



Razonamiento, Actuación y Aprendizaje bajo Incertidumbre

Inteligencia Artificial e Ingeniería del Conocimiento.

**Escuela Técnica Superior de Ingeniería Informática
Universidad de La Laguna**

Información Básica

- **Profesor: Fernando Pérez Nava**

- Teléfono: 922845048
- e-mail: fdoperez@ull.es
- Despacho: Edificio de la ETSII. Segunda Planta
- Tutorías: Lunes 9.30-13.30

- **Algunas Referencias Bibliográficas:**

- Español:

- **S. Russel y P. Norvig**, *Inteligencia Artificial. Un enfoque moderno*, 2004, Prentice-Hall. Cap 14-17
- **N. J. Nilsson**, *Inteligencia Artificial. Una nueva síntesis*, 2000, McGraw Hill. Cap 19-20
- **F.J. Díez**, *Introducción al Razonamiento Aproximado*. Dpto. Inteligencia Artificial, UNED, 2001.
<http://ia-serv.dia.uned.es/~fjdiez/libros/razaprox.zip>
- **E. Castillo, J.M. Gutiérrez y A.S. Hadi**, *Sistemas Expertos y Modelos de Redes Probabilísticos*, Monografías de la Academia Española de Ingeniería, Madrid, 1998.
<http://personales.unican.es/gutierjm/papers/BookCGH.pdf>

- Inglés:

- **F.V. Jensen**, *Bayesian Networks and Influence Diagrams*, Aalborg University, 2001
<http://www.cs.auc.dk/~fvj/BSS99/book99.ps>
- **H. Bengtsson**, *Bayesian Networks*, Lund Institute of Technology, 1999.
<http://www.maths.lth.se/matstat/staff/hb/hbbn99.pdf>

- Cursos:

- K.B. Laskey Computational Models for Probabilistic Inference (George Mason Univ.)
<http://ite.gmu.edu/~klaskey/CompProb/>
- N. Friedman Probabilistic Methods in AI (Hebrew University)
<http://www.cs.huji.ac.il/~pmai/index.html>



Contenidos

- **Tema 4**
 - La Incertidumbre y su Representación.
- **Tema 5**
 - Redes Bayesianas
- **Tema 6**
 - Decisiones y Aprendizaje en Redes Bayesianas

4

Incertidumbre

- En los temas anteriores se han descrito técnicas de representación del conocimiento y razonamiento para un **modelo del mundo**:
 - Completo
 - Consistente
 - Inalterable
- Sin embargo, en muchos dominios de interés no es posible crear tales modelos debido a la presencia de **incertidumbre**:
 - “Falta de conocimiento seguro y claro de algo”. (Diccionario RAE)
- En los próximos temas se revisarán las principales aproximaciones para tratar con la incertidumbre haciendo incapié en las Redes Bayesianas

Fuentes de Incertidumbre: Hechos

- **Con respecto a los hechos:**

- **Ignorancia**

- Puede que en un determinado campo el **conocimiento sea incompleto**.
 - Por ejemplo en el campo de las Ciencias Médicas.
 - Aunque se pudiera completar el conocimiento, puede ser necesario **tomar decisiones con información incompleta**.
 - Un paciente llega con gravedad a urgencias y es necesario proponer un tratamiento sin que sea posible realizar todos los tests necesarios para saber con total exactitud su enfermedad.
 - En otros campos la **ignorancia es irreducible**
 - Presente en modelos físicos
 - » ¿Cuál será el resultado del lanzamiento de una moneda?
 - Presente en la vida real
 - » ¿Es la otra persona sincera?

- **Vaguedad e Imprecisión**

- Algunos conceptos son vagos o imprecisos.
 - Las personas altas, guapas, felices etc.

Fuentes de Incertidumbre: Reglas

- **Con respecto a las reglas:**
 - Las reglas son generalmente heurísticas que utilizan los expertos en determinadas situaciones.
 - En el mundo real utilizamos habitualmente reglas que son :
 - **Inexactas o incompletas**
 - “Si es un ave entonces vuela”, ¿y los pingüinos?
 - “Si te duele la cabeza tienes gripe” ¿y si te diste un golpe?
 - **Imprecisas**
 - “Si el agua está caliente añade un poco de sal”
 - **Inconsistentes**
 - “Al que madruga Dios le ayuda”
 - “No por mucho madrugar amanece más temprano”

Razonamiento con Incertidumbre

- **Objetivo:**

- Ser capaz de **razonar sin tener todo el conocimiento relevante** en un campo determinado utilizando lo mejor posible el conocimiento que se tiene.

- **Implementación**

- Es difícil cumplir estos requerimientos utilizando la lógica de primer orden
- Deben de introducirse modelos para manejar información vaga, incierta, incompleta y contradictoria.
- **Crucial para un sistema funcione en el “mundo real”**

4

Actuar con Incertidumbre

- El propósito último de un sistema inteligente es **actuar de forma óptima** utilizando el conocimiento del sistema y un conjunto de percepciones.
- Para actuar se necesita **decidir que hacer**.
- ¿Cuál es la forma correcta de decidir?
 - La **decisión racional**:
 - Cuando se tienen distintas opciones un sistema debe decidirse por aquella acción que le proporcione el mejor resultado.
 - Cuando hay incertidumbre para poder decidir racionalmente se requiere:
 - La importancia de los distintos resultados de una acción
 - La certidumbre de alcanzar esos resultados cuando se realiza la acción.

