -precisión para cada medida de d -rugosidad, simple-

mente componiéndola con una función decreciente, que no ha de ser necesariamente la función g . La nueva medida de precisión vendrá entonces expresada en términos de la composición de una función creciente y una medida de “desigualdad” entre pares de conjuntos. Algunos de los axiomas propuestos en la literatura con el fin de medir el grado de similaridad o igualdad entre dos conjuntos son “duales” a algunos de los axiomas enumerados en la sección 1.2.3, (tal dualidad es expresada en términos de una aplicación decreciente). Sin embargo, otros axiomas de “similaridad” considerados en la literatura no pueden ser vistos como el homólogo de ningún otro axioma de “desigualdad” contenido en ese listado (recordemos que una comparación formal detallada puede encontrarse en la sección 1.5 de esta memoria). Una posible ampliación de este trabajo consiste en revisar todos los axiomas de medidas de “igualdad” o “similaridad” entre conjuntos en este contexto particular, donde comparamos pares de conjuntos anidados, y entonces estudiar que propiedades satisfacen las correspondientes medidas de precisión derivadas de esas medidas de similaridad.

Agracecimientos

Este trabajo ha sido apoyado conjuntamente por el Ministerio Español de Educación y Ciencia bajo los proyectos MTM2010-17844 y TIN2011-24302, la Universidad de Oviedo y el Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática de dicha universidad.

Bibliografía

- [1] J. Aumann. Integral of set valued functions. J. Math. Analysis and Applications 12 (1965) 1-12.
- [2] V. Balopoulos, A. G. Hatzimichailidis, B. K. Papadopoulos. Distance and similarity measures for fuzzy operators. Information Sciences 177 (2007) 2336-2348.
- [3] M. Banerjee, S. K. Pal. Roughness of a fuzzy set. Information Sciences 93 (1996) 235-246.
- [4] O. Barndorff-Nielsen, W.S. Kendall, M.N.M. van Lieshout. (Eds.), Stochastic Geometry, Likelihood and Computation. CRC Press/Chapman and Hall (1999) Boca Raton
- [5] C. Baudrit, I. Couso, D. Dubois. Joint propagation of probability and possibility in risk analysis: towards a formal framework. International Journal of Approximate Reasoning 45 (2007) 82-105.
- [6] T. Beaubouef, F.E. Petry, G. Arora. Information-theoretic measures of uncertainty for rough sets and rough relational databases. Information Sciences 109 (1998) 535-563.
- [7] C. Bertoluzza, N. Corral, A. Salas. On a new class of distances between fuzzy numbers. Mathware & Soft Computing 2 (1995) 71-84.

- [8] K. R. Bhutani, A. Rosenfeld. Dissimilarity measures between fuzzy sets or fuzzy structures. *Information Sciences* 152 (2003) 313-318.
- [9] K. Bosteels, E. E. Kerre. A triparametric family of cardinality-based fuzzy similarity measures. *Fuzzy Sets and Systems* 158 (2007) 2466-2479.
- [10] B. Bouchon-Meunier, G. Coletti, M.-J. Lesot, M. Rifqi. Towards a conscious choice of a fuzzy similarity measure: a qualitative point of view. In: E. Hullermeier, R. Kruse, and F. Hoffmann (Eds.), *IPMU 2010, LNAI 6178*, Springer-Verlag, Berlin Heidelberg (2010) pp.1-10.
- [11] B. Bouchon-Meunier, M. Rifqi, S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems* 84 (1996) 143-153.
- [12] I. Cascos, I. Molchanov. Multivariate risks and depth-trimmed regions. *Finance Stoch* 11 (2007) 373-397.
- [13] C. Castaing, M. Valadier. *Convex analysis and measurable multi-functions*. Springer, Berlín (1977).
- [14] A. Castaldo, F. Maccheroni, M. Marinacci. Random correspondences as bundles of random variables. *Sankhya* 66 (2004) 409-427.
- [15] S-M. Chen, M-S. Yeh, P-Y. Hsiao. A comparison of similarity measures of fuzzy values. *Fuzzy Sets and Systems* 72 (1995) 79-89.
- [16] G. Choquet. Theory of capacities. *Annales de l'institut Fourier* 5 (1954) 131-295.

