TEORÍA DE DEMPSTER-SHAFER

Kyburg/Halpern

Paradoja de Ellsberg
(Halpern, 2003, con modificaciones)

En una caja hay **100** bolillas. Saben que **30** son **azules**, y las **70** restantes **rojas** o **verdes** en cualquier proporción y no se tiene más información. Se extraerá una bola de la caja, revolviendo y sin mirar. Ustedes deben decir el color de la bola que saldrá, y si erran, les sucederá algo desagradable. ¿Qué color dirían?

Solución probabilística

Antes que nada, ¿qué filosofías de la probabilidad nos indican que la probabilidad provee una respuesta unívoca al problema?

Si hay una respuesta unívoca, solo puede ser porque en algún punto se aplicó el principio de indiferencia. Esto puede ser apoyado por la teoría clásica, por la teoría lógica, y por el bayesianismo objetivo.

Bayesianismo objetivo
(no visto en el seminario)

Si le interesa estudiar bayesianismo objetivo, una teoría de gran importancia y muy en boga, puede encontrar una amplia exposición en: Williamson, J. (2010). *In Defense of Objective Bayesianism* que se encuentra disponible en la plataforma EVA.

Principios del bayesianismo objetivo:

1.Los grados de creencia de un agente racional deben conformar los axiomas de la teoría de la probabilidad.

2. Los grados de creencia de un agente racional deben calibrarse con al evidencia disponible.

3. En todo lo que no quede determinado por 1 y 2, se debe maximizar la entropía de los grados de creencia (esto es equivalente a aplicar el principio de indiferencia, aunque algo más general).

Teoría de la decisión probabilista sobre la paradoja de Ellsberg

Volviendo a nuestro problema, y decididos a aplicar el principio de indiferencia, llamando al suceso de la salida de una bola de un color por la inicial del nombre de ese color, tenemos:

$P(A)=0,3$; $P(R \vee V)=0,7$.

Como $P(R \wedge V)=0$, tenemos $P(R)+P(V)=0,7$. Aplicando principio de indiferencia (¿qué otra cosa se puede hacer?) llegamos a $P(R)= P(V)= 0,35$.

Premio por acierto:0 Error:-1

	Sale A	Sale R	Sale V
Digo A	**0**	**-1**	**-1**
Digo R	**-1**	**0**	**-1**
Digo V	**-1**	**-1**	**0**

Calculemos el valor esperado de cada una de las acciones:

E(digo A) = P(A)Premio(digo A/A) +

+P(R)Premio(digo R/A) +

+P(V)Premio(digo V/A)=

=0(0,03)+(-1)0,35+(-1)0,35= -0,7

Análogamente:

E(digo R) = (-1)0,3+0(0,35)+(-1)0,35= -0,65

E(digo V) = (-1)0,3+(-1)0,35+0(0,35)= -0,65

Por el principio de la maximización del valor esperado deberíamos elegir decir Rojo o Verde indistintamente, pero no Azul.

«A pesar de su amplia aceptación, hay algunos problemas al usar probabilidades para representar incertidumbre. Tres de los más serios son (1) la probabilidad no es buena para representar ignorancia, (2) mientras un agente puede estar preparado para asignar probabilidades a algunos conjuntos, puede no estar preparado para asignar probabilidades a todos los conjuntos y (3) mientras un agente puede desear en principio asignar probabilidades a todos los conjuntos en algún álgebra, computar esas probabilidades requiere un esfuerzo computacional; puede no disponer de los recursos computacionales requeridos para hacerlo».

Halpern

Teoría de funciones de creencia de Dempster-Shafer

«Una de las singularidades del principio de indiferencia –el principio según el cual si tenemos información similar sobre dos alternativas, debe asignarse a ambas la misma probabilidad- es que otorga las mismas grandes probabilidades a un par de alternativas acerca de las cuales no sabemos nada igualmente que a los resultados posibles de una moneda rigurosamente balanceada y testeada. Una forma de capturar esa diferencia es ofrecida por Glen Shafer en A Mathematical Theory of Evidence (1976). 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 T pero también un pequeño grado de apoyo, digamos 0,1, a su negación $\neg T$.»

Kyburg & Choh Man Teng

Funciones de creencia y funciones de masa

Marco de discernimiento Θ: es un conjunto de mundos posibles, o modelos con un dominio dado.

El nombre fue elegido por Shafer «para enfatizar la naturaleza epistémica del conjunto de posibilidades».

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 $\mathscr{P}(\Theta)$.