Cuestiones a Resolver por las Aproximaciones a la Incertidumbre

- **¿Cómo evaluar la aplicabilidad de las condiciones de las reglas?**
 - Si X es mayor necesita gafas.
 - ¿Se puede aplicar la regla si X tiene 40 años?
- **¿Cómo combinar hechos imprecisos ?**
 - X es alto, X es rubio
 - ¿Con que certidumbre puedo afirmar que X es alto y rubio?
- **Dada un hecho imprecisa y una regla imprecisa: ¿qué confianza se puede tener en las conclusiones?**
 - X estudia mucho
 - Si X estudia mucho aprobará
 - ¿Con que certidumbre puedo afirmar que X aprobará?

Cuestiones a Resolver por las Aproximaciones a la Incertidumbre

- Dada la **misma conclusión incierta de varias reglas**: ¿qué confianza se puede tener en la conclusión?
 - Juan llega tarde, Luis llega tarde
 - Si Juan llega tarde la carretera está cortada
 - Si Luis llega tarde la carretera está cortada
 - ¿Cuál es la certidumbre de que la carretera esté cortada?

4

Algo de Historia

- **Inicialmente la mayoría de los investigadores en IA enfatizaban la importancia del razonamiento simbólico y evitaban la utilización de números.**
 - Los sistemas expertos no deben usar números puesto que los expertos humanos no lo hacen.
 - Los expertos no pueden suministrar los números requeridos.
- **Sin embargo los ingenieros que desarrollaban las aplicaciones se dieron cuenta pronto de la necesidad de representar la incertidumbre**
 - El sistema experto MYCIN (años 70) para el tratamiento de infecciones bacterianas fue el primer éxito en este campo.
- **Los métodos numéricos (especialmente los basados en probabilidad) son actualmente una herramienta aceptada en IA**
 - Debido a los éxitos prácticos
 - A la complejidad de las teorías alternativas

Principales Modelos de Representación de la Incertidumbre

- **La lógica de primer orden (LPO) no es adecuada para modelar la incertidumbre por lo que son necesarios nuevos modelos, entre ellos destacan:**
 - **Modelos Simbólicos**
 - Lógicas por Defecto
 - Lógicas basadas en Modelos Mínimos
 - La asunción del mundo cerrado
 - Terminación de predicados
 - **Modelos Numéricos**
 - Probabilidad
 - Teoría de Dempster-Shaffer
 - Lógica difusa

Representación Simbólica de la Incertidumbre: LPO

- **La LPO asume que el conocimiento:**
 - Es exacto.
 - Los hechos son ciertos o falsos
 - Es completo.
 - Se conoce todo acerca del campo de trabajo.
 - Es consistente.
 - No tiene contradicciones.
- **Por tanto, con la LPO :**
 - No se puede expresar incertidumbre.
 - No puede hacer deducciones lógicamente incorrectas pero probables
 - No se puede trabajar con información contradictoria

El Razonamiento no Monótono


- Como la LPO asume que el conocimiento es **completo y consistente**, **una vez que un hecho se asume/es cierto permanece así (Monotonía)**
 - Si de una Base de Conocimiento (BC) se deduce una expresión s , y se tiene otra Base de conocimiento BC' de forma que $BC \subset BC'$, entonces de BC' también se deduce s .
 - Por tanto **el añadir nuevo conocimiento siempre incrementa el tamaño de la Base de Conocimiento**.
- La presencia de **conocimiento incompleto** nos lleva a modelos **no monótonos**:
 - El conocimiento de nuevos hechos puede hacer que nos **retractemos de algunas de nuestras creencias**.
 - Ejemplo:
 - Generalmente voy a la Universidad en guagua
 - Veo en el periódico que hay huelga
 - Tengo que retractar que voy a la Universidad en guagua

Representación Simbólica de la Incertidumbre

- **Lógica por defecto**
 - Propuesta por Reiter para **solucionar el problema del conocimiento incompleto** (1980).
 - Para ello se introducen una serie de reglas por defecto.
 - Intuitivamente:
 - “Las reglas por defecto expresan características comunes a un conjunto de elementos que se asumen ciertas salvo que se indique lo contrario”.
- **Asunción del mundo cerrado**
 - Sirve para manejar **conocimiento incompleto**.
 - Intuitivamente:
 - “Lo que no se puede probar a partir de mi Base de Conocimiento es falso”
 - Utilizado en las B.D. y Prolog.
- **Inconvenientes**
 - Teorías complejas y a veces inconsistentes.

Representación Numérica de la Incertidumbre: Factores de Certeza

- **Los Factores de certeza aparecieron en el sistema experto MYCIN:**
 - desarrollado en la Universidad de Stanford (década de los 70) para el diagnóstico y consulta de enfermedades infecciosas.
- **Factores de certeza**
 - La Base de Conocimiento de MYCIN consistía en reglas de la forma:
Evidencia \rightarrow Hipótesis $FC(H|E)$
 - El **factor de certeza FC** representa la certidumbre en la Hipótesis cuando se observa la Evidencia.
 - Los FC varían entre -1 (creencia nula) y 1 (creencia total)



Representación Numérica de la Incertidumbre: Factores de Certeza

- **Grados de creencia**

- Los FC se calculan a partir de los **grados de creencia GC** y no creencia en la hipótesis
- Los GC varían entre 0 (creencia nula) y 1 (creencia total)
- La relación entre FC y GC es: $FC(H|E) = GC(H|E) - GC(\neg H|E)$

- **Propiedades**

- A diferencia de los grados de creencia probabilísticos
 $GC(H|E) + GC(\neg H|E) \neq 1$

4

Las Reglas en Mycin

- **Ejemplo:**

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

- **Lo que significa:**

SI el organismo es gram-negativo

Y tiene forma de bastón

Y es anaerobio

ENTONCES el organismo es bacteriodes (con certeza 0.6)

