Teoría de la Evidencia

Evidencia según el diccionario de la RAE es:

1. f. Certeza clara y manifiesta de la que no se puede dudar.
2. f. Prueba determinante en un proceso.

Evidencia para una proposición es lo que apoya esta proposición. Suele entenderse como una indicación de que la proposición apoyada es verdadera.

El papel que desempeña una determinada evidencia y cómo se concibe qué constituye una evidencia significativa varía de un campo a otro. En la epistemología, rama de la filosofía que estudia el conocimiento científico, evidencia es lo que justifica creencias o lo que hace que sea racional mantener una cierta actitud doxástica. La lógica doxástica es un sistema lógico que se ocupa del razonamiento acerca de las creencias. Típicamente, una lógica doxástica utiliza la expresión $B_c(p)$ para significar "el razonador c cree que p es verdadero", y el conjunto B_c se refiere al conjunto de creencias de c.

Por ejemplo, una experiencia perceptiva de un árbol puede actuar como evidencia que justifica la creencia de que hay un árbol. En este papel, la evidencia suele entenderse como un estado mental privado. Los temas importantes en este campo incluyen las preguntas de cuál es la naturaleza de estos estados mentales, por ejemplo, si tienen que ser proposicionales, y si los estados mentales engañosos todavía deben considerarse como evidencia.

En otros campos, como las ciencias, la medicina y el sistema legal, se tiende a enfatizar el carácter público de las evidencias. En la filosofía de la ciencia, evidencia se entiende como lo que confirma o desconfirma hipótesis científicas.

Las mediciones de la órbita "anómala" de Mercurio, por ejemplo, se consideran evidencia que confirma la teoría de la relatividad general de Einstein. Para desempeñar el papel de árbitro neutral entre teorías en competencia, es importante que la evidencia científica sea pública y no controvertida, como objetos o eventos físicos observables, para que los proponentes de las diferentes teorías puedan acordar sobre cuál es la evidencia. Esto se asegura siguiendo el método científico y tiende a conducir a un consenso científico emergente a través de la acumulación gradual de evidencia.

Relación Evidencial

El término "relación evidencial" se refiere a la relación entre una evidencia y la proposición apoyada. El tema de la naturaleza de la relación evidencia tiene que ver con la cuestión de cómo debe ser esta relación para que una cosa justifique una creencia o confirme una hipótesis. Las teorías importantes en este campo son el enfoque probabilístico, el hipotético-deductivo y el enfoque de la instancia positiva.

Los enfoques probabilísticos, también llamados teoría bayesiana de la confirmación, explican la relación evidencial en términos de probabilidades. Asumen que el único aspecto necesario de las evidencias es que su existencia aumenta la probabilidad de que la hipótesis sea verdadera.

El enfoque hipotético-deductivo es un enfoque no probabilístico que caracteriza la relación evidencial con respecto a las consecuencias deductivas de la hipótesis. Según este punto de vista, "la evidencia de una hipótesis es una consecuencia observable real de esa hipótesis". Un problema con la caracterización anterior es que las hipótesis suelen contener relativamente poca información y, por lo tanto, tienen pocas o ninguna consecuencia deductiva observable.

Por ejemplo, la hipótesis de que hay un fuego no conduce en sí misma a que se observe humo. En cambio, hay que hacer varias suposiciones adicionales sobre la ubicación del humo, el fuego, el observador, las condiciones de iluminación, las leyes químicas, etc. De este modo, la relación evidencial se convierte en una relación ternaria entre evidencia, hipótesis y suposiciones adicionales. Esto significa que la cuestión de si una cosa es evidencia para una hipótesis depende de las suposiciones adicionales que se tiene. Este enfoque encaja bien con varias prácticas científicas.

Según el enfoque de la instancia positiva, una oración de observación es evidencia para una hipótesis universal si la oración describe una instancia positiva de esa hipótesis. Por ejemplo, la observación de que "este cisne es blanco" es una instancia de la hipótesis universal de que "todos los cisnes son blancos". Este enfoque puede formularse precisamente en la lógica de primer orden: una oración es evidencia de una hipótesis si el "desarrollo de la hipótesis" se deriva de ella

Teoría de Dempster–Shafer

La teoría Dempster–Shafer, también llamada teoría de la evidencia o teoría de funciones de creencia es un marco general para razonar con incertidumbre, que incluye conexiones con otros marcos como la probabilidad, la posibilidad y las teorías de probabilidad imprecisa. Fue introducida por primera vez por Arthur P. Dempster en el contexto de la inferencia estadística, y posteriormente desarrollada por Glenn Shafer como un marco general para modelar la incertidumbre epistémica, en la forma de una teoría matemática de la evidencia. La teoría permite combinar evidencias de diferentes fuentes para llegar a un grado de creencia (representado por un objeto matemático llamado función de creencia) que tiene en cuenta toda la evidencia disponible.

