SCALAB

Universidad Carlos III de Madrid

# Incertidumbre

Confianza que tenemos de un suceso.

# En este tema

## Incertidumbre

# En este tema

## Incertidumbre

# Incertidumbre en IA

- En lógica clásica expresamos certeza absoluta: verdadero o falso
  - *Maria trabaja en Google* $\rightarrow$ *trabaja(Maria, Google)*
  - En los Sistemas de Producción los hechos en la BH son ciertos
- Incertidumbre
  - *tiempo de mañana, tráfico esta tarde, posición de un vehículo*
- Soluciones simples
  - Omitir la incertidumbre
  - Representar los diferentes resultados con una disyunción.
    - *buen tiempo OR mal tiempo*
    - Necesitaríamos nuevos algoritmos
    - *SI buen tiempo ENTONCES jugar al tenis* $\rightarrow$ ???
- Es ésto suficiente?

# Representación de la incertidumbre

- **Probabilidades**
  - Con una probabilidad del $90\%$ mañana lloverá
  - Razonamiento probabilístico, redes bayesianas
- **Conceptos vagos**
  - Juan es bajo, hace frío
  - Lógica borrosa

# Razonamiento con incertidumbre

## Historia

- Teorema de Bayes (1763)
- Lógica borrosa (Zadeh, 1965)
- Factores de certeza en sistemas expertos: MYCIN (1976), PROSPECTOR (1979)
- Teoría Dempster-Shafer (1976)
- Redes bayesianas (Pearl, 1986)
- Secuencias de decisiones: MDPs, HMMs, POMDPs

# En este tema

## Incertidumbre

# En este tema

## Incertidumbre

# Interpretación de las probabilidades

La probabilidad puede interpretarse como una medida de:

- **la proporción de veces en que algo es cierto.**
  - *De 22 estudiantes, 20 aprobaron el examen*
  - Suele referirse a propiedades físicas
  - que se pueden medir experimentalmente
- **el grado de creencia en que algo es cierto**
  - *Tengo una certeza del* $80\%$ *de que el Real Madrid ganará la liga*
  - Puede variar dependendiendo de la persona o sistema inteligente

# Interpretación de las probabilidades

La probabilidad puede interpretarse como una medida de:

- **la proporción de veces en que algo es cierto.**
  - *De 22 estudiantes, 20 aprobaron el examen*
  - Suele referirse a propiedades físicas
  - que se pueden medir experimentalmente
- **el grado de creencia en que algo es cierto**
  - *Tengo una certeza del* $80\%$ *de que el Real Madrid ganará la liga*
  - Puede variar dependendiendo de la persona o sistema inteligente

El cálculo de probabilidades no depende de la interpretación

- Una probabilidad toma valores en [0,1]
- Probabilidad = 0 $\rightarrow$ falso
- Probabilidad = 1 $\rightarrow$ verdadero

# Variables aleatorias

## Lógica proposicional

- Los estados se describen como conjuntos de variables booleanas: $p, q, r$ $\ldots$
- Una interpretación es una asignación de valores de verdad (verdadero, falso) a las variables $p = V, q = F, r = V$.

# Variables aleatorias

## Lógica proposicional

- Los estados se describen como conjuntos de variables booleanas: $p, q, r$ $\ldots$
- Una interpretación es una asignación de valores de verdad (verdadero, falso) a las variables $p = V, q = F, r = V$.

## Teoría de probabilidades

- Usamos un conjunto definido de variables aleatorias $X, Y \ldots$, que pueden tomar valores en un dominio:

$$X \in \{1, 2, 3\}, Y \in \{\text{azul}, \text{rojo}\}$$

- El valor asociado a una variable aleatoria es desconocido
- Podemos asignar una probabilidad a cada valor posible
  - $P(X = 1) = 0.2, P(X = 2) = 0.5, P(X = 3) = 0.3$
- Estas probabilidades definen una distribución de probabilidad

# Ejemplo

*Se tienen dos dados equilibrados. Indique las variables aleatorias, con sus dominios, y la distribución de probabilidad. ¿Cuál es la probabilidad de que la suma de ambos dados sea 3?*

# Ejemplo

*Se tienen dos dados equilibrados. Indique las variables aleatorias, con sus dominios, y la distribución de probabilidad. ¿Cuál es la probabilidad de que la suma de ambos dados sea 3?*