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

$$f_{\text{comb}}(x,y) = \begin{cases} x+y-xy & x,y > 0 \\ (x+y)/(1-\min(|x|,|y|)) & xy \leq 0 \\ x+y+xy & x,y < 0 \end{cases}$$

- Encadenado de Reglas**

Si A entonces B con $FC(B|A)$

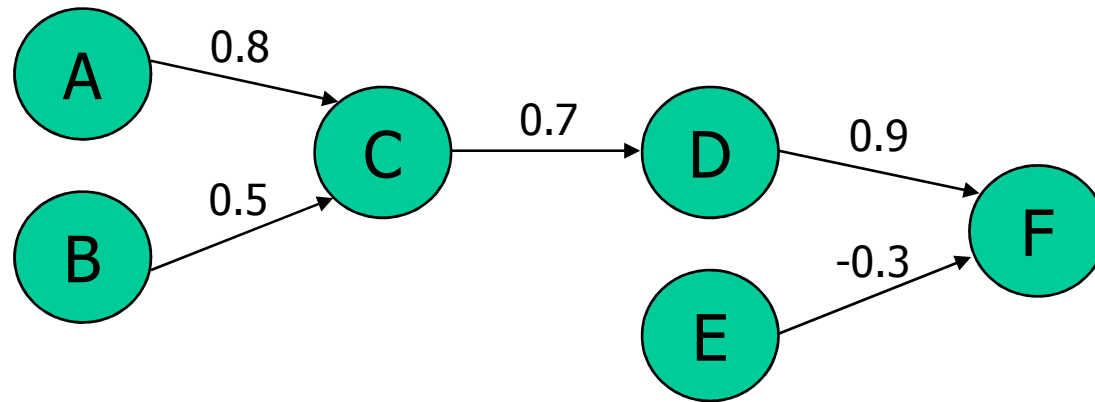
Si B entonces C con $FC(C|B)$

Si A entonces C con $FC(C|A)$

$$\text{Con: } FC(C|A) = \begin{cases} FC(C|B) FC(B|A) & FC(B|A) \geq 0 \\ 0 & FC(B|A) < 0 \end{cases}$$

Ejemplo de Combinación de Factores de Certeza

- Dadas siguientes reglas calcular el factor de certeza de la proposición $A \vee B \vee E \rightarrow F$



- $A \rightarrow C, B \rightarrow C$ $FC(C|A,B) = 0.8 + 0.5(1-0.8) = 0.9$ (Convergencia)
- $A \vee B \rightarrow C, C \rightarrow D$ $FC(D|A,B) = 0.9 \times 0.7 = 0.63$ (Encadenado)
- $A \vee B \rightarrow D, D \rightarrow F$ $FC(F|A,B) = 0.63 \times 0.9 = 0.567$ (Encadenado)
- $A \vee B \rightarrow F, E \rightarrow F$ $FC(F|A,B,E) =$
 $(0.567 - 0.3) / (1 - \min(0.567, 0.3))$
 $= 0.38$ (Convergencia)

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

¿Cómo era el rendimiento de Mycin?

- **El sistema experto Mycin proporcionaba diagnósticos y recomendaciones terapéuticas que eran al menos tan buenas como los mejores expertos en la especialidad**
- **Sin embargo los factores de certeza tienen graves incoherencias, por ejemplo:**
 - De Sarampión→Ronchas (0.8) y Ronchas→Alergia (0.5) obtenemos (encadenado) Sarampión→Alergia (0.4)!
- **Los factores de certeza no se utilizan ya en aplicaciones reales.**

Representación Numérica de la Incertidumbre: Lógica difusa

- **Desarrollada a partir de los trabajos de Zadeh (mediados de los 60)**
- **Asigna a cada proposición A un “grado de verdad” V entre 0 y 1 donde:**
 - $V(A)=0$ indica que la proposición es completamente falsa
 - $V(A)=1$ indica que la proposición es totalmente verdadera
 - Valores intermedios de $V(A)$ reflejan diferentes grados de verdad de la proposición
- **Es una generalización de la lógica clásica (que aparece tomando los valores de verdad de todas las proposiciones como 0 ó 1)**
- **Relacionada con la descripción de la **vaguedad** en vez de la incertidumbre**

4

Ejemplo

- Dada la proposición “La temperatura del enfermo es alta”

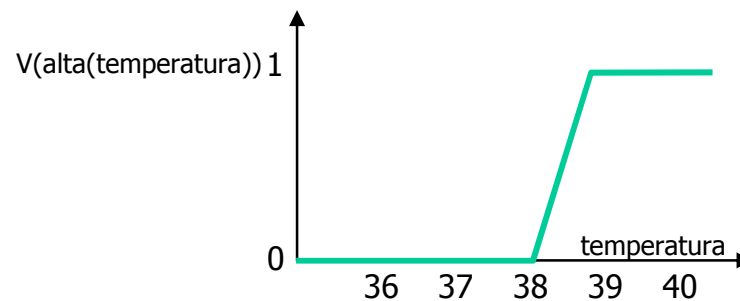
- En la lógica clásica

- Diríamos por ejemplo $\text{alta}(\text{temperatura}) \Leftrightarrow \text{temperatura} > 38$
- Por tanto si la temperatura es 37.99 diríamos que la temperatura no es alta. Nosotros no solemos razonar así

- En la lógica difusa

- Se da un grado de verdad V a alta en función de temperatura
- Por ejemplo:

