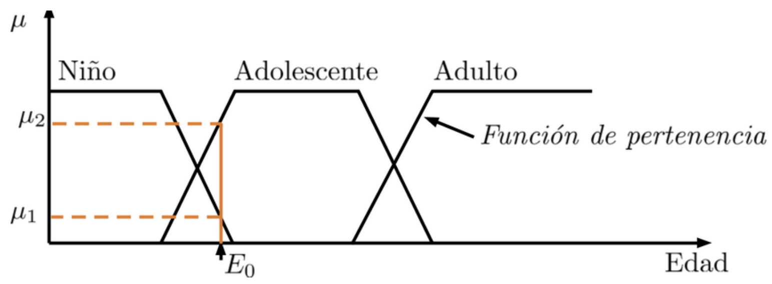
RCRA: Lógica Difusa



Índice

Pagina

.

1. Conjuntos Difusos

1.1. Definición y ejemplos

La diferencia entre los conjuntos ordinarios y los conjuntos difusos es la siguiente:

- **Conjuntos ordinarios**: Dado un universo U y un subconjunto A de ese universo, existen una función que determina exactamente si un elemento x está (1) o no está (0) en el subconjunto A.

$$U, A \subseteq U / \exists \mu_A(x) : U \to \{0, x \notin A : 1, x \in A\}, \forall x \in U \tag{1}$$

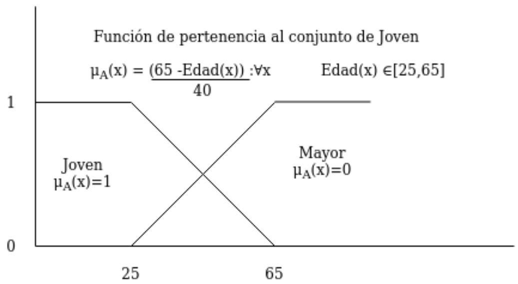
- **Conjuntos difusos (Lofti Zadeh, 1965)**: Dado un universo U y un subconjunto A de ese universo, existe una función $\mu_A(x)$ que determina el grado de pertenencia de un elemento x a dicho subconjunto A, siendo 1 pertenecer totalmente al conjunto y siendo 0 no pertenecer en absoluto y siendo (0,1) pertenecer y no pertenecer al conjunto.

$$\begin{aligned} U, A \subseteq U / \exists \mu_A(x) : U \to [0,1], \forall x \in U \\ \mu_A(x) = 1 \to x \in A \\ \mu_A(x) = 0 \to x \notin A \\ 0 < \mu_A(x) < 1 \to (x \notin A) \wedge (x \in A) \end{aligned} \tag{2}$$

Aclaraciones:

- La función difusa no tiene porque ser lineal (Puede exponencial p.e).
- La escala lingüística es arbitraria (Se explica más adelante).
- El número de elementos de la escala es arbitrario, hasta cierto punto, porque se demostró que una clasificación más fina que 7 elementos lingüísticos es inútil (Insertar quien lo demostró 1984).
- Podemos definir conjuntos complementarios (Muy joven - Poco viejo).
- Cualquier conjunto es difuminable: Los conjuntos ordinarios son un subconjunto de los conjuntos difusos (Si hacemos la parte difusa infinitamente pequeña, se volverá un conjunto ordinario).

Ej: Eres 100 % joven si tienes menos de 25 años (Renfe Joven) y 100 % mayor si tienes más de 65 (Renfe Oro).



- $Edad(Marisa) = 31 \to \mu_{joven}(Marisa) = 0{,}85$ joven
- $Edad(Ana) = 57 \to \mu_{joven}(Ana) = 0{,}20$ joven
- $Edad(Alex) = 73 \to \mu_{joven}(Ana) = 0{,}00$ joven

Semánticamente no hablamos así, si no que usamos escalas lingüísticas arbitrarias como las siguientes:

- No es $\mu_{joven}(x) = 0{,}00$
- Es poco $0{,}00 < \mu_{joven}(x) \leq 0{,}40$
- Es algo $0{,}40 < \mu_{joven}(x) < 0{,}80$
- Es bastante $0{,}80 \leq \mu_{joven}(x) < 1{,}00$
- Es $\mu_{joven}(x) = 1{,}00$

1.2. Caracterización de Conjuntos Difusos

- **Caracterización de Conjuntos Ordinarios:**
  - Implícitamente: En función del dominio (Ej. : Naturales pares menores a 10).
  - Explícitamente: Identificando sus elementos (Ej. : $\{2, 4, 6, 8\}$).
  - Mediante una función booleana que de 0 o 1.

  Existe una simetría total en los sistemas ordinarios entre aritmética, teoría de conjuntos, lógica. Ej. : Lógica (AND, OR), aritmética (X,+) y teoría de conjuntos (Intersección, unión)

- **Caracterización de Conjuntos Difusos (Importante):**

$$\forall A \subset U : A_{difuso} \leftrightarrow \exists \mu_A(x) : U \to [0, 1], \forall x \in U \tag{3}$$

  Para todo A incluido en U, A es difuso si y solo si, existe una función de grado de pertenencia que aplicada sobre x en el universo U le asoció un valor entre $[0, 1]$, si no, no es difuso.

  ¿Cumplen los conjuntos difusos la estructura de un álgebra de Boole?

  - **Conjuntos vacío**: Dado un conjunto difuso Z, Z es igual al vacío si y solo si la función de pertenencia a Z para cualquier x es igual a 0.

$$\begin{aligned} Z \subset U / \exists \mu_z(x) : U \to [0, 1], \forall x \in U \\ Z = \emptyset \leftrightarrow \exists \mu_z(x) = 0, \forall x \in U \end{aligned} \tag{4}$$

  - **Identidad**: Dado dos conjuntos difusos A y B pertenecientes al universo U, A y B son iguales si su función de pertenencia da el mismo resultado para toda $x \in U$.

$$\begin{aligned} A \subset U / \exists \mu_A(x) : U \to [0, 1], \forall x \in U \\ B \subset U / \exists \mu_B(x) : U \to [0, 1], \forall x \in U \\ A = B \leftrightarrow \mu_B(x) = \mu_A(x) : \forall x \in U \end{aligned} \tag{5}$$

    Pregunta metafísica random: Si coinciden los elementos del conjuntos, ¿Son el mismo? ¿Esta clase es el mismo conjunto que si estamos todos en la cafetería porque coinciden todos sus elementos? Pues para los matemáticos sí, para el profesor de RCRA no. (Creo que el punto es que la nomenclatura es importante)

  - **Complementariedad**: Un conjunto A' es complementario a otro A si su función de grado de pertenencia cumple que $\mu_{A'}(x) = 1 - \mu_A(x)$ para cualquier x del universo.

$$\begin{aligned} A \subset U / \exists \mu_A(x) : U \to [0, 1], \forall x \in U \\ \neg A = A' \leftrightarrow \mu_{A'}(x) = 1 - \mu_A(x) : \forall x \in U \end{aligned} \tag{6}$$

  - **Inclusión:** Cuando la función de pertenencia de un subconjunto difuso B es totalmente menor o igual la de un subconjunto difuso A, se dice que B está incluido en A.

$$\begin{aligned} A \subset U / \exists \mu_A(x) : U \to [0, 1], \forall x \in U \\ B \subset U / \exists \mu_B(x) : U \to [0, 1], \forall x \in U \\ B \subset A \leftrightarrow \mu_B(x) \leq \mu_A(x) : \forall x \in U \end{aligned} \tag{7}$$

    - Ej. 1: Al conjunto muy alto Pepe pertenece un 0.5, al conjunto alto Pepe pertenece un 0.7. Muy alto está incluido en alto porque para todo x (Pepe, María, ...) es menor o igual.

    - Ej. 2:

$$\begin{aligned} U = \{1, 2, 3, 4, 5, 6\}, A \subset U / A = \{1, 2, 3, 4\}, B \subset U / B = \{1, 3\} \\ \mu_A(x) = 1 + 1/2 + 1/3 + 1/4 + 0/5 + 0/6 \\ \mu_B(x) = 1 + 0/2 + 1/3 + 0/4 + 0/5 + 0/6 \\ 1/1 = 1/1 : 1/2 > 0/2 > 1/3 = 1/3 : 1/4 > 0/4 : 0/5 = 0/5 : 0/6 = 0/6 \end{aligned} \tag{8}$$

- **Unión:** Decimos que C es igual a la unión A y B si para todo $x \in U$ está entre el máximo de $\mu_A(x)$ y $\mu_B(x)$.

$$\begin{aligned} A \subset U / \exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U / \exists \mu_B(x) : U \to [0,1], \forall x \in U \\ C \subset U / \exists \mu_C(x) : U \to [0,1], \forall x \in U \\ C = A \cup B \leftrightarrow \mu_c(x) = max\{\mu_A(x), \mu_B(x)\} : \forall x \in U \end{aligned} \tag{9}$$

- **Intersección:** Decimos que C es igual a la intersección A y B si para todo $x \in U$ está entre el mínimo de $\mu_A(x)$ y $\mu_B(x)$.

$$\begin{aligned} A \subset U / \exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U / \exists \mu_B(x) : U \to [0,1], \forall x \in U \\ C \subset U / \exists \mu_C(x) : U \to [0,1], \forall x \in U \\ C = A \cap B \leftrightarrow \mu_c(x) = min\{\mu_A(x), \mu_B(x)\} : \forall x \in U \end{aligned} \tag{10}$$

  Tanto la intersección como la unión son **asociativas**: $A \cap (B \cap C) = (A \cap B) \cap C$

- **Leyes DeMorgan:**

  - 1º Ley:

$$\neg(A \cup B) = \neg A \cap \neg B \equiv \neg(\mu_A \vee \mu_B) = \neg\mu_A \wedge \neg\mu_B : \forall x \in U \tag{11}$$

  - 2º Ley:

$$\neg(A \cap B) = \neg A \cup \neg B \equiv \neg(\mu_A \wedge \mu_B) = \neg\mu_A \vee \neg\mu_B : \forall x \in U \tag{12}$$

  Demostración:

$$\begin{aligned} \neg(A \cup B) \to \mu_{\neg(A\cup B)} = 1 - max\{\mu_A, \mu_B\} : \forall x \in U \\ \neg A \cap \neg B \to \mu_{\neg A \cap \neg B}(x) = min\{1-\mu_A, 1-\mu_B\} : \forall x \in U \end{aligned} \tag{13}$$

$$\begin{aligned} \mu_A(x) \geq \mu_B(x) : \forall x \in U \\ 1 - max\{\mu_A(x), \mu_B(x)\} = 1 - \mu_A(x) : \forall x \in U \\ min\{1-\mu_A(x), 1-\mu_B(x)\} = 1 - \mu_A(x) : \forall x \in U \\ \mu_A(x) \leq \mu_B(x) : \forall x \in U \\ 1 - max\{\mu_A(x), \mu_B(x)\} = 1 - \mu_B(x) : \forall x \in U \\ min\{1-\mu_A(x), 1-\mu_B(x)\} = 1 - \mu_B(x) : \forall x \in U \end{aligned} \tag{14}$$

