Tema 5 - Incertidumbre

Profesor Juan Luis Castro Peña, Curso 20-21

- Hasta el momento se han descrito técnicas de representación del conocimiento y razonamiento para un modelo del mundo:
  - **Completo**: Se conocen todos sus conocimientos y sus datos.
  - **Consistente**: El conocimiento no da lugar a conclusiones contradictorias.
  - **Inalterable**: El conocimiento siempre es válido y no cambia.
- Sin embargo, en la mayoría de los dominios de interés no es posible crear tales modelos debido a la presencia de incertidumbre; lo que es falta de conocimiento seguro y claro de algo.

Fuentes de Incertidumbre:

Con respecto a los hechos

- **Ignorancia o Desconocimiento**
  - Puede que en un campo el conocimiento sea incompleto.
  - A veces la falta de conocimiento es irreducible o inevitable.
  - Aunque el conocimiento pueda llegar a ser completo, a veces conviene tomar decisiones con conocimiento incompleto. (En un ejemplo médico, conviene tratar un paciente aunque no se conozcan todos los resultados de las pruebas)
- **Vaguedad o Imprecisión**: Algunos conceptos son vagos e imprecisos, por ejemplo: persona alta, deuda pequeña, asignatura difícil.

Con respecto a las reglas

- Muchas veces las reglas que usan los expertos en determinadas situaciones son heurísticas.
- El ser humano razona y utiliza reglas que son:
  - Inexactas o Incompletas
    - "Si es un ave, entonces vuela" ¿Y los pingüinos?
    - "Si está en España habla español" ¿Y los turistas?
  - Imprecisas
    - Cuando haya mucho calor, encender el aire acondicionado.
  - Inconsistentes
    - "Al que madruga Dios le ayuda"
    - "No por mucho madrugar amanece más temprano"
- Esto hace que el razonamiento humano sea extremadamente potente en el sentido que permite con pocas reglas describir conocimiento muy complejo, ya que si solo se utilizara conocimiento preciso harían falta una cantidad inimaginable de reglas.

Razonamiento con Incertidumbre:

- **Objetivo**: Razonar con conocimiento incompleto, impreciso e incluso parcialmente inconsistente.
- **Implementación**:
  - Es difícil cumplir estos requerimientos utilizando los modelos de razonamiento clásico, la lógica de primer orden.
  - Se debe introducir modelos para manejar información vaga, incierta, incompleta y contradictoria, esto es crucial para que el sistema funcione en el mundo real.

Cuestiones a resolver por las aproximaciones a la incertidumbre:

- ¿Cómo evaluar si se da una condición imprecisa?
  - "Si tiene fiebre, beber agua" ¿Es 37.8º fiebre? ¿Cuánta agua?
- ¿Cómo combinar hechos imprecisos?

- "Con un grado de incertidumbre sé que X es alto, y con otro grado que X es delgado" ¿Con que grado de incertidumbre puedo afirmar que X es alto y delgado?
- ¿Qué confianza se puede tener en las conclusiones?
  - "Si X estudia bastante, aprobará; Juan estudia mucho" ¿Con qué incertidumbre puedo afirmar que Juan aprobará?
- ¿Cómo combinar la incertidumbre de un mismo hecho que han sido obtenidos por distintas deducciones?
  - "Si duele cabeza y garganta, es amigdalitis con certeza. Si hay puntos blancos en las amígdalas, es amigdalitis con otra certeza" ¿Si son esos tres síntomas juntos con que certeza se puede afirmar la amigdalitis?

Enfoque histórico

- Al principio se utilizaba razonamiento simbólico y evitaban el uso de números.
  - Se entendía que si los humanos no hacen uso de números, tampoco lo debían utilizar los Sistemas Expertos.
  - Los expertos tampoco podían suministrar números.
- Al crear aplicaciones más concretas se cayó en cuenta de la necesidad de representar y manejar la incertidumbre
  - El S.E. MYCIN que se utilizó para el tratamiento de infecciones bacterianas fue el primer éxito de este campo.
- Los métodos numéricos, medir la incertidumbre mediante números, son actualmente la herramienta más utilizada en IA para manejar la incertidumbre debido a los éxitos prácticos.

<u>Principales Modelos de Representación de la Incertidumbre:</u>

- La Lógica de Primer Orden (LPO) asume que el conocimiento es:
  - Exacto, los hechos son ciertos o falsos.
  - Completo, se conoce todo acerca del campo de trabajo.
  - Consistente, no tiene contradicciones.
- Por lo tanto la LPO:
  - **No** puede expresar incertidumbre.
  - **No** puede hacer deducciones lógicamente incorrectas pero probables.
  - **No** se puede trabajar con información contradictoria.
- Como la LPO asume que el conocimiento es completo y consistente, una vez que un hecho se asume es cierto, permanece así, lo cual es un razonamiento monótono.
  - Por lo tanto, añadir nuevo conocimiento siempre incrementa el tamaño de la Base de Conocimiento.
  - Si se tienen una base de conocimiento BS y BS' tal que $BC \subset BC'$, si con la primera se deduce una expresión s, también con BS' se deducirá s.
- La LPO no es adecuada para modelar la incertidumbre por lo que se necesitan dos nuevos modelos:
  - **Modelos Simbólicos**
    - **Lógica por Defecto**
    - **Lógica basada en Modelos Mínimos (La asunción del Mundo Cerrado)**
  - **Modelos Numéricos**
    - **Factores de Certeza**
    - **Lógica Difusa**
    - **Probabilidad**
    - Teoría de Dempster-Shaffer

Representación Simbólica de la Incertidumbre:

- **Lógica por Defecto**
  - Propuesta por Reiter para solucionar el problema del conocimiento incompleto.
  - Para ello, se introducen una serie de reglas por defecto
    - "Las reglas por defecto expresan características comunes a un conjunto de elementos que se asumen ciertas salvo que se indique lo contrario".

        - Se tiene reglas con elementos que han sido deducidos pero no son del todo seguros, si aparece un hecho que contradice alguno, se retracta.
            - Reglas:
                - Si se dan $Condiciones\ Positivas$ y no se sabe de $Condiciones\ Negativas$ entonces asumir $Consecuente$. "Si se sabe que X es un ave y no se sabe si X no vuela, asumir que X vuela"
                - Se deben de poder retractar: Si $Consecuente$, $Condiciones\ Positivas$ y $Condiciones\ Negativas$ entonces retractar $Consecuente$. "Si X vuela y es una ave y X no vuela, entonces retractar que X vuela"
    - **Asunción del Mundo Cerrado**
        - Sirve para manejar el conocimiento incompleto.
            - "Lo que no se puede probar a partir de mi Base de Conocimiento, es falso"
        - Utilizado en Bases de Datos y en PROLOG, y es lo que se ha estudiado y lo que se ha utilizado en CLIPS hasta ahora.
    - **Inconvenientes**:
        - Teorías complejas y a veces inconsistentes.
        - Si se realiza una serie de deducciones las cuales parten de una deducción la cual en un momento se creía era cierta, puede llegar un punto que se encuentre una inconsistencia con esa deducción inicial y deba retractarse, lo que hace que el resto de lo inferido deba también retractarse, lo cual aumenta mucho la complejidad del sistema.
        - O puede suceder que los hechos deducidos a partir de hechos asumidos por defecto no se retractan si posteriormente los hechos asumidos son retractados.

Representación Numérica de la Incertidumbre: Factores de Certeza

- **Grados de Creencia**
    - Los Factores de Certeza se calculan a partir de los Grados de Creencia y no Creencia en la hipótesis.
    - Varían entre 0 (creencia nula) y 1 (creencia total)
    - La relación de los FC y GC es $FC(H|E) = GC(H|E) - GC(\neg H|E)$
        - Se quiere saber el grado de confianza cuando dada una hipótesis H se tiene una evidencia E.
- Propiedades
    - A diferencia de los grados de creencia probabilísticos, $GC(H|E) + GC(\neg H|E)$ no necesariamente tienen que ser igual a 1, esto se debe a que los grados de creencia de la hipótesis se le preguntan al o los Expertos y estos no tienen que dar un grado de creencia que sume a 1, además que las diferentes respuestas de los Expertos se pueden promediar.
- Los FC aparecieron en el Sistema Experto MYCIN
    - Fue desarrollado en la Universidad de Stanford en la década de los 70 para el diagnóstico y consulta de enfermedades infecciosas.
    - Los factores de certeza de MYCIN consistían en reglas de la forma:
        - Evidencia→Hipótesis ≡ $FC(H|E)$
        - El factor de certeza representa la certidumbre de la Hipótesis cuando se observa la Evidencia
        - Los FC varían entre -1 (creencia nula) y 1 (creencia total)
    - Reglas de MYCIN: Ejemplo

```
( $AND  (SAME CNTXT GRAM GRAMNEG)
        (SAME CNTXT MORPH ROD)
        (SAME CNTXT AIR ANAEROBIC)
    (CONCLUDE CNTXT IDENTITY BATEROIDES TALLY .6)
)
```

        - Significado: Si el organismo es gram-negativo, tiene forma de bastón y es anaeróbico, entonces con una certeza de 0.6 el organismo es un bacteoride
        - Los factores de certidumbre se introducían a mano por el diseñador.