- Variables aleatorias
  - $D_1 \in \{1, 2, 3, 4, 5, 6\}$
  - $D_2 \in \{1, 2, 3, 4, 5, 6\}$

# Eventos

- El espacio muestral, $\Omega$, es el conjunto de todos los posibles resultados de un experimento aleatorio

$$\Omega = \{(1,1), (1,2), (1,3), \ldots, (2,1), (2,2), \ldots (6,6)\}$$

Dado 1 Dado 2

- Un evento (o suceso) es cualquier subconjunto del espacio muestral, $e \subseteq \Omega$:

*Los dos dados suman 3* $\rightarrow \{(1,2),(2,1)\}$

Como hay 2 no es atomico

- Un evento atómico $e \in \Omega$ es un evento de un único elemento y define completamente de un estado del mundo

*El primer dado tiene un 1 y el segundo dado tiene un 2* $\rightarrow (1,2)$

Como solo hay 1 atomico.

# Distribución de probabilidad

Distribución de probabilidad conjunta $\Longrightarrow$ $P(D_1, D_2)$ $D_1$ y $D_2$ toma un valor. Representa la prob. de que ocurra.

- Una distribución de probabilidad asigna a cada evento $e \in \Omega$, un valor que representa la probabilidad, $P(e)$, de que ese evento ocurra

Hay una probabilidad para cada evento del espacio muestral.
Para hallar $P(e)$ debemos tener datos de $e$.

- Propiedades
  - $0 \leq P(e) \leq 1$
  - $\sum_{e \in \Omega} P(e) = 1$

# Distribución de probabilidad

- Una distribución de probabilidad asigna a cada evento $e \in \Omega$, un valor que representa la probabilidad, $P(e)$, de que ese evento ocurra
- Propiedades
  - $0 \leq P(e) \leq 1$
  - $\sum_{e \in \Omega} P(e) = 1$
- Ejemplo (dados equilibrados)

$$P(D_1 = 1, D_2 = 1) = P(D_1 = 1, D_2 = 2) = \ldots = \\ P(D_1 = 1, D_2 = 6) = P(D_1 = 2, D_2 = 1) = \ldots = \\ P(D_1 = 6, D_2 = 6) = \frac{1}{36}.$$

# Distribución de probabilidad

- La probabilidad de un evento no atómico es la suma de las probabilidades de todos los eventos atómicos que lo componen

Se suman los posibles.

$$P(A) = \sum_{e \in A} P(e)$$

Teorema de la probabilidad total:

$$P(A) = \sum_{b_1 \dots b_n} P(A, B_1, \dots, B_n)$$

la suma de la prob que tienen a A.

Regla del producto:

$$P(A/B) = \frac{P(A, B)}{P(B)}$$

# Distribución de probabilidad

- La probabilidad de un evento no atómico es la suma de las probabilidades de todos los eventos atómicos que lo componen

$$P(A) = \sum_{e \in A} P(e)$$

- Probabilidad de que los dos dados sumen 3

$$P(D_1 + D_2 = 3) = P(D_1 = 1, D_2 = 2) + P(D_1 = 2, D_2 = 1) = \frac{2}{36} = \frac{1}{18}$$

Probabilidad a posteriori: Sobre una observación (evidencia)

P( lluvia 10 mar / nubes negras este 10 mar )

Tras ver nubes negras

# Probabilidad a priori

Probabilida que algo ocurra cuando no tenemos información.

- La probabilidad a priori representa el grado de creencia en algo en cuando no se cuenta con ninguna otra información

  P(lluvia 10 marzo) = 0'3

  - Ejemplo (dado)

$$P(D_1 = 1) = P(D_1 = 2) = \ldots = P(D_1 = 6) = \frac{1}{6}$$

- La distribución de probabilidad de una variable se suele representar con un vector

P(D, E, C) Dist. conjunta

P(si, no, si)