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



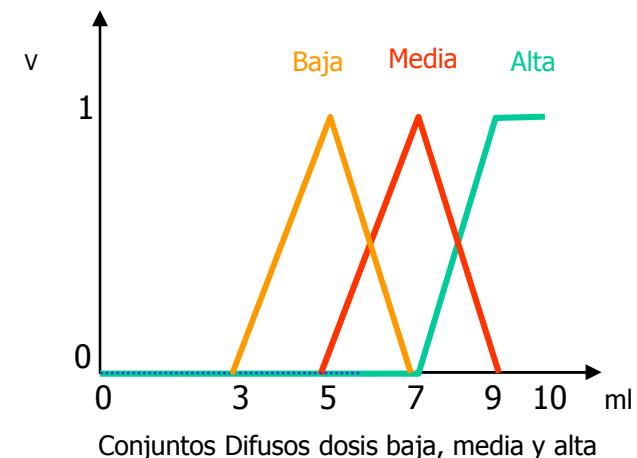
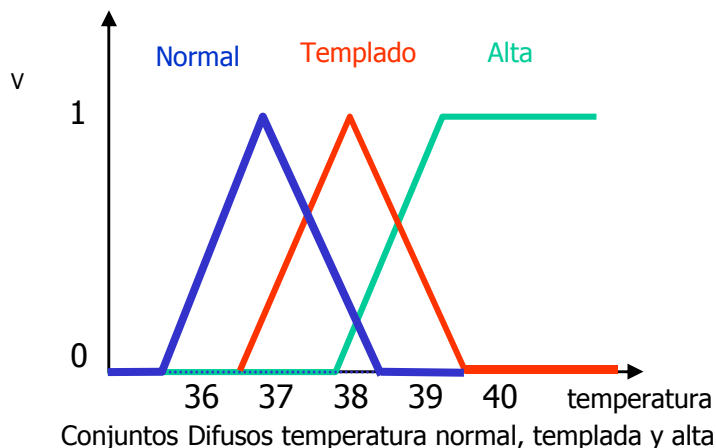
Grados de verdad del predicado $\text{alta}(\text{temperatura})$

Proposiciones Compuestas

- En la lógica difusa al igual que en la lógica clásica **el valor de verdad de una proposición compuesta se calcula a través del valor de verdad de las proposiciones individuales**
- **Existen varias formas de calcular estos valores de verdad. Los más usuales son:**
 - $V(A \wedge B) = \min(V(A), V(B))$
 - $V(A \vee B) = \max(V(A), V(B))$
 - $V(\neg A) = 1 - V(A)$
 - $V(A \rightarrow B) = \max(1 - V(A), V(B))$
- **Nótese que la lógica difusa no cumple:**
 - No contradicción: $V(A \wedge \neg A) \neq 0$ en general
 - Tercio excluso: $V(A \vee \neg A) \neq 1$ en general

Razonamiento Difuso basado en Reglas

- **Lo mostraremos con un ejemplo:**
 - Se le toma la temperatura a un paciente y se quiere saber la dosis apropiada de un medicamento.
 - Hechos:
 - temperatura=38
 - Reglas:
 - normal(temperatura)→baja(dosis)
 - templada(temperatura) →media(dosis)
 - alta(temperatura) →alta(dosis)



Razonamiento Difuso basado en Reglas

- **Generalmente el proceso de razonamiento difuso consta de 4 pasos**

- **Difusión: Obtener los grados de verdad de los antecedentes.**

- Se obtienen los grados de verdad de los antecedentes utilizando los hechos observados.

- En el ejemplo:

- Hechos:

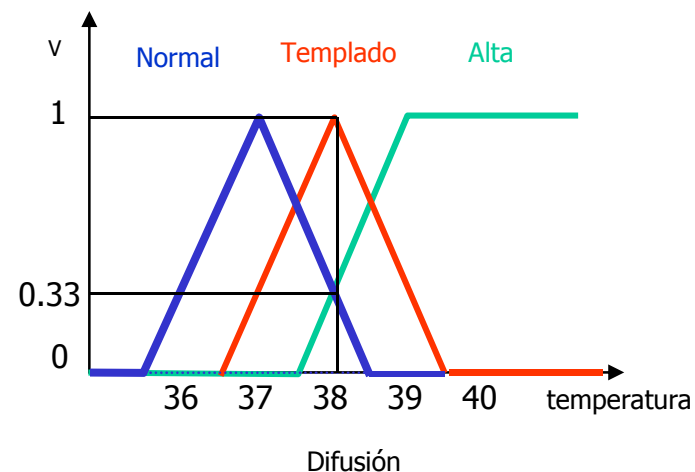
- temperatura=38

- Grados de verdad:

- » normal(temperatura): 0.33

- » templado(temperatura): 1

- » alta(temperatura): 0.33



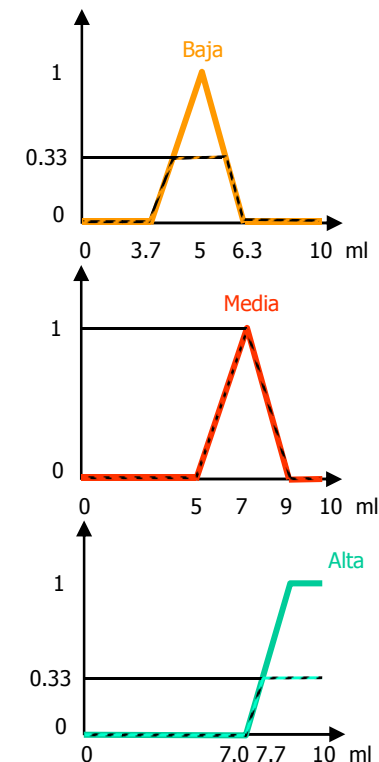
Razonamiento Difuso: Inferencia

– Inferencia: Obtener los grados de verdad de los consecuentes

- Una vez calculados los grados de verdad de la premisa de cada regla se recalculan los grados de verdad de los consecuentes mediante:
 - **Min**: los grados de verdad del consecuente se cortan a la altura del grado de verdad de la premisa, o
 - **Producto**: se multiplican los grados de verdad de consecuente y premisa

