RAZONAMIENTO BAYESIANO

José Manuel Molina López
Grupo de Inteligencia Artificial Aplicada

PROBABILIDAD: INTRODUCCIÓN

- 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 (Lógica de Primer Orden):
  - En la LPO las proposiciones son ciertas o falsas (o desconocidas)
  - Con la Teoría de la Probabilidad las proposiciones son ciertas o falsas y se tiene un grado de creencia en la certeza o falsedad.

ENFOQUES DE LA PROBABILIDAD

- Frecuentista: las probabilidades provienen de experimentos y estadísticas
- Objetivista: la probabilidad como aspecto de la realidad que refleja la propensión de los objetos a comportarse de una determinada manera
- Subjetivista: la probabilidad como caracterización del grado de creencia de un agente, sin significación externa
- Teoría de juegos: Es una forma de asignar certidumbres de forma óptima a eventos en juegos en los que hay recompensas en función de predicciones y resultados

ENFOQUES DE LA PROBABILIDAD

- Probabilidad de que el sol exista mañana? (Inquiry, Hume)
  - Indefinida, no se han realizado experimentos
  - 1, en el pasado siempre se ha cumplido
  - 1-ε, donde ε es la proporción de estrellas en el universo que cambian a supernova cada día y explotan
  - Puede derivarse del tipo, edad, tamaño, y temperatura del sol, …

SUCESOS

<u>Suceso</u>: cualquier subconjunto del espacio muestral de un experimento aleatorio

<u>Sucesos elementales o atómicos</u>: conjunto de resultados posibles de un experimento que verifican:

- Siempre ocurre alguno de ellos (conjunto exhaustivo, la unión de todos ellos es igual al espacio muestral)
- Son mutuamente excluyentes (no hay ninguna intersección distinta de $\varnothing$)

<u>Suceso compuesto</u>: construido a partir de la unión de sucesos elementales

EJEMPLO DE SUCESOS

Lanzar un dado y observar el número resultante

- Suceso elemental: Sale un 2 {2}
- Suceso compuesto: Sale un número par {2,4,6}

Lanzar una moneda tres veces y observar el número total de caras

- Suceso elemental: el número total de caras es 3 {CCC}
- Suceso compuesto: el número total de caras es 2 {CC+, C+C, +CC}

Encender un foco eléctrico hasta que se funda. Medir el tiempo transcurrido

- Suceso: el foco dura mas de 100 horas $\{t \mid t > 100\}$

SUCESOS Y Tª DE CONJUNTOS

Suceso seguro: siempre se verifica, E

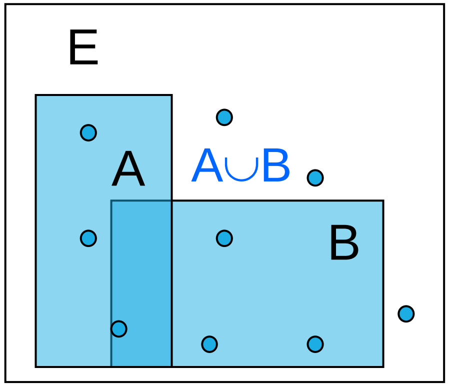
Suceso imposible: nunca se verifica, $\varnothing$

Se llama suceso complementario de un suceso A al suceso que se verifica si no se verifica A, y se representa $\neg A$ o A^c. Está formado por los elementos del espacio muestral (E) que no están en A

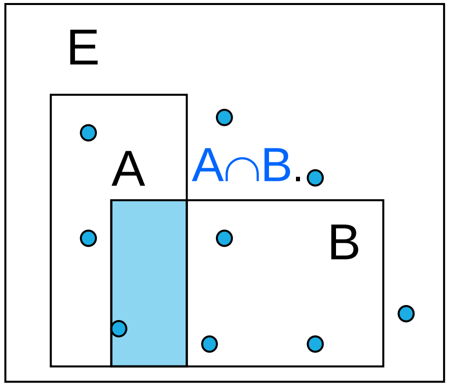
Un suceso A está contenido en otro suceso B, representado $A \subseteq B$, si al verificarse A necesariamente se verifica B.

SUCESOS Y Tª DE CONJUNTOS

Se llama <u>suceso unión</u> de A y B, $A \cup B$, al que verifica A o B. Está formado por los resultados experimentales que están en A o en B.



Se llama <u>suceso intersección</u> de A y B, $A \cap B$, al que verifica tanto A como B. Está formado por los resultados experimentales que están simultáneamente en A y B



PROBABILIDAD COMO FRECUENCIA

Sea *r* el numero de resultados obtenidos mediante un experimento y r_A el número de veces que el resultado fue el suceso A. La frecuencia relativa f_A de A se define como:

$$f_A = \frac{r_A}{r}$$

Y cumple

- $0 \leq f_A \leq 1$
- $f_A = 0 \Leftrightarrow r_A=0$; $f_A = 1 \Leftrightarrow r_A=0$
- Si A y B son eventos mutuamente excluyentes, entonces $f_{A \cup B} = f_A + f_B$
- Cuando $r \rightarrow \infty$ entonces f_A = Probabilidad(A)

REGLA DE LAPLACE

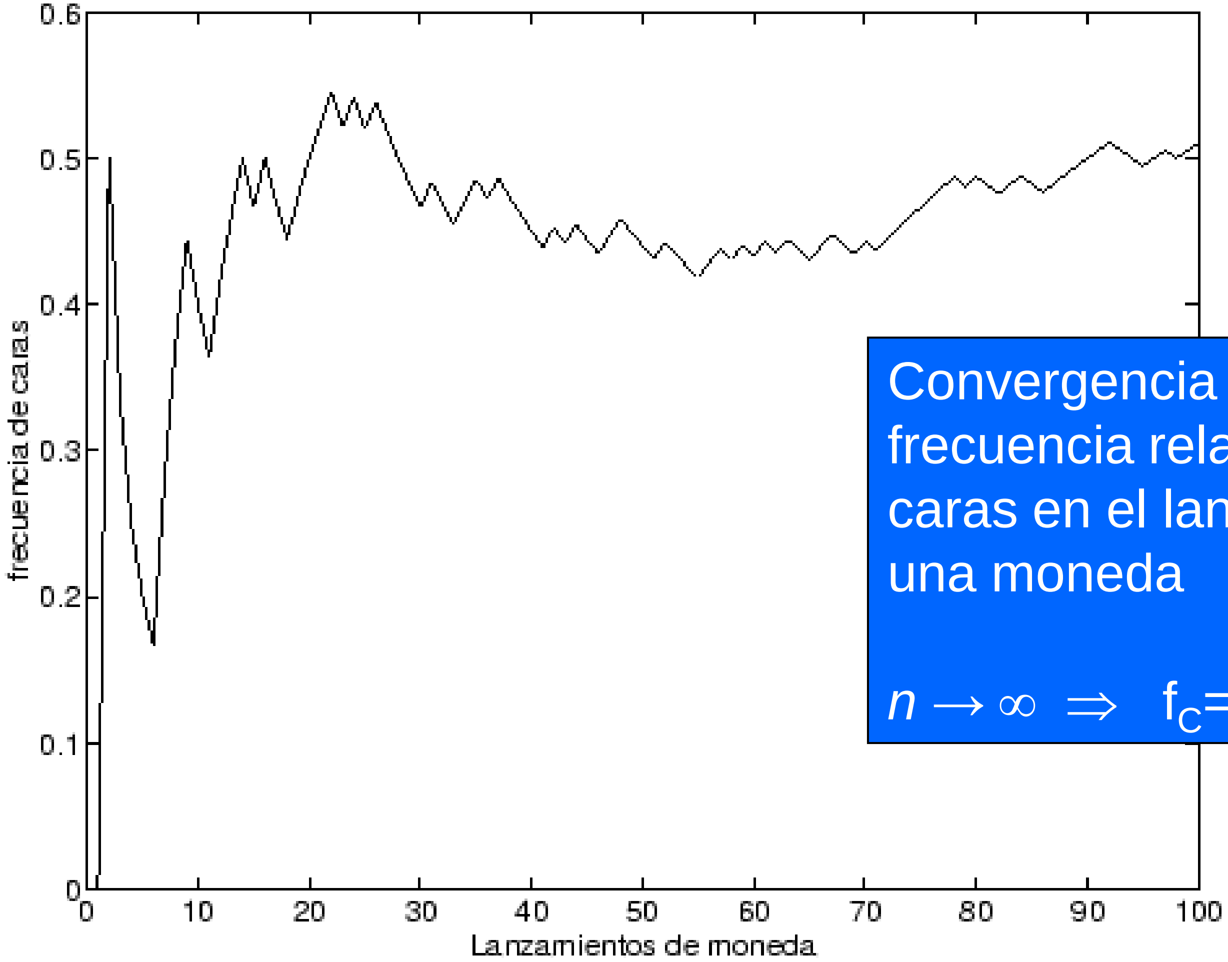
Dado un experimento aleatorio con un espacio muestral finito (de cardinalidad n) y sucesos equiprobables.

Si A es un suceso compuesto de n_A sucesos elementales, entonces la probabilidad de que se observe A al realizar un experimento es:

$$P(A) = \frac{\text{num. de casos favorables a A}}{\text{num. de casos posibles}} = \frac{n_A}{n}$$

$\frac{1}{n}$ Probabilidad de cada suceso elemental

PROBABILIDAD COMO FRECUENCIA



Convergencia de la frecuencia relativa de caras en el lanzamiento de una moneda

$n \longrightarrow \infty \;\Rightarrow\; f_C = 1/2 = P(C)$

DEFINICIÓN DE PROBABILIDAD (AXIOMÁTICA)

Se llama <u>probabilidad</u> a cualquier función P, que asigna a cada suceso A_i de un espacio muestral E un valor numérico $P(A_i)$, que verifica los axiomas siguientes:

- $0 \leq P(A_i) \leq 1$
- $P(E) = 1$
- Para cualquier número finito k de sucesos mutuamente excluyentes $A_1, A_2, \ldots, A_k$ se cumple:
  $P(\cup_{i=1..k} A_k) = \sum_{i=1..k} P(A_i)$

ALGUNAS PROPIEDADES ÚTILES

$P(\neg A) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

…

$P(A \cap B)$?

Lo veremos después

VARIABLES ALEATORIAS

- Muchas veces tenemos un suceso con un conjunto de resultados mutuamente excluyente
  - 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 caso 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=cara y M=Cruz y podemos hablar de
- P(M=cara) y P(M=cruz) que representan la probabilidad de obtener una cara y una cruz respectivamente

DISTRIBUCIÓN DE PROBABILIDAD

- $X = x$ denota que la variable aleatoria X toma el valor $x \in \Omega_X$.
- La función $P(X=x)$ es una <u>función de probabilidad</u> para X en Ω_X si se cumple:
  $0 \leq P(X=x) \leq 1, \quad \forall x \in \Omega_X$
  $1 = \sum_{x \in \Omega x} Pr(X=x)$
- La colección de pares $[(x, P(X=x)), \forall x \in \Omega_X]$ se llama <u>distribución de probabilidad</u> de X
  - Ejemplo: M VA representa lanzamiento de una moneda DP(M)=((0,1/2),(1,1/2))

NOTACIÓN

Variable aleatoria

- Numéricas: X, Y, Z
- Booleanas (Proposicional): $A, \neg A$

Probabilidad sobre variables aleatorias

- En general: $P(x_i) \equiv P(X=x_i)$
- Caso Booleano: $P(A) \equiv P(A = True)$, $P(\neg A) \equiv P(A = False)$
- $P(A \wedge B) \equiv P(A,B) \equiv P(A=True \cap B=True)$
- $P(A \vee B) \equiv P(A=True \cup B=True)$

PROPOSICIONES MÁS COMPLEJAS

Podemos estar interesados en estudiar varias variables en conjunto

- Por ejemplo
  - P(Sarampión=verdadero ^ Fiebre=verdadero) que es la probabilidad de que el paciente tenga sarampión y fiebre
- Generalmente lo escribiremos como:
  - P(Sarampión ^ Fiebre) o P(Sarampión, Fiebre)

Para ello se necesita la distribución conjunta del conjunto de variables

Recuerda a la tabla de la verdad lógica excepto que:

- Describe los valores de probabilidad para cada combinación de valores de las variables
- Generalmente dichos valores no se pueden calcular a partir de sus componentes

PROBABILIDAD CONJUNTA Y MARGINAL

- La distribución conjunta contiene todo lo que se necesita saber acerca de un conjunto de VA
- Dado el conjunto de VA {X,Y,Z}
- <u>Probabilidad conjunta</u>:

  $P(x_i, y_j, z_k) \equiv P(X=x_i \cap Y=y_j \cap Z=z_k)$

$$\sum_{\forall i,j,k} P(x_i, y_j, z_k) = 1$$

- <u>Probabilidad marginal</u>

$$P(x_i) = \sum_{\forall j,k} P(x_i, y_j, z_k)$$

$$1 = \sum_{\forall i} P(x_i) = \sum_{\forall j} P(y_j) = \sum_{\forall k} P(z_k)$$

EJEMPLO

	Dolor dental (D)	$\neg$ Dolor dental
Caries (C)	0.04	0.06
$\neg$Caries	0.01	0.89

Probabilidades conjuntas

$P(C) = P(C,D) + P(C,\neg D) = 0.06 + 0.04 = 0.1$

$P(D)= P(C,D) + P(\neg C,D) = 0.04+0.01 = 0.05$

Probabilidades marginales

Aplicando propiedades…

$P(\neg C) = 1 – P(C)$

$P(C \vee D) = P(C) + P(D) – P(C \wedge D) = 0.05 + 0.1 - 0.04 = 0.11$

PROBABILIDADES INCONDICIONALES

La interpretación de las probabilidades utilizadas hasta ahora nos dan el grado de creencia en una proposición ***asumiendo que no se sabe nada más***.

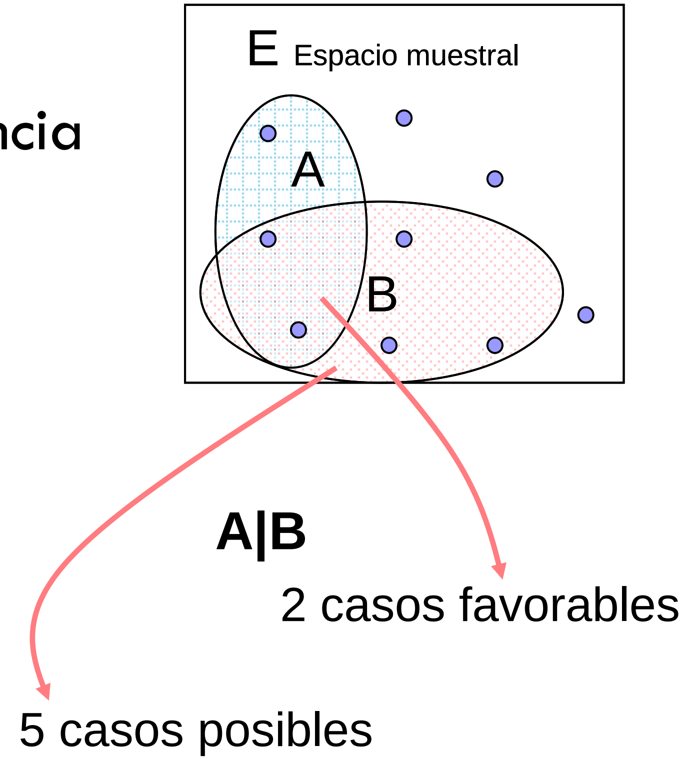
- Así P(Sarampión) significa la probabilidad de que el paciente tenga sarampión asumiendo que no se sabe nada más.
- A estas probabilidades se las llama probabilidades incondicionales o probabilidades a priori
- Cuando aprendemos algo nuevo (por ejemplo el paciente tiene manchas) deberíamos cambiar nuestro grado de creencia en que el paciente tenga sarampión
- Sin embargo, el valor de P(Sarampión) no se altera cuando se aprende algo nuevo puesto que expresa “asumiendo que no se sabe nada más”.
- En lugar de esto se representa el impacto de la nueva información mediante una proposición distinta

PROBABILIDAD CONDICIONADA

Con P(A | B) indicamos que el Espacio muestral de interés se ha "reducido" a aquellos resultados que definen la ocurrencia del suceso B

Entonces, P(A | B) "mide" la probabilidad relativa de A con respecto al espacio muestral reducido a B

$$\boxed{P(A|B)=\frac{2}{5}=\frac{2/9}{5/9}=\frac{P(A\cap B)}{P(B)}}$$

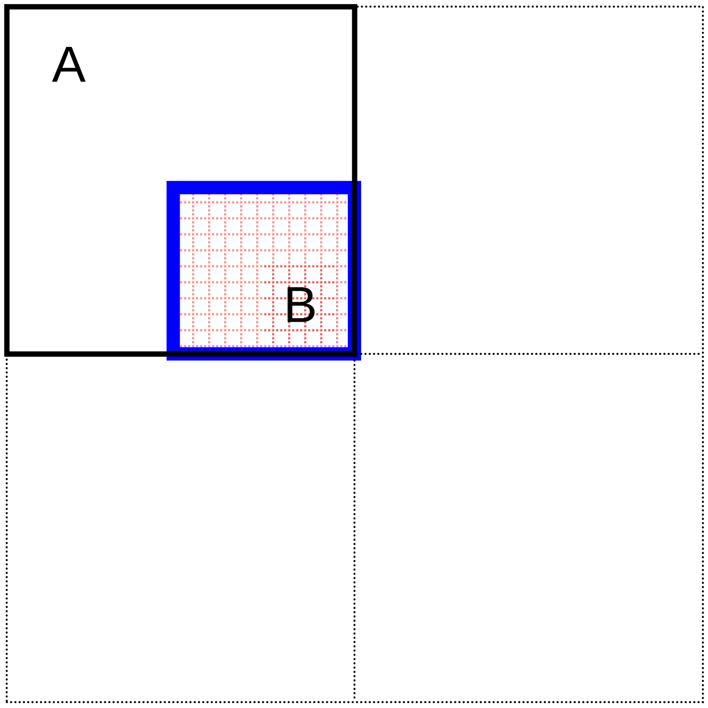


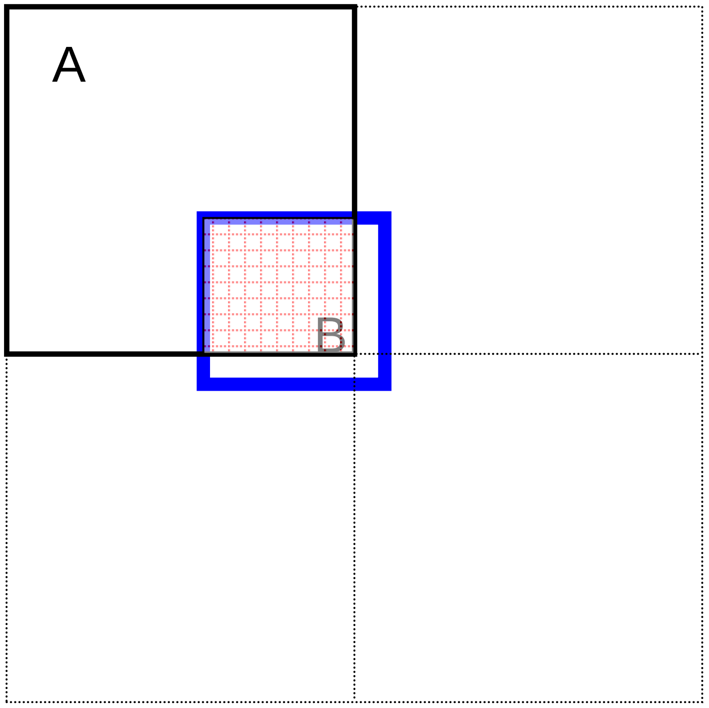
PROBABILIDAD CONDICIONADA

- Se llama <u>probabilidad</u> de A <u>condicionada</u> a B, representado P(A|B), a la probabilidad de que haya ocurrido A sabiendo que ha ocurrido B.

$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}}$$

Intuir la probabilidad condicionada





$P(A) = 0{,}25$
$P(B) = 0{,}10$
$P(A,B) = 0{,}10$

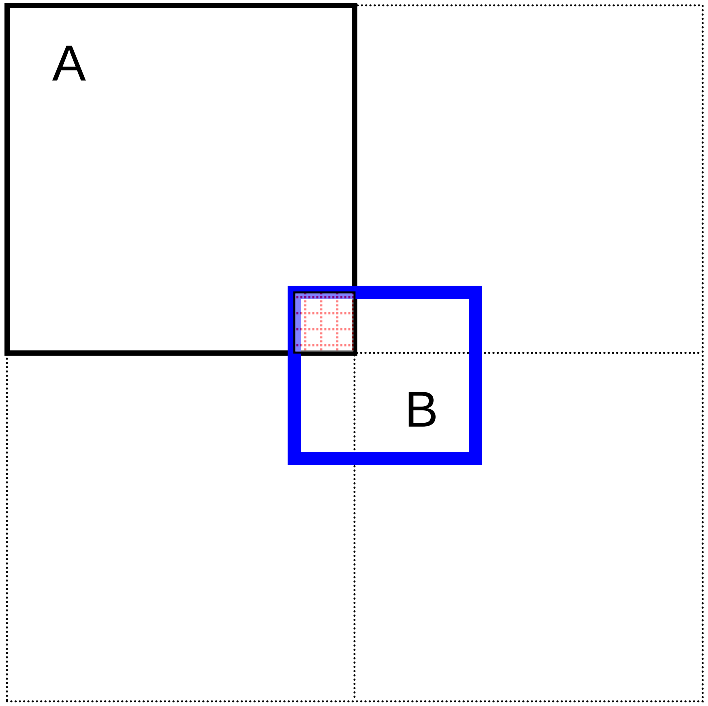
$P(A) = 0{,}25$
$P(B) = 0{,}10$
$P(A,B) = 0{,}08$

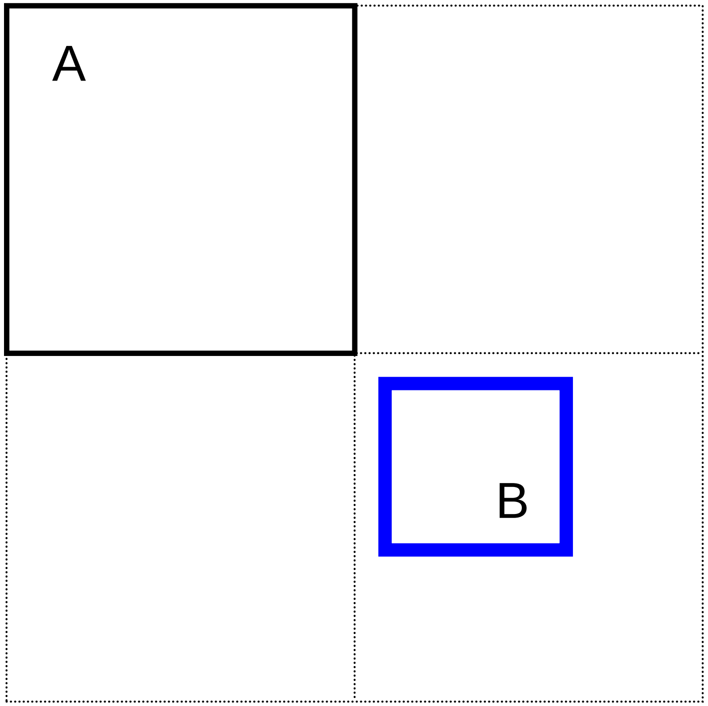
¿Probabilidad de A sabiendo que ha pasado B?

P(A|B)=1

P(A|B)=0,8

Intuir la probabilidad condicionada





P(A) = 0,25
P(B) = 0,10
P(A,B) = 0,005

P(A) = 0,25
P(B) = 0,10
P(A,B) = 0

¿Probabilidad de A sabiendo que ha pasado B?

P(A|B)=0,05

P(A|B)=0

EJEMPLO

	Dolor dental	$\neg$ Dolor dental
Caries	0.04	0.06
$\neg$Caries	0.01	0.89

Probabilidades conjuntas

$P(C) = P(C,D) + P(C,\neg D) = 0.06 + 0.04 = 0.1$

$P(D)= P(C,D) + P(\neg C,D) = 0.04 + 0.01 = 0.05$

$P(C \mid D)= P(C,D) / P(D) = 0.04/(0.04+0.01)=0.80$

MÁS PROPIEDADES

Si A y B son mutuamente excluyentes: $P(A,B) = 0 \Leftrightarrow P(A \mid B) = 0 = P(B \mid A)$.

Si $A \subset B$ entonces $P(B \mid A) = 1$

Regla del producto

- $P(A,B) = P(A \mid B)\, P(B)$, si $p(B) > 0$
- $P(A,B) = P(B \mid A)\, P(A)$, si $p(A) > 0$

Regla del producto condicional

- $P(A,B \mid C) = P(A \mid B,C)\, P(B \mid C)$
- $P(A,B \mid C) = P(B \mid A,C)\, P(A \mid C)$

MÁS PROPIEDADES

Regla de la cadena (factorización de la probabilidad conjunta)

- $P(x_1,x_2,..,x_n)=$ $P(x_1)$
  $P(x_2|x_1)$
  $P(x_3|x_1,x_2)$
  …
  $P(x_n|x_1,..,x_{n-1})$

EJEMPLO2

- 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

- **P(llueve)?, P(¬llueve)?, P(llueve v enCalle)?, P(llueve|enCalle)?**

EJEMPLO2

- 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

- **P(llueve)?, P(¬llueve)?, P(llueve v enCalle)?, P(llueve|enCalle)?**

P(llueve)= P(llueve ^ enCalle)+ P(llueve ^ ¬enCalle)= 0.01+0.09 =0.1.
De forma similar P(¬llueve)= 0.9 (El valor que debe tener)
P(llueve v enCalle)= 0.01+ 0.09+ 0.2 = 0.3
P(llueve|enCalle)=P(llueve ^ enCalle)/P(enCalle)= 0.01/(0.01+0.2) = 0.05

SUCESOS INDEPENDIENTES

Dos <u>sucesos</u> A,B son <u>independientes</u> si el hecho de que ocurra uno no modifica la probabilidad de que ocurra el otro

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

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

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

$$P(A,B) = P(A) \cdot P(B)$$

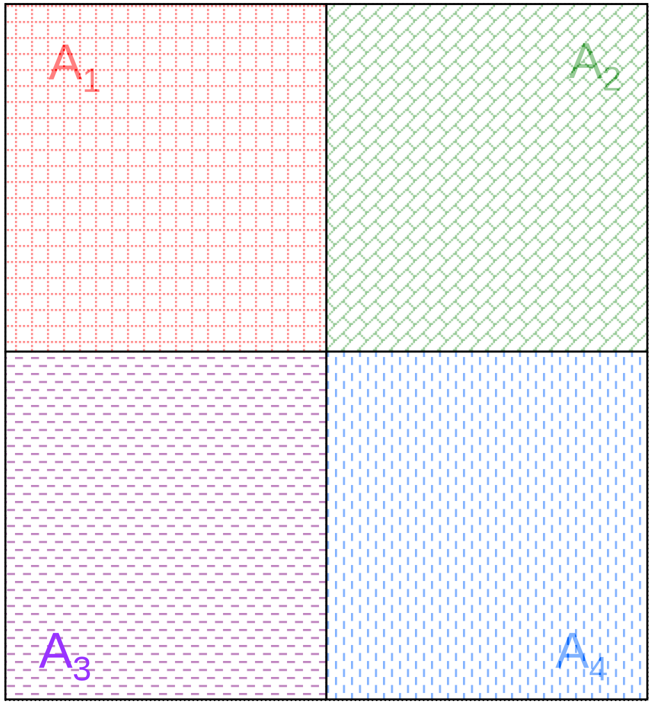
SUCESOS INDEPENDIENTES

Los sucesos A y B son independientes dado el evento C con $P(C)>0$ si y solo si

$$P(A,B \mid C) = P(A|C)\, P(B|C)$$

Un número cualquiera de sucesos son independientes si la probabilidad conjunta de todos los subconjuntos que pueden formarse con ellos es el producto de las probabilidades individuales

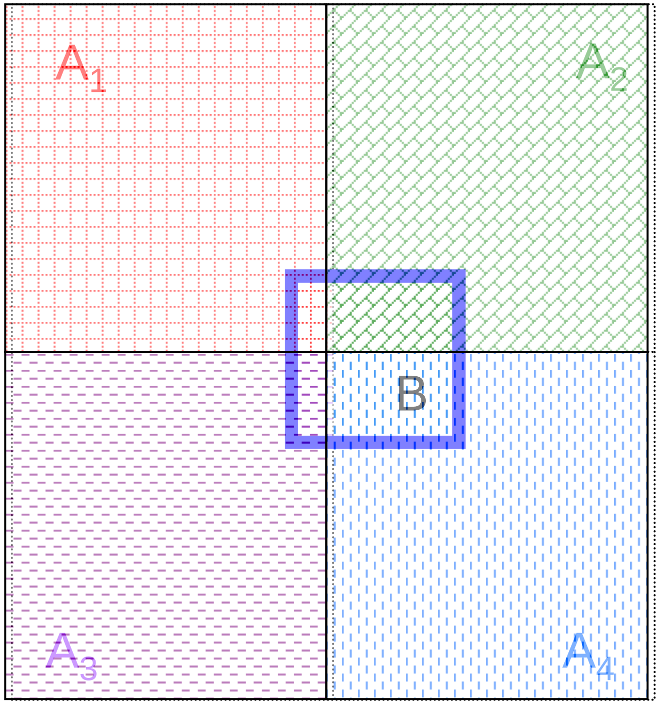
PARTICIONES



Una partición del Espacio muestral es un conjunto de sucesos
A_1, A_2, A_3, A_4… que cumplen:

- Son mutuamente excluyentes (sus intersecciones son disjuntas).
- La unión de todos ellos es el Espacio muestral (exhaustivos)

TEOREMA DE LA PROBABILIDAD TOTAL



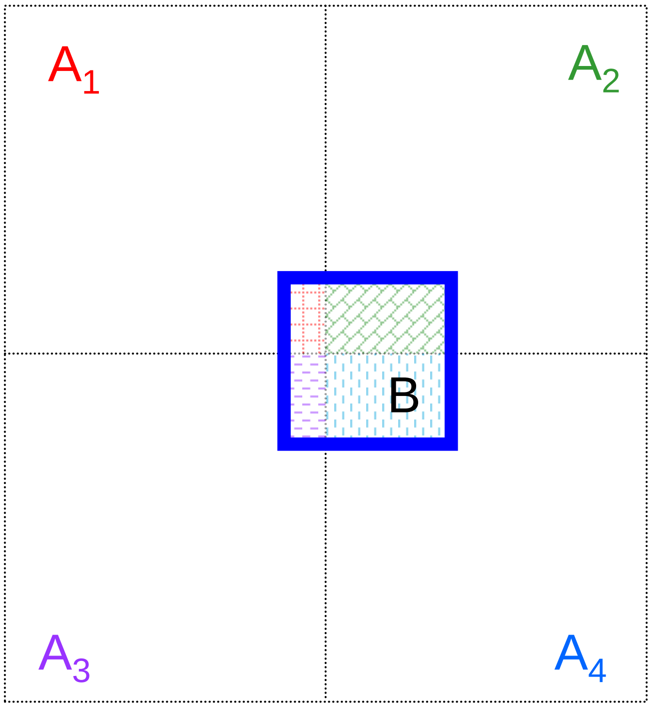
Todo suceso B, puede ser descompuesto en componentes de una partición.

$B = (B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3) \cup (B \cap A_4)$

Si conocemos la probabilidad de B en cada uno de los componentes de una partición, podemos calcular la probabilidad total de B como:.

$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + (B \cap A_4)$
$= P(B|A_1)\, P(A_1) + P(B|A_2)\, P(A_2) + \ldots$

TEOREMA DE BAYES



Si conocemos la probabilidad de B en cada uno de los componentes de una partición de sucesos, entonces…
…si ocurre B, podemos calcular la probabilidad (*a posteriori*) de ocurrencia de cada A_i.

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)}$$

P(B) se puede calcular usando el teorema de la probabilidad total:

$$P(B)=P(B\cap A1) + P(B\cap A2) + P(B\cap A3\ ) + (B\cap A4)$$

REGLA DE BAYES

Se obtiene a partir de las 2 expresiones de la regla del producto

- $P(A,B) = P(A \mid B) \cdot P(B)$
- $P(A,B) = P(B \mid A) \cdot P(A)$,

Igualamos los términos derechos

- $P(A \mid B) \cdot P(B) = P(B \mid A) \cdot P(A)$,

Y despejamos $P(A|B)$ ó $P(B|A)$

EJEMPLO

En un aula el 70% de los alumnos son mujeres. De ellas el 10% son fumadoras. De los hombres, son fumadores el 20%.

¿Qué proporción de fumadores hay en total?

$P(F) = P(F,M) + P(F,H) = P(F \mid M)\,P(M) + P(F \mid H)\,P(H)$

$= 0{,}1 \times 0{,}7 + 0{,}2 \times 0{,}3 = 0{,}13$

T. Prob. Total.
H y M forman una partición

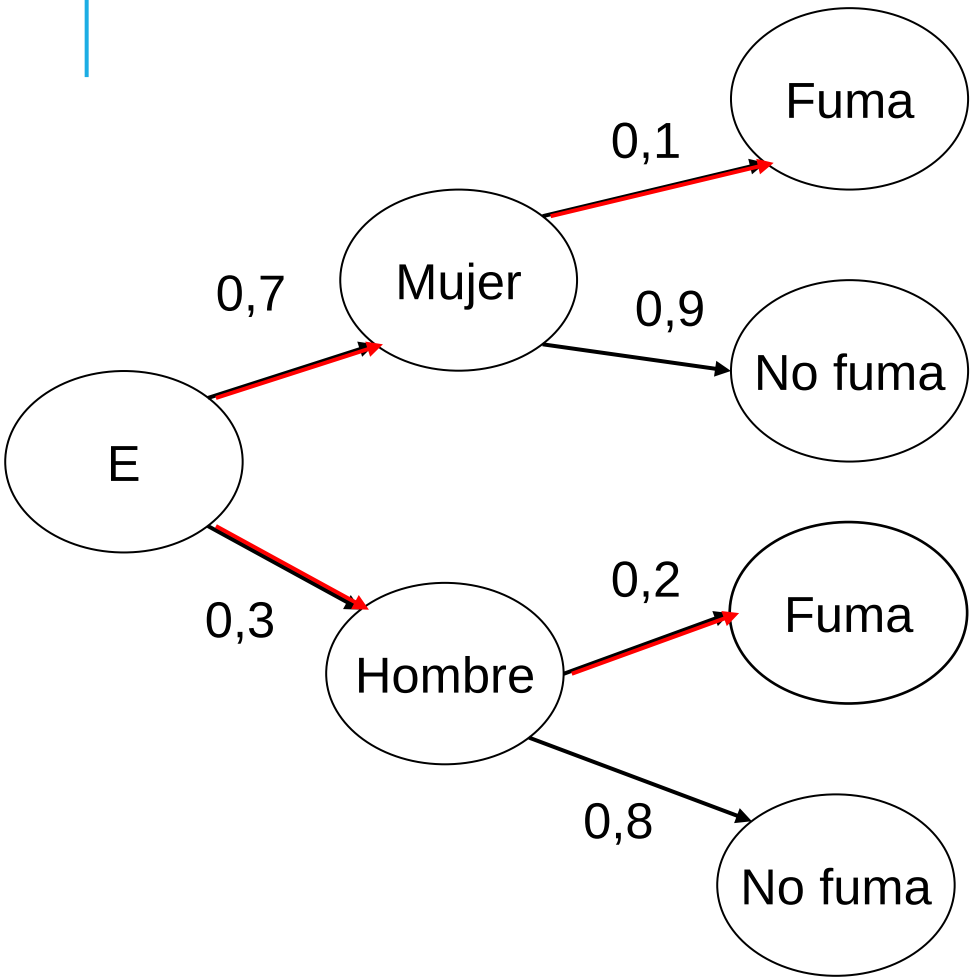
¿Se elije a un individuo al azar y resulta fumador. ¿Cuál es la probabilidad de que sea un hombre?

$P(H \mid F) = P(F, H) / P(F) = P(F \mid H)\,P(H) / P(F)$

$= 0{,}2 \times 0{,}3 / 0{,}13 = 0{,}46$

T. Bayes

EXPRESIÓN DEL PROBLEMA EN FORMA DE ÁRBOL



- Los caminos representan intersecciones, y cada arco es una probabilidad condicional
- Las bifurcaciones representan uniones disjuntas.

P(F)?
= P(F,M) + P(F,H)
= 0,7 x 0,1 + 0,3 x 0,2 = 0,13
= P(M) P(F | M) + P(H) P(F | H)

P(H | F)?
= P(H,F) / P(F)
= 0,3 x 0,2 / 0,13 = 0,46

REGLA DE BAYES GENERALIZADA O CONDICIONAL

$$\boxed{P(A \mid B, C) = \frac{P(B \mid A, C)P(A \mid C)}{P(B \mid C)}}$$

PROBLEMAS DE DIAGNÓSTICO

Partimos de la estimación de P(Enfermo) (*prob. a priori*)

- Prevalencia: % de la población que presenta una enfermedad.

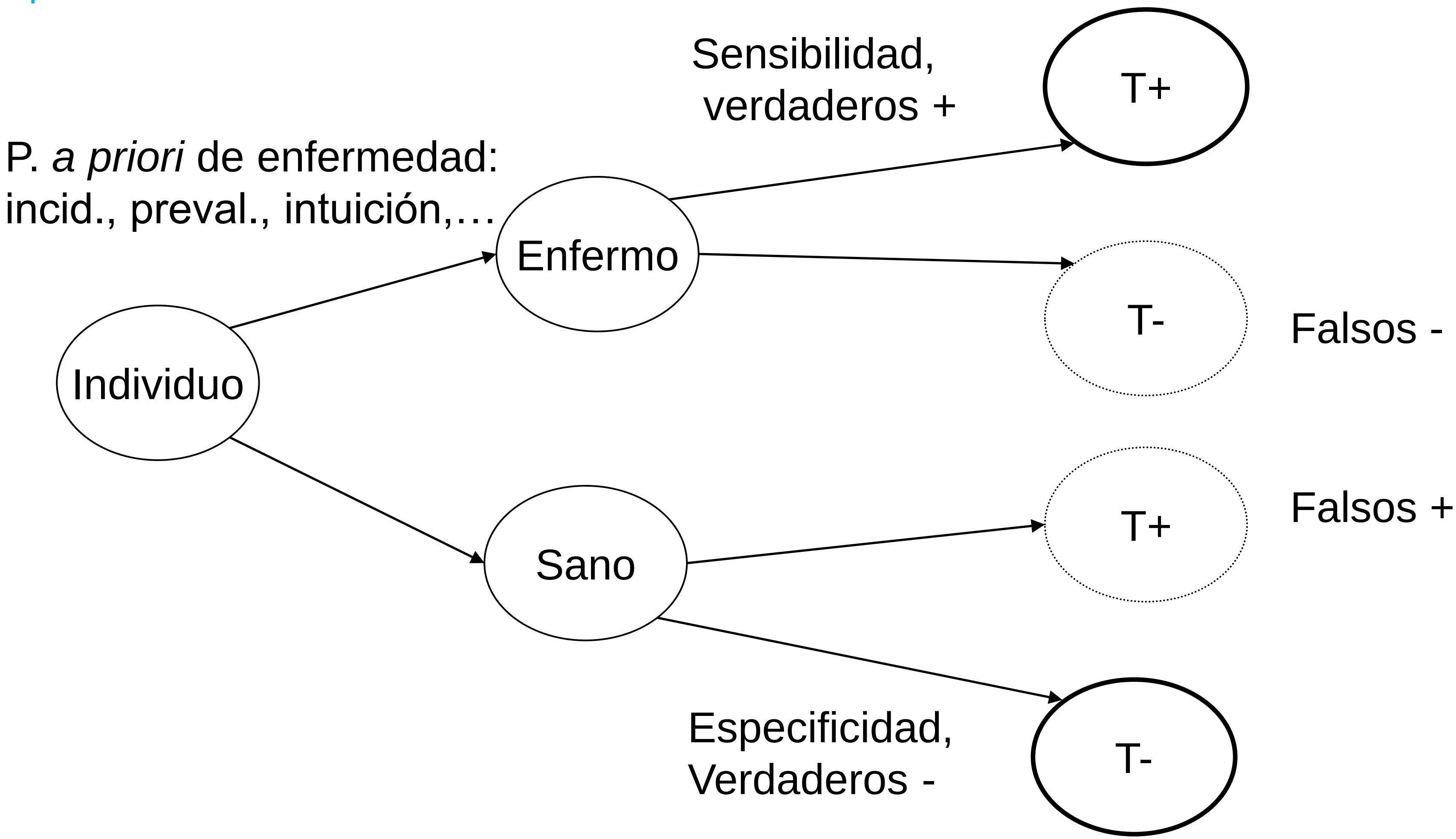
Aplicamos una prueba diagnóstica, de la cual sabemos:

- Sensibilidad (verdaderos +)= Tasa de acierto sobre enfermos.
- Especificidad (verdaderos -)= Tasa de acierto sobre sanos.

Usando T. de Bayes, podemos calcular las probabilidades *a posteriori* (en función de los resultados del test): Índices predictivos