P(d, e, f) Valores concretos

Por definir

$$P(D_1) = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$$

# Distribución de probabilidad conjunta

- La distribución de probabilidad conjunta define la distribución de probabilidad de un conjunto de variables
- Ejemplo (2 dados)

$$P(D_1, D_2)$$

- Dada una distribución conjunta, la probabilidad de una variable se ~~obtiene marginando~~

$$P(D_1 = 1) = \sum_{\{e \text{ tal que } D_1=1\}} P(D_1, D_2)$$

$$P(D_1 = 1) = P(D_1 = 1, D_2 = 1) + \ldots + P(D_1 = 1, D_2 = 6)$$

$$P(D_1 = 1) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$$

# Ejemplo distribución conjunta

D: dolor de dientes si/no
E: engancha si/no $P(D, E, C)$
C: caries si/no

De los registros de pacientes de un odontólogo obtenemos la siguiente tabla, sobre un total de 1000 casos, que relaciona síntoma observado (si tiene o no dolor de dientes) y resultado de la exploración (la sonda dental se engancha o no), con el diagnóstico final del mismo (tiene o no caries).

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | 108 | 12 | 72 | 8 |
| ¬ caries | 16 | 64 | 144 | 576 |

- ¿Cuál es la distribución de probabilidad conjunta?
- ¿Cuál es la probabilidad de que un paciente tenga caries, $P(\text{caries})$?
- ¿Cuál es la probabilidad de que la sonda se enganche pero el paciente no tenga dolor de dientes $P(\neg\text{dolorDeDientes}, \text{engancha})$?

# Ejemplo distribución conjunta

De los registros de pacientes de un odontólogo obtenemos la siguiente tabla, sobre un total de 1000 casos, que relaciona síntoma observado (si tiene o no dolor de dientes) y resultado de la exploración (la sonda dental se engancha o no), con el diagnóstico final del mismo (tiene o no caries).

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- Cuál es la distribución de probabilidad conjunta? P( D, E, C)
  $P(\text{Caries}, \text{DolorDeDientes}, \text{Engancha})$
- ¿Cuál es la probabilidad de que un paciente tenga caries, $P(\text{caries})$? P( C = si)
- ¿Cuál es la probabilidad de que la sonda se enganche pero el paciente no tenga dolor de dientes $P(\neg\text{dolorDeDientes}, \text{engancha})$?

# Ejemplo distribución conjunta

De los registros de pacientes de un odontólogo obtenemos la siguiente tabla, sobre un total de 1000 casos, que relaciona síntoma observado (si tiene o no dolor de dientes) y resultado de la exploración (la sonda dental se engancha o no), con el diagnóstico final del mismo (tiene o no caries).

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- ¿Cuál es la distribución de probabilidad conjunta?
  $\mathbf{P}(\text{Caries}, \text{DolorDeDientes}, \text{Engancha})$

  Al no ser atómico

- ¿Cuál es la probabilidad de que un paciente tenga caries, $P(\text{caries})$?

  La suma de todas las que tienen caries = si.

  $$P(\text{caries}) = \sum_{caries} P(\text{Caries}, \text{DolorDeDientes}, \text{Engancha}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

  Teorema prob. total.

- ¿Cuál es la probabilidad de que la sonda se enganche pero el paciente no tenga dolor de dientes $P(\neg\text{dolorDeDientes}, \text{engancha})$?

# Ejemplo distribución conjunta

De los registros de pacientes de un odontólogo obtenemos la siguiente tabla, sobre un total de 1000 casos, que relaciona síntoma observado (si tiene o no dolor de dientes) y resultado de la exploración (la sonda dental se engancha o no), con el diagnóstico final del mismo (tiene o no caries).

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- ¿Cuál es la distribución de probabilidad conjunta?
  $P(\text{Caries}, \text{DolorDeDientes}, \text{Engancha})$
- ¿Cuál es la probabilidad de que un paciente tenga caries, $P(\text{caries})$?

$$P(\text{caries}) = \sum_{caries} P(\text{Caries}, \text{DolorDeDientes}, \text{Engancha}) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

- ¿Cuál es la probabilidad de que la sonda se enganche pero el paciente no tenga dolor de dientes $P(\neg\text{dolorDeDientes}, \text{engancha})$? Suma de las prob. que tienen en D = no y E = si.

$$P(\neg\text{dolorDeDientes}, \text{engancha}) = 0.072 + 0.144 = 0.216$$

# Probabilidad a posteriori (condicional)

- Se utiliza cuando se ha obtenido alguna evidencia que modifica el conocimiento sobre el dominio, de manera que la probabilidad a priori no es aplicable.
- Se calcula en términos de probabilidades condicionadas

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \text{ siempre que } P(B) > 0$$

Comparar probabilidades condicionadas.