– Ejemplo (continuación)

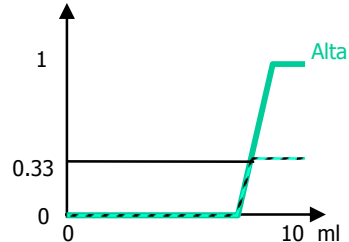
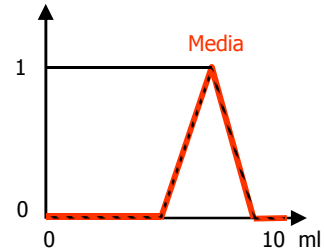
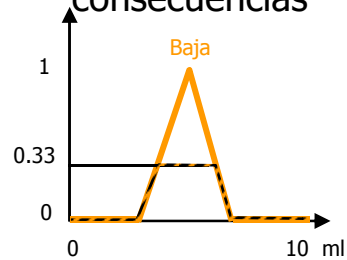
- Reglas:
 - $\text{normal(temperatura)} \rightarrow \text{baja(dosis)} \quad v=0.33$
 - $\text{templada(temperatura)} \rightarrow \text{media(dosis)} \quad v=1$
 - $\text{alta(temperatura)} \rightarrow \text{alta(dosis)} \quad v=0.33$



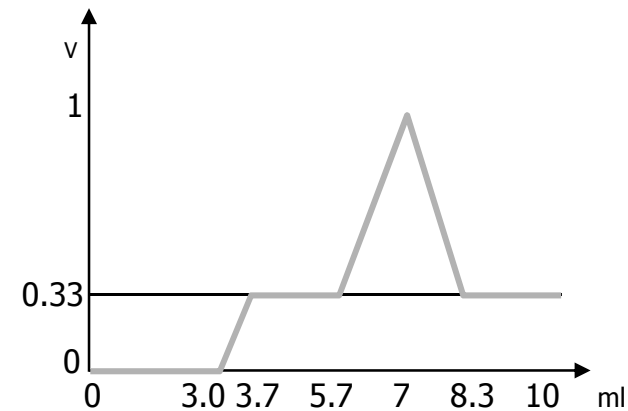
Razonamiento Difuso: Composición

– Composición de consecuentes

- Todos los grados de verdad difusos correspondientes a reglas con el mismo consecuente se combinan para dar lugar a los grados de verdad de la conclusión de las reglas mediante:
 - **Max:** Se toma el máximo de los grados de verdad correspondientes a las distintas consecuencias, o
 - **Sum:** Se toma la suma de los grados de verdad correspondientes a las distintas consecuencias



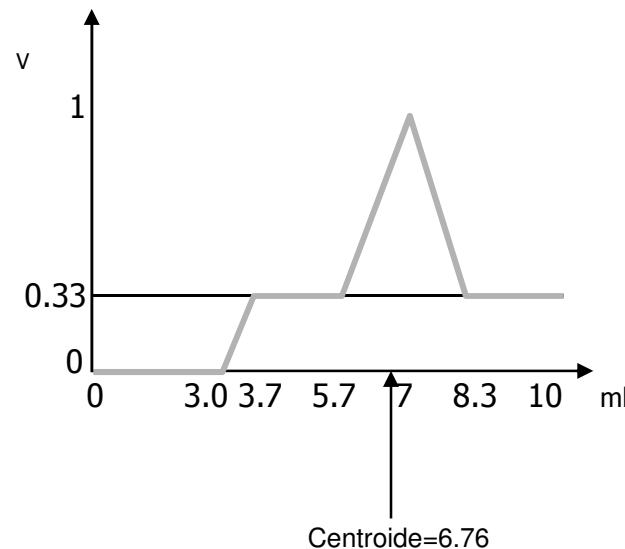
Composición (max) →



Razonamiento Difuso: Concisión

– Concisión (Opcional)

- Se utiliza cuando se necesita convertir una conclusión difusa en concreta.
- Generalmente se utilizan los métodos:
 - **Centroide**: Se calcula el centro de gravedad de los grados de verdad de la conclusión difusa, o
 - **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.

Representación Numérica de la Incertidumbre: Probabilidad

- **La Teoría de la Probabilidad (TProb)**
 - Es un área de las Matemáticas que ha sido aplicada a problemas de razonamiento con incertidumbre
 - Es una **teoría elegante, bien entendida y con mucha historia** (formalizaciones a partir de mediados del siglo XVII)
 - Asigna **valores numéricos** (llamados probabilidades) a las proposiciones.
 - Nos dice, dadas las probabilidades de ciertas proposiciones, y algunas relaciones entre ellas como asignar probabilidades a las proposiciones relacionadas
 - Relación con la LPO:
 - En la LPO las proposiciones son ciertas o falsas.
 - Con la Tprob las proposiciones son también ciertas o falsas pero se tiene un grado de creencia en la certeza o falsedad.

¿Qué son las Probabilidades?