Combinación de Factores de Certeza

- Combinación de Reglas Convergentes

> Si E_1 entonces H con $FC(H|E_1)$
> Si E_2 entonces H con $FC(H|E_2)$
> ---
> Si $E_1 \vee E_2$ entonces H con $FC(H|E_1 \vee E_2)$

  - Con $FC(H|E_1 \vee E_2)$=$f_{comb}\left(FC(H|E_1), FC(H|E_2)\right)$ definida como:

$$f_{comb}(x,y) = \begin{cases} x+y-xy & si\ x,y > 0 \\ \dfrac{x+y}{1-\min(|x|,|y|)} & si\ \left(x > 0\ \wedge y \leq 0\right) \vee (x \leq 0 \wedge y > 0) \\ x+y+xy & si\ x,y \leq 0 \end{cases}$$

- Encadenado de Reglas

> Si A entonces B con $FC(B|A)$
> Si B entonces C con $FC(C|B)$
> --------------------------------------
> Si A entonces C con $FC(C|A)$

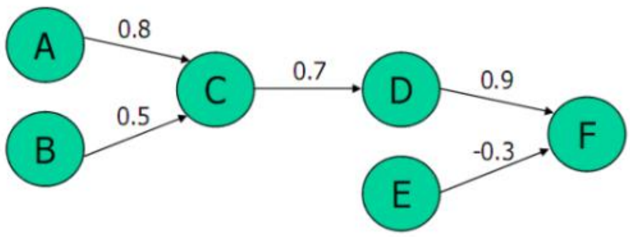
  - Con $FC(C|A)$ definido como

$$FC(C|A) = \begin{cases} FC(C|B) \times FC(B|A) & si\ FC(B|A) \geq 0 \\ 0 & si\ FC(B|A) < 0 \end{cases}$$

    - "Si de A no puedo inferir B, entonces C es 0"

Ejemplo de Combinación de Factores de Certeza

- **Dadas siguientes reglas calcular el factor de certeza de la proposición A ∨ B ∨ E→F**



  - A→C , B→C FC(C|A,B)=0.8+0.5(1-0.8)=0.9 (Convergencia)
  - A ∨ B →C, C →D FC(D|A,B)=0.9x0.7=0,63 (Encadenado)
  - A ∨ B →D, D →F FC(F|A,B)=0.63x0.9=0,567 (Encadenado)
  - A ∨ B →F, E →F FC(F|A,B,E)= (0.567-0.3)/(1-min(0.567,0.3)) =0.38 (Convergencia)

- **Por tanto, si observamos A,B y E podemos concluir F con certidumbre 0.38**

¿Cómo era el rendimiento de MYCIN?

- El sistema MYCIN proporcionaba diagnósticos y recomendaciones terapéuticas que eran al menos tan buenas como los mejores expertos en la especialidad.
- Si embargo, los factores de certeza pueden producir incoherencias, por ejemplo:
  - Si se deduce con certeza 0.8 que sarampión implica ronchas y que ronchas con 0.5 implica alergias, se puede obtener encadenando que el sarampión con 0.4 implica una alergia.

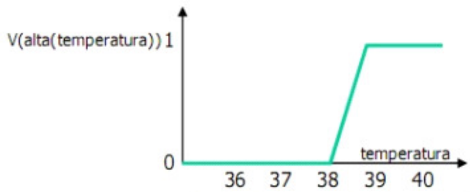
Lógica Difusa:

- Fue propuesta a mediados de 1960
- El ser humano se expresa y razona comúnmente con términos vagos en lugar de precisos, se propone una lógica que funcione de esta manera, que tienen como un continuo, pueden tener una cierta verdad.
  - Fiebre alta, persona delgada, velocidad alta, …
- Se pueden representar numéricamente los términos vagos utilizando la lógica multivaluada entre $[0,1]$.
  - Por ejemplo gradoDeVerdad("fiebre alta")∈ $[0,1]$, en función de la temperatura del paciente se obtendría un valor entre 0 y 1 que indicaría el grado de verdad de eso.
- Las proposiciones, los términos imprecisos, se interpretan y representan como restricciones de los valores que puede tomar una variable.

Interpretación de proposiciones lingüísticas

- ¿Cómo se representa, por ejemplo, "La temperatura del enfermo es alta"?
- Si se utiliza la lógica clásica sería que la temperatura fuese > 38.5 pero si un enfermo tuviera una temperatura de 38.4 entonces no se denominaría "alta".
- Con lógica difusa, que sea "alta" la temperatura es una función que indica el grado de verdad entre $[0,1]$ según el valor de la temperatura. Se mide la **posibilidad** no **probabilidad**. Se habla de posible porque la suma de valores de la función no tiene que sumar entre 0 y 1.

$$V(\text{alta}(\text{temperatura})) = \begin{cases} 0 & \text{temperatura} < 38 \\ \text{temperatura} - 38 & 38 \leq \text{temperatura} \leq 39 \\ 1 & \text{temperatura} > 39 \end{cases}$$



Grados de verdad del predicado alta(temperatura)

Proposiciones Simples

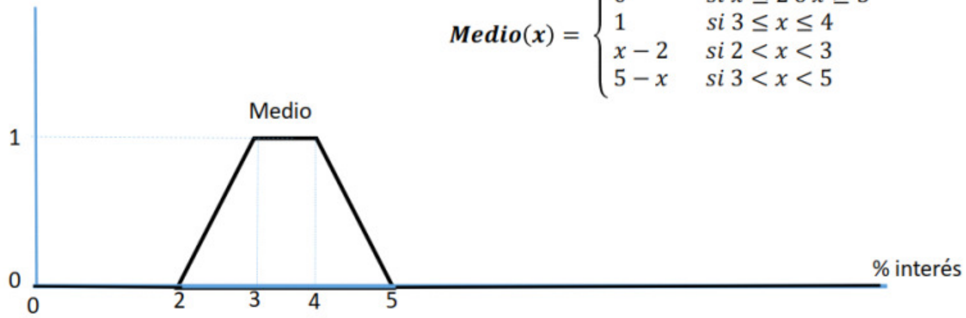
- Una proposición simple P será de la forma $\boldsymbol{X}$ **es** $\boldsymbol{A}$.
  - *Temperatura de Enfermo* es *alta*;
  - X es una variable.
  - A es una etiqueta lingüística.
  - Se interpreta como una función $A: Dominio(X) \rightarrow [0,1]$
    - La función va de los posibles valores que puede tomar X a $[0,1]$.
- Cuando $X = x$, el grado de verdad $V(P) = A(x)$, de esta manera $V(P) \in [0,1]$. Es decir, el rango de la función indicará el grado de verdad P para x.
- P se interpreta como una restricción graduada al conjunto de valores que puede tomar X.
- Si $A(x) = 0$, X no podrá tomar valor x, si $A(x) > 0$ puede tomarlo. Cuanto más grande sea $A(x)$ es más probable que tome ese valor.

Ejemplo:

- Un conjunto difuso común es el conjunto de números difusos trapezoidal. Medio es un conjunto difuso en este ejemplo.