- **Distributiva**

  - 1º Ley:

$$\begin{aligned} C \cup (A \cap B) = (C \cup A) \cap (C \cup B) \\ \mu_C(x) \vee [\mu_B(x) \wedge \mu_A(x)] = [\mu_B(x) \vee \mu_C(x)] \wedge [\mu_A(x) \vee \mu_C(x)] \end{aligned} \tag{15}$$

  - 2º Ley:

$$\begin{aligned} \cap(A \cup B) = (C \cap A) \cup (C \cap B) \\ \mu_C(x) \wedge [\mu_B(x) \vee \mu_A(x)] = [\mu_B(x) \wedge \mu_C(x)] \vee [\mu_A(x) \wedge \mu_C(x)] \end{aligned} \tag{16}$$

Conjuntos difusos no son álgebra del Boole porque no cumplen las siguientes leyes:

- **Ley del tercero excluido:**
  - Conjuntos ordinarios: Unión de un conjunto con su complementario es todo el universo $A \cup A' = U$.
  - En conjuntos difusos: La función de pertenencia de un conjunto unida a su complementario es mayor o igual a 1/2 ya que te quedas con el mayor en la operación de unión. En el caso que los dos conjuntos pertenezcan por igual, será 1/2.
- **Ley de no contradicción:**
  - Conjuntos ordinarios: La intersección de un conjunto con su complementario es el vacío $A \cap A' = \emptyset$.
  - En conjuntos difusos: La función de pertenencia de un conjunto intersecada a su complementario es menor o igual a 1/2 ya que te quedas con el menor en la operación de unión.
- **Relación producto - Intersección - AND**: El producto rompe la correspondencia entre lógica, aritmética y conjuntos. En conjuntos ordinarios $A \times B = A \cap B$ sin embargo, con conjuntos difusos el producto de dos número menores a cero siempre es menor a la intersección.

$$\begin{aligned} A \subset U/\exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U/\exists \mu_B(x) : U \to [0,1], \forall x \in U \\ A \times B \to \mu_{AB}(x) = \mu_A(x) \times \mu_B(x) : \forall x \in U \end{aligned} \tag{17}$$

Si un la función de pertenencia de un conjunto B es siempre menor a la de otro está contenido A, se dice que $A \cap B$. Es decir, el producto está contenido en la intersección:

$$\begin{aligned} A \subset U/\exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U/\exists \mu_B(x) : U \to [0,1], \forall x \in U \\ A \times B \to \mu_{AB}(x) = \mu_A(x) \times \mu_B(x) : \forall x \in U \\ A \cap B \leftrightarrow \mu_c(x) = min\{\mu_A(x), \mu_B(x)\} : \forall x \in U \\ \mu_{AB} < min\{\mu_A(x), \mu_B(x)\} : \forall x \in U \\ A \times B \subset A \cap B \end{aligned} \tag{18}$$

Ej: ¿Qué es más cauteloso usando el producto $A \times B = C$ o la intersección $A \cap B = C$? Salvo que A y B siempre sean uno, cuyo caso da igual, siempre se es más cauteloso con el producto.

- **Relación suma - Unión - OR**: Si definimos la suma como $\mu_A(x) + \mu_B(x)$ puede ser mayor que uno. Por lo que **Lofti Zadeh** definió la suma acotada: trunca los valores mayores a uno.

$$\begin{aligned} A \subset U/\exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U/\exists \mu_B(x) : U \to [0,1], \forall x \in U \\ A + B \to \mu_{A+B}(x) = \mu_A(x) + \mu_B(x) : \forall x \in U \\ A \oplus B \to \mu_{A\oplus B}(x) = min\{1, \mu_A(x) + \mu_B(x)\} : \forall x \in U \end{aligned} \tag{19}$$

- **Diferencia y diferencia absoluta**: Para que no quede fuera con valores negativos.

$$\begin{aligned} A \subset U/\exists \mu_A(x) : U \to [0,1], \forall x \in U \\ B \subset U/\exists \mu_B(x) : U \to [0,1], \forall x \in U \\ A - B \to \mu_{A-B}(x) = \mu_A(x) - \mu_B(x) : \forall x \in U \\ |A - B| \to \mu_{|A-B|}(x) = |\mu_A(x) - \mu_B(x)| : \forall x \in U \end{aligned} \tag{20}$$

- **Núcleo de un conjunto difuso**: Conjunto ordinario asociado al conjunto difuso, aquellos que perteneces totalmente. Se dice que un conjunto está normalizado si su núcleo es diferente conjunto vacío.

$$\begin{aligned} A \subset U/\exists \mu_A(x) : U \to [0,1], \forall x \in U \\ Núcleo : N_A = \{x \in U/\mu_A(x) = 1\} \\ A = Normalizado \leftrightarrow N_A \neq \{\emptyset\} \end{aligned} \tag{21}$$

1.3. Relación difusa

1.3.1. Definición

Dado un referencial U, definimos una relación difusa de orden 'n' en U, como un conjunto difuso A en el espacio $U \times U \times ... \times U$ (n veces), caracterizado por una función de grado de pertenencia del tipo:

$$\mu_A(x_1, ...x_n) : \forall x \in U \tag{22}$$

Ejemplo:

- Referencial U = $\{1, 2, 3, 4\}$
- Relación difusa de orden 2
- A = Números aproximadamente iguales
- Tabla dimensión dos que habla del grado de pertenencia a cada una de las opciones a las otras:

	U2			
U1	**1**	**2**	**3**	**4**
1	1	0.8	0.2	0
2	0.8	1	0.8	0.8
3	0.2	0.8	1	0.8
4	0.0	0.2	0.8	1

1.3.2. Representación del conocimiento

En conjuntos difusos se utilizan:

- **Predicados difusos**: Tráfico ligero, ...
- **Cuantificadores difusos**: La mayoría, algunos, ...
- **Probabilidades difusas**: Muy/algo/poco/ ... probable.

En lógica convencional solo existen dos cuantificadores: existe y para todo.

El conocimiento derivado del sentido común es léxicamente impreciso y no categórico por lo que las lógicas de 1^{er} orden o las teorías clásicas de la probabilidad no son suficientes para representar el sentido común.

Ejemplos (en negrita la parte difusa):

- **Normalmente** se tarda **alrededor** de 45 en llegar desde Oleiros a Santiago, por la autopista, y con **tráfico ligero**.
- **No es previsible** que el paro disminuya en España al menos de **manera drástica**, en los **próximos** meses.
- **La mayoría** de los expertos opinan que **la probabilidad** de un terremoto serio en el Caurel es **pequeña** en un futuro **inmediato.**

1.4. Diferencia entre la imprecisión y la incertidumbre

La imprecisión tiene que ver con el hecho en sí.
Ejemplos:

- La costa gallega es más imprecisa que la levantina debido al efecto de la marea: Una señora va de vacaciones al mediterráneo, deja la toalla en la playa y se va a tomar unas cervezas, cuando vuelve: la toalla sigue igual. Una año más tarde, repite lo mismo pero en Galicia, cuando vuelve la toalla ya no está por culpa de la marea.

La incertidumbre es un aspecto relacionado con las relaciones de implicación ($A \rightarrow B$, en conjuntos difusos no siempre es así), se produce un efecto de arrastre de la imprecisión por medio del razonamiento.

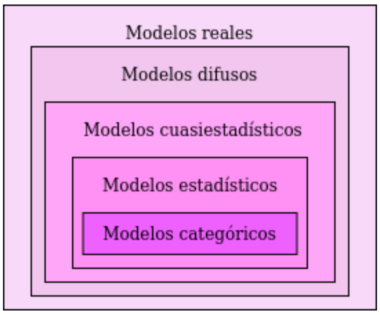
2. Razonamiento Clásico

2.1. Aspectos generales del razonamiento

Punto de partida:

- Debemos formalizar mecanismos y procesos inferenciales, derivar conocimiento nuevo a través del ya existente.
- Establecer y diseñar modelos de razonamiento.

Cada modelo funciona como capas de cebolla ya que se contienen unos a otros:



Los condicionantes del problema del razonamiento son:

- Características del dominio:
  - Dominios categóricos (D1): Aquel en el que a priori no hay ninguna incertidumbre/imprecisión pero en el que van a aparecer de forma natural.
  - Dominios estadísticos (D2): Puramente estadísticos.
  - Dominios cuasiestadísticos (D3): Dominios que siendo tratados de forma estadística, no podemos/queremos obtener datos objetivos.
  - Dominios difusos (D4)
  - Dominios reales: Pueden ser al mismo tiempo cualquiera de los anteriores, por lo que serán descritos como una combinación lineal de los anteriores ($\lambda_1(D1) + \lambda_2(D2) + \lambda_3(D3) + \lambda_4(D4)$).
- Características del problema

2.2. Interpretación diferencial

En dominios de naturaleza puramente simbólica, ¿Cómo podemos definir un procedimiento encadenado y lógico para discriminar entre posibles soluciones candidatas, obtenidas a partir de datos y de verdades demostradas?

El proceso pasa por los siguientes pasos:

- 1 - Recopilación de información
- 2 - Análisis de la importancia relativa de las manifestaciones. Ej: Caída -¿Rotura de cadera.
- 3 - Análisis de posibles causas (Establecimiento tentativo de relaciones causa-efecto). Ej: ¿La caída sucedió porque se rompió la cadera o se rompió la cadera por la caída?
- 4 - Exclusión una a una de las posibles interpretaciones.
- 5 - Fin del proceso: No hay solución, hay solo una o hay varias (Lo que produce incertidumbre).

2.3. Razonamiento Categórico

2.3.1. Elementos del razonamiento categórico

Elementos que constituyen el razonamiento categórico junto con ejemplo meteorológico:

- 1 - Descripción del dominio de discurso, donde tendremos que considerar todas las manifestaciones posibles que dependen del dominio. Ej: Color del crepúsculo, ...
- 2 - Considerar todas las interpretaciones posibles. Ej: Posibilidad de lluvia, de sol, ...
- 3 - Relaciones causa-efecto. Ej: Si el crepúsculo ... Y ... Entonces

Formalmente esto se podría esquematizar como:

X = {manifestaciones individuales} = $\{x_1, ..., x_n\}$
Y = {interpretaciones} = $\{y_1, ..., y_m\}$

En un caso concreto:

- Debemos poder definir una función booleana f del conjunto de manifestaciones $X = f(x_1, ..., x_n)$ tal que dada una i, $f(x_i) = 0 \leftrightarrow x_i$ si no es una manifestación del problema y tal que, $f(x_i) = 1 \leftrightarrow x_i$ si sí es una manifestación del problema.
- Debemos poder definir una función booleana g del conjunto de interpretaciones $Y = g(y_1, ..., y_m)$ tal que dada una j, $g(y_j) = 0 \leftrightarrow y_j$ si no es una interpretación del problema y tal que, $g(y_j) = 1 \leftrightarrow x_j$ si sí es una interpretación del problema.