- **A pesar de su larga historia los valores numéricos que representan las probabilidad no tiene una interpretación única.**
- **Algunas Interpretaciones:**
 - **Frecuentista:** Es el valor, cuando el número de pruebas tiende a infinito, de la frecuencia de que ocurra algún evento
 - **Subjetiva:** Es un grado de creencia acerca de un evento incierto
- **Aún así:**
 - Existe un consenso sobre el modelo matemático que soporta la Teoría

Los Valores Numéricos de la Probabilidad

- **Denotaremos por $P(A)$ a la probabilidad de la proposición A**
 - A ="El paciente tiene sarampión"
 - A ="Mañana saldrá el sol" ...
- **Los valores de la Probabilidad satisfacen un conjunto de **axiomas**:**
 - $0 \leq P(A) \leq 1$
 - $P(\text{Proposición Verdadera})=1$
 - $P(A \vee B)=P(A)+P(B)$
 - Siempre que A y B sean disjuntos, es decir $\neg(A \wedge B)$
- **A partir de ellos se puede demostrar por ejemplo:**
 - $P(\neg A)=1-P(A)$
 - $P(\text{Proposición Falsa})=0$
 - $P(A \vee B)=P(A)+P(B)-P(A \wedge B)$

4

Variables Aleatorias

- **Muchas veces tenemos un evento con un conjunto de resultados:**

- **Completo**

Se conocen **todos** los posibles resultados

- **Mutuamente excluyente**

No se pueden dar dos resultados distintos **simultáneamente**.

Ejemplos

- Si tiramos una moneda, el resultado es cara o cruz
 - Si tiramos un dado, se producen seis resultados distintos
 - La temperatura de un paciente puede estar en un conjunto de intervalos: <36.5 , $36.5-37.4$, $37.5-38.4$, $38.5-39.4$, >39.4
- **En lugar de tener una proposición para cada resultado se introduce el concepto de **Variable aleatoria****
 - **Se permiten proposiciones de la forma **Variable = resultado****
 - Por ejemplo, si M ="Resultado de tirar una moneda con valores posibles cara y cruz" se permiten las proposiciones:
 - $M=\text{cara}$ y $M=\text{Cruz}$ y podemos hablar de
 - $P(M=\text{cara})$ y $P(M=\text{cruz})$ que representan la probabilidad de obtener una cara y una cruz respectivamente

Variables Aleatorias cont.

- **Por consistencia, se puede considerar que **todas las proposiciones son variables aleatorias** que toman dos valores verdadero o falso**
 - Por ejemplo, dada la proposición “Tiene Sarampión”
 - Construimos la variable aleatoria Sarampión que toma los valores verdadero y falso
 - Y representamos la probabilidad de que un paciente tenga Sarampión $P(\text{“Tiene Sarampión”})$ como $P(\text{Sarampión} = \text{verdadero})$
- **Abreviaturas**
 - Se suele escribir $P(M=\text{cara})$ como $P(\text{cara})$, cuando está claro que nos referimos a la variable aleatoria M .
 - Si una variable aleatoria como Sarampión toma únicamente los valores verdadero o falso se suele escribir $P(\text{Sarampión} = \text{verdadero})$ como $P(\text{sarampión})$ y $P(\text{Sarampión} = \text{falso})$ como $P(\neg \text{sarampión})$

4

Distribuciones de Probabilidad

- Dada una Variable Aleatoria nos gustaría conocer la probabilidad para **cada** valor que pueda 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	→	Llueve	P(Llueve)	
		Verdadero	0.1	
Valores	↙ ↘	Falso	0.9	↔ Probabilidades

4

Proposiciones más Complejas

- **Podemos estar interesados en estudiar varias variables en conjunto.**
 - Por ejemplo
 - $P(\text{Sarampión}=\text{verdadero} \wedge \text{Fiebre}=\text{verdadero})$ que es la probabilidad de que el paciente tenga sarampión y fiebre
 - Generalmente lo escribiremos como:
 - $P(\text{sarampión} \wedge \text{fiebre})$ o $P(\text{sarampión}, \text{fiebre})$
- **Para ello se necesita asignar probabilidades a cada posible combinación de los valores de las variables.**
- **El listado de todos esos valores se llama la distribución conjunta del conjunto de variables**

4

Ejemplo de distribución conjunta

- **Distribución conjunta** de las variables Llueve y EnCalle $P(\text{Llueve}, \text{EnCalle})$:

Llueve	EnCalle	$P(\text{Llueve}, \text{EnCalle})$
Verdadero	Verdadero	0.01
Verdadero	Falso	0.09
Falso	Verdadero	0.2
Falso	Falso	0.7

- **También se puede escribir como:**

Llueve	EnCalle	$P(\text{Llueve}, \text{EnCalle})$
llueve	encalle	0.01
llueve	¬encalle	0.09
¬ llueve	encalle	0.2
¬ llueve	¬encalle	0.7

- **Recuerda a la tabla de la verdad lógica excepto que:**
 - Describe las probabilidad para cada combinación de valores de las variables
 - Generalmente dichos valores no se pueden calcular a partir de sus componentes

La Importancia de la Distribución Conjunta

- La distribución conjunta contiene **todo lo que se necesita saber acerca** de un conjunto de variables aleatorias.
- En particular, la distribución de cada variable individual se puede calcular a partir de la distribución conjunta (y se llama **distribución marginal**)
 - Ejemplo: Supongamos las variables aleatorias: Llueve y EnCalle con distribución conjunta $P(\text{Llueve}, \text{EnCalle})$