% de interés es Medio

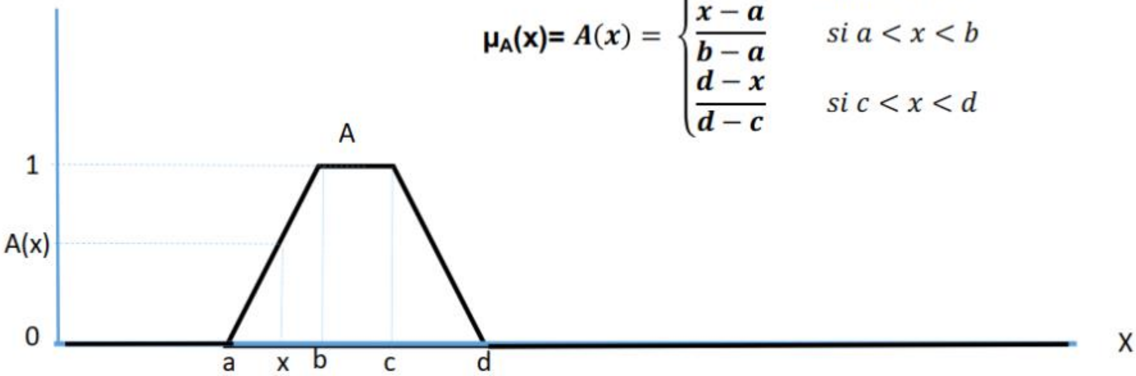
$$Medio(x) = \begin{cases} 0 & si\ x \leq 2\ ó\ x \geq 5 \\ 1 & si\ 3 \leq x \leq 4 \\ x-2 & si\ 2 < x < 3 \\ 5-x & si\ 3 < x < 5 \end{cases}$$



- Existe una relación clara entre la teoría de subconjuntos y la lógica proposicional, ya que dada una proposición se tiene un conjunto de elementos que hacen verdad dicha proposición.
- Además del concepto de proposición difusa se tiene el concepto de subconjunto difuso, si un elemento está en 1 quiere decir que seguro está en el subconjunto, 0 es que no está y entre 0 y 1 aquellos elementos que estarán en un cierto grado en ese subconjunto.
- Se define ese subconjunto como cualquier función que va desde ese subconjunto al intervalo $[0,1]$.

A: D → [0,1]
A(x)="grado en que x pertenece a A"

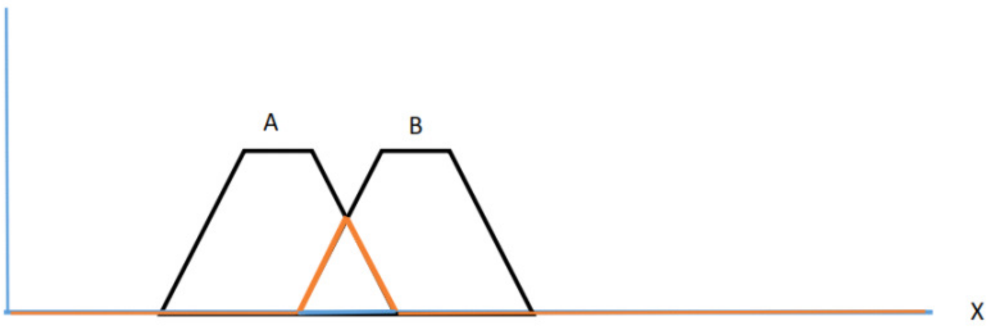
$$\mu_A(x) = A(x) = \begin{cases} 0 & si\ x \leq a\ ó\ x \geq d \\ 1 & si\ b \leq x \leq c \\ \dfrac{x-a}{b-a} & si\ a < x < b \\ \dfrac{d-x}{d-c} & si\ c < x < d \end{cases}$$



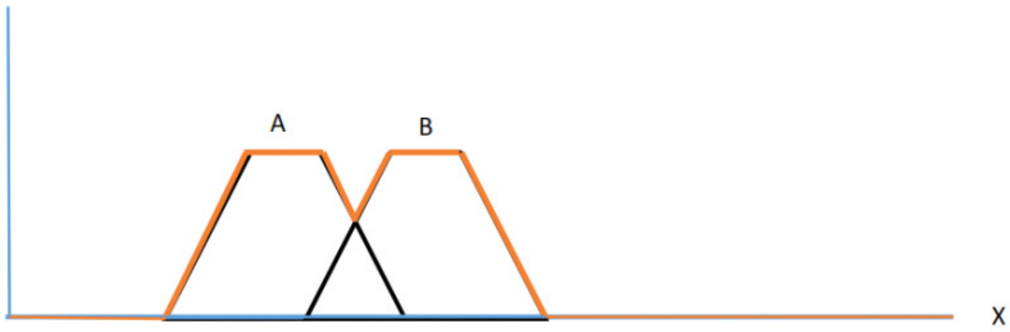
- Un subconjunto difuso A de un dominio D es cualquier función $A(x)$ que vaya de los elementos de D hacia $[0,1]$. Se interpreta $A(x)$ como el grado en que x pertenece a ese subconjunto A.
- $\mu_A(x)$ es otra manera de representar la función de pertenencia.

Operaciones con Subconjuntos Difusos

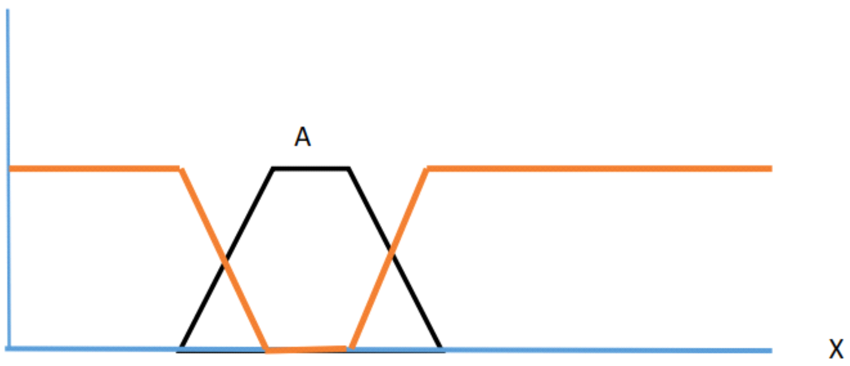
- **Intersección**: $A(x) \cap B(x) = Min(A(x),\ B(x))$
  - Si se tienen dos subconjuntos difusos que se intersecan, el valor de esta intersección es el mínimo de los valores de la función de pertenencia de esos subconjuntos para un valor x dado.



- **Unión**: $A(x) \cup B(x) = Max(A(x),\ B(x))$



- **Complemento**: $D(x) - A(x) = 1 - A(x)$; con D siendo el conjunto al que A pertenece.
    - $A(x) \cap 1 - A(x)$ no necesariamente es un conjunto vacío.



Proposiciones Compuestas

- El valor de verdad de las proposiciones compuestas se calcula a partir del valor de verdad de las proposiciones que se combinan, como en lógica clásica.
    - $V(P \wedge Q) = Min\big(V(P), V(Q)\big)$
    - $V(P \vee Q) = Max\big(V(P), V(Q)\big)$
    - $V(\neg P) = 1 - V(P)$
    - $V(P \rightarrow Q) = Min\big(1, 1 - V(P) + V(Q)\big)$
- No se verifica…
    - Principio de No Contradicción, puede ocurrir que $V(P \wedge \neg P) \neq 0$ pero si que $V(P \wedge \neg P) \leq \frac{1}{2}$
    - Principio del Tercero Excluido, puede ocurrir que $V(P \vee \neg P) \neq 1$ pero si que $V(P \vee \neg P) \geq \frac{1}{2}$

Interpretación de Relaciones Vagas

- Si se tiene que la proposición implica más de una variable esto también implica que la función de pertenencia ya pasará a ser también de múltiples variables, por ejemplo, si se tratan de dos variables, el espacio se convierte en una superficie de posibilidades.

- Ejemplo: “Interés del banco A es mucho mas grande que el del banco B”

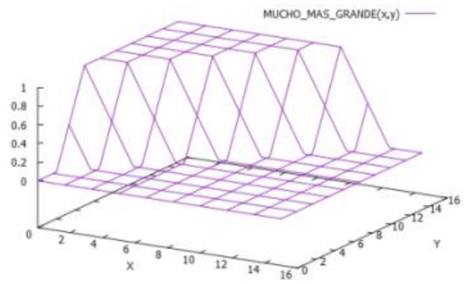
- Se representa como un X esta en relación MUCHO_MAS_GRANDE que Y

MUCHO_MAS_GRANDE(x,y) =

=“grado en que x es mucho mas grande que y” ∈ [0,1]

(X,Y) es MUCHO_MAS_GRANDE

W es Mucho_MAS_GRANDE



Relaciones Difusas

- Una relación n-aria de conjuntos será el producto cartesiano de n dominios de cada una de las variables.
  - $R: D_1 \times \cdots \times D_n \to [0,1]$
  - $R(x_1, \dots, x_n) =$ "Grado en que $x_1, \dots, x_n$ se verifican en la relación R"
- Operadores
  - No se profundizará en ello, pero se deben conocer.