En un sentido estricto, el término teoría de Dempster-Shafer se refiere a la concepción original de la teoría de Dempster y Shafer. Sin embargo, es común referirse a la teoría Dempster-Shaferel adaptándola a distintos tipos de situaciones específicas. En particular, muchos autores han propuesto varias reglas para combinar diferentes evidencias, a menudo con el fin de manejar mejor los conflictos entre ellas. Las primeras contribuciones también han sido puntos de partida para muchos desarrollos importantes, incluyendo el modelo de creencias transferibles y la teoría de las pistas.

La teoría de Dempster-Shafer es una generalización de la teoría bayesiana de probabilidad subjetiva. Las funciones de creencia basan grados de creencia (o confianza) para una pregunta sobre las probabilidades de una pregunta relacionada. Los grados de creencia en sí mismos pueden o no tener las propiedades matemáticas de las probabilidades; cuánto difieren depende de qué tan estrechamente estén relacionadas las dos preguntas. Dicho de otra manera, es una forma de representar las plausibilidades epistémicas, pero puede dar respuestas que contradicen a las que se obtienen utilizando la teoría de la probabilidad.

Su objetivo es modelar la incertidumbre del conocimiento eliminando algunos de los puntos flacos del enfoque probabilista o Bayesiano. En particular hace énfasis en que la suma de la creencia en un hecho y en su contrario no tiene por qué ser uno. En probabilidad el aumento de la probabilidad de una hipótesis

A menudo utilizada como método de fusión de datos, la teoría Dempster-Shafer se basa en dos ideas: obtener grados de creencia para una pregunta a partir de probabilidades subjetivas para una pregunta relacionada, y la regla de Dempster para combinar tales grados de creencia cuando se basan en elementos independientes de evidencia. En esencia, el grado de creencia en una proposición depende principalmente del número de respuestas (a las preguntas relacionadas) que contienen la proposición y la probabilidad subjetiva de cada respuesta. También contribuyen las reglas de combinación que reflejan supuestos generales sobre los datos.

La teoría considera como punto de partida una serie de entornos $q_1, q_2, \cdots, q_n$ que representan el universo. Un entorno cualquiera $H = h_1, h_2, \cdots, h_m$ consta de un conjunto de hipótesis mutuamente exclusivas y exhaustivas que cubren todas las posibles situaciones en el entorno. Este conjunto de hipótesis se llama marco de discernimiento.

Se define conjunto potencia del marco H al formado por todos los subconjuntos posibles de hipótesis h_i. (Es algo así como todas las posibles respuestas a las posibles preguntas).

$$P(H) = [\{\varnothing\}\{h_1\}\{h_2\}\ldots\{h_1, h_2\}\ldots\{h_1, h_2, \ldots, \}$$

Un marco H con M hipótesis tendrá 2^M subconjuntos representando su conjunto potencia.

En este formalismo, un grado de creencia (también conocido como masa) se representa como una función de creencia más que como una distribución de probabilidad bayesiana. Los valores de probabilidad se asignan a conjuntos de posibilidades en lugar de eventos individuales: su atractivo se basa en el hecho de que codifican naturalmente la evidencia a favor de las proposiciones.

La teoría de Dempster-Shafer asigna sus masas a todos los subconjuntos de las proposiciones que componen un sistema, o en términos de teoría de conjuntos, al conjunto potencia de las proposiciones. Por ejemplo, suponga una situación en la que hay dos preguntas relacionadas, o proposiciones, en un sistema. En este sistema, cualquier función de creencia asigna masa a la primera proposición, a la segunda, a ambas o a ninguna.

Según el principio de indiferencia si tenemos información similar sobre dos alternativas, debe asignarse a ambas la misma probabilidad. Una de las singularidades de dicho principio es que otorga las mismas probabilidades a dos alternativas de las cuales no sabemos nada, al igual que a los resultados posibles de una moneda rigurosamente balanceada y testeada.