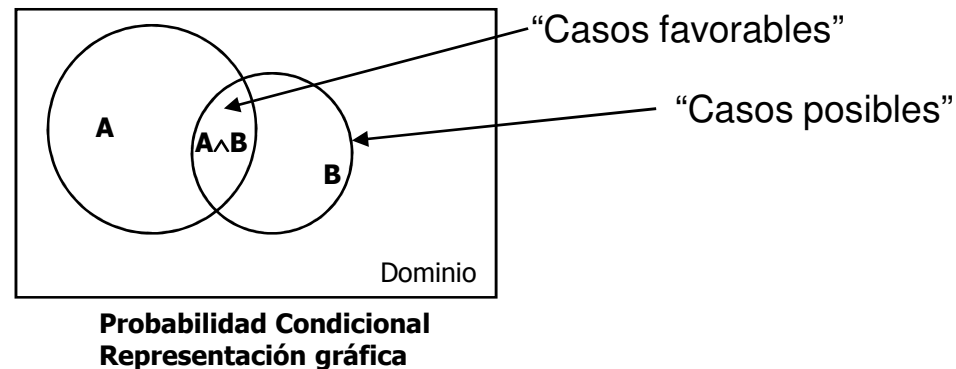
llueve	encalle	0.01
llueve	¬encalle	0.09
¬ llueve	encalle	0.2
¬ llueve	¬encalle	0.7

- Entonces $P(\text{llueve}) = P(\text{llueve} \wedge \text{enCalle}) + P(\text{llueve} \wedge \neg \text{enCalle}) = 0.01 + 0.09 = 0.1$.
- De forma similar $P(\neg \text{llueve}) = 0.9$
- También podemos calcular la probabilidad de disyunciones:
 $P(\text{llueve} \vee \text{enCalle}) = 0.01 + 0.09 + 0.2 = 0.3$

4

Probabilidad Condicional

- Escribiremos $P(A|B)$ para representar la probabilidad de A dado B. Esta probabilidad se llama **probabilidad condicional**.
- Lo podemos interpretar como **mi grado de creencia en A cuando todo lo que sé es B**.
 - O de forma alternativa, de los casos en los que se da B, ¿en que proporción se da A?



- Se define como:
 - $P(A|B) = P(A \cap B) / P(B)$ (Asumiendo $P(B) \neq 0$) o equivalentemente
 - $P(A \cap B) = P(A|B)P(B)$ (Regla del Producto)

4

Distribución Condicional

- **Nos 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: $P(\text{Llueve}|\text{enCalle})$

Llueve	$P(\text{Llueve} \text{enCalle})$
llueve	0.05
\neg llueve	0.95

- Ejemplo: $P(\text{Llueve}|\neg \text{enCalle})$

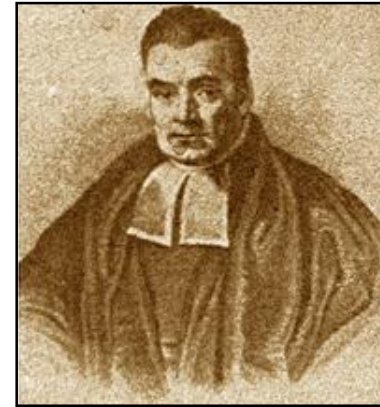
Llueve	$P(\text{Llueve} \neg \text{enCalle})$
llueve	0.11
\neg llueve	0.89

- Nótese que $\text{Llueve}|\text{enCalle}$ y $\text{Llueve}|\neg \text{enCalle}$ son variables aleatorias

Razonamiento con Probabilidades: La Regla de Bayes

- **Propuesta en 1763 por el Reverendo T. Bayes**

- $P(A|B) = P(B|A) P(A) / P(B)$
- Es una consecuencia de la regla del producto:
 - $P(A|B)P(B) = P(A,B) = P(B|A)P(A)$



Thomas Bayes

- **De forma intuitiva:**

- La probabilidad de una hipótesis A dada una evidencia B: $P(A|B)$ es proporcional a probabilidad de la hipótesis $P(A)$ multiplicada por el grado en que la hipótesis predice los datos $P(B|A)$

- **Aplicabilidad**

- En muchos problemas dado un conjunto de datos (evidencia) B tenemos que seleccionar la hipótesis A más probable mediante $P(A|B)$

4

Regla de Bayes: Forma General

- **Forma general de la Regla de Bayes**

- Si se tiene un conjunto de proposiciones $\{A_1, A_2, \dots, A_m\}$ completas y mutuamente excluyente se tiene:

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{P(B|A_1) P(A_1) + \dots + P(B|A_n) P(A_m)}$$

O lo que es lo mismo, si tiene una variable aleatoria A con valores a_1, a_2, \dots, a_m

$$P(a_i|B) = \frac{P(B|a_i) P(a_i)}{P(B|a_1) P(a_1) + \dots + P(B|a_n) P(a_m)}$$

4

La Regla de Bayes: Ejemplo

- **Intentemos resolver un caso real con probabilidades:**
 - Se pretende determinar la presencia o no de una enfermedad con un test.
 - En este caso:
 - Hipótesis (A): Enfermedad (variable aleatoria con dos valores verdadero y falso)
 - Evidencia (B): Test (variable aleatoria con dos valores positivo y negativo)
 - Se tiene:
 - $P(a)=1/10000$ (Prevalencia)
 - $P(b|a)=0.99$ (Sensibilidad)
 - $P(\neg b | \neg a)=0.99$ (Especificidad)
 - Aplicando la Regla de Bayes:

$$P(a|b) = \frac{P(b|a) P(a)}{P(b|a) P(a) + P(b | \neg a) P(\neg a)} =$$

$$\frac{0.99 \times 0.0001}{0.99 \times 0.0001 + (1-0.99) \times (1-0.0001)} =$$

$$(1/102) = 0.0098$$

$$P(\neg a | b) = 1 - 0.0098 = 0.9902$$
 - Al elegir la hipótesis más probable debemos concluir que con este test si el resultado es positivo lo más probable es que el paciente no esté enfermo!