- Extensión cilíndrica: $\text{Ext}(R): D_1 \times \dots \times D_n \times D \rightarrow [0,1]$

  $\text{ExtR}(x_1,\dots,x_n,y) = R(x_1,\dots,x_n)$
- Proyección: $\text{Proy}(R) : D_1 \rightarrow [0,1]$

  $\text{Proy}(R)(x_1) = \sup \{R(x_1,\dots,x_n) \,/\, x_2 \in D_2, \dots, x_n \in D_n\}$

Composición de Relaciones

- Si $R_1(x_1, \dots, x_n)$ y $R_2(y_1, \dots, y_m)$ son dos relaciones, entonces…

  - $(R1 \vee R2)(x_1,\dots,x_n, y_1,\dots,y_m) = \max(R1(x_1,\dots,x_n), R2(y_1,\dots,y_m))$
  - $(R1 \wedge R2)(x_1,\dots,x_n, y_1,\dots,y_m) = \min(R1(x_1,\dots,x_n), R2(,y_1,\dots,y_m))$
  -
  - $\neg R(x_1,\dots,x_n) = 1 - R(x_1,\dots,x_n)$

  - X es A y Y es B → (X,Y) es A^B $\qquad$ A^B(x,y)= Min(A(x),B(y))
    - X es A y X es B → X es A∩B

  - X es A o Y es B → (X,Y) es A˅B $\qquad$ A˅B(x,y)= Max(A(x),B(y))
    - X es A o X es B → X es AUB

  - X no es A → X es ¬A $\qquad$ ¬A(x) = 1 – A(x)

Inferencia Difusa

- Exista una única regla que se utliza en lógica difusa: Regla composicional de inferencia.

(X,Y) es R

X es A

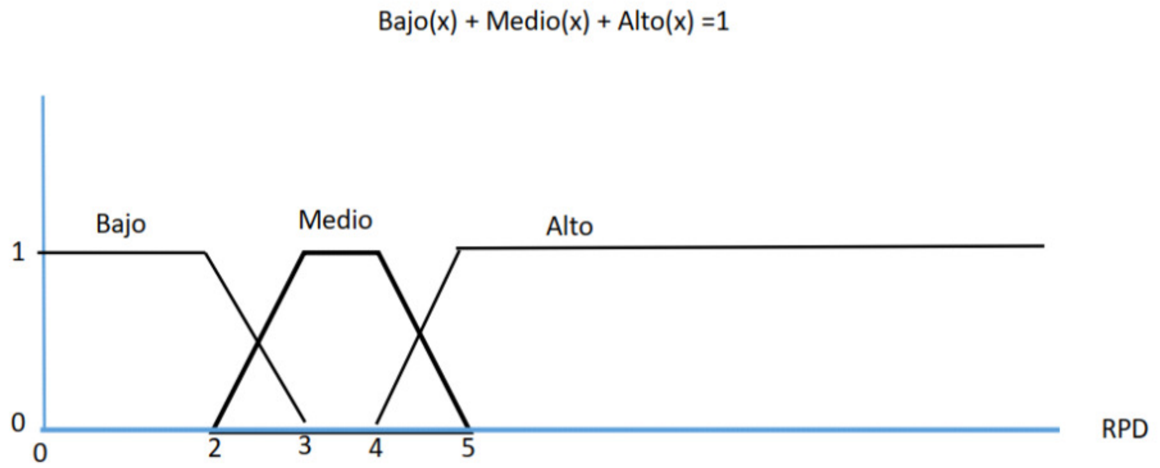
Y es AoR

$$A \vee R(y) = \sup\{\min(A(x), R(x,y) \mid x \in Dominio(x)\}$$

- Cuando existe una relación entre dos variables que es imprecisa, y se tiene una afirmación sobre una de las variables imprecisa, se puede deducir otra afirmación imprecisa sobre la otra variable.

Partición Difusa

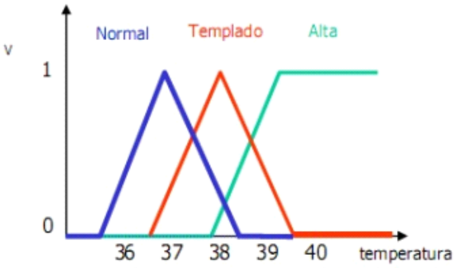
- Para describir ciertas cosas se utilizan etiquetas lingüísticas, por lo tanto, se puede particionar el valor de una variable en diferentes subconjuntos difusos.
- Para cada x, el valor en un punto de las funciones de pertenencia de las particiones siempre debe sumar 1.



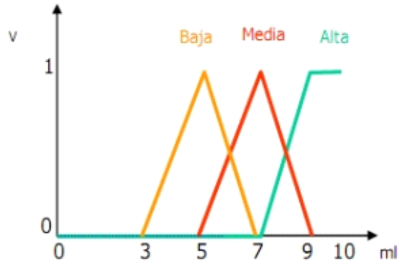
Razonamiento Difuso basado en Reglas

- El razonamiento difuso consta de 4 pasos.
    - Dado por ejemplo, lo siguiente:

- Se le toma la temperatura a un paciente y se quiere saber la dosis apropiada de un medicamento.
- Hechos:
  - temperatura=38
- Reglas:
  - Si la temperatura en Normal entonces dosis es Baja
  - Si la temperatura es Templada entonces dosis es Media
  - Si la temperatura es Alta entonces dosis es Alta



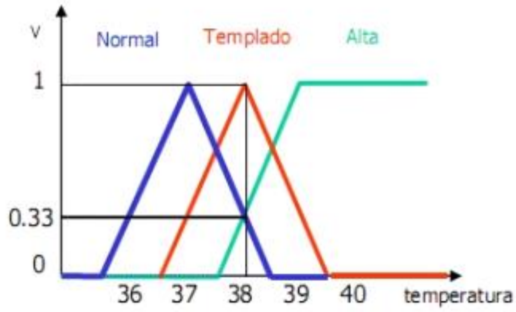
Conjuntos Difusos temperatura normal, templada y alta



Conjuntos Difusos dosis baja, media y alta

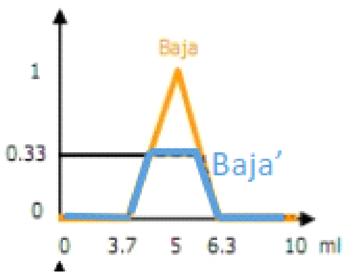
- **Paso 1: Matching**
  - Es la obtención del grado de verdad de los antecedentes utilizando los hechos observados. Se aplica la función de pertenencia dado un valor concreto.
    - En el ejemplo…

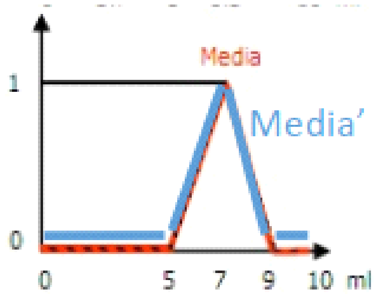
- Hechos:
  - temperatura=38
  - Grados de verdad:
    - » normal(temperatura): 0.33
    - » templado(temperatura): 1
    - » alta(temperatura): 0.33

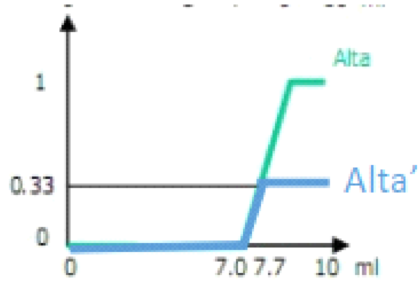


Difusión

- **Paso 2: Inferencia Difusa**
  - Se obtienen las consecuencias deducidas como conjuntos derivados de los consecuentes de la regla; se infiere una variante del consecuente dado los valores de la premisa. Esto se calcula…
    - Con el **mínimo**: Los grados de verdad del consecuente se cortan a la altura del grado de verdad de la premisa, esto es lo más común.
    - Con el **producto**: Se multiplican los grados de verdad del consecuente y la premisa.