Debemos hacer compatible las manifestaciones con las interpretaciones. Para ello, vamos a usar las relaciones causales. Estas relaciones causales entre manifestaciones e interpretaciones se formalizan a través de la función de conocimiento E:

$$E = E(X, Y) \tag{23}$$

¿Cual es nuestro problema lógico? Dadas unas manifestaciones caracterizadas por una función f tendremos que encontrar una función g que satisfaga cualquiera de las siguientes:

$$\begin{aligned} E(f) &= g \\ E(\neg g) &= \neg f \\ E &= E(x_1, ...x_n, y_1, ..., y_m) \end{aligned} \tag{24}$$

2.3.2. Ejemplo de razonamiento categórico

A continuación pondremos un **ejemplo del razonamiento categórico sin usar el razonamiento sistemático**:

- Manifestaciones: M = {m(1), m(2)}
- Interpretaciones: I = {i(1), i (2)}
- Relaciones causales:
  - Para que i(2) sea cierta, m(1) debe estar presente: $i(2) \rightarrow m(1)$ [R1]
  - Para que i(1) sea cierta e i(2) sea falsa, m(2) debe estar presente: $i(1) \wedge \neg i(2) \rightarrow m(2)$ [R2]
  - Para que i(2) sea cierta e i(1) sea falsa, m(2) no debe estar presente: $\neg i(1) \wedge i(2) \rightarrow \neg m(2)$ [R3]
  - Determinación/Coherencia. Si algunas de las manifestaciones está presente es porque se puede establecer alguna interpretación: $m(2) \vee m(2) -> i(1) \vee i(2)$ [R4]
- E = {R1, R2, R3, R4}

- En el dominio D aparece una situación en la que m(2) está presente, m(1) está ausente... ¿Cuál es la interpretación lógica?
  - f: f([m(1),m(2)]) = $\neg m(1) \wedge m(2)$
  - E: $(f \rightarrow g)$

    Eliminamos implicaciones:
  - R1: $i(2) \rightarrow m(1) \equiv \neg i(2) \vee m(1)$
  - R2: $i(1) \wedge \neg i(2) \rightarrow m(2) \equiv \neg i(1) \vee i(2) \vee m(2)$
  - R3: $\neg i(1) \wedge i(2) \rightarrow \neg m(2) \equiv i(1) \vee \neg i(2) \vee \neg m(2)$
  - R4: $m(2) \vee m(2) -> i(1) \vee i(2) \equiv [\neg m(1) \wedge \neg m(2)] \vee i(1) \vee i(2)$

  - Teniendo f y E, lo único que nos falta es g, para calcularlo usaremos E (Reglas) y f (manifestaciones):
    - R1: Debemos hacer esta regla cierta y para ello, tenemos que cumplir al menos una de las dos posibilidades del $\vee$, como m(1) no lo tenemos $\neg i(2)$ debe ser cierta. $[\neg i(2)]$
    - R4: Debemos hacer esta regla cierta, el $\wedge$ de esta regla no lo cumplimos, nos queda o bien cumplir i(1) o bien cumplir i(2).Para que R1 sea cierta llegamos a la conclusión de que $\neg i(2)$ por lo que necesitamos que $i(1)$ sea cierta. $[i(1)]$

    Resultado: g: g([i(1),i(2)]) = $\neg i(2) \wedge i(1)$

Si el dominio implicara sólo 30 manifestación, 600 interpretaciones y 125 relaciones causales: el tratamiento lógico convencional sería poco eficiente. Necesitamos otra solución: **Procedimiento sistemático**.

2.3.3. Procedimiento sistemático en el razonamiento categórico

Pasos a seguir para llevarlo a cabo:

- 1 - Identificación de M
- 2 - Identificación de I
- 3 - Construcción de E
- 4 - Construcción del conjunto completo de complejos de manifestaciones
- 5 - Construcción del conjunto completo de complejos de interpretaciones
- 6 - Construcción del conjunto completo de complejos de manifestaciones-interpretaciones

Suponiendo que tenemos n manifestaciones $M = \{m(1), ..., m(n)\}$ y t interpretaciones $I = \{i(1), i(2), ..., i(t)\}$:

- El núm. de complejos de manifestaciones es el producto cartesiano de las posibilidades de las manifestaciones: 2^n posibilidades.
- El núm de complejos de las interpretaciones: 2^t posibilidades.
- El núm. de manifestaciones-interpretaciones es: 2^{n+t} posibilidades.

Conjunto complejo de manifestaciones-interpretaciones: Este conjunto es de carácter exhaustivo y, sus elementos son mutuamente excluyentes. Representa el total de situaciones que son idealmente posibles en el dominio: iremos eliminando posibilidades de este conjunto a través del conocimiento del sistema.

Recuperando el ejemplo anterior pero aplicando el procedimiento sistemático:

- Paso 1 - Manifestaciones: M = {m(1), m(2)}
- Paso 2 - Interpretaciones: I = {i(1), i (2)}
- Paso 3 - Relaciones causales:
  - R1: $i(2) \to m(1)$
  - R2: $i(1) \wedge \neg i(2) \to m(2)$
  - R3: $\neg i(1) \wedge i(2) \to \neg m(2)$
  - R4: $m(2) \vee m(2) - > i(1) \vee i(2)$
- Paso 4 y 5 - Si representamos la situación en términos de complejos (sensible al criterio), podemos construir **M** e **I** que contienen todas las combinaciones posibles de manifestaciones y de interpretaciones, respectivamente.

M = {m1,m2,m3,m4}
I = {i1,i2,i3,i4}

m(1)	0	0	1	1
m(2)	0	1	0	1
	m1	**m2**	**m3**	**m4**

i(1)	0	0	1	1
i(2)	0	1	0	1
	i1	**i2**	**i3**	**i4**

Cada mn es un complejo de manifestaciones que es exhaustivo y excluyente pues sabemos exactamente que manifestaciones se presentan y cuales no. Por ejemplo, con m1 sabemos que es $\neg m(1) \wedge m(2)$ no como en el caso de tener únicamente m(1). Y lo mismo ocurre con las interpretaciones.

Paso 6 - Construir el conjunto completo de complejos manifestación-interpretación (Base Lógica Expandida).

$$BLE = \mathbf{M} \times \mathbf{I} = \begin{pmatrix} m1i1 & m1i2 & m1i3 & m1i4 \\ m2i1 & m2i2 & m2i3 & m2i4 \\ m3i1 & m3i2 & m3i3 & m3i4 \\ m4i1 & m4i2 & m4i3 & m4i4 \end{pmatrix} \tag{25}$$

Dado el carácter exhaustivo de la misma la solución de cualquier problema está siempre en la Base Lógica Expandida pero, hay muchas combinaciones absurdas: el papel del conocimiento E es eliminarlas y pasar a una Base Lógica Reducida ($BLE \to BLR$).

Como podemos ver en la tabla la solución será un par $mn \times it$.

	m1	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**
m(1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m(2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i(1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i(2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	**i1**	**i1**	**i1**	**i2**	**i2**	**i2**	**i2**	**i3**	**i3**	**i3**	**i3**	**i4**	**i4**	**i4**	**i4**

Para descartar opciones deberemos usar el conocimiento E:

- R1 : $i(2) \rightarrow m(1)$. Elimina m1i2, m1i4, m2i2 y m2i4. Ya que estamos en una situación que si está presente i(2) debe estarlo m(1). (Rojo)
- R2: $i(1) \wedge \neg i(2) \rightarrow m(2)$. Elimina m1i3 y m3i3. (Azul)
- R3: $\neg i(1) \wedge i(2) \rightarrow \neg m(2)$. Elimina m2i2 y m4i2. (Verde)
- R4: $m(2) \vee m(2) -> i(1) \vee i(2)$. Elimina m2i1, m3i1 y m4i1. (Amarillo)

	m1	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**
m(1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m(2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i(1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i(2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	**i1**	**i1**	**i1**	**i2**	**i2**	**i2**	**i2**	**i3**	**i3**	**i3**	**i3**	**i4**	**i4**	**i4**	**i4**

En términos de complejos: el problema planteado inicialmente $f = \neg m(1) \wedge m(2) = m2$, ¿Qué complejos de la BLR contienen a m2? m2i3. Por lo que $g = i3 = i(1) \wedge \neg i(2)$. Por lo que sabemos que i(1) es correcta y, además sabemos que i(2) es falsa. Aquí tenemos una única solución.

¡Cuidado! Hay que entender que se está haciendo y no simplemente usar el método: Intentar añadir por ejemplo esta regla R5 $\neg i(1) \wedge \neg i(2) \rightarrow m(1) \vee m(2)$ sería un error porque es incompatible con las anteriores. Elimina m1i1 (Naranja).

	m1	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**	**m1**	**m2**	**m3**	**m4**
m(1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
m(2)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
i(1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
i(2)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	i1	**i1**	**i1**	**i1**	**i2**	**i2**	**i2**	**i2**	**i3**	**i3**	**i3**	**i3**	**i4**	**i4**	**i4**	**i4**

Además, si no tenemos manifestaciones para que queremos interpretaciones. ¿Qué ocurriría ahora si encontramos un problema caracterizado por $f = \neg m(1) \wedge \neg m(2)$? ¿Puede existir este problema? Ese problema, si no hay implicaciones temporales no puede existir. Aún así, con esta BLR este problema si fuera real no tendría solución: habría que pensar que en términos generales ...

- Las manifestaciones no son realmente esas.
- El conocimiento no es correcto.
- El dominio no está bien construido.

Eliminado R5 (Naranja) y volviendo a la BLR inicial. ¿Qué ocurriría si $f'' = m(1) \wedge \neg m(2) = m3$?

En este caso, los complejos de BLR son m3i2 y m3i4. La función $g = i2 \vee i4 = [\neg i(1) \wedge i(2)] \vee [i(1) \wedge i(2)]$, sabemos que i(2) es cierta pero de i(1) no sabemos nada.

2.3.4. Problemas del razonamiento categórico

A pesar de que nuestras reglas de conocimiento E son categóricas $(A \rightarrow B)$, cuando aplicamos el modelo sobre un problema la incertidumbre aparece de forma espontánea: ¿Por qué sucede esto? Porque operamos con conocimiento y que las cosas no son deterministas.

Otro 'problema' es la cantidad de pasos a dar: con 7 manifestaciones y 24 interpretaciones se obliga a discriminar entre 2.147.483.600 complejos. Aunque ahora sí, antes esto no se podía resolver.

2.4. Razonamiento Bayesiano

Antes de este punto es bueno que sepas la diferencia entre probabilidad y estadística: Sección 3 apartado 1. Las interpretaciones categóricas son pocos frecuentes en el mundo real por varias razones:

- La existencia de una determinada causa no siempre conlleva la presencia de una manifestación.
- Ante una manifestación dada no podemos afirmar siempre y de forma categórica que existe una determinada causa.

Por ello usaremos el razonamiento bayesiano.

¿Qué diferencia hay entre probabilidad total y probabilidad condicional?