- [17] A. Colubi. Statistical inference about the means of fuzzy random variables: Applications to the analysis of fuzzy- and real-valued data. *Fuzzy Sets and Systems* 160 (2009) 344-356.
- [18] I. Couso. Teoría de la Probabilidad para datos imprecisos. Algunos aspectos. Tesis doctoral. Departamento de Estadística e I.O. y Didáctica de la Matemática. Universidad de Oviedo (1999).
- [19] I. Couso, D. Dubois. On the variability of the concept of variance of a fuzzy random variable. *IEEE Transactions on Fuzzy Systems* 17 (2009) 1070 - 1080.
- [20] I. Couso, L. Garrido, S. Montes, L. Sánchez. Expected Pair-Wise Comparison of the Outcomes of a Fuzzy Random Variable. *Combining Soft Computing and Statistical Methods in Data Analysis, Advances in Intelligent and Soft Computing* 77 (2010) 105-113.
- [21] I. Couso, L. Garrido, L. Sánchez. Similarity and dissimilarity measures between fuzzy sets: a formal relational study. *Information Sciences* 229 (2013) 122-141.
- [22] I. Couso, L. Garrido, L. Sánchez. Measures of roughness of a set. *Eurofuse* 2013.
- [23] I. Couso, L. Garrido, L. Sánchez. Random rough sets. *Eurofuse* 2013.
- [24] I. Couso, L. Garrido, L. Sánchez. General measures of roughness of fuzzy sets. *Information Sciences* (bajo revisión)
- [25] I. Couso, L. Garrido, L. Sánchez. Ill perception of random fuzzy sets by means of random rough fuzzy sets. *Internatio-*

nal Journal of Uncertainty, Fuzziness and Knowledge-Based Systems (bajo revisión).

- [26] I. Couso, S. Montes. An Axiomatic definition of fuzzy divergence measures. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 16 (2008) 1-17.
- [27] I. Couso, S. Montes, P. Gil. The necessity of the strong alpha-cuts. *Internat. J. Unc., Fuzz. and Knowledge-Based Systems* 9 (2001) 249-262.
- [28] I. Couso, S. Montes, P. Gil. Second order possibility measure induced by a fuzzy random variable. *Statistical Modeling, Analysis and Management of Fuzzy Data Studies*. In *Fuzziness and Soft Computing* 87 (2002) 127-144.
- [29] I. Couso, L. Sánchez. Higher order models for fuzzy random variables. *Fuzzy Sets Systems* 159 (2008) 237-258.
- [30] I. Couso, L. Sánchez. Inner and outer fuzzy approximations of confidence intervals. *Fuzzy Sets and Systems* 184 (2011) 68-83.
- [31] I. Couso, L. Sánchez. Mark-recapture techniques in statistical tests for imprecise data. *Int. J. Approx. Reasoning* 52 (2011) 240-260.
- [32] I. Couso, L. Sánchez. Upper and lower probabilities induced by a fuzzy random variable. *Fuzzy Sets and Systems* 165 (2011) 1-23.
- [33] I. Couso, L. Sánchez, P. Gil. Imprecise distribution function associated to a random set. *Information Sciences* 159 (2004) 109-123.