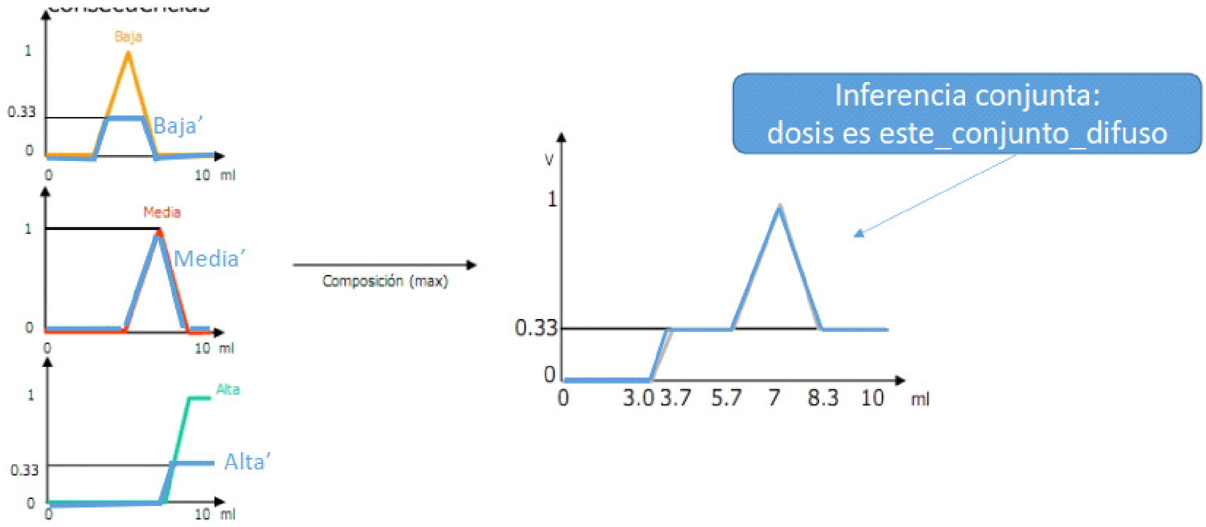




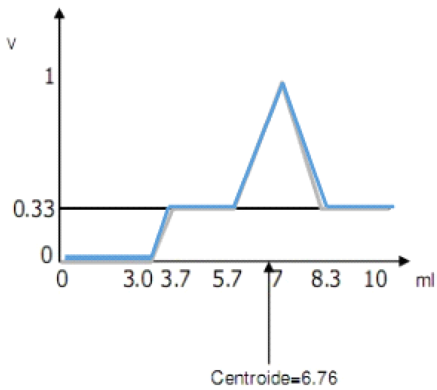


- **Paso 3: Agregación de Consecueintes**

- Todas las inferencias relativas a la misma variable se combinan para producir una salida difusa, una inferencia difusa conjunta, se calcula…
    - Con el **máximo**: se toma el máximo de los grados de verdad correspondientes a las distintas consecuencias, esto es lo más común.
    - Con la **suma**: se toma la suma de los grados de verdad correspondientes a distintas consecuencias.



- **Paso 4: Concisión o Defuzzification**
    - Convertir la salida difusa en un valor concreto para la variable, se calcula…
        - Con el **centroide**: Se calcula el centro de gravedad de los grados de verdad de la conclusión difusa, es decir, obtener el punto que divide exactamente el área en dos de la salida.
        - Con el **máximo**: Se elige el máximo valor de los grados de verdad.



$$\text{centroide} = \frac{\int x \cdot v(x)\, dx}{\int v(x)\, dx}$$

– Por tanto con 38 grados la dosis sería de 6.76 ml.

- Con estos simples pasos, hasta descripciones que no poseen modelos matemáticos que lo describan, con la lógica difusa se puede obtener un modelo aproximado que funciona muy bien.

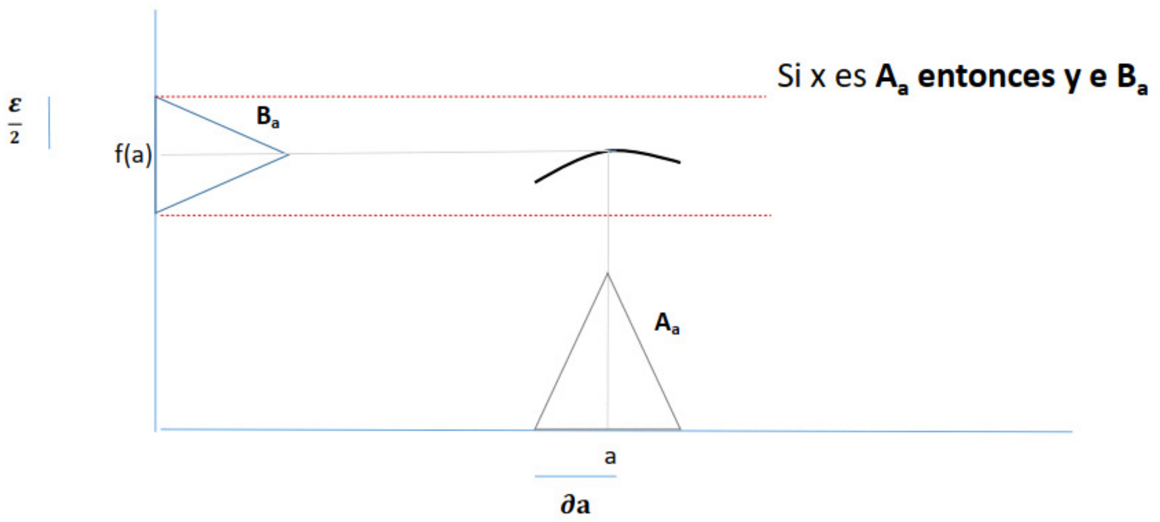
Capacidad de la Lógica Difusa

- Teorema de aproximación universal de los Sistemas Basados en Reglas Difusas:
    - Los SBRD son aproximadores universales, es decir, si se tienen un conjunto de funciones definidas sobre un ***compacto***, estas funciones se pueden aproximar tanto como se quiera por un sistema basado en reglas difusas.
        - Por ejemplo, los polinomios. Una función continua cualquiera se puede aproximar por un

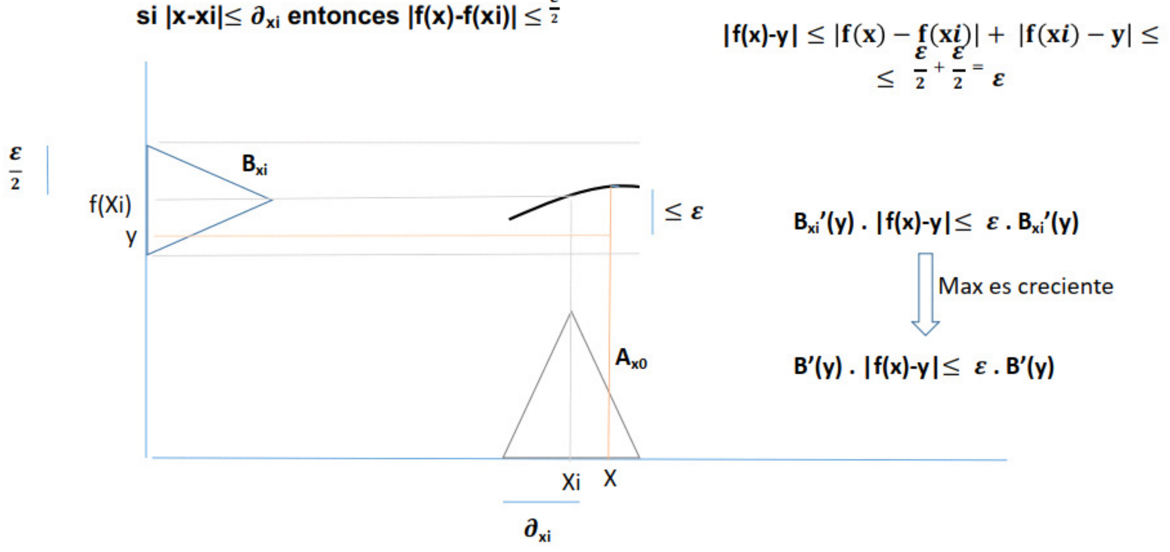
polinomio.
- En las técnicas de Inteligencia Artificial, el tener un aproximador universal asegura que, al menos a nivel teórico, utilizar esa técnica existiría al menos un elemento que se puede definir con esa herramienta que puede resolver el problema al nivel necesitado.