Dado un universo (BLE) N y una lista de atributos (manifestaciones) $f(x, y, ..., t)$, siendo $N(f) =$ nº elementos del universo N que presentan los atributos f. ¿Cuál sería la probabilidad total $P(f)$ de que un determinado elemento de este universo presente tales atributos?

$$P(f) = \frac{N(f)}{N} \tag{26}$$

Ej: Con nuestro lenguaje categórico, supongamos que tenemos n interpretaciones $(i1, ..., in)$ que definen el complejos de interpretaciones Dj y, N(Dj) es el número de elemento sdel universo cuyo 'problema' es Dj. Sea N el número de elementos en el universo. La probabilidad total del complejo de interpretaciones $P(Dj) = \frac{N(Dj)}{N}$

Sin embargo, no podemos usar únicamente la probabilidad total porque nos faltan las relaciones causales por lo que usaremos la probabilidad condicional, concepto propuesto por el reverendo Bayes.

Probabilidad condicional

En la probabilidad condicional aparecen involucrados dos sucesos, en donde la ocurrencia del segundo depende de la ocurrencia del primero. m12(C)$_{normalizada} = \frac{\sum_{C=A_i \cap B_j} m1(A_i) \times m2(B_j)}{1-K}$

$Consideremosdosaccionesposiblesydosresultadosquedependendelasacciones...$

Acciones: A, B

Resultados: Existo (E), Fracaso (F)

Elegimos la acción por el método de la moneda (Lanzar una moneda y si sale cara opción A y si sale cruz hacemos B). La probabilidad total para cada una de ellas es P(A) = 0.5 y P(B) = 0.5.

Si hacemos estadística:

- La acción A genera el resultado E el 20 % de las veces.
- La acción B genera el resultado E el 60 % de las veces.

Si intentamos un nuevo caso, podemos estimar la probabilidad a priori de un resultado E:

$$\begin{aligned} P(A) = 0{,}5 - P(E/A) = 0{,}2 \\ P(B) = 0{,}5 - P(E/B) = 0{,}6 \\ P(E) = P(E/A) \times P(A) + P(E/B) \times P(B) = 0{,}4 \end{aligned} \tag{27}$$

Recomendación de como leer P(E/A) en la sección 3 apartado 3.

P(E) es la definición estricta de la media estadística, es decir, es un valor medio: la función que usamos para calcular la media estadística es un caso particular de esta expresión.

2.4.1. Teorema de Bayes

Hasta ahora, hemos visto razonamiento a priori, pero Bayes introduce una forma de razonamiento a posteriori. Por lo que el planteamiento deja de ser cuál es la probabilidad de E dado A y pasa a ser, cual es la probabilidad de que E haya sido causado por A o, mejor dicho cuale s la probabilidad de A dado E $P(A/E)$.

$$P(A/E) = \frac{P(E/A) \times P(A)}{P(E)} \tag{28}$$

Esto no es más que una ecuación particularizada del teorema de Bayes.

Debemos obtener la ecuación elemental del teorema de Bayes:

- En términos lingüísticos: Sea una población donde sobre parte de sus elementos se ha realizado una acción A de un número de posibles acciones. Todos los elementos de la población fueron tratados, con A o cualquier otra de las acciones permitidas, registrándose un cierto número de respuestas E.
- En términos matemáticos:
  - N = Población
  - Acciones posibles: $A, \neg A$
  - Resultados posibles: $E, \neg E$
  - $n(A)$: Número de casos A
  - $n(\neg A)$: Número de casos que no A
  - $n(E)$: Número de casos E
  - $n(\neg E)$: Número de casos que no E
  - $n(A \cap E) \equiv n(E \cap A)$

Por definición de probabilidad condicional:

$$P(E/A) = \frac{n(A \cap E)}{n(A)} = \frac{n(A \cap E)/N}{n(A)/N} = \frac{P(A \cap E)}{P(A)} \rightarrow P(A \cap E) = P(A) \times P(E/A) \tag{29}$$

$$P(A/E) = \frac{n(A \cap E)}{n(E)} = \frac{n(A \cap E)/N}{n(E)/N} = \frac{P(A \cap E)}{P(E)} \rightarrow P(A \cap E) = P(E) \times P(A/E) \tag{30}$$

$$P(A) \times P(E/A) = P(E) \times P(A/E) \tag{31}$$

Para obtener una ecuación generalizada del teorema de Bayes:

- X: Característica cualquiera
- A: Número de casos de una población estadística en los que X está presente.
- $\neg A$: Número de casos de la misma población donde X no está presente.
- P: Prueba potencialmente resolutiva para investigar la característica X
- E: Número de casos en los que la prueba da resultados positivos
- $\neg E$: Número de casos en los que la prueba da resultados negativos

Pasado a tabla:

	A	$\neg A$	**TOTAL**
E	a	b	a+b
$\neg E$	c	d	c+d
TOTAL	a+c	b+d	N

Siendo:

- a = Positivos reales
- b = Falsos Positivos
- c = Falsos negativos
- d = Verdaderos negativos

Relaciones:

- $P(E/A) = a/a + c$ (Sensibilidad: Explicado en 3 apartado 4)
- $P(\neg E/A) = c/a + c$
- $P(A/E) = a/a + b$
- $P(\neg A/E) = b/a + b$
- $P(E/\neg A) = b/b + d$
- $P(\neg E/\neg A) = d/b + d$ (Especificidad: Explicado en 3 apartado 4)
- $P(A/\neg E)$ = c/ c+d
- $P(\neg A/\neg E)$ = d/ c+d

Prevalencia, probabilidades totales:

- $P(E) = (a + b)/N$
- $P(\neg E) = (c + d)/N$
- $P(A) = (a + c)/N$
- $P(\neg A) = (b + d)/N$

Tras conocer las probabilidades a priori, ¿Cómo se plantea la probabilidad condicional a posteriori de la característica X dado un resultado positivo de la prueba?

$$\begin{gathered}
Bayes \rightarrow P(A/E) = \frac{P(E/A) \times P(A)}{P(E)} : P(E) = \frac{a+b}{N} \\
a = (a + c) \times P(E/A) : b = (b + d) \times P(E/\neg A) \\
a + c = N \times P(A) : (b + d) = N \times P(\neg A) \\
a = N \times P(A) \times P(E/A) : b = N \times P(\neg A) \times P(E/\neg A) \\
P(E) = P(A) \times P(E/A) + P(\neg A) \times P(E/\neg A) \\
P(A/E) = \frac{P(E/A) \times P(A)}{P(E/A)(A) + P(E/\neg A)P(\neg A)}
\end{gathered} \tag{32}$$

En la última fórmula se relacionan las probabilidades a priori con las probabilidades a posteriori. Es directamente generalizable si consideramos todos los Ai posibles obtenemos:

$$P(A0/E) = \frac{P(E/A0) \times P(A0)}{\sum_i P(E/Ai) \times P(Ai)} \tag{33}$$

Ejemplo anterior categórico:

- BLR = {m1i1, m3i2, m2i3, m4i3, m4i4}
- Dado $f = m(1) \wedge \neg m(2) =$ m3i1, m3i2, m3i3, m3i4.
- ¿Cuál es g?: $g = i2 \vee i4 = [\neg i(1) \wedge i(2)] \vee [i(1) \wedge i(2)]$

Planteándonoslo desde el punto de vista bayesiano, queremos saber cual de las dos respuestas es más probable:

- $P(i2/m3) = \frac{P(m3/i2)\times P(i2)}{P(m3/i1)\times P(i1)+P(m3/i2)\times P(i2)+P(m3/i3)\times P(i3)+P(m3/i4)\times P(i4)}$
- $P(i4/m3) = \frac{P(m3/i4)\times P(i4)}{P(m3/i1)\times P(i1)+P(m3/i2)\times P(i2)+P(m3/i3)\times P(i3)+P(m3/i4)\times P(i4)}$
- $\frac{P(i2/m3)}{P(i4/m3)} = \frac{P(m3/i2)\times P(i2)}{P(m3/i4)\times P(i4)}$

En cualquier caso, un tratamiento exhaustivo del problema obliga a conocer todas las posibilidades de la BLE: las prevalencias (Probabilidades absolutas) y, las correspondientes probabilidades condicionales a priori.

Sin embargo, hay otro modo de plantearse este problema que no sea un análisis estadístico crudo y duro de todos los datos: usando la BLR. Así eliminamos un montón de probabilidades no reales (ya que no están en la BLR) que pasarán su probabilidad a valer 0 y el resto, deberán sumar entre sí 1.

2.4.2. Necesidad de independencia en el esquema bayesiano

Este es uno de los problemas del esquema bayesiano. Vamos a usar los datos de la tabla para evaluar $P(A1/m1)$ desde dos métodos distintos:

- Utilizando el teorema de bayes
- Utilizando directamente los datos de la tabla

	Solo A1	**A1 y A2**	**Solo A2**	**A3**	**Total**
Solo m1	B1	C1	D1	E1	N1
m1 y m2	B2	C2	D2	E2	N2
Solo m2	B3	C3	D3	E3	N3
m3	B4	C4	D4	E4	N4
Total	B	C	D	E	N

Podemos observar que tanto m1 y m2 no hay independencia y tampoco en A1 y A2.

Recordemos que $P(X/Y)$ probabilidad de X dado Y $\equiv$ probabilidad de la intersección X e Y entre la probabilidad total de Y:

- $P(A1) = (B + C)/N$
- $P(A2) = (C + D)/N$
- $P(A3) = E/N$
- $P(m1/A1) = (B1 + B2 + C1 + C2)/(B + C)$
- $P(m1/A2) = (D1 + D2 + C1 + C2)/(C + D)$
- $P(m1/A3) = (E1 + E2)/E$

- Si usamos el teorema de Bayes y después sustituimos el conocimiento a priori:

$$P(A1/m1) = \frac{P(m1/A1) \times P(A1)}{P(m1/A1) \times P(A1) + P(m1/A2) \times P(A2) + P(m1/A3) \times P(A3)}$$

$$P(m1/A1) \times P(A1) = \frac{(B1 + B2 + C1 + C2)}{N}$$

$$P(m1/A2) \times P(A2) = \frac{(D1 + D2 + C1 + C2)}{N} \tag{34}$$

$$P(m1/A3) \times P(A3) = \frac{E1 + E2}{N}$$

$$P(A1/m1) = \frac{B1 + B2 + C1 + C2}{N1 + N2 + C1 + C2}$$

- Directamente de la tabla:

$$P(A1/m1) = \frac{B1 + B2 + C1 + C2}{N1 + N2} \tag{35}$$

En Bayes dividimos entre C1 y C2 porque estos son los casos donde no hay independencia: lo que nos lleva a una respuesta errónea.

Hay soluciones, como agrupar A1 y A2 en otro cluster distinto y llamarle A4, por ejemplo: pero ya aparece una fisura en el planteamiento Bayesiano.

2.4.3. Otros problemas del esquemas bayesiano

Este planteamiento esta pensado para **trabajar de forma estática** pero las evidencias van apareciendo de forma temporal. Por ejemplo:

- Sea un conjunto E1 el conjunto de toda la información en un momento dado y sea S1 un nuevo elemento de información. Llamaremos E al nuevo conjunto formado por E1 y S1.
- El teorema de Bayes se complicaría de la siguiente forma para una interpretación Ii:

$$P(Ii/E) = \frac{P(S1/Ii \wedge E1) \times P(Ii/E1)}{\sum_j P(S1/Ij \wedge E1) \times P(Ij/E1)} \tag{36}$$

  Y cada vez que aparece una nueva pieza de información debemos rehacer todos los cálculos.