Glen Shafer argumenta que nuestra creencia en una proposición particular debe estar basada explícitamente en la evidencia que tenemos a favor de esa proposición. Si tenemos muy poca evidencia a favor o en contra, podríamos querer asignar un pequeño grado de apoyo (digamos 0.2) a la proposición T pero también un pequeño grado de apoyo, digamos 0.1, a su negación $\neg T$.

Shafer denomina marco de discernimiento Θ a un conjunto de mundos posibles, o modelos con un dominio dado para enfatizar la naturaleza epistémica del conjunto de posibilidades. Θ es el universo de las proposiciones a las que se evaluará su "confianza". Es análogo al espacio muestral Ω en la teoría de la probabilidad. Al igual que en la teoría de la probabilidad, y por las mismas razones, estamos interesados en el conjunto $P(\Theta)$.

Las fuentes de evidencia son las que generan el "grado de confianza" a subconjuntos del marco de discernimiento. Dos fuentes de evidencia son evidencialmente Independientes si los errores que ocurren en cada fuente son independientes de la otra.

Sea S un conjunto finito de fuentes evidencialmente independientes que generan las Asignaciones Probabilísticas Básicas (APB) para varios subconjuntos de Θ. Sea $n = |S|$ el cardinal de S. Cada fuente generará un conjunto finito de APBs. La fuente S_i puede ser interpretada como un mapeo de un subconjunto del conjunto potencia de Θ al intervalo $[0, 1]$, es decir,

$$S_i : P(\Theta) \to [0, 1]$$

Las APBs generadas por la fuente S_i se definen como funciones de masa, donde, dado Θ, se define la función de masa m con dominio $P(\Theta)$ y codominio $[0, 1]$. La cantidad $m(X)$ representa la masa asignada al subconjunto $X \subset \Theta$ que, a su vez, mide el apoyo colocado exactamente a ese subconjunto X y no a algún subconjunto propio de X. La función m se conoce también como asignación de probabilidad básica.

La función de asignación de probabilidad básica o masa de probabilidad es una función $m : 2^M \to [0, 1]$ (del conjunto potencia de H al intervalo real $[0, 1]$) definida como:

- $m(\emptyset) = 0$
- $\sum_{S_i \subset H} m(S_i) = 1$

donde:

$0 < m(S_i) < 1$

S_i cualquier subconjunto del marco de discernimiento (o lo que es lo mismo, del conjunto potencia)

$m(S_i)$ es *asignada por el experto según su juicio*

La creencia en un conjunto de hipótesis S será:

$$BEL(S) = \sum_{X \subset S} m(X)$$

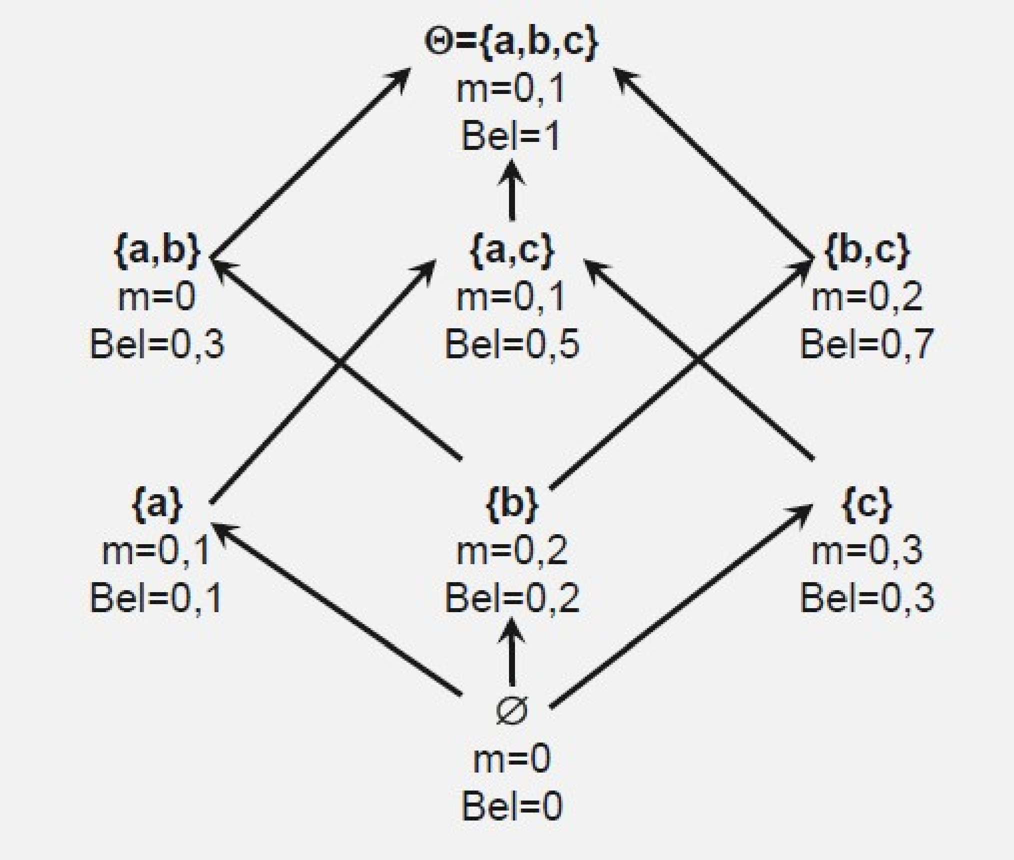
Es la suma de todas las masas de probabilidad de los subconjuntos de S.