Función de masa o asignación de probabilidad básica

Dado un Θ, nos interesa la función m de dominio $\mathscr{P}(\Theta)$ y codominio [0,1]. La cantidad m(X) representa la *masa asignada al subconjunto X de Θ* 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 llama *asignación de probabilidad básica*.

Definición. Si Θ es un marco de discernimiento, la función m: $\mathscr{P}(\Theta) \rightarrow [0,1]$ es una asignación de probabilidad básica si y solo si

1. $m(\emptyset)=0$
2. $\sum_{X \subseteq \Theta} m(X) = 1$

Los subconjuntos a los que se le asigna masa positiva se llaman «elementos focales».

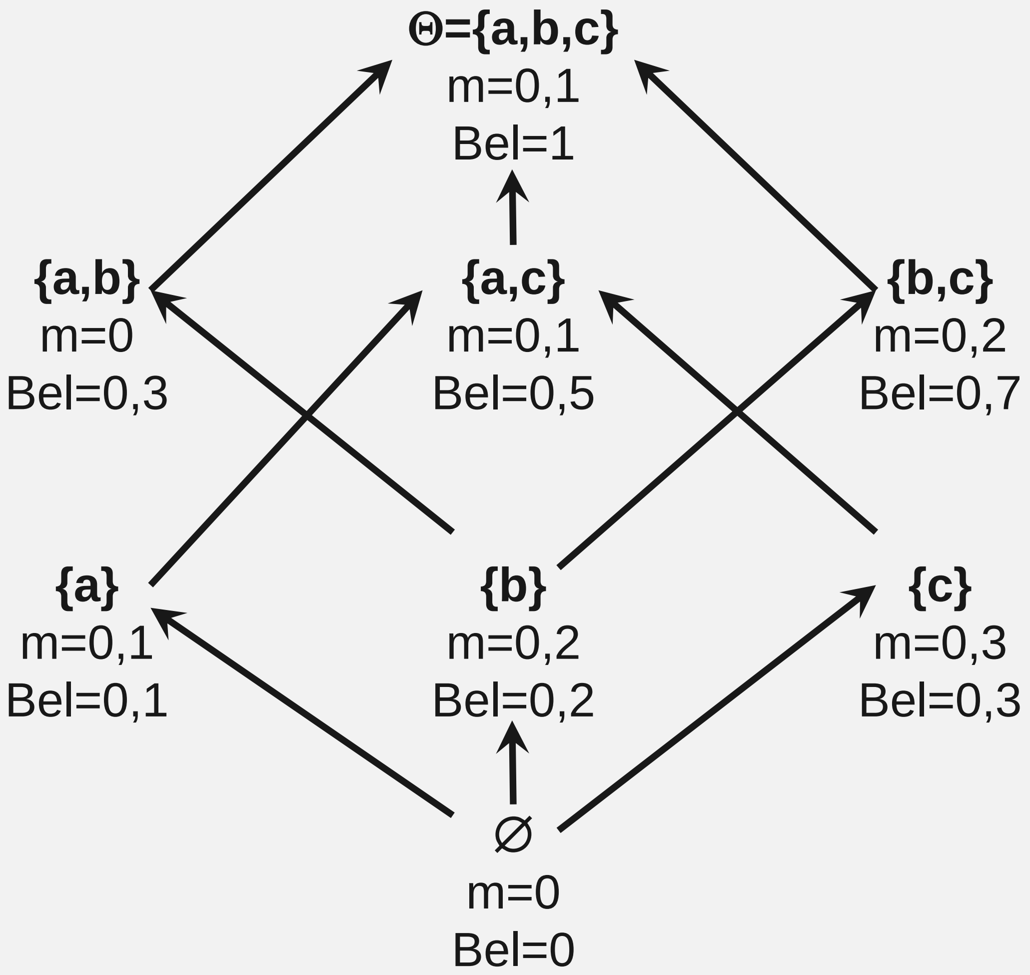
Función de creencia

La creencia asignada a un conjunto X es la suma de todas las masas asignadas a X y a sus subconjuntos

Bel(X)= $\sum_{A \subseteq X} m(A)$

Bel(X) se llama «grado de creencia en X».

No se puede representar la relación entre masa y grado de creencia con diagramas de Venn. Una representación adecuada surge de los diagramas de Hasse (que representa la inclusión). Supongamos Θ = {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) $Bel(A_1 \cup \ldots \cup A_n) \geq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right)$

Obviamente, dada una función de creencia, se puede deducir la función de masa asociada a ella.