Además otro problema del esquema bayesiano es la **aplicación poco cuidadosa del modelo**. Ejemplo:

- Existe una relación entre visión borrosa y glaucoma.
- P(visión borrosa) = 231/15315
- P(glaucoma) = 320/15315
- P(visión borrosa/glaucoma) = 150/320
- $Bayes \rightarrow P(glaucoma/visionborrosa) = 0{,}67$
- Pero... ¿Y si esto es porque tiene las gafas sucias?

Hay que tener clara la relación causal entre los elementos que estamos utilizando para hacer la probabilidad condicional.

Otro problema, es la **consistencia matemática del modelo**.

- Dada una hipótesis y una evidencia si calculamos la probabilidad condicionada de la hipótesis dada la evidencia saldría un número entre 0 y 1. $P(H/E) = x, (0 \leq x \leq 1)$.
- Entonces la probabilidad de la negación de la hipótesis debería ser la siguiente: . $P(\neg H/E) = 1-x$ Lo cual es incongruente porque la probabilidad de la negación de la hipótesis también es positiva.
- La misma evidencia apoya positivamente a una hipótesis y a la negación de la misma: incongruencia. No es matemática, que es debido a que estamos usando conocimiento y no estadística pura.

2.4.4. Ejemplo de problema categórico-bayesiano

Una madre observa que su hija de 5 semanas tiene distress respiratorio. Esta situación se mantiene durante 5 días, y el proceso no parece remitir. La niña no presentaba este problema cuando nació. El examen médico muestra que la niña está bien nutrida aunque tiene hemangiomas en la parte trasera del cuello bajo la oreja izquierda. El examen del tórax por rayos X muestra una masa en el mediastino, que desplaza la tráquea. Acompañan a esta masa pequeños gránulos de calcio.

A la vista de esta información, el clínico que atiende el caso debe discriminar entre las siguientes posibilidades para poder establecer un tratamiento adecuado:

- 1: La paciente tiene un timo anormalmente grande. Es bien conocido que un timo anormalmente grande puede producir distress respiratorio.
- 2: La paciente tiene un gran hemangioma en el mediastino, que se manifiesta a través de los hemangiomas superficiales
- 3: La paciente puede tener un quiste en el mediastino. La presencia de calcio junto a la masa detectada a través de rayos X sugiere esta posibilidad

Además el clínico sabe que:

Si un paciente tiene un timo anormalmente grande, y además tiene o bien un hemangioma en el mediastino, o bien un quiste, entonces debe presentar distress respiratorio, y una masa en el mediastino.

Si un paciente no tiene un hemangioma en el mediastino, entonces no debe presentar hemangiomas superficiales Si un paciente no tiene un timo anormalmente grande, pero tiene un hemangioma en el mediastino, y además un quiste, entonces debe presentar una masa en el mediastino.

Al resolverlo, sale una base lógica expandida de 64 elementos (2^6). Además la evidencia gránulos de cálcicos no sirve para nada.

2.5. Razonamiento Cuasi-Estadísticos

Se llaman cuasi-estadísticos porque usan de alguna manera los razonamientos estadísticos para obtener soluciones pero, sin un amplio muestreo estadístico: utilizan impresiones, lo que se llama probabilidad condicional subjetiva que es distinta de la habitual porque se basa en la experiencia del experto.

Aquí veremos dos modelos, cuya bibliografía en la que se basa estés apartados: sección 3 apartado 5.

- **Factores de Certidumbre**: Es un modelo establecido para resolver un problema que luego se generalizó.
- **Teoría Evidencial**: Contiene al modelo de factores de incertidumbre que a su vez contiene al modelo de bayes que a su vez contiene al categórico.

La idea de estos modelos es evitar inconvenientes de modelos anteriores, como por ejemplo, la necesidad de utilizar amplios estudios estadísticos utilizando conocimiento heurístico. De este modo introducimos un concepto nuevo: la probabilidad condicional subjetiva.

La **probabilidad condicional subjetiva** relaciona dos sucesos y, se distingue de la probabilidad condicional de Bayes en que la ocurrencia del primero está condicionada por la ocurrencia del segundo. Además, es subjetiva porque la relación no está avalada por grandes estudios estadísticos sino que se utiliza conocimiento heurístico.

Si hablamos de probabilidad condicional tenemos que si la probabilidad de I_i dado S_k es $P(I_i/S_k) = x$ significa que "Si la manifestación S_k está presente, entonces hay una probabilidad subjetiva x de que la interpretación sea I_i".

Para que la probabilidad condicional tenga una consistencia matemática a pesar de ser subjetiva necesitamos que: $\sum_i P(I_i/S_k) = 1$. Esto puede no darse ya que usamos conocimiento heurístico (no matemático) el cual no es determinista. Para solucionarlo debemos normalizarlo de la siguiente manera: $P(I_i/S_k) = \frac{x}{\sum_i P(I_i/S_k)}$.

Además de todo esto debemos recordar otros problemas como la incertidumbre, la imprecisión, la falta de confianza, credibilidad, ... Esto nos lleva a usar nuevos mecanismo para tratar el razonamiento.

2.5.1. Conceptos nuevos

$P(I_i/S_k) = x$ puede interpretarse en términos de implicación como $S_k \xrightarrow{x} I_i$

"x"define la potencia evidencial de la implicación, es decir, define la "la fuerza"de la flecha.

Si es $x = 1$ será una implicación sólida

Si es menor que uno, la flecha se difumina, es decir, la implicación viene afectada por la incertidumbre.

Ej: $S_k \xrightarrow{0{,}7} I_i$ esto se leería como siempre que se de S_k es muy probable que se de I_i.

2.5.2. Modelo de factores de Certidumbre

El modelo de factores de Certidumbre es un modelo y no una teoría porque le falta base teórica. Este modelo nace 1975 de la mano de Shortliffe y Buchanan al construir un sistema experto llamado Mycin para enfermedades infecciones.

Ellos se dieron cuenta de la inconsistencia física del modelo de Bayes y, llegaron a la conclusión de que una evidencia no puede apoyar simultáneamente a una hipótesis y a la negación de la misma.

Dada una hipótesis h, y una evidencia e, dicha evidencia no puede simultáneamente incrementar la confianza en h y, disminuir la confianza en h. Para reflejar todo esto dada una hipótesis h la potencia evidencial de una declaración se debe representar a través de dos medidas diferentes:

- Medida de confianza creciente $MB(h, e)$: incremento de confianza en h dada la evidencia e.
- Medida de desconfianza creciente $MD(h, e)$: incremento de la desconfianza en h dada la evidencia e.

Estos índices son dinámicos y se actualizan cada vez que hay una evidencia nueva.

Formalmente:

- Sea $P(h)$ la confianza previa en h antes de e
- Sea $P(h/e)$ la confianza en h tras aparecer e
- Sea $1 - P(h)$ la desconfianza en h antes de e
- Caso 1 - Si $P(h/e) > P(h)$, la nueva evidencia produce un aumento de confianza en la hipótesis considerada

$$\begin{aligned} MB(h, e) &> 0 \\ MD(h, e) &= 0 \\ MB(h, e) &= \frac{P(h/e) - P(h)}{1 - P(h)} \end{aligned} \tag{37}$$

  $P(h/e) - P(h)$ es el incremento que trae e a h pero, ¿Por qué dividimos entre $1 - P(h)$? Si e confirma totalmente a la hipótesis $P(h/e) = 1$, el numerador y el denominador deben ser el mismo porque no se pude crecer más en confianza, lo que nos dejaría con: $MB(h, e) = \frac{1-P(h)}{1-P(h)} = 1$. (Compatible con la teoría de la probabilidad)

- Caso 2 - Si $P(h/e) < P(h)$, la nueva evidencia produce una disminución de confianza en la hipótesis.

$$\begin{aligned} MB(h, e) &= 0 \\ MD(h, e) &> 0 \\ MD(h, e) &= \frac{P(h) - P(h/e)}{P(h)} \end{aligned} \tag{38}$$

  En este caso, si la evidencia descarta por completo la hipótesis h P(h/e)=0, y para que la desconfianza sea máxima sea 1 hay que dividir entre P(h): $MD(h, e) = \frac{P(h)-0}{P(h)} = 1$.

- Caso 3 - Si p(h) = p(h/e), la nueva evidencia es independiente de la hipótesis considerada, ya que ni aumenta la confianza ni la desconfianza. O bien, no tenemos conocimiento sobre ello.

$$\begin{aligned} MB(h, e) &= 0 \\ MD(h, e) &= 0 \end{aligned} \tag{39}$$

Otra forma de verlos ya que p(h) es una probabilidad a priori en el sentido clásico y igual las expresiones de MB y MD pero discerniendo entre mínimo (MD) y máximo (MB):

$$\begin{aligned} MB(h,e) = 1 &\Longleftrightarrow p(h) = 1 \\ MB(h,e) = \frac{max[p(h/e), p(h)] - p(h)}{max[0,1] - p(h)} &\Longleftrightarrow p(h) \neq 1 \\ MD(h,e) = 1 &\Longleftrightarrow p(h) = 0 \\ MD(h,e) = \frac{min[p(h/e), p(h)] - p(h)}{min[0,1] - p(h)} &\Longleftrightarrow p(h) \neq 0 \end{aligned} \tag{40}$$

Con todo esto ya podemos definir el **factor de certidumbre CF(h,e)**: $CF(h,e) = MB(h,e) - MD(h,e)$.

Esta expresión es de carácter formal ya que una misma e nunca puede incrementar simultáneamente la confianza entre potencias evidénciales de hipótesis alternativas. Sin embargo, CF(h,e) sirve para facilitar las comparaciones entre potencias evidenciales de hipótesis alternativas $(h_1, ..., h_n)$ en relación con una misma evidencia e.

Cuyos rangos son:

$$\begin{aligned} 0 \leqslant MB(h,e) \leqslant 1 \\ 0 \leqslant MD(h,e) \leqslant 1 \\ -1 \leqslant CF(h,e) \leqslant 1 \end{aligned} \tag{41}$$

Esto nos permite interpretar -1 como no, 1 como si y únicamente 0.5 quiere decir no se.

CF(h,e) en casos extremos e hipótesis mutuamente excluyentes.

- Caso 1 - Si h es cierta $P(h/e) = 1$

$$\begin{aligned} MB(h,e) = \frac{1 - p(h)}{1 - p(h)} = 1 \\ MD(h,e) = 0 \\ CF(h,e) = 1 \end{aligned} \tag{42}$$

- Caso 2 - Si la negación de h es cierta $P(\neg h/e) = 1$

$$\begin{aligned} MD(h,e) = \frac{p(h) - 0}{p(h)} = 1 \\ MB(h,e) = 0 \\ CF(h,e) = -1 \end{aligned} \tag{43}$$

Lo que nos lleva a que los casos anteriores conducen a que si $MD(\neg h/e) = 1$ entonces $MB(h/e) = 1$.