El valor m asignado a S_i indica la creencia en el conjunto S_i, mientras que la función $Bel(S_i)$ indica la creencia en S_i y todos sus subconjuntos:

$Bel(H) = 1$ (La creencia del marco H es 1)

La función de creencia asignada a un conjunto X es la suma de todas las masas asignadas a X y a sus subconjuntos. $Bel(X)$ se denomina grado de creencia en X.

La relación entre masa y grado de creencia no se puede representar con diagramas de Venn. Una representación adecuada surge de los diagramas de Hasse (que representa la inclusión). Supongamos $\Theta = \{a, b, c\}$.



Se puede definir Bel por medio de axiomas, sin usar la función m. Si Θ es un marco de discernimiento, entonces una función Bel es una función de creencia si y solo si:

1. $Bel(\emptyset) = 0$
2. $Bel(\Theta) = 1$
3. $Bel(\cup_{i=1}^{n} A_i) \geq \sum_{I \subset \{1,\ldots,n\}} (-1)^{i+1} Bel(\cap_{i \in I} A_i)$

Obviamente, dada una función de creencia, se puede deducir la función de masa asociada a ella. En general se cumple que, para todo $A \subset \Theta$:

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

Se define como intervalo de creencia o certidumbre de un subconjunto S al intervalo $[BEL(S), 1 - BEL(\neg S)]$ donde $\neg S$ es el conjunto complementario de S con respecto al conjunto potencia del marco de discernimiento.

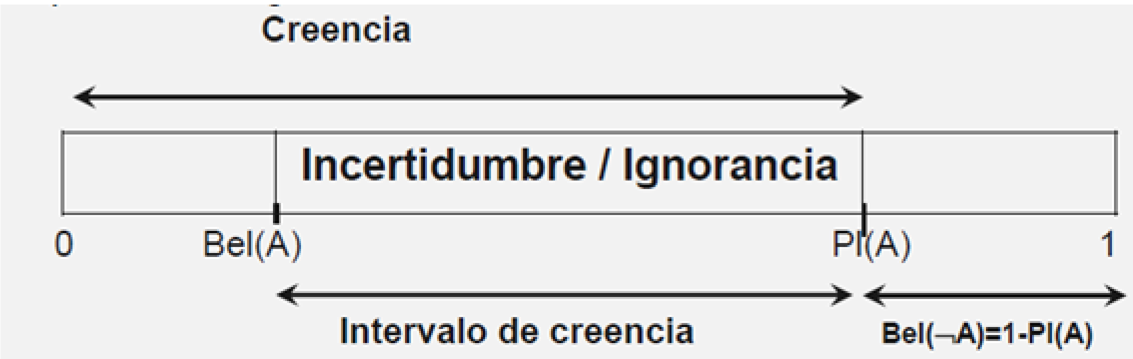
A su vez, se define la plausibilidad como una medida del grado en que la evidencia no puede refutar S (de donde S es admisible, pausible) dada por:

$$PL(S) = 1 - BEL(\neg S)$$

de donde el intervalo de creencia o certidumbre queda como $[BEL(S), PL(S)]$.