- [34] A.P. Dempster. Upper and lower Probabilities Induced by a Multi-valued Mapping. *Ann. Math. Statistics* 38 (1967) 325-339.
- [35] L. Dengfeng, C. Chuntian. New similarity measures of intuitionistic fuzzy sets and application to pattern recognition. *Pattern Recognition Letters* 23 (2002) 221-225.
- [36] T. Denoeux, M.H. Masson, P.A. Hébert. Nonparametric rank-based statistics and significance tests for fuzzy data. *Fuzzy Sets and Systems* 153 (2005) 1-28.
- [37] D. Dubois. Statistical Reasoning with Set-Valued Information: Ontic vs. Epistemic Views. Volume 285 of *Studies in Fuzziness and Soft Computing*, Springer (2013) pages 119-136.
- [38] D. Dubois, H. Prade. Rough Fuzzy Sets and Fuzzy Rough Sets. *International Journal of General Systems* 17 (1990) 191-209.
- [39] D. Dubois, H. Prade. The three semantics of fuzzy sets. *Fuzzy Sets and Systems* 90 Vol.2 (1997) 141-150.
- [40] D. Dubois, H. Prade. Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems* 192 (2012) 3-24.
- [41] R. Féron. Ensembles aléatoires flous. *C.R. Acad. Sci. Paris Ser. A* 282 (1976) 903-906.
- [42] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, L. Ginzburg. Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty, Technical Report. SAND2007-0939, Unlimited Release (2007).

- [43] M. Fréchet. Les éléments aléatoires de nature quelconque dan un espace distancié. *Ann Inst Henri Poincaré* 10 (1948) 215-310.
- [44] L. Garrido. Observaciones imprecisas: algunos resultados para variables aleatorias difusas. Trabajo para la obtención de la suficiencia investigadora en el doctorado de Matemáticas de la Universidad de Valladolid. 2006.
- [45] L. Garrido. On the extension of the scalar variance of a fuzzy random variable. 1st International Workshop on Trends in Uncertainty and Imprecision, TUIM 2010, Gijón (2010).
- [46] L. Garrido. Disimilitud esperada de una variable aleatoria difusa. *Actas del XVI Congreso Español sobre Tecnologías y Lógica Fuzzy* (2012) 331-335.
- [47] L. Garrido, T. Brezmes. Fuzzy envelope of a probability distribution. *Actas del congreso Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'08* (2008) 79-85.
- [48] G. Gediga, I. Düntsch. Rough approximation quality revisited. *Artificial Intelligence* 132 (2001) 219-234.
- [49] G. González-Rodríguez, Á. Blanco, A. Colubi, M.A. Lubiano. Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets and Systems* 160 (2009) 357-370.
- [50] G. González-Rodríguez, M. Montenegro, A. Colubi, M.A. Gil. Bootstrap techniques and fuzzy random variables: Sy-

- nergy in hypothesis testing with fuzzy data. *Fuzzy Sets and Systems* 157 (2006) 2608-2613.
- [51] I.R. Goodman, H.T. Nguyen. *Uncertainty Models for Knowledge-Based Systems*. Elsevier Science Publishers. Amsterdam. 1985.
- [52] D. Guha, D. Chakraborty. A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers. *Applied Soft Computing* 10 (2010) 90-99.
- [53] S.R. Hejazi, A. Doostparast, S.M. Hosseini. An improved fuzzy risk analysis based on a new similarity measures of generalized fuzzy numbers. *Expert Systems with Applications* 38 (2011) 9179-9185.
- [54] C. J. Himmelberg. Measurable relations. *Fundamenta Mathematicae* 87 (1975) 107-114.
- [55] V-N Huynh, Y. Nakamori. A roughness measure for fuzzy sets. *Information Sciences* 173 (2005) 255-275.
- [56] I. Jenhani, S. Benferhat, Z. Elouedi. Possibilistic Similarity Measures, *Foundations of Reasoning under Uncertainty. Studies in Fuzziness and Soft Computing* 249 (2010) 99-123.
- [57] D.G. Kendall. Foundations of a theory of random sets. . In: E.F. Harding and D.G. Kendall, eds. *Stochastic geometry*. New York: Wiley, Nueva York (1974).
- [58] D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science* 4 (1989) 87-99.