- Demostración "por encima"
  - **Compacto**: Cada recubrimiento por una familia de conjuntos abiertos tiene un subrecubrimiento finito.
    - Sobre R^n el compacto es equivalente a "cerrado" o "acotado", a efectos prácticos, un compacto es que, por ejemplo, una variable tenga por dominio un intervalo cerrado.
    - Otro ejemplo es de una bola cerrada en R^3 o un círculo cerrado en R^2.
  - Formalmente, $\forall K\ compacto\ de\ R^n, \forall f \in C(K), \forall \epsilon > 0, \exists S \in SBRD(K)\ tal\ que\ \left\|f - S\right\| \leq \epsilon$
  - Visualmente, se define un conjunto difuso para cada espacio del dominio. Este conjunto se hace tan pequeño como para asegurar que todos los puntos que forman ese conjunto difuso no varíen más de $\frac{\epsilon}{2}$ en el espacio del dominio. Como esto lo que consigue es un recubrimiento abierto, entonces existirá un subrecubrimiento finito cerrado y ese es el que se necesita. Ahora se utiliza una función como consecuente para asegurar que el valor no se sale de ese espacio por debajo de $\frac{\epsilon}{2}$ y finalmente se desfuzzifica. Y se desmuestra que, entonces, se puede aproximar esa función a un nivel ϵ.

f continua en a → $\exists \partial a$ tal que si $|x-a| \leq \partial a$ entonces $|f(x)-f(a)| \leq \frac{\varepsilon}{2}$



si $|x-xi| \leq \partial_{xi}$ entonces $|f(x)-f(xi)| \leq \frac{\varepsilon}{2}$



Defuzzificando

$$|\mathbf{f(x)\text{-}S(x)}| = \left| f(x) - \frac{\int y.B'(y)dy}{\int B'(y)dy} \right| = \left| \frac{\int f(x).B'(y)dy - \int y.B'(y)dy}{\int B'(y)dy} \right| =$$

$$= \left| \frac{\int (f(x)-y)B'(y)dy}{\int B'(y)dy} \right| \leq \frac{\int |f(x)-y| B'(y)dy}{\int B'(y)dy} \leq \frac{\int \varepsilon\, B'(y)dy}{\int B'(y)dy} = \varepsilon$$

Razonamiento Probabilístico:

Recordatorio de estadística

- La teoría de la probabilidad (TProb) es el área de las matemáticas que ha sido aplicada a problemas de razonamiento con incertidumbre.
  - Asigna valores numéricos a las proposiciones y permite que dadas las probabilidades de ciertas proposiciones y algunas relaciones entre ellas, se pueden asignar probabilidades a las proposiciones relacionadas.
  - En relación con la Lógica de Primer Orden, en la **LPO** las proposiciones son ciertas o falsas mientras que en TProb también pueden ser ciertas o falsas pero se tiene un grado de creencia de la certeza o falsedad.

¿Qué son las probabilidades?

- Definición **frecuentista**: Es el valor cuando el número de pruebas tiende a infinito de la frecuencia de que ocurra algún evento.
- Definición **subjetiva**: Es el grado de creencia acerca de un evento incierto.
- Aun así, existe un consenso sobre el modelo matemático que soporta la teoría.

Valores numéricos de la probabilidad

- Se denota con $P(A)$ la probabilidad de la proposición A.
  - $A =$ "El paciente tiene Covid"
  - $A =$ "Kayne está en París"
- Los valores satisfacen una serie de axiomas
  - $0 \leq P(A) \leq 1$
  - $P(Proposición\ Verdadera) = 1$
  - $P(A \vee B) = P(A) + P(B)$, siempre que A y B sean disjuntos
- A partir de los mismos se puede demostrar, por ejemplo
  - $P(\neg A) = 1 - P(A)$
  - $P(Proposición\ Falsa) = 0$
  - $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ si A y B no son disjuntos.

Variables Aleatorias

- Muchas veces se tiene un evento con un conjunto de resultados
  - **Completo**: Se conocen todos los posibles resultados,
  - **Mutuamente excluyente**: No se pueden dar dos resultados distintos simultáneamente.
- En lugar de tener una proposición para cada resultado se introduce el concepto de la variable aleatoria.
- Se permiten proposiciones de la forma "Variable = Resultado".
  - Si se tiene por ejemplo $M =$"Resultado de tirar una moneda con valores posibles cara y cruz" se permiten las proposiciones

    - $M = Cara$ y $M = Cruz$ y se pueden representar así, $P(M = Cara)$ y $P(M = Cruz)$, las probabilidades de obtener una cara y una cruz respectivamente.
- Por consistencia, se puede considerar que todas las proposiciones son variables aleatorias que toman dos valores: verdadero y falso.

Distribuciones de Probabilidad

- Dada una variable aleatoria, se desea saber la probabilidad de cada valor que puede tomar. Esta descripción se llama distribución de probabilidad (**Dprob**) de la variable aleatoria y consiste en listar los valores de probabilidad para cada valor de la variable.

Ejemplo:

- Distribución de probabilidad de la variable Llueve

Variable
Valores
Probabilidades

Llueve	P(Llueve)
Verdadero	0.1
Falso	0.9

Proposiciones más complejas

- Se puede estar interesado en estudiar varias variables en conjunto, para ello se debe asignar probabilidades a cada posible combinación de los valores de las variables.
- El listado de todos esos valores se llaman la distribución conjunta del conjunto de variables.
    - Por ejemplo, $P(Sarampión = Verdad \vee Fiebre = Verdad)$ es la probabilidad de que el paciente tenga sarampión y fiebre. Se escribe como $P(sarampión \vee fiebre)$ o $P(sarampión, fiebre)$

Distribución conjunta de las variables Llueve y EnCalle P(Llueve, EnCalle):

Llueve	EnCalle	P(Llueve,EnCalle)
Verdadero	Verdadero	0.01
Verdadero	Falso	0.09
Falso	Verdadero	0.2
Falso	Falso	0.7

También se puede escribir como:

Llueve	EnCalle	P(Llueve,EnCalle)
llueve	encalle	0.01
llueve	¬encalle	0.09
¬ llueve	encalle	0.2
¬ llueve	¬encalle	0.7

- Esto recuerda las tablas de vedad de lógica, excepto que describe la probabilidad para cada combinación de valores de las variables y generalmente dichos valores no se pueden calcular a partir de sus componentes.
- La distribución conjunta contiene todo lo que se necesita saber de un conjunto de variables aleatorias, en particular, la distribución de cada variable individual se puede calcular a partir de la distribución conjunta. Esto se llama la **distribución marginal**.

– Ejemplo: Supongamos las variables aleatorias: Llueve y EnCalle con distribución conjunta P(Llueve,EnCalle)

llueve	encalle	0.01
llueve	¬encalle	0.09
¬ llueve	encalle	0.2
¬ llueve	¬encalle	0.7

- Entonces P(llueve)= P(llueve ∧ enCalle)+ P(llueve ∧¬ enCalle)=0.01+0.09=0.1.
- De forma similar P(¬llueve)= 0.9
- También podemos calcular la probabilidad de disyunciones: P(llueve ∨ enCalle)=0.01+0.09+0.2= 0.3

Probabilidad Condicional

- Se denota como $P(A|B)$ a la probabilidad de que suceda A tal que haya sucedido B. Se puede interpetar como el grado de creencia en A cuando todo lo que se sabe es B. O alternativamente, de los casos en los que se da B, ¿en qué proporción se da A? También se conoce como **Probabilidad A Posteriori.**
- Se define como $P(A|B) = \frac{P(A \wedge B)}{P(B)}$ asumiendo que $P(B) \neq 0$ o equivalentemente $P(A \wedge B) = P(A|B) \times P(B)$
- Permite conocer la probabilidad de que se tomen unos determinados valores por un conjunto de variables aleatorias cuando se saben los valores que han tomado otras.

– Ejemplo: P(Llueve|enCalle)

Llueve	P(Llueve|enCalle)
llueve	0.05
¬ llueve	0.95

– Ejemplo: P(Llueve| ¬ enCalle)

Llueve	P(Llueve| ¬enCalle)
llueve	0.11
¬ llueve	0.89

– Nótese que Llueve|enCalle y Llueve| ¬ enCalle son variables aleatorias

Razonamiento con Probabilidades: La Regla de Bayes

- $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$, de manera intuitiva, la probabilidad de una hipótesis A dada una evidencia B es proporcional a la probabilidad de la hipótesis A multiplicada por el grado en que la hipótesis predice los datos. Es una especie de Modus Tollens para probabilidad.
- En muchos problemas dado un conjunto de datos evidencia B, se tiene que seleccionar la hipótesis A más probable mediante la fórmula.

Regla de Bayes: Forma General

- Si se tienen un conjunto de proposiciones $A_1, \dots, A_n$ completas y mutuamente excluyentes se tiene que:
  - $P(A|B) = \dfrac{P(B|A_i) \times P(A_i)}{P(B|A_1) \times P(A_1) + \cdots + P(B|A_n) \times P(A_n)}$
- O lo que es lo mismo, si se tiene una variable aleatoria A con valores $a_1, a_2, \dots, a_m$
  - $P(a_i|B) = \dfrac{P(B|a_i) \times P(a_i)}{P(B|a_1) \times P(a_1) + \cdots + P(B|a_n) \times P(a_n)}$

- La regla de Bayes tiene ciertas limitaciones
  - Para poder deducir utilizando Bayes se tiene que tener problemas que poseen mucha evidencia.
  - Aplicar el Teorema de Bayes a hipótesis muy poco probables aumenta sus probabilidades, por lo tanto, es mejor tener un indicio de que la probabilidad será alta; o bien tener más evidencia para aumentar la probabilidad.
  - Se necesitará todas las posibles combinaciones de probabilidades conjuntas, si fuera una variable binaria utilizaría 2^n muestras.
  - Esto es inviable excepto en el caso de que las evidencias sean independientes entre sí.