A su vez, se define el grado de duda en S como el grado de creencia en la negación de S tal que $D(S) = Bel(\neg S)$

Y la incertidumbre de S o ignorancia es: $PL(S) - BEL(S)$



$Bel(A)$ es el grado mínimo de creencia que aporta la evidencia para A, mientras $Pl(A)$ es el grado máximo de creencia que podemos tener en A, dado que se obtiene sumando las masas de toda la evidencia que no es incompatible con A. Entre ambos valores se encuentra un intervalo que representa la incertidumbre o ignorancia que tenemos en la situación. La representación probabilística no admite esto ya que $P(A)$ determina $P(\neg A)$ y ambas suman 1.

En el enfoque de Bayes, la probabilidad es un número, mientras que en la teoría de Dempster-Shafer las probabilidades se expresan por intervalos de creencia. Cuando los intervalos se estrechan, ambas teorías tienden a coincidir.

Ejemplo: Un proceso industrial sólo puede tener cuatro modos de fallo denominados A, B, C y D. Las masas de probabilidad asignadas por el experto son:
$m(\emptyset) = 0, m(A) = 0.7, m(B) = 0.05, m(C) = 0.02, m(D) = 0.03,$
$m(A, B) = 0.1, m(A, C) = 0.0, m(A, D) = 0.05, m(B, C) = 0, m(B, D) = 0, m(C, D) = 0,$
$m(A, B, C) = 0, m(A, B, D) = 0.05, m(A, C, D) = 0, m(B, C, D) = 0, m(A, B, C, D) = 0$

Creencias:

$Bel(A) = m(A) = 0.7,$

$Bel(AB) = m(A) + m(B) + m(AB) = 0.85,$

$Bel(AD) = m(A) + m(D) + m(AD) = 0.78$

Plausibilidad:

$Pl(A) = 1 - m(B) - m(C) - m(D) - m(BC) - m(BD) - m(CD) - m(BCD)$
$= 1 - 0.05 - 0.02 - 0.03 = 0.9$
$Pl(B) = 1 - m(A) - m(C) - m(D) - m(AC) - m(AD) - m(CD) - m(ACD)$

$= 1 - 0.7 - 0.02 - 0.03 - 0.05 = 0.2$
$Pl(BC) = 1 - m(A) - m(D) - m(AD) = 1 - 0.7 - 0.03 - 0.05 = 0.22$

Incertidumbres:

$Pl(A) - Bel(A) = 0.9 - 0.7 = 0.2$

$Pl(B) - Bel(B) = 0.2 - 0.05 = 0.15$

Intervalos de creencia de $A : [0.7, 0.9]$

Intervalo de creencia de $ACD : [0.8, 0.95]$

Conjunto	Creencia	Plausibilidad
A	0.7	0.9
B	0.05	0.2
AB	0.85	0.95
C	0.02	0.02
AC	0.72	0.92
BC	0.07	0.22
ABC	0.87	0.97
D	0.03	0.13
AD	0.78	0.93
BD	0.08	0.28
ABD	0.98	0.98
CD	0.05	0.15
ACD	0.8	0.95
BCD	0.1	0.3
ABCD	1	1

Ejemplos: Supongamos que estamos frente a un dado que va a ser lanzado. En ese caso $\Theta = O_1, \ldots, O_6$ donde O_i representa que sale el número i.

Caso 1. Si consideramos que el dado es «normal», asignamos una masa igual a $1/6$ a cada O_i y queda totalmente determinada la función de creencia, que en este caso coincide con la probabilidad aplicando principio de indiferencia.

Caso 2. El dado es sospechoso, y tenemos la certeza de que tiene un sesgo por el cual el 1 sale más frecuentemente que el 2, o al revés, pero que los otros números salen aproximadamente un sexto de las veces que se tira. Nuestro estado de incertidumbre se puede representar asignando una masa de $1/6$ a cada uno de O_3, O_4, O_5 y O_6, una masa de Θ a O_1 y O_2 (porque no tenemos absolutamente ninguna evidencia sobre estos valores particulares) y una masa de $1/3$ a O_1, O_2. Se puede pensar que esta última masa está disponible para ser repartida potencialmente entre O_1 y O_2 en cualquier proporción, que corresponda al sesgo.

Caso 3. Creemos que el dado tiene un sesgo desconocido, pero que no llega a reducir la frecuencia de ningún resultado por debajo de $1/10$. Podemos asignar una masa de $1/10$ a cada singleton $O_1, \ldots, O_6$ y una masa de $2/5$ a Θ.

La teoría de Dempster-Shafer cuenta, entre sus principales novedades, una forma de combinar evidencia. Shafer llama «regla de combinación de Dempster» a una forma de combinar distintas evidencias a través de su «suma ortogonal». Esa operación se denota con $\bigoplus$.