A veces no es necesario saber el P(B), solo con el numerador nos vale, (si es el mismo denominador)

$$P(A=si/B=b_1) = \frac{P(A=si, B=b_1)}{P(B=b_1)} = \overset{\frac{1}{P(B=b_1)}}{\alpha} P(A=si, B=b_1)$$

Constante de normalización.

Cuál es mayor? El que es mayor $\alpha P(A,B)$

$$P(A=no/B=b_1) = \frac{P(A=no, B=b_1)}{P(B=b_1)} = \alpha P(A=no, B=b_1)$$

Constantes

$P(A=si,B) \propto 0'6$

$P(A=no,B) \propto 0'7$

$$\alpha 0'6 + \alpha 0'7 = 1 \; ; \; \alpha = \frac{1}{P(B=b_1)} = \frac{1}{0'6+0'7} \Rightarrow P(B=b_1) = 0'6+0'7$$

# Probabilidad a posteriori (condicional)

- Se utiliza cuando se ha obtenido alguna evidencia que modifica el conocimiento sobre el dominio, de manera que la probabilidad a priori no es aplicable.
- Se calcula en términos de probabilidades condicionadas

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} \text{ siempre que } P(B) > 0$$

- $P(A|B)$ puede interpretarse como la probabilidad actualizada de $A$, una vez que $B$ ha sido observado
- **Regla del producto**

$$P(A \wedge B) = P(A/B)P(B)$$

$$P(A \wedge B) = P(B/A)P(A)$$

# Probabilidad Condicional (a posteriori)

- Una distribución de probabilidad condicional se puede representar como una matriz o tabla n-dimensional (n es el número de variables)

$$P(X \mid Y) = \begin{pmatrix} P(X = x_1 \mid Y = y_1) & P(X = x_1 \mid Y = y_2) & \dots & P(X = x_1 \mid Y = y_n) \\ P(X = x_2 \mid Y = y_1) & P(X = x_2 \mid Y = y_2) & \dots & P(X = x_2 \mid Y = y_n) \\ \dots & & & \\ P(X = x_m \mid Y = y_1) & P(X = x_m \mid Y = y_2) & \dots & P(X = x_m \mid Y = y_n) \end{pmatrix}$$

- Las columnas de esta matriz suman uno (son distribuciones de probabilidad!)

# Ejemplo probabilidad condicional

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- Se observa que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

caries sabiendo dolor

$$\underbrace{P(C/D) = \frac{P(C,D)}{P(D)}}_{\text{Regla del producto}} = \frac{\text{caries y dolor}}{\text{dolor}} = \frac{0'108+0'012}{0'108+0'012+0'016+0'064} = 0'6$$

# Ejemplo probabilidad condicional

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- Se observa que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\text{caries}/\text{dolorDeDientes}) = \frac{P(\text{caries}, \text{dolorDeDientes})}{P(\text{dolorDeDientes})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

# Ejemplo probabilidad condicional

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- Se observa que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\text{caries}/\text{dolorDeDientes}) = \frac{P(\text{caries}, \text{dolorDeDientes})}{P(\text{dolorDeDientes})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalización: no es necesario conocer $P(\text{dolorDeDientes})$!!

# Ejemplo probabilidad condicional

| | dolorDeDientes | | ¬ dolorDeDientes | |
|---|---|---|---|---|
| | engancha | ¬ engancha | engancha | ¬ engancha |
| caries | **0.108** | **0.012** | **0.072** | **0.008** |
| ¬ caries | **0.016** | **0.064** | **0.144** | **0.576** |

- Se observa que un paciente tiene dolor ¿Cuál es la probabilidad de que tenga caries?

$$P(\text{caries}/\text{dolorDeDientes}) = \frac{P(\text{caries}, \text{dolorDeDientes})}{P(\text{dolorDeDientes})} = \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

- Normalización: no es necesario conocer $P(\text{dolorDeDientes})$!!

$$\begin{aligned} P(\text{Caries}/\text{dolorDeDientes}) &= \alpha P(\text{Caries}, \text{dolorDeDientes}) \\ &= \alpha[P(\text{Caries}, \text{dolorDeDientes}, \text{engancha}) + \\ &\quad P(\text{Caries}, \text{dolorDeDientes}, \neg\text{engancha})] \\ &= (0.6, 0.4) \end{aligned}$$

# Distribución conjunta: razonamiento

- ¡La distribución conjunta en forma tabular permite contestar en términos probabilísticos a cualquier pregunta sobre variables discretas!

# Distribución conjunta: razonamiento

- ¡La distribución conjunta en forma tabular permite contestar en términos probabilísticos a cualquier pregunta sobre variables discretas!

- **¡Pero... no escala!**
  - En problemas reales puede haber cientos o miles de variables
  - Es imposible definir todas las probabilidades y trabajar con ellas

Hay que tener en cuenta el coste y tiempo de respuestas

# Distribución conjunta: razonamiento

- ¡La distribución conjunta en forma tabular permite contestar en términos probabilísticos a cualquier pregunta sobre variables discretas!

- **¡Pero... no escala!**
  - En problemas reales puede haber cientos o miles de variables
  - Es imposible definir todas las probabilidades y trabajar con ellas
- **¡Pero es la base teórica para otras aproximaciones que escalan mejor!**

# Probabilidad Condicional (Regla de la cadena)

Mirar

- Regla del producto

$$P(A \wedge B) = P(A|B)P(B)$$

- **Regla de la cadena**: aplicación sucesiva de la regla del producto

$$P(X_1, \dots, X_n) = P(X_1, \dots, X_{n-1})P(X_n|X_1, \dots, X_{n-1})$$

$$= P(X_1, \dots, X_{n-2})P(X_{n-1}|X_1, \dots, X_{n-2})P(X_n|X_1, \dots, X_{n-1})$$

$$= \dots$$

$$P(X_1, \dots, X_n) = \prod_{i=1}^{n} P(X_i|X_1, \dots, X_{i-1})$$

# Probabilidad Condicional (Regla de la cadena)

- Regla del producto

$$P(A \wedge B) = P(A|B)P(B)$$

- **Regla de la cadena**: aplicación sucesiva de la regla del producto

$$\begin{aligned} P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})P(X_n|X_1, \ldots, X_{n-1}) \\ = \quad & P(X_1, \ldots, X_{n-2})P(X_{n-1}|X_1, \ldots, X_{n-2})P(X_n|X_1, \ldots, X_{n-1}) \\ = \quad & \ldots \end{aligned}$$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})$$

$P(\text{Caries}, \text{DolorDeDientes}, \text{Engancha}) =$
$P(\text{Caries}/\text{DolorDeDientes}, \text{Engancha})P(\text{DolorDeDientes}, \text{Engancha}) =$
$P(\text{Caries}/\text{DolorDeDientes}, \text{Engancha})P(\text{DolorDeDientes}/\text{Engancha})P(\text{Engancha})$

# Probabilidad Condicional (Regla de la cadena)

- Regla del producto

$$P(A \wedge B) = P(A|B)P(B)$$

- **Regla de la cadena**: aplicación sucesiva de la regla del producto

$$\begin{aligned} P(X_1, \ldots, X_n) = P(X_1, \ldots, X_{n-1})P(X_n|X_1, \ldots, X_{n-1}) \\ = \quad P(X_1, \ldots, X_{n-2})P(X_{n-1}|X_1, \ldots, X_{n-2})P(X_n|X_1, \ldots, X_{n-1}) \\ = \quad \ldots \end{aligned}$$

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1})$$

**¡Podemos definir la distribución conjunta mediante probabilidades condicionales!**

# Teorema de Bayes

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)} = \alpha P(B/A)P(A)$$

Thomas Bayes 1702 - 1761

# Teorema de Bayes

$$P(A/B) = \frac{\overbrace{P(B/A)P(A)}^{P(A\cap B)}}{P(B)} = \overbrace{\alpha}^{\frac{1}{P(B)}} P(B/A)P(A)$$

**¡Podemos calcular unas condicionales a partir de otras!**

Thomas Bayes 1702 - 1761

- A veces obtener $P(B/A)$ es más sencillo que obtener $P(A/B)$

$$P(Causa \mid Efecto) = \frac{P(Efecto \mid Causa)P(Causa)}{P(Efecto)}$$

# Ejemplo

- Probabilidades a priori
  - $P(\text{Enfermo} = \text{si}) = 0.008$
  - $P(\text{Enfermo} = \text{no}) = 0.992$
- Probabilidades condicionadas: $P(\text{Test\_laboratorio} \mid \text{Paciente\_enfermo})$

| | Enfermo = si | Enfermo = no |
|---|---|---|
| Test = $\oplus$ | 0.98 | 0.03 |
| Test = $\ominus$ | 0.02 | 0.97 |

- Observación: un nuevo paciente tiene un resultado positivo en el test de laboratorio
  - ¿$P(\text{Enfermo} = \text{si}/\text{Test} = \oplus)$? $= \dfrac{P(\oplus, si)P(si)}{P(\oplus)} = \alpha\, 0'98 \cdot 0'008$
  - ¿$P(\text{Enfermo} = \text{no}/\text{Test} = \oplus)$? $= \dfrac{P(\oplus, no)P(no)}{P(\oplus)} = \alpha\, 0'03 \cdot 0'992$

Desconocida

$P(\oplus) = \dfrac{1}{\alpha} = 0'03 \cdot 0'992 + 0'98 \cdot 0'008 =$

# Ejemplo

- Probabilidades a priori
  - $P(\text{Enfermo} = \text{si}) = 0.008$
  - $P(\text{Enfermo} = \text{no}) = 0.992$
- Probabilidades condicionadas: $P(\text{Test\_laboratorio} \mid \text{Paciente\_enfermo})$

| | Enfermo = si | Enfermo = no |
|---|---|---|
| Test = $\oplus$ | 0.98 | 0.03 |
| Test = $\ominus$ | 0.02 | 0.97 |

- Observación: un nuevo paciente tiene un resultado positivo en el test de laboratorio
  - ¿$P(\text{Enfermo} = \text{si}/\text{Test} = \oplus)$?
  - ¿$P(\text{Enfermo} = \text{no}/\text{Test} = \oplus)$?

$$P(\text{Enfermo} = \text{si}/\text{Test} = \oplus) = \alpha P(\text{Test} = \oplus/\text{Enfermo} = \text{si}) P(\text{Enfermo} = \text{si}) = \alpha\, 0.0078 = 0.2$$
$$P(\text{Enfermo} = \text{no}/\text{Test} = \oplus) = \alpha P(\text{Test} = \oplus/\text{Enfermo} = \text{no}) P(\text{Enfermo} = \text{no}) = \alpha\, 0.0298 = 0.8$$

# Independencia

- $A$ y $B$ son independientes si y sólo la ocurrencia de uno de ellos no afecta a la ocurrencia del otro
  - $P(A|B) = P(A)$, ó
  - $P(B|A) = P(B)$, ó
  - $P(A, B) = P(A)P(B)$
- Ejemplo

$$P(\mathrm{DolorDeDientes}, \mathrm{Caries}, \mathrm{Engancha}, D_1) =$$
$$P(\mathrm{DolorDeDientes}, \mathrm{Caries}, \mathrm{Engancha})P(D_1)$$

- **Reducción de tamaño de distribución**

de $2^3 \cdot 6 = 48$ probabilidades a $8 + 6 = 14$

# Independencia

- $A$ y $B$ son independientes si y sólo la ocurrencia de uno de ellos no afecta a la ocurrencia del otro
  - $P(A|B) = P(A)$, ó
  - $P(B|A) = P(B)$, ó
  - $P(A, B) = P(A)P(B)$
- Ejemplo

$$P(\text{DolorDeDientes}, \text{Caries}, \text{Engancha}, D_1) = P(\text{DolorDeDientes}, \text{Caries}, \text{Engancha})P(D_1)$$

- **Reducción de tamaño de distribución**

  de $2^3 \cdot 6 = 48$ probabilidades a $8 + 6 = 14$

  3 binarias   6 valores dados

  Mejor suma que producto.
- Menor tamaño implica
  - Algoritmos más eficientes
  - Menos datos (probabilidades) a especificar

# En este tema

## Incertidumbre

# Uso de probabilidades en Inteligencia Artificial

Tareas típicas: toma de decisiones, clasificación, predicción, ...

- ¿Que es cierto?
  - Uso de lógica clásica: satisfacción proposicional, verificación de circuitos, sistemas de producción, ...
- vs. ¿Qué es más probable?
  - Uso de probabilidades: redes bayesianas, predicción de secuencias (reconocimiento de voz), clasificación (de idioma), ...

# Uso de probabilidades en Inteligencia Artificial

Tareas típicas: toma de decisiones, clasificación, predicción, ...

- ¿Que es cierto?
  - Uso de lógica clásica: satisfacción proposicional, verificación de circuitos, sistemas de producción, ...
- vs. ¿Qué es más probable?
  - Uso de probabilidades: redes bayesianas, predicción de secuencias (reconocimiento de voz), clasificación (de idioma), ...
- ¿Qué pasa si el modelo elegido es incorrecto?
  - En lógica clásica
    - Modelo incompleto $\rightarrow$ ok
    - Modelo con errores $\rightarrow$ problemático
  - En probabilidades
    - Interesa más la relación entre probabilidades que los números exactos: ¿$P(e) > P(e')$?
    - Podría ser más robusto

# Inferencia

Principal tarea:

*Calcular la probabilidad de eventos dada cierta evidencia*

# Tareas de inferencia

- Calcular la distribución a posteriori dada cierta evidencia $P(X|e)$

# Tareas de inferencia

- Calcular la distribución a posteriori dada cierta evidencia $P(X|e)$
  - **Predicción**: ej. determinar si una zona se inundará según la situación meteorológica actual

$$P(\text{Inundacion}/\text{Meteorologia})$$

# Tareas de inferencia

- Calcular la distribución a posteriori dada cierta evidencia $P(X|e)$
  - **Predicción**: ej. determinar si una zona se inundará según la situación meteorológica actual

    $$P(\text{Inundacion}/\text{Meteorologia})$$

  - **Diagnosis**: ej. determinar si una persona tiene una enfermedad según unos resultados de pruebas

    $$P(\text{Enfermedad}/\text{Resultados\_pruebas})$$

# Tareas de inferencia

- Calcular la distribución a posteriori dada cierta evidencia $P(X|e)$
  - **Predicción**: ej. determinar si una zona se inundará según la situación meteorológica actual

$$P(\text{Inundacion}/\text{Meteorologia})$$

  - **Diagnosis**: ej. determinar si una persona tiene una enfermedad según unos resultados de pruebas

$$P(\text{Enfermedad}/\text{Resultados\_pruebas})$$

  - **Clasificación**: determinar a qué clase pertenecen unas observaciones. Ej. dada una imagen determinar si contiene un gato o un perro:

$$P(\text{clase} = \text{gato}/\text{caracteristicas\_imagen})$$

$$P(\text{clase} = \text{perro}/\text{caracteristicas\_imagen})$$

# Tareas de inferencia

- Calcular la distribución a posteriori dada cierta evidencia $P(X|e)$
  - **Predicción**: ej. determinar si una zona se inundará según la situación meteorológica actual
    $$P(\text{Inundacion}/\text{Meteorologia})$$
  - **Diagnosis**: ej. determinar si una persona tiene una enfermedad según unos resultados de pruebas
    $$P(\text{Enfermedad}/\text{Resultados\_pruebas})$$
  - **Clasificación:** determinar a qué clase pertenecen unas observaciones. Ej. dada una imagen determinar si contiene un gato o un perro:
    $$P(\text{clase} = \text{gato}/\text{caracteristicas\_imagen})$$
    $$P(\text{clase} = \text{perro}/\text{caracteristicas\_imagen})$$
- **Toma de decisiones**: elegir acciones más útiles (maximizan la utilidad esperada)

$$\text{UtilidadEsperada}(\text{Accion} = j)$$
$$= \sum P(\text{Salida} = i|\text{Accion} = j, \text{evidencia}) \times \text{Utilidad}(\text{Salida} = i)$$

# Resumen

- Las probabilidades son un formalismo adecuado para el manejo de incertidumbre
- La distribución de probabilidad conjunta es adecuada para tareas de razonamiento, pero no escala
- Se necesitan mecanismos que permitan un razonamiento eficiente
- Es este aspecto los siguientes conceptos son cruciales
  - Regla de la cadena
  - Teorema de Bayes
  - Concepto de independencia

# Créditos

Material basado en:

- Material de años anteriores en la UC3M.
- Libro y notas docentes de *Artificial Intelligence: A Modern Approach*. Por Russell&Novig. 2da edición.
- Material docente por Héctor Geffner.

# En este tema

## Incertidumbre

# En este tema

## Incertidumbre

# En este tema

## Incertidumbre