Independencia

- Se dice que dos afirmaciones son independientes si el conocimiento de una no cambia la probabilidad de la otra.

  - Formalmente A_1,A_2 son independientes si $P(A_1|A_2)=P(A_1)$
    o de forma equivalente: $P(A_2|A_1)=P(A_2)$
    o utilizando la regla del producto $P(A_1 \wedge A_2)= P(A_1)\ P(A_2)$
  - Entonces $P(A_1 \wedge A_2 \wedge \ldots \wedge A_n)= P(A_1)\ P(A_2) \ldots P(A_n)$
    Para especificar la distribución conjunta de n variables se necesitan o(n) números en lugar de $o(2^n)$

  - Ahora si es algo más manejable el cálculo del Teorema de Bayes con evidencia independiente, pero el problema es que es muy raro que exista evidencia independiente una de otra. Es muy restrictivo.

Independencia Condicional

- Es mucho más posible que existan evidencias que sean solamente independientes de otras, se dice que dos proposiciones A_1, A_2 son independientes dada una tercera B si cuando está presente B el conocimiento de una no influye a la otra: $P\left(A_1 \middle| A_2, B\right) = P(A_1|B)$

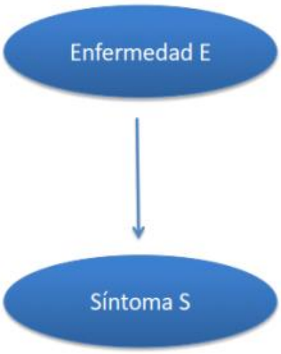
  o de forma equivalente: $P(A_2|A_1,B)=P(A_2|B)$
  o de forma equivalente: $P(A_1 \wedge A_2\ |B)= P(A_1|B)\ P(A_2|B)$
  - Ejemplo:
    - A_1="Tengo congestión nasal" A_2="Tengo fiebre" A_3="Tengo gripe"
    - A_1 y A_2 son dependientes pero son independientes si se conoce A_3.
  - Ahora se tiene: $P(A_1 \wedge A_2 \wedge \ldots \wedge A_n\ |B)=P(A_1|B)\ P(A_2|B) \ldots P(A_n|B)$
    - Tenemos o(n) números en lugar de $o(2^n)$

Redes Bayesianas

- La clave que hace factible la inferencia con probabilidades es la introducción explícita de independencia de las variables
- El modelo más extendido de representación de independencias lo constituyen las Redes Bayesianas, en este modelo se representa de forma explícita la dependencia entre variables mediante un grafo.
- Los nodos del grafo se corresponden con variables y las dependencias se representan mediante arcos entre ellas.

Aplicación de Redes Bayesianas

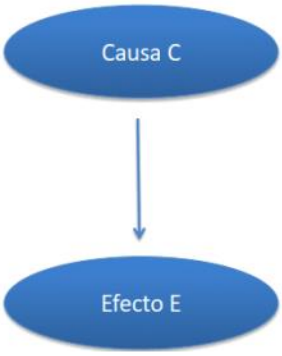
- Ejemplo simple



Tenemos un conocimiento sobre la probabilidad de E (P(E), probabilidad a priori). Indago y descubro que se da S. Conociendo P(S/E) y P(S/¬E), ¿que probabilidad hay ahora (P(E/S) probabilidad a posteriori) de que se de E?

$$P(E|S) = \frac{P(S|E).P(E)}{P(S|E).P(E)+P(S|\neg E).P(\neg E)}$$

- **Caso simple: Probabilidad inducida por un efecto**



Se conoce P(C), P(E|C) y P(E|¬C), ¿Puedo calcular P(C|E)?

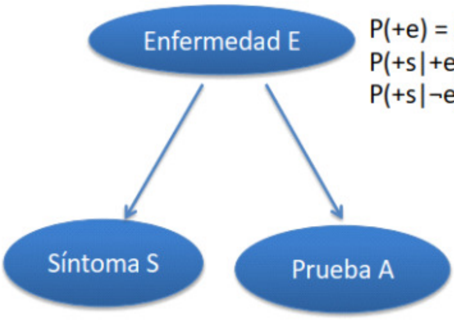
$$P(C|E) = \frac{P(E|C)*P(C)}{P(E|C)*P(C)+P(E|\neg C)*P(\neg C)}$$

- Se necesitan de 3 datos, la propiedad a priori de la causa, la probabilidad que se dé el **efecto dada la causa** y que **se dé sin la causa** para calcular la probabilidad que se dé la **causa dado el efecto**.

- **Probabilidad inducida por varias efectos** (suponiendo independencia)

ENFERMEDAD (E): presente (+e), ausente (¬e)
SÍNTOMA (S): presente (+s), ausente (¬s)
PRUEBA ANALÍTICA (A): positivo (+a), negativo (¬a)

Grafo dirigido acíclico



P(+e) = 0’002 → probabilidad a priori
P(+s|+e) = 0’93 P(+a|+e) = 0’995 → prob de los efectos si +e
P(+s|¬e) = 0’01 P(+a|¬e) = 0’003 → prob de los efectos si -e

P(+e|+s,+a)= P(+e,+s,+a) / P(+s,+a) ecuación cálculo
P(+e,+s,+a) = P(+e).P(+s|+e).P(+a|+e) = 0,00185
P(¬e,+s,+a) = P(¬e).P(+s|¬e).P(+a|¬e) = 0,00003
P(+s,+a) = P(+e,+s,+a) + P(¬e,+s,+a) = 0,00188
P(+e|+s,+a)= P(+e,+s,+a) / P(+s,+a) = 0,984

- Cuando hay varios efectos, se asume independencia entre los mismos. Es recomendable buscar efectos que tengan una probabilidad alta de que se den dada una causa y que sean de probabilidad baja si no es por la causa para mejorar la precisión.

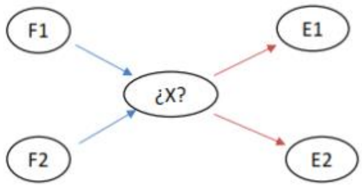
Redes Causales

- Probabilidad inducida por varios factores independientes y varios efectos independientes.

- Probabilidad inducida por varios factores independientes y varios efectos independientes.

Paludismo (X): presente +x, ausente –x
Zona de origen (F1): alto riesgo f_1^+, riesgo medio f_1^0, riesgo bajo f_1^-
Tipo sanguíneo (F2): mayor inmunidad f_2^+, menor inmunidad f_1^-
Gota gruesa (E1): positivo e_1^+ , negativo e_1^-
Fiebre (E2): presente e_2^+ , ausente e_2^-

- Grafo dirigido acíclico



  - El conocimiento que se necesita para las Redes Bayesianas:
    - Para cada uno de los factores, la probabilidad de que se dé.
    - La probabilidad condicionada de la causa con respecto a los factores, en sus combinaciones.
    - La probabilidad de cada uno de los efectos dada la causa y no dada la causa.

Distribución de probabilidad de los factores

$P(f_1+) = 0,1$
$P(f_10) = 0,1$
$P(f_1-) = 0,8$

$P(f_2+)=0,6$
$P(f_2-)=0,4$

Distribución de probabilidad condicionada de la variable con respecto a los factores

$P(+x/f_1,f_2)$	f_1+	f_10	f_1-
f_2+	0,015	0,003	0,0003
f_2-	0,022	0,012	0,0008

Distribución de probabilidad condicionada de los efectos con respecto a la variable

$P(+e_1\|+x) = 0,992$
$P(+e_2\|+x) = 0,98$

$P(+e_1\|-x) = 0,006$
$P(+e_2\|-x) = 0,017$

  - No se necesita tener $P(-e_i| \pm x)$ porque se puede obtener haciendo $1-\left(e_i \middle| \pm x\right)$, igual con $P\left(-x \middle| f_1, f_2\right)$.

Inferencia

- La inferencia se realiza con la fórmula de Redes Bayesianas dependiendo de los valores para la probabilidad que se desea obtener.