Definición. Supongamos que Bel_1 y Bel_2 son funciones de creencia sobre el mismo marco Θ con asignaciones de probabilidad básica m_1 y m_2 y elementos focales $A_1, \ldots, A_k$ y $B_1, \ldots, B_n$ respectivamente. La asignación de probabilidad básica de la función de creencia combinada $Bel_1 \bigoplus Bel_2$ es:

$$m_1 \bigoplus m_2(X) = \frac{\sum_{A_i \cap B_j = X} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = 0} m_1(A_i) m_2(B_j)}$$

donde el denominador debe ser mayor que 0.

Supongamos que una persona perdida en un bosque encuentra unas frutas y quiere evaluar si son buenas para comer. Las frutas son rojas y tienen pelusa. El sujeto sabe que el 60% de las frutas rojas son buenas para comer, así como el 70% de las frutas que tienen pelusa. Representemos «ser bueno para comer» como C. Una evidencia se representa con una función de masa m_r que asigna masa 0.6 a C y 0.4 a $\neg C$. La otra evidencia se representa con una función de masa m_p que asigna masa 0.7 a C y 0.3 a $\neg C$. (Porque el sujeto piensa en términos de probabilidades). La combinación de la evidencia de acuerdo a la regla de Dempster nos da $m_r \bigoplus m_p = 0.78$.

¿Cómo se justifica la función de creencia? Así como las filosofías de la probabilidad justifican los axiomas de la probabilidad interpretando esta de diversas maneras, del mismo modo se puede justificar la función de creencia de Dempster-Shafer. Veamos una justificación posible.

Supongamos que un agente sabe que recibirá un mensaje de la forma «p es verdadero» para algún p no contradictorio. Sea $P(p^*)$ la probabilidad (posiblemente subjetiva) de que el mensaje sea «p^* es verdadero» (o una sentencia lógicamente equivalente a p^*). Entonces la probabilidad de que el agente aprenda que q es verdadero es igual a la probabilidad de que reciba un mensaje «p^* es verdadero» con p^* que implica lógicamente a q. Entonces, si interpretamos la creencia en p^* como la probabilidad de aprender la verdad de p^* obtenemos una función de creencia Dempster-Shafer.

Probabilidad de que el agente aprenda que q es verdadero $= \sum_{p^* |= q} P(p*)$

Entonces, si interpretamos la creencia en p^* como la probabilidad de aprender la verdad de p^* obtenemos una función de creencia Dempster-Shafer.

Sea m_1 una función de asignación básica de probabilidad sobre un marco de discernimiento H, y m_2 otra función de asignación básica de probabilidad también sobre H pero con evidencia añadida o actualizada.

El resultado de combinar ambas funciones m_1 y m_2 es:

$$m_3(X) = m_1 \bigoplus m_2(X) = \frac{\sum_{A_i \cap B_j = X} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

donde X, A_i, B_j son subconjuntos del marco de discernimiento. El denominador es un factor de normalización para que m_3 esté en $[0, 1]$.

A medida que hay más evidencia común se da mayor soporte a las posibles hipótesis (numerador). Por otra parte, la cantidad de creencia medida por

$$C = \sum_{A_i \cap B_j = \varnothing} m_1(A_i) m_2(B_j)$$

mide el conflicto existente entre evidencias que soportan las mismas hipótesis. Si $C = 1$, entonces $m_3(X)$ no se puede definir. Es una función de asignación de probabilidad contradictoria

Ejemplo: Sean cuatro personas en una habitación cerrada: Bob, Jim, Sally y Karen. De repente se va la luz y menos de un minuto después se recupera. En ese momento los miembros de la sala ven que Karen ha sido asesinada víctima de una puñalada. Watson llega a la escena del crimen y deduce que uno de los siguientes hechos debe de ser cierto o su combinación:

Bob es culpable

Jim es culpable

Sally es culpable

Watson aplica la teoría de DS y tiene el siguiente marco de discernimiento: $H = \{B, J, S\}$

Su conjunto potencia será: $H = [\{\varnothing\}, \{B\}, \{J\}, \{S\}, \{B, J\}, \{B, S\}, \{J, S\}, \{B, J, S\}]$,

Supongamos que Watson asigna las siguientes masas de probabilidad a las diferentes hipótesis:

$m(\{\varnothing\}) = 0, m(\{B\}) = 0.1, m(\{J\}) = 0.2, m(\{S\}) = 0.1,$

$m(\{B, J\}) = 0.1, m(\{B, S\}) = 0.1, m(\{J, S\}) = 0.3, m(\{B, J, S\}) = 0.1$

Creencias:

		Bel
{∅}	0	0
{B}	0.1	0.1
{J}	0.2	0.2
{S}	0.1	0.1
{B,J}	0.1	0.4
{B,S}	0.1	0.3
{J, S}	0.3	0.6
{B,J,S}]	0.1	1.0

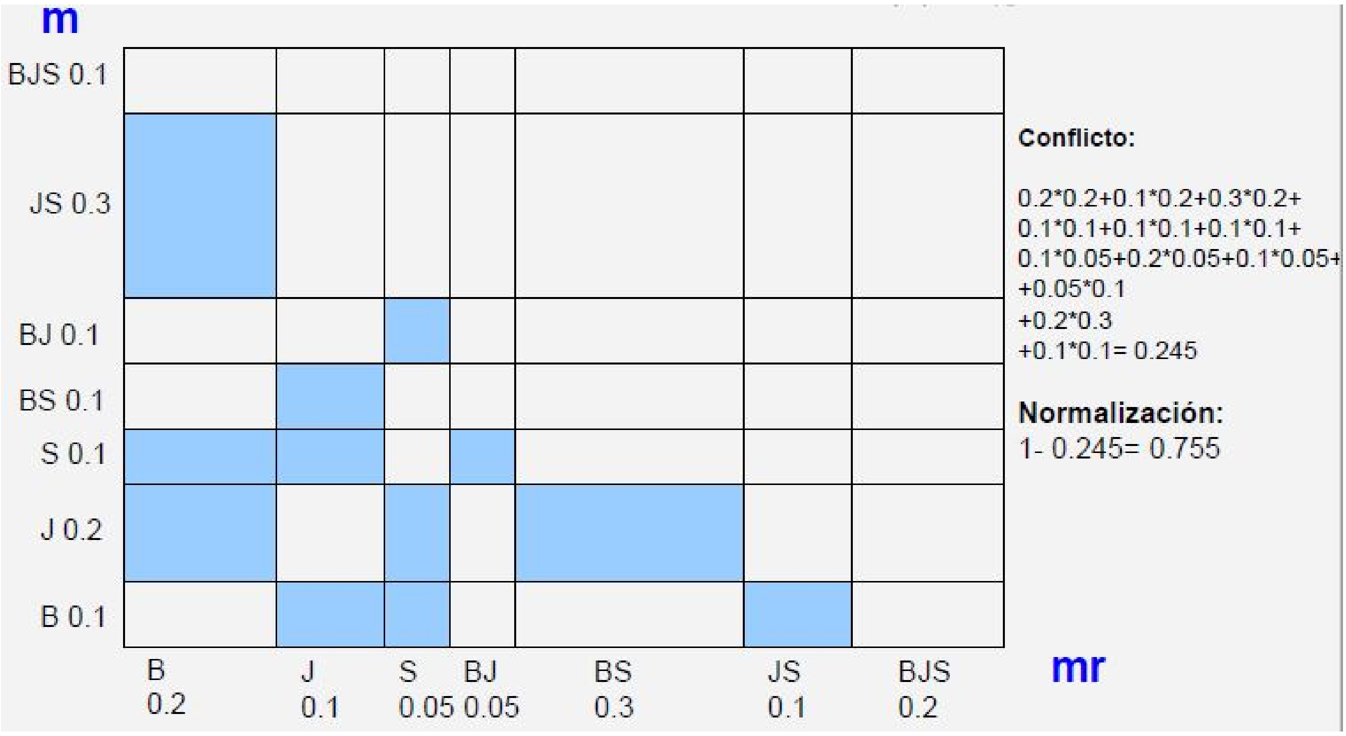
Watson encuentra unas colillas en un cenicero y unas facturas que antes no había visto y revisa el caso, por lo que reasigna las masas de probabilidad ante las nuevas evidencias:

$mr(\{\varnothing\}) = 0, mr(\{B\}) = 0.2, mr(\{J\}) = 0.1, mr(\{S\}) = 0.05,$

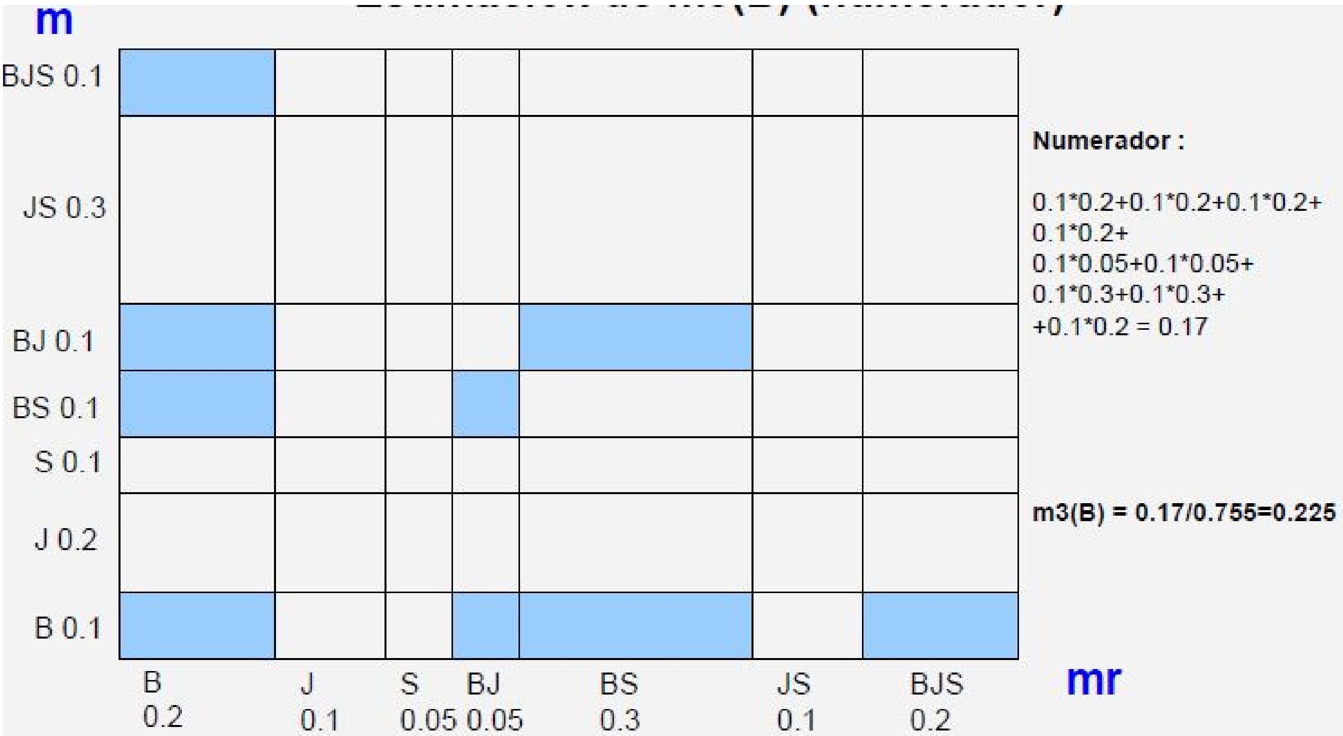
$mr(\{B, J\}) = 0.05, mr(\{B, S\}) = 0.3, mr(\{J, S\}) = 0.1, mr(\{B, J, S\}) = 0.2$

¿Cuál sería la nueva masa de probabilidad combinada de ambas?

Estimación del conflicto $A_i \cap B_j = \varnothing$



Estimación del conflicto (numerador)



Nuevas masas y creencias combinadas

	m	Bel
{∅}	0	0
{B}	0.225	0.225
{J}	0.219	0.219
{S}	0.25	0.25
{B,J}	0.105	0.58
{B,S}	0.04	0.484
{J, S}	0.131	0.6
{B,J,S}]	0.03	1.0

Bibliografía

- Evidencia (https://es.wikipedia.org/wiki/Evidencia)
- Evidence (https://en.wikipedia.org/wiki/Evidence)
- Teoría de Dempster-Shafer (https://es.wikipedia.org/wiki/Teor%C3%ADa_de_Dempster-Shafer)
- Teoría de Dempster-Shafer (https://www.fhce.edu.uy/images/Instituto_de_filosofia/Departamento_logica/cursos/seminario2012/slides shafer.pdf)
- Teoría De Dempster y Shafer (https://www.u-cursos.cl/ingenieria/2008/1/EM755/1/material_docente/bajar?id_material=176243)