Si h1 y h2 son dos hipótesis mutuamente excluyentes y, sabemos que MB(h1,e) = 1, entonces podemos afirmar que MD(h2,e)=1. Ya que si una hipótesis si apoya totalmente a una, hace que la desconfianza en la otra sea plena.

CF(h,e) en casos donde las evidencias sean independientes

$$\begin{aligned} p(h,e) = p(h)MB(h,e) = 0 \\ MD(h,e) = 0 \\ CF(h,e) = 0 \end{aligned} \tag{44}$$

2.5.2.1. Conceptos importantes

- Los CFs de la hipótesis h y $\neg h$ no son complementarios en la unidad si no que son opuestos entre sí.
- Si el apoyo que una evidencia presta a una hipótesis es bajo, no debería ser alto el apoyo a la negación de la hipótesis. Sobre todo en caso de información incompleta.
- Ejemplo: Un gorrión pasa al lado de un avión, ¿Tiene mucha influencia el vuelo del gorrión para que el avión se caiga o no se caiga? No. Si la presencia del gorrión coincide con que se cae o no el avión tiene muy poco que ver y su relación causal tiene que ser como mínimo muy pequeña.
- Demostración usando casos extremos y tiendo en cuenta que $p(\neg h) = 1 - p(h)$ y viceversa:

$$\begin{aligned} p(h/e) > p(h) \\ CF(h,e) = MB(h,e) > 0 \\ CF(\neg h,e) = -MD(h,e) < 0 \end{aligned} \quad (45)$$

$$MB(h,e) = \frac{p(h/e) - p(h)}{1 - p(h)} = \frac{1 - p(\neg h/e) - 1 + p(\neg h)}{p(\neg h)} = \frac{p(\neg h) - p(\neg h/e)}{p(\neg h)} = MD(\neg h,e)$$

  Conclusión: $CF(h,e) = -CF(\neg h,e)$

  Y lo mismo ocurre si p(h) > p(h/e)

- Esto demuestra que si algo apoya poco a la hipótesis también apoya poco a la negación de la hipótesis.

2.5.2.2. Significado de los CFs

- $0 < CF(h,e) \leq 1$: La evidencia apoya a la hipótesis
- $-1 \leq CF(h,e) < 0$: La evidencia va en contra de la hipótesis
- $CF(h,e) = 0$: La evidencia es independiente de la hipótesis o no se tiene información sobre la relación causal h/e.

2.5.2.3. Combinación de evidencias CF

- Situación: Hay más de una evidencia relativa a la misma hipótesis.

  IF e_1 then h with $CF(h,e_1)$

  IF e_2 then h with $CF(h,e_2)$

  ....

  IF e_n then h with $CF(h,e_n)$

  $E = \{e_1, e_2, ..., e_n\}$

- Problema: ¿Cómo evaluamos CF de la hipótesis dada toda la evidencia disponible $CF(h,E)$?

- **Primera aproximación:**
  - Caso A, e1 y e2 contribuyen positivamente a la veracidad de la hipótesis.

$$\begin{gathered} CF(h,e_1) > 0, CF(h,e_2) > 0 \\ CF(h,e_1 \wedge e_2) = CF(h,e_1) + CF(h,e_2) - (CF(h,e_1) \times CF(h,e_2)) \end{gathered} \tag{46}$$

    En la operación hacemos la unión (+) menos la intersección (x).
  - Caso B, e1 y e2 contribuyen negativamente a la veracidad de la hipótesis.

$$\begin{gathered} CF(h,e_1) < 0, CF(h,e_2) < 0 \\ CF(h,e_1 \wedge e_2) = CF(h,e_1) + CF(h,e_2) + (CF(h,e_1) \times CF(h,e_2)) \end{gathered} \tag{47}$$

    En este caso como los dos tienen CF tienen valor negativo, será una suma (+).
  - Caso C, e1 y e2 contribuyen una positivamente y otra negativamente a la veracidad de la hipótesis.

$$\begin{gathered} CF(h,e_1) \times CF(h,e_2) < 0 \\ CF(h,e_1 \wedge e_2) = CF(h,e_1) + CF(h,e_2) \end{gathered} \tag{48}$$

    En este caso, el mayor tendrá más peso.

  Ventajas de esta primera aproximación:
  - Nos previene de la hipotética situación de que ambas evidencias no fueran completamente independientes.
  - Las expresiones son generalizables directamente. Ene l caso de evidencias todas positivas tendríamos:

$$CF(h,E) = \sum_{i}^{n} CF_i - \sum_{i,i<j}^{n} CF_i \times CF_j + \sum_{i,i<j<k}^{n} CF_i \times CF_j \times CF_k - ... \tag{49}$$

    Existe una alternancia de símbolos, esto nos asegura que va ser menor que 1. El papel de $i < j < k < ...$ es evitar duplicidades.

  Desventajas de esta primera aproximación:
  - Falta de asociatividad de la formulación
  - Sensibilidad de la formulación ante la aparición de evidencias contradictorias en estados avanzados del proceso de razonamiento.
  - Ejemplo:
    - Tenemos 8 reglas que apoyan a la hipótesis con $CF_i \in [0{,}4, 0{,}8]$ y con ellas, $CF_{12345678} = 0{,}99$.
    - Si apareciera una novena evidencia tardía: $CF_9 = -0{,}6$.
    - $CF_{123456789} = 0{,}39 <<< CF_{192345678} = 0{,}99$, por lo que no es asociativa. ¿Tiene que ser asociativa? Más información en sección 3 apartado 6.

- **Segunda aproximación:**
  - Caso A: Igual
  - Caso B: Igual
  - Caso C: Modificación para volverlo asociativo.

$$\begin{gathered} CF(h,e_1) \times CF(h,e_2) < 0 \\ CF(h,e_1 \wedge e_2) = \frac{CF(h,e_1) + CF(h,e_2)}{1 - min\{|CF(h,e1)|, |CF(h,e2)|\}} \end{gathered} \tag{50}$$

Ventajas: Las evidencias se pueden considerar en cualquier orden sin alterar el resultado final.

Desventajas: Si $e1 \rightarrow e2$, entonces $CF(h, e1 \wedge e2) \equiv CF(h, e1)$, lo que no se deduce de la aplicar de este modelo. ¿Posibles soluciones? Estructurar bien las bases del conocimiento o agrupar en una sola reglas cláusulas con evidencias condicionalmente dependientes.

2.5.2.4. Conocimiento inexacto

- Imprecisión: Afecta a entidades, hechos y datos del dominio. ¿Por qué aparece?
  - Porque la propia evidencia es imprecisa
  - Porque la evidencia considerada es una consecuencia de otra reglas, y forma parte de un proceso de razonamiento que supone varias inferencias: el conocimiento inexacto se va propagando hasta una situación de $CF = 0$
- Incertidumbre: Ligada a los procesos de razonamiento.

Ambas pueden darse aislada o conjuntamente.

Entropía de la información: La información se degrada a medida que se utiliza.

Ejemplo:

- Reglas: Cuando llueve mucho casi siempre me quedo en casa y cuando me quedo en casa suelo leer.
- Representación del conocimiento:
  - E1 = llover mucho
  - H1 = Quedarse en casa
  - $CF(H1, E1) = 0{,}95$ (Indica incertidumbre en la $\rightarrow$, ya que hemos dicho casi siempre)
  - E2 = Quedarse en casa
  - H2 = Leer
  - $CF(H2, E2) = 0{,}75$ (Indica incertidumbre en la $\rightarrow$, ya que hemos dicho suelo)
- Imprecisión:
  - En vez de llover mucho (E1), esta lloviendo bastante (e1) y consideramos arbitrariamente que $CF(E1, e1) = 0{,}75$, así convertimos la imprecisión en incertidumbre.
  - Por lo que: $e \xrightarrow{0,75} E1 \xrightarrow{0,95} H1(= E2) \xrightarrow{0,75} H2$

¿Cómo se propaga la incertidumbre?

Si E' entonces E con $CF(E, E') = x$ y, si E entonces H con $CF(H, E) = y$ ($E' \xrightarrow{x} E \xrightarrow{y} H$).

Shortliffe y Buchanan definen $CF(H, E') = CF(H, E) \times max\{0, CF(E, E')\}$, añadiendo ese máximo para evitar situaciones contradictorias (negativas): este proceso se denomina como **arrastre de la inexactitud y propagación**.

Inconveniente: Si $CF(E, E') \rightarrow CF(H, E') = 0$

Solución propuesta por Heckerman: Recordando 2.5.2.1 donde se llegaba a la conclusión de que $CF(\neg E, E') = -CF(E, E')$ es decir, no es complementario si no, opuesto. Podemos establecer que:

$$\begin{gathered} E' \rightarrow \neg E \rightarrow H \\ CF(H, E') = CF(H, \neg E) \times max\{0, CF(\neg E, E')\} \end{gathered} \tag{51}$$

Pero para esto necesitamos conocer $CF(H \neg E)$

2.5.2.5. Combinación lógica de evidencias

Para tener un modelo totalmente completo necesitamos combinar evidencias. Las cláusulas de los antecedentes de las reglas pueden estar relacionadas a través de los operadores lógico ($\neg, \wedge \vee$). Por lo que conociendo $CF(H_1, E)$ y $CF(H_2, E)$ y sabiendo que E es toda la evidencia del antecedente, ¿Cómo podemos obtener $CF(H_1 \vee H_2, E)$ o $CF(H_1 \wedge H_2, E)$?