Probabilidad de que una persona tenga Covid si vive en una zona de alta incidencia, no está vacunada, y tiene fiebre alta pero no tiene tos

Esta probabilidad se representa como...

$$P(+x|f_1^+, f_2^-, e_1^+, e_2^-) = \frac{P(+x, f_1^+, f_2^-, e_1^+, e_2^-)}{P(f_1^+, f_2^-, e_1^+, e_2^-)}$$

Para el numerador...

$$\begin{aligned} P(+x, f_1^+, f_2^-, e_1^+, e_2^-) &= P(+x|f_1^+, f_2^-) \times P(f_1^+) \times P(f_2^-) \times P(e_1^+| + x) \times P(e_2^-| + x) \\ &= 0,045 \times 0,7 \times 0,8 \times 0,004 \times (1 - 0,09) \\ &= 0,000091728 \end{aligned}$$

Y para denominador...

$$\begin{aligned} P(-x, f_1^+, f_2^-, e_1^+, e_2^-) &= P(-x|f_1^+, f_2^-) \times P(f_1^+) \times P(f_2^-) \times P(e_1^+| - x) \times P(e_2^-| - x) \\ &= (1 - 0,045) \times 0,7 \times 0,8 \times 0,0025 \times (1 - 0,3) \\ &= 0,37286 \\ P(f_1^+, f_2^-, e_1^+, e_2^-) &= P(+x, f_1^+, f_2^-, e_1^+, e_2^-) + P(-x, f_1^+, f_2^-, e_1^+, e_2^-) \\ &= 0,000091728 + 0,37286 = 0,372951728 \end{aligned}$$

Finalmente...

$$\begin{aligned} P(+x|f_1^+, f_2^-, e_1^+, e_2^-) &= \frac{0,000091728}{0,372951728} = 0,000245951 \\ &\approx 0,0246\,\% \text{ de que padezca de Covid.} \end{aligned}$$

Deducciones

- Dados los datos de entrada se pueden deducir más cosas a partir de ellos y de operaciones aritméticas

$P(+x, f_1, f_2)$ surge multiplicando la distribución de probabilidad con los factores.

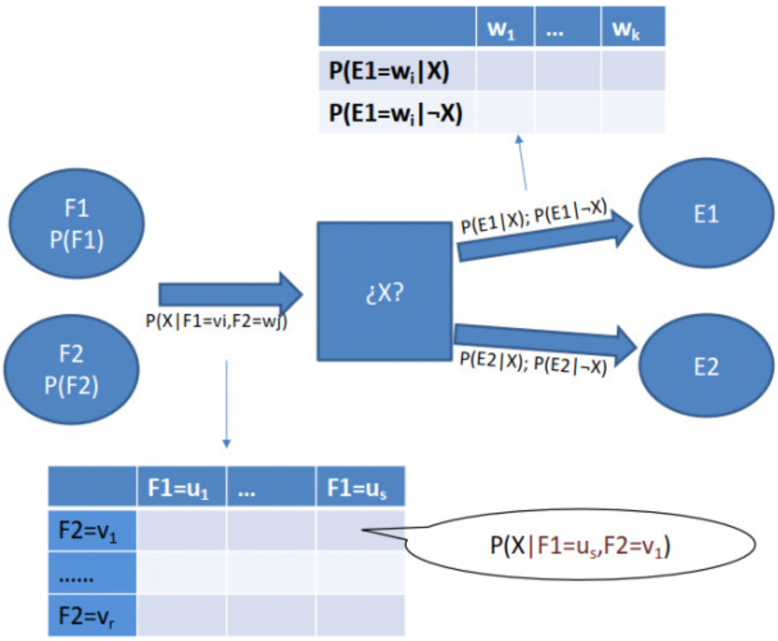
Probabilidad de $+x$ y f_2^+ es la suma marginal de f_2^+

$P(+x,f_1,f_2)$	f_1^+	f_1^o	f_1^-				
f_2^+	0,0009	0,00018	0,000144	0,001224	$P(+x, f_2^+)$	0,00204	$P(+x\|f_2^+)$
f_2-	0,00088	0,00048	0,000256	0,001616	$P(+x, f_2^-)$	0,00404	$P(+x\|f_2^-)$
	0,00178	0,00066	0,0004	0,00284	$P(+x)$		
	$P(+x, f_1^+)$	$P(+x, f_1^o)$	$P(+x, f_1^-)$				
	0,0178	0,0066	0,0005				
	$P(+x\|f_1^+)$	$P(+x\|f_1^o)$	$P(+x\|f_1^-)$				

La probabilidad condicional es $\frac{P(+x,f_2^+)}{P(f_2^+)}$

- **Probabilidad a priori:** $P(+x) = 0,00284$
- **Probabilidad tras conocer los factores:** $P(+x|f_1^o,f_2^-) = 0,012$
  $P(+x|f_1^o) = 0,0066$ si desconozco el valor de F2
  $P(+x|f_2^-) = 0,0016$ si desconozco el valor de F1
- **Probabilidad tras añadir conocimiento sobre los efectos:** $P(+x|f_1^o,f_2^-,\neg e_1,e_2) = 0,0056$
  Si desconozco el valor de F1 y de E2, ¿$P(+x|f_2^-,\neg e_1)$?

- Esto permite también inferir con información desconocida.
- La representación en un Sistema Basado en el Conocimiento sería muy similar al grafo...



- Probabilidad a priori

  P(X a_priori) $= P(X) = \sum P(X, F1 = u_i, F2 = v_j)$

  $P(X,F1=u_i,F2=v_j) = P(F1=u_i)*P(F2=v_j)*P(X|F1=u_i,F2=v_j)$

- Probabilidad a posteriori tras factores *F1=u,F2=v*

  P(X a_posteriori_factores) = **P(X|F1=u,F2=v)**

- Probabilidad a posteriori tras factores y efectos E1=w,E2=z

  P(X a_posteriori_factores_efectos) = P(X|F1=u,F2=v,E1=w,E2=z)

Cx= P(F1=u)*P(F2=v)*P(X|F1=u,F2=v)*P(E1=w|X)*P(E2=z|X) = P(X,F1=u,F2=v,E1=w,E2=z)
C¬x=P(F1=u)*P(F2=v)*P(¬X|F1=u,F2=v)*P(E1=w|¬X)*P(E2=z|¬X)=P(¬X,F1=u,F2=v,E1=w,E2=z)

C= Cx + C¬x = P(F1=u,F2=v,E1=w,E2=z)

P(X a_posteriori_factores_efectos) = Cx /C

Resumen de Representaciones Numéricas:

- **Grados de Incertidumbre de Mycin:**
  - Asigna un número entre 1 y -1 a cada regla.
  - Mide la incertidumbre asociada a cada regla.
  - Se aplica en Sistemas Expertos.
  - **Ventajas**
    - Número de parámetros es bajo.
    - Funciona bien y rápido en sistemas sencillos.
  - **Desventajas**
    - Débil representación de la independencia ya que no tiene justificación teórica clara.
    - Pueden surgir incoherencias si se permiten ciclos.

- **Lógica Difus**a
  - Asigna un número entre 0 y 1 a cada proposición.
  - Mide la verdad asociada a cada proposición.
  - Se aplica en Sistemas Expertos y Sistemas de Control.

    - Ventajas
        - Proporciona una manera de razonar con la vaguedad asociada al lenguaje natural.
    - Desventajas
        - Tiene muchas elecciones arbitrarias (Combinación de grados de creencia, como se calcula la agregación, la defuzzificación, inferencia, etc.)

- **Probabilidad**
    - Asigna un número entre 0 y 1 a cada proposición.
    - Mide la incertidumbre asociada a cada proposición.
    - Se aplica en Sistemas Expertos y en Sistemas de Clasificación.
    - Ventajas
        - Sistema formalmente probado y robusto.
    - Desventajas
        - Necesita de mucha información.

Imprecisión en las afirmaciones vs Imprecisión entre la veracidad del conocimiento.

Conocimiento verdad / Afirmaciones	PRECISO	IMPRECISO: RETRACTABLE	IMPRECISO: BASADA EN ESTADISTICA	IMPRECISO: BASADO EN CREENCIAS
PRECISAS	LÓGICA CLASICA	- LÓGICA POR DEFECTO - HIPOTESIS DEL MUNDO CERRADO	PROBABILIDAD	FACTORES DE CERTEZA
IMPRECISAS	LÓGICA DIFUSA	LÓGICA DIFUSA POR DEFECTO	LÓGICA DIFUSA PROBABILISTICA	LÓGICA DIFUSA CON FACTORES DE CERTEZA

- *"En Sierra Nevada está lloviendo"*: Conocimiento preciso, afirmación Precisa. Lógica clásica.
- *"En Sierra Nevada está lloviendo mucho"*: Conocimiento preciso, afirmación imprecisa. Lógica difusa.