4

La Regla de Bayes: Ejemplo

- **Continuamos con el ejemplo:**

- ¿Y si hay varios tests B_1, B_2, \dots, B_m ?
 - Supondremos que cada test B_1, B_2, \dots, B_m es una variable aleatoria con dos resultados: positivo y negativo.
- Entonces si queremos calcular la probabilidad de que el paciente esté enfermo necesitamos calcular:

$$P(A | B_1, B_2, \dots, B_m) = P(B_1, B_2, \dots, B_m | A) P(A) / P(B_1, B_2, \dots, B_m)$$

- Si al paciente se le hace un conjunto de 30 pruebas y por simplificar se supone que cada una da como resultado sí o no.
 - Entonces para almacenar la tabla de probabilidad conjunta $P(B_1, B_2, \dots, B_m | A)$ se necesitan guardar unos 2^{30} números reales (unos 8 DVD's por paciente).
 - ¿De donde sacamos los números ? ¿Cómo estimar los números a partir de casos (en la Tierra hay 2^{32} personas aproximadamente)?
 - ¿Cómo hacemos los cálculos computacionalmente eficientes?

Independencia: ¿Una Solución?

- **Independencia**

- Decimos que dos proposiciones A_1 y A_2 son **independientes** si el conocimiento de una no cambia la probabilidad de la otra
 - Por ejemplo si
 - A_1 ="Es rubio" , A_2 ="Tiene la piel clara" , A_3 ="Lloverá mañana"
 - A_1 y A_3 son independientes A_1 y A_2 no.
- 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 $P(A_1 \wedge A_2 \wedge \dots \wedge A_n) = P(A_1) P(A_2) \dots P(A_n)$
Para especificar la distribución conjunta de n variables se necesitan $O(n)$ números en lugar de $O(2^n)$
- Dos **variables aleatorias son independientes** si el conocimiento del valor que toma una no cambia la probabilidad de los valores de la otra: $P(A_1=c|A_2=d) = P(A_1=c)$

4

Independencia Condicional

- **Pero...**

- La condición de independencia es muy restrictiva.
- Por ejemplo, los resultados de los tests en medicina no suelen ser independientes.

- **Independencia condicional**

- Se dice que dos proposiciones A_1, A_2 son independientes dada una tercera B si cuando B está presente el conocimiento de una no influye en la probabilidad de la otra: $P(A_1|A_2, B) = P(A_1|B)$
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: $P(A_1 \wedge A_2 \wedge \dots \wedge A_n | B) = P(A_1|B) P(A_2|B) \dots P(A_n|B)$
 - Tenemos $O(n)$ números en lugar de $O(2^n)$

4

Independencia Condicional

- **Finalizamos el ejemplo:**

- ¿Y si hay varios tests B_1, B_2, \dots, B_m ?
- Como vimos, para calcular la probabilidad de que el paciente esté enfermo hay que calcular:

$$P(A | B_1, B_2, \dots, B_m) = P(B_1, B_2, \dots, B_m | A) P(A) / P(B_1, B_2, \dots, B_m)$$

- Si los tests B_1, B_2, \dots, B_m son independientes dada la enfermedad A (aproximación que suele dar buenos resultados):

$$P(B_1, B_2, \dots, B_m | A) = P(B_1 | A) P(B_2 | A) \dots P(B_m | A)$$

- El problema a resolver ya es **abordable**:

Basta calcular:

$$P(A | B_1, B_2, \dots, B_m) = P(B_1 | A) P(B_2 | A) \dots P(B_m | A) P(A) / P(B_1, B_2, \dots, B_m)$$

$$P(\neg A | B_1, B_2, \dots, B_m) = P(B_1 | \neg A) P(B_2 | \neg A) \dots P(B_m | \neg A) P(\neg A) / P(B_1, B_2, \dots, B_m)$$

$$\text{con } P(B_1, B_2, \dots, B_m) = P(B_1 | A) P(B_2 | A) \dots P(B_m | A) P(A) + P(B_1 | \neg A) P(B_2 | \neg A) \dots P(B_m | \neg A) P(\neg A)$$

Representación de la Independencia: Redes Bayesianas

- **La clave hacer factible la inferencia con probabilidades es la introducción explícita de la independencia entre variables**
- **El modelo más extendido de representación de independencias lo constituye 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**

Resumen de representaciones numéricas (1)

- **Grados de certidumbre en Mycin**
 - Asigna: Un número entre -1 y 1 a cada regla
 - Mide: La incertidumbre asociada a cada regla
 - Aplicaciones: Sistemas Expertos
 - Ventajas: El número de parámetros necesario es razonable
 - Inconvenientes: Débil representación de la independencia, Incoherencias
- **Lógica difusa**
 - Asigna: Un número entre 0 y 1 a cada proposición
 - Mide: La verdad asociada a cada proposición
 - Aplicaciones: Sistemas Expertos, Control
 - Ventajas: Proporciona una forma de razonar con la vaguedad asociadas al lenguaje natural
 - Inconvenientes: Tiene muchas elecciones arbitrarias (combinación de grados de creencia, inferencia, etc.)

Resumen de representaciones numéricas (2)

- **Probabilidad**

- Asigna: Un número entre 0 y 1 a cada proposición
- Mide: La incertidumbre asociada a dicha proposición
- Aplicaciones: Sistemas Expertos, Clasificación
- Ventajas: Sistema formalmente probado y robusto
- Inconvenientes: Se necesita mucha información