- $CF(H_1 \vee H_2, E) = max\{CF(H_1, E), CF(H_2, E)\}$
- $CF(H_1 \wedge H_2, E) = min\{CF(H_1, E), CF(H_2, E)\}$

Ejemplo: $(E_2 \vee E_3) \wedge E_1 \wedge E_4 \xrightarrow{CFr} H$

$$\begin{aligned} CF(H, E) &= CFr \times CF(antecedente, E) = CFr \times CF\{(E_2 \vee E_3) \wedge E_1 \wedge E_4, E\} \\ &= CFr \times min\{CF(E_1, E), max[CF(E_2, E), CF(E_3, E)], CF(E_4, E)\} \end{aligned} \tag{52}$$

2.5.2.6. Tabla resumen

- Combinación de evidencias sobre la misma hipótesis
  - $+, +$: $CF(H, E) = CF1 + CF2 - CF1 \times CF2$
  - $-, -$: $CF(H, E) = CF1 + CF2 + CF1 \times CF2$
  - $+, -$: $CF(H, E) = \frac{CF1+CF2}{1-min\{|CF1|, |CF2|\}}$
- Propagación de la incertidumbre
  - $E' \rightarrow E \rightarrow H$ con $CF(H, E)$ y $CF(H, E')$:
    $CF(H, E') = CF(H, E) \times max\{0, CF(E, E')\}$
- Combinación lógica de evidencias
  - $CF(H_1 \vee H_2, E) = max\{CF(H_1, E), CF(H_2, E)\}$
  - $CF(H_1 \wedge H_2, E) = min\{CF(H_1, E), CF(H_2, E)\}$

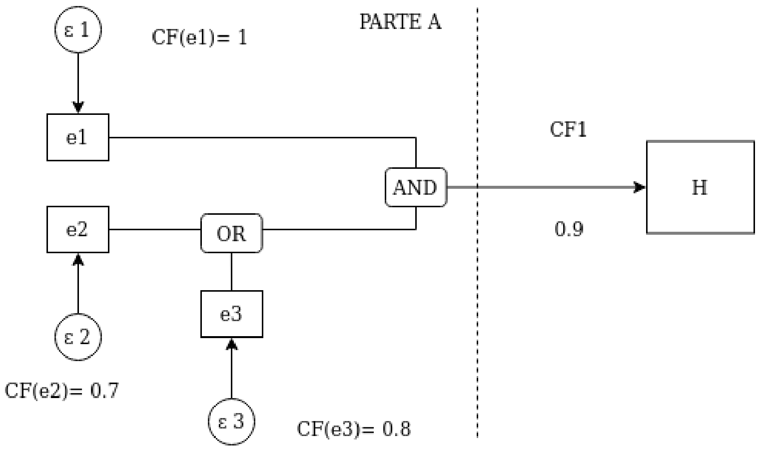
2.5.2.7. Ejemplo exhaustivo Durante un proceso regresivo (de hipótesis a datos) y exhaustivo (busqueda completa) centrado en la hipótesis H, se activaron las reglas, y se obtuvieron las evidencias, que aparecen a continuación. Evaluar el factor de certidumbre final de la hipótesis H dada toda la evidencia disponible.

- R1: $e1 \wedge (e2 \vee e3) \rightarrow H$, CF(H/R1) = 0.9
- R2: $e4 \rightarrow H$, CF(H/R2) = 0.8
- R3: $e5 \wedge e6 \rightarrow e4$, CF(H/R3) = 0.6
- R4: $e7 \rightarrow H$, CF(H/R4) = 0.8
- R5: $e8 \rightarrow e7$, CF(H/R4) = 0.6
- R6: $e9 \rightarrow e7$, CF(H/R4) = 0.6

  $CF(e_1) = 1$, $CF(e_2) = 0{,}7$, $CF(e_3) = 0{,}8$

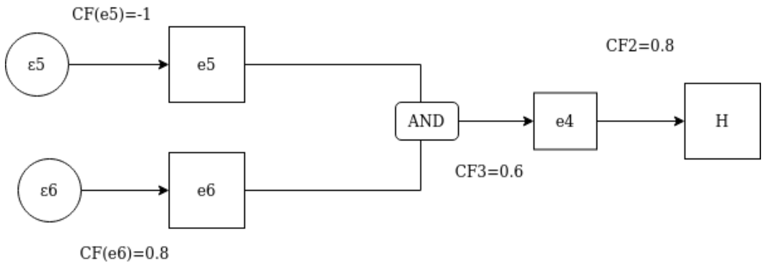
  $CF(e_5) = -1$, $CF(e_6) = 0{,}8$, $CF(e_8) = 0{,}6$, $CF(e_9) = 0{,}71$

Contexto: En una fábrica se detecta un problema que puede estar relacionado con la parte mecánica (A), con la eléctrica (B) y con la parte de mantenimiento (C). Respecto a la rama A, la evidencia A



apoya a la hipótesis un 0.72:

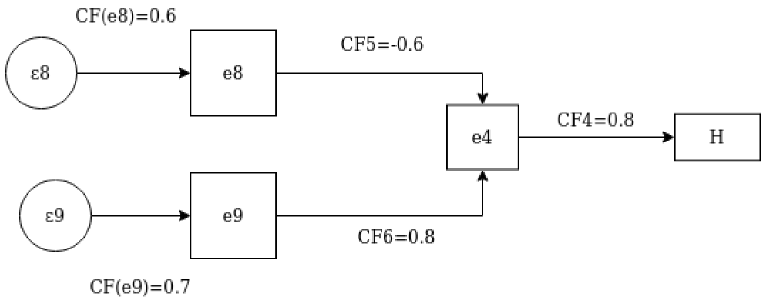
$$\begin{aligned} E_A &= \epsilon_1, \epsilon_2, \epsilon_3 \\ Antecedente_A &= e1 \wedge (e2 \vee e3) \\ CF(H, E_A) &= CF(H/R1) \times max\{0, CF(Antecedente_A, E_A)\} \\ CF(Antecedente_A, E_A) &= CF\{e1 \wedge (e2 \vee e3)\} = \\ min\{CF(e1, \epsilon_1), max[CF(e_2, \epsilon_2), CF(e_3, \epsilon_3)]\} &= \\ min1, max(0{,}7, 0{,}8) &= 0{,}8 \\ CF(H, E_A) &= 0{,}9 \times max\{0{,}0, 0{,}8\} = 0{,}72 \end{aligned} \tag{53}$$



Respecto a la rama B, la evidencia no se sabe si apoya a la hipótesis 0.0:

$$\begin{aligned} E_B &= \epsilon_5, \epsilon_6 \\ Antecedente_B &= e5 \wedge e6 \\ CF(H, E_B) &= CF(H/R2) \times max\{0, CF(e4, E_B)\} \\ CF(e4, E_B) &= CF(e4/R3) \times max\{0, CF(Antecedente_B, E_B)\} \\ CF(Antecendente_B, E_B) &= CFe5 \wedge e6, E_B = \\ &= min\{CF(e5, \epsilon_5), CF(e6, \epsilon_6)\} = min\{-1, 0{,}8\} = -1{,}0 \\ CF(e4, E_B) &= CF(e4/R3) \times max\{0, -1\} \\ CF(H, E_B) &= CF(H/R2) \times max\{0, 0\} = 0 \end{aligned} \tag{54}$$

Respecto a la rama C, la evidencia apoya a la hipótesis un 0.25:



$$\begin{aligned} E_C &= \epsilon_8, \epsilon_9 \\ Antecedente_C &= 2 - reglas \\ CF(H, E_C) &= CF(H/R4) \times max\{0, CF(e7, E_C)\} \\ CF(e7, E_C) &= 2 - reglas \rightarrow Evidencias - independientes - sobre - e7 \\ & RAMA_C 1 \\ CF(e7, \epsilon_8) &= CF(e7/R5) \times max\{0, CF(e8, \epsilon_8)\} = -0{,}6 \times max\{0{,}0, 0{,}6\} = -0{,}36 \\ & RAMA_C 2 \\ CF(e7, \epsilon_9) &= CF(e7/R6) \times max\{0, CF(e9, \epsilon_9)\} = 0{,}8 \times max\{0{,}0, 0{,}7\} = 0{,}56 \\ Luego: CF(e7, E_C) &= \frac{0{,}56 - 0{,}36}{1 - min\{|0{,}56|, |-0{,}36|\}} = 0{,}2/0{,}64 \\ CF(H, E_C) &= CF(H/R4) \times max\{0, CF(e7, E_C)\} = 0{,}8 \times \frac{0{,}2}{0{,}64} = 0{,}25 \end{aligned} \tag{55}$$

Tenemos que: $CF(H, E_A) = 0{,}72$, $CF(H, E_B) = 0{,}0$, $CF(H, E_C) = 0{,}25$,
Finalmente, dada la evidencia positiva: $CF(H, E) = 0{,}72 + 0{,}25 - 0{,}72 \times 0{,}25 = 0{,}79$

2.5.3. Teoría evidencial de Dempster y Shafer

Características:

- Tiene una fuerte base teórica.
- Permite modelar de forma sencilla la incertidumbre asociada a evidencias e hipótesis.
- Permite considerar conjuntos de hipótesis sin que la confianza depositada en cada uno de ellos tenga que ser distribuida de ningún modo entre las hipótesis individuales. Es decir, trabaja con conjuntos de hipótesis como un todo.
- Permite reflejar la falta de conocimiento.
- Contiene a la teoría de la probabilidad y a su vez, contiene algunas de las funciones combinatorias de evidencias del modelo Shortliffe y Buchanan.

2.5.3.1. Marco de discernimiento θ :

Se representa con θ y se refiere al conjunto finito de todas las hipótesis (mutuamente excluyentes) que se pueden estableces en el dominio del problema.

El efecto de la evidencia sobre el conjunto global de hipótesis no viene determinado por la contribución de confianza depositada en hipótesis individuales. Por ej: Si tenemos tos que se asocia a gripe o a catarro, por si sola no podemos discernir cual de las dos es.

2.5.3.2. Elementos de la teoría

- Marco de discernimiento $\theta = \{h_1, h_2, ..., h_n\}$
- Elemento focal A: $A \subseteq \theta$.
- Evidencia (Parámetro cualitativo): e
- Función básica de asignación de verosimilitud (Parámetro cuantitativo): m. Es la potencia evidencial del modelo anterior dado que e apoya a un subconjunto A de hipótesis.
- $e : A \subseteq \theta \rightarrow m(A) = s$, $0 \leq s \leq 1$
- El resto de la confianza no apoya al complementario de A si no al propio marco de discernimiento $m(\theta) = 1 - m(A) = 1 - s$
- Para cualquier otro subconjunto del marco de discernimiento: $m(B) = 0, \forall \subseteq \theta, B \neq \theta, B \neq A$

Ejemplo: Se sortea m trozos de una tarta ($m>n$). Antes de empezar una sorteo de una tarta, el desconocimiento es total y le puede tocar a cualquiera $m(\theta) = 1$. Saco la paleta de la urna y sale el número 7 $m(A) = 1/n$, n número de personas, el resto tiene una probabilidad de $m(\theta) = 1 - m(A)$ para la siguiente ronda. Hacemos una segunda rifa con todos, y sacamos el 6. Uno de los que tenia el número 7 antes ahora tiene el 6, pues le vuelve a tocar un trozo de tarta. Es decir, el concepto es que le quitas a la situación inicial m(subconjunto A conocido).

Para asegurar la consistencia matemática del modelo se deben cumplir estas dos condiciones:

- A es un elemento focal solo si $m(A) \neq 0$, es decir existe una evidencia asociada. Y para ello debe cumplir que:

  1.- $\sum_{A \subseteq \theta} = 1$, es decir, la solución está en el marco.

  2.- $m(\phi) = 0$, es decir que la solución también está en el marco.

Mientras que en el planteamiento probabilístico si $p(A) = s$ entonces $p(\neg A) = 1 - s.$,en el planteamiento evidencial, con el siguiente marco de discernimiento $\theta = \{h1, h2, h3, h4\}$ y un subconjunto de θ $A = h1, h2$ con la siguiente evidencia apoyándolo $e : A \xrightarrow{m} (A) = m(\{h1, h2\}) = s$ entonces $m(\theta) = m(\{h1, h2, h3, h4\}) = 1 - s$. Es decir, a medida que aparecen evidencial vamos quitándole ignorancia al marco.