En general se cumple

Para todo A⊆Θ, $m(A) = \sum_{X \subseteq A} (-1)^{|A-X|} Bel(X)$

Grado de duda –
Grado de Plausibilidad

Se define el grado de duda en A como el grado de creencia en la negación de A

- **D(A) = Bel (¬A)**

Se define el grado de plausibilidad de A como

- **Pl(A) = 1 – D(A)**

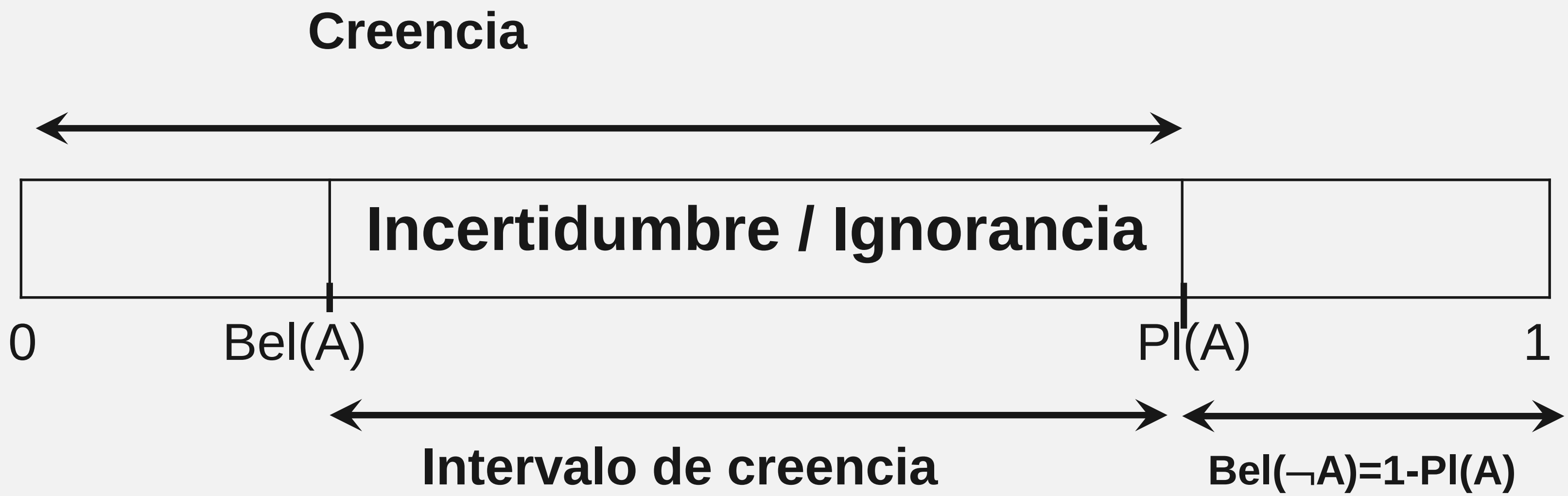
Se tiene que **Bel (A)** ≤ **Pl(A)** para todo A.

Demostración: Por Axioma 3 de las funciones de creencia, se tiene que 1 ≥ Bel(A)+Bel(¬A), de donde

1-Bel (¬A) ≥ Bel(A) y de aquí aplicando las definiciones de D(A) y Pl(A) llegamos a la tesis.

Intervalo de creencia

Como Bel (A) ≤ Pl(A), tiene sentido definir el intervalo [Bel (A) , Pl(A)]. Este intervalo será llamado intervalo de creencia. Representación gráfica:



Bel(A) es el grado mínimo de creencia que aporta la evidencia para A. 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(¬A) y ambas suman 1. A diferencia de en la probabilidad, la creencia en los singleton no determina la función Bel.

Ejemplos

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

Caso 1. Si consideramos que el dado es «normal», asignamos una masa igual a 1/6 a cada Oi 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 0 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.

Combinación de evidencia

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 $\oplus$.

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 $Bel_1 \oplus Bel_2$ es

$$m_1 \oplus m_2(X) = \frac{\sum_{A_i \cap B_j = X} m_1(A_i) m_2(B_j)}{1 - \sum_{A_1 \cap B_j = 0} m_1(A_i) m_2(B_j)}$$

siendo el denominador mayor que 0.

La condicionalización de Dempster es prácticamente incompatible con la interpretación de grados de creencia como relacionada con frecuencias observadas. 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 $m_r \oplus m_p = 0{,}78$.

Otra consecuencia contraintuitiva de la condicionalización de Dempster es que si tenemos dos fuentes de evidencia sobre el mismo marco de discernimiento de forma que las funciones de masa sean idénticas, al combinarlas obtenemos –en general- una función de masa distinta a la original.

¿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.

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

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