- [59] E. Klein, A. Thompson. Theory of correspondences. In: Including applications to mathematical economics. Wiley, Nueva York(1984).
- [60] R. Körner. On the variance of fuzzy random variables. Fuzzy Sets and Systems. 92 (1997) 83-93.
- [61] R. Körner. An asymptotic α -test for the expectation of random fuzzy. Journal of Statistical Planning and Inference 83 (2000) 331-346.
- [62] V. Krätschmer. Limit theorems for fuzzy-random variables. Fuzzy Sets and Systems 126 (2002) 253-263.
- [63] R. Kruse. On the variance of random sets. Journal of Mathematical Analysis and Applications 122 (1987) 469-473.
- [64] R. Kruse, K.D. Meyer. Statistic with Vague Data. D. Reidel Publishing Company, Dordrecht (1987).
- [65] Z. Liang, P. Shi. Similarity measures on intuitionistic fuzzy sets. Pattern Recognition Letters 24 (2003) 2687-2693.
- [66] A. Lubiano, M.A. Gil, M. López-Díaz, M.T. López. The $\overrightarrow{\lambda}$ -mean squared dispersion associated with a fuzzy random variable. Fuzzy Sets and Systems 111 (2000) 307-317.
- [67] E. Marczewski, H. Steinhaus. On a certain distance of sets and the corresponding distance of functions. Colloquium Mathematicum 6 (1958) 319-327.
- [68] G. Matheron. Random sets and integral geometry. Editorial: John Wiley and Sons (1975).

- [69] E. Miranda, I. Couso, P. Gil. Random intervals as a model for imprecise information. *Fuzzy Sets and Systems* 154 (2005) 386-412.
- [70] E. Miranda, I. Couso, P. Gil. Random sets as imprecise random variables. *J. Math. Anal. Appl.* 307 (2005) 32-47.
- [71] E. Miranda, I. Couso, P. Gil. Approximations of upper and lower probabilities by measurable selections. *Information Sciences* 180 (2010) 1407-1417.
- [72] H.B. Mitchell. On the Dengfeng-Chuntian similarity measure and its application to pattern recognition. *Pattern Recognition Letters* 24 (2003) 3101-3104.
- [73] I.S. Molchanov. Averaging of Random Sets and Binary Images. *CWI Quarterly* 11 (1998) 371-384.
- [74] B. Mondal, D. Mazumdar, S. Raha. Similarity in approximate reasoning. *International Journal of Computational Cognition* 4 (2006) 46-56.
- [75] S. Montes. Particiones y medidas de divergencia en modelos difusos. Tesis doctoral. Universidad de Oviedo (1998).
- [76] S. Montes, I. Couso, P. Gil, C. Bertoluzza. Divergence measure between fuzzy sets, *International Journal of Approximate Reasoning* 30 (2002) 91-105.
- [77] H.T. Nguyen. On fuzziness and linguistic probabilities. *J. Math. Analysis and Applications* 61 (1977) 658-671.
- [78] H.T. Nguyen. On random sets and belief functions. *J. Math. Analysis and Applications* 63 (1978) 531-542.

- [79] C. P. Pappis, N. I. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems* 56 (1993) 171-174.
- [80] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences* 11 (1982) 341-356.
- [81] Z. Pawlak. Rough probability. *Bull. Polish Acad. Sci. Math.*, Vol. 32 n° 9-10 (1984) 607-612.
- [82] Z. Pawlak. *Rough Sets - Theoretical Aspects of Reasoning about Data*. Kluwer Academic, Dordrecht (1991).
- [83] Z. Pawlak, A. Skoron. Rudiments of rough sets. *Information Sciences* 177 (2007) 2-27.
- [84] M.L. Puri, D. Ralescu. Fuzzy Random Variables. *J. Math. Analysis and Applications* 114 (1986) 409-422.
- [85] A.M . Radzikowska, E.E. Kerre. A comparative study of fuzzy rough sets. *Fuzzy Sets and Systems* 126 (2002) 137-155.
- [86] M. Rifqi, V. Berger, B. Bouchon-Meunier. Discrimination power of measures of comparison. *Fuzzy Sets and Systems* 110 (2000) 189-196.
- [87] J.F. Omhover, M. Rifqi, M. Detyniecki. Ranking invariance based on similarity measures in document retrieval. In: *Adaptive Multimedia Retrieval AMR'05*. Springer LNCS (2006) 55-64.
- [88] L. Sánchez, I. Couso. Advocating the use of imprecisely observed data in genetic fuzzy systems. *IEEE Transactions on Fuzzy Systems* 15 (2007) 551-562.

- [89] S. Santini, R. Jain. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999) 871-883.
- [90] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press (1976).
- [91] B. Sinova, M.A. Gil, A. Colubi, S. Van Aelst. The median of a random fuzzy number, The 1-norm distance approach. *Fuzzy Sets and Systems* 200 (2012) 99-115.
- [92] W. Trutschnig, G. González-Rodríguez, A. Colubi, M.A. Gil. A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. *Information Sciences* 179 (2009) 3964-3972.
- [93] P.Z. Wang. *Fuzzy Sets and Its Applications*. Shanghai Science and Technology Press, Shanghai (1983) en Chino.
- [94] D. Wu, J. M. Mendel. A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets. *Information Sciences* 178 (2008) 381-402.
- [95] D. Wu, J. M. Mendel. A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. *Information Sciences* 179 (2009) 1169-1192.
- [96] Z. S. Xu. An overview of distance and similarity measures of intuitionistic fuzzy sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 16 (2008) 529-555.
- [97] Z. Xu, M. Xia. Distance and similarity measures for hesitant fuzzy sets. *Information Sciences* 181 (2011) 2128-2138.

- [98] L. Xudong, C. Zhang. An axiom foundation for uncertain reasonings in rule based expert systems: NT-Algebra. Knowledge and Information Systems 1 (1999) 415-433.
- [99] L. Xuecheng. Entropy, distance measure and similarity measure of fuzzy sets and their relations. Fuzzy Sets and Systems 52 (1992) 305-318.
- [100] J. Ye. Cosine similarity measures for intuitionistic fuzzy sets and their applications. Mathematical and Computer Modelling 53 (2011) 91-97.
- [101] Y.Y Yao. Combination of rough and fuzzy sets based on α -level sets. Rough Sets and Data Mining: Analysis for Imprecise Data, Lin, T.Y., Cercone, N. (eds.), Kluwer Academic Publishers (1997) 301-321.
- [102] Y.Y. Yao. Information granulation and rough set approximation. International Journal of Intelligent Systems 16 (2001) 87-104.
- [103] Y.Y. Yao. Notes on Rough Set Approximations and Associated Measures. Journal of Zhejiang Ocean University (Natural Science) Vol. 29, No. 5, (2010) 399-410.
- [104] L.A. Zadeh. The concept of a linguistic variable and its application in approximate reasoning, part II. Information Science 8 (1975) 301-357.
- [105] W. Zeng, H. Li, Relationship between similarity measure and entropy of interval valued fuzzy sets, Fuzzy Sets and Systems 157 (2006) 1477-1484.

- [106] H. Zhang, W. Zhang, C. Mei, Entropy of interval-valued fuzzy sets based on distance and its relationship with similarity measure, *Knowledge-Based Systems* 22 (2009) 449-454
- [107] P. Zhu. An axiomatic approach to the roughness measure of rough sets. *Fundamenta Informaticae* 109 (2011) 463-480.
- [108] R. Zwick, E. Carlstein, D.V. Budeanu, Measures of similarity among fuzzy concepts: a comparative analysis, *International Journal of Approximate Reasoning* 1 (1987) 221-242.