Sin embargo, las evidencias no tienen porque aparecer solas ni refererirse a los mismo elementos focales. Por ejemplo:

- $\theta = \{h1, h2, h3, h4\}$
- e1: A1 = (h1,h2)
  - m1(h1,h2) = x
  - m1(h1,h2,h3,h4) = 1-x
- e2: A2 = (h2,h3,h4)
  - m2(h2,h3,h4) = y
  - m2(h1,h2,h3,h4) = 1-y
- e3: A3 = (h1)
  - m3(h1) = z
  - m3(h1,h2,h3,h4) = 1-z
- ¿Cómo combinamos las evidencias?
  - $m12(C) = m1(A1) \otimes m2(A2) = m1(A1) \times m2(A2)$ siendo C = $A1 \cap A2$

2.5.3.3. Combinación de evidencial

$$m12(C) = \sum_{Ai \cap Bj} m1(A_i) \times m2(B_j) \tag{56}$$

Coincide estrictamente con la asignación de probabilidad a la intersección de dos sucesos independientes por lo que, la teoría evidencial asume implícitamente independencia entre evidencias. Por lo que se cumple la primera condición exigida a la función de verosimilitud dada anteriormente en la sección 2.5.3.2:

$$\sum_{C = Ai \cap Bj} m12(C) = 1 \tag{57}$$

Sin embargo puede ocurrir que la intersección de elementos focales, sea nula. Es decir, no haya ninguna hipótesis en común entre los dos elementos focales. Entonces se introduce una perculiaridad en el modelo:

- Si e1 : A1 / m1(A1) = x ($0 \leqslant x \leqslant 1$)
- Si e2 : A2 / m2(A2) = y ($0 \leqslant x \leqslant 1$)
- $C = A1 \cap A2 = \phi$
- $M12(C) = m12(\phi) = m1(A1) \times m2(A2) = xy \neq 0$

Vulneramos la segunda condición de la función de verosimilitud para que el modelo sea consistente 2.5.3.2. Cuando esto ocurre, debemos normalizar el resultado para lo cual definimos el grado de conflicto y el factor de normalización.

Grado de conflicto

$$K = \sum_{A_i \cap B_j = \phi} m1(A_i) \times m2(B_j)$$

Factor de normalización FN:

$$FN = \frac{1}{1-K} \tag{58}$$

Función de asignación verosimilitud normalizada:

$$m12(C)_{normalizada} = \frac{\sum_{C=A_i \cap B_j} m1(A_i) \times m2(B_j)}{1-K} \tag{59}$$

Cuando la intersección $A_i \cap B_j$ no es nula (k = 0) tendremos la fórmula inicial.

¿Y si la intersección es el vacío? K toma un valor que hay que repartir entre las demás funciones de combinación $1/1-K$, siendo $1>K>0$ nos dará un resultado mayor a 1.

¿Y K puede valer 1? Solamente en un caso, caso de conflicto total, por ejemplo, si dos evidencias apoyan totalmente a diferentes hipótesis, la intersección es el vació: $FN = \infty$.

Para resolver este tipo de ejercicios usaremos tablas para combinar las en forma de filas y columnas elementos focales junto a marcos de discernimiento. Cada casilla será la intersección de columna y fila y, su valor será la multiplicación de la función de verosimilitud de la fila multiplicada por la función de verosimilitud de la columna.

Cuando se combinen con una tercera evidencia se cogerá todos los resultados de la tabla como filas para formar otra tabla con la tercera evidencia.

Veremos un ejemplo de esto a continuación:

2.5.3.4. Ejemplo de teoría evidencia :

Hay consejos para que se cumplan las reglas de la función de verosimilitud en la sección 3 apartado 7.

- $\theta = \{h1, h2, h3, h4\}$
- $e1 : A1 = \{h1, h2\}$, $m1(A1) = 0{,}6$, $m1(\theta 1) = 0{,}4$
- $e2 : A2 = \{h2, h3, h4\}$, $m2(A2) = 0{,}7$, $m2(\theta 2) = 0{,}3$
- $e3 : A3 = \{h1\}$, $m3(A3) = 0{,}8$, $m3(\theta 3) = 0{,}2$
- $\theta = \theta 1 = \theta 2 = \theta 3 = \{h1, h2, h3, h4\}$

Con la evidencia e1 y e2:

Intersecciones	**A1= {h1,h2}**	$\boldsymbol{\theta 1}$
A2 = {h2,h3,h4}	{h2} $m12(h2) = m1(A1) \times m2(A2) = 0{,}42$	{h2,h3,h4} $m12(h2, h3, h4) = m1(\theta 1) \times m2(A2) = 0{,}28$
$\theta 2$	{h1,h2} $m12(\{h1, h2\}) = m1(A1) \times m2(\theta 2) = 0{,}18$	$\theta 12$ $m12(\theta 12) = m1(\theta 1) \times m2(\theta 2) = 0{,}12$

La suma total es 0.42+0.18+0.12+0.28 = 1, lo que cumple las reglas de la función de verosimilitud:

$$\sum_{i1}^{4} m12(C_i) = m12(h2) + m12(h1, h2) + m12(h2, h3, h4) + m12(\theta 12) = 1$$

$$m12(\phi) = 0$$

Con la evidencia (e1 y e2) y e3 se deben considerar los resultados de la tabla anterior (e1 y e2):

	{h2}	**{h1,h2}**	**{h2,h3,h4}**	$\theta 12$
{h1}	$m123(\phi) =$ $m12(\{h2\}) \times m3(\{h1\})$ $= 0{,}336$	$m123(\{h1\}) =$ $m12(\{h1,h2\}) \times m3(\{h1\})$ $= 0{,}144$	$m123(\phi) =$ $m12(\{h2,h3,h4\})$ $\times m3(\{h1\}) = 0{,}224$	$m123(\{h1\}) =$ $m12(\{\theta 3\}) \times m3(\{h1\})$ $= 0{,}096$
$\phi 3$	m123({h2}))= $m12(\{h2\}) \times m3(\theta 3)$ 0.084	m123({h1,h2})= m12({h1,h2}) $\times m3(\theta 3)$ =0.036	m123({h2,h3,h4})= m12({h2,h3,h4}) $\times m3(\theta 3) = 0.056$	$m123(\theta 123) =$ $m12(\theta 12) \times m3(\theta 3)$ $= 0.024$

A pesar, de que sumar todos los valores no vacíos dan cero, hay un grado de conflicto $K = 0{,}336 + 0{,}224 = 0{,}560$. Ahora debemos agrupar (por ejemplo el foco h1 sale repetido dos veces y habría que sumarlo) y posteriormente normalizar (Dividir entre $1 - K$):

- $m123(h1) = (0{,}144 + 0{,}096)/(1 - 0{,}56) = 0{,}545$
- $m123(h2) = 0{,}084/(1 - 0{,}56) = 0{,}191$
- $m123(h1, h2) = 0{,}036/(1 - 0{,}56) = 0{,}082$
- $m123(h2, h3, h4) = 0{,}127$
- $m123(h1, h2, h3, h4) = 0{,}055$

$\sum = 1{,}000$ (Dejar tantos decimales como evidencias haya).

2.5.3.5. Como tratar con la incertidumbre :

La incertidumbre en este modelo se basa en los siguientes parámetros:

- **Credibilidad**: Indicador de la mínima confianza que podemos depositar en un elemento focal dado.

  De tal forma que si tenemos un marco de discernimiento θ y un elemento focal $A \subseteq \theta$, la credibilidad de A será igual a:
  $$Cr(A) = \sum_{B \subseteq A} m(B) \tag{60}$$

- **Plausibilidad**: Indicador de la máxima confianza que podemos depositar en un elemento focal dado.

  De tal forma que si tenemos un marco de discernimiento θ y un elemento focal $A \subseteq \theta$, la plausibilidad de A será igual a:
  $$Pl(A) = \sum_{B \cap A} m(B) \tag{61}$$

  En esta formula no solo consideramos A si no también todos aquellos elementos focales que tienen algo que ver con A aunque no sean subconjuntos

- **Intervalo de confianza**: Segmento del espacio numérico [0,1] que tiene como valor mínimo la credibilidad y cómo valor máximo la plausibilidad. En esta teoría, la incertidumbre no es un número, si no un un intervalo dinámico que va cambiando en función de las evidencias que vayan apareciendo.
  $$IC(A) = [Cr(A), PL(A)] \tag{62}$$

  ¿Cómo interpretar este intervalo?

  - $Cr(A) = 0 \wedge PL(A) = 1 \rightarrow$ No sé

- $Cr(A) = 1 \wedge PL(A) = 1 \rightarrow$ Sí
- $Cr(A) = 0 \wedge PL(A) = 0 \rightarrow$ No
- $Cr(A) \leq Pr(A) \leq PL(A)$
- $Cr(A) = PL(A) \neq 0 \neq 1 \rightarrow$ Probabilidad estadística del elemento focal

Esto demuestra que de forma natural y, esta teoría contiene a la teoría de la probabilidad.

2.5.3.6. Teoría evidencial vs Factores de Certidumbre :

Sean dos evidencias independientes que apoyan al mismo elemento focal $A \subseteq \theta$:

- $e1 : A1/m1(A) = s1$, $m(\phi) = 1 - s1$
- $e2 : A2/m2(A) = s2$, $m(\phi) = 1 - s2$
- m12(A) = s1s2 +s1(1-s2)+s2(1-s1)=s1+s2-s1s2

En este caso, considerando A como hipótesis en el modelo de Shortliffe y Buchanan estamos ante dos evidencias independientes que apoyan a la misma hipótesis $CF(H, E) = s1 + s2 - s1s2$

3. Curiosidades

- La diferencia entre estadística y probabilidad: La probabilidad es una teoría, por ejemplo, si tenemos un dado y usamos la fórmula anterior para averiguar cual es la probabilidad de que salga un 2 el resultado sería $\frac{N(f)}{N} = \frac{1}{6}$. Sin embargo, la estadística usaría la experiencia: registrando en x tiradas cuantas veces sale un 1, lo que se irá aproximando a $\frac{1}{6}$.
- En estadística es lo mismo tirar un dado un millón de veces que una vez un millón de dados, esto no es así en cuántica (Principio de Heisenberg).
- Recomendación: Siempre que digamos dado poner una '/'. Ej: $P(E/A) = 0{,}2$ P de E dado A quiere decir que hacemos A y obtenemos E el 20 % de las veces.
- Sensibilidad y especificidad:
  - Sensibilidad: TODO
  - Especificidad: TODO
- Bibliografía Razonamiento Cuasi-Estadístico:
  - Factores de centidumbre:
    - Heckerman, Probabilistic interpretation for MYCIN's certainty factors, Uncertainty in Artificial Intelligence, 1986
    - Shortliffe Buchanan, A model of inexact reasoning in medicine, Mathematical Biosciences, vol.23, 1975
  - Teoría evidenciaL: Shafer, A mathematical theory of evidence, Princeton University press, eds., 1976
- FC: A pesar de la "no deseabilidad"de que un problema no sea asociativo, en el mundo real existen muchos problemas que no lo son, no es lo mismo se rompió la cadera y se cayó que, se cayó y se rompió la cadera.
- Usar el mismo número de decimales que evidencial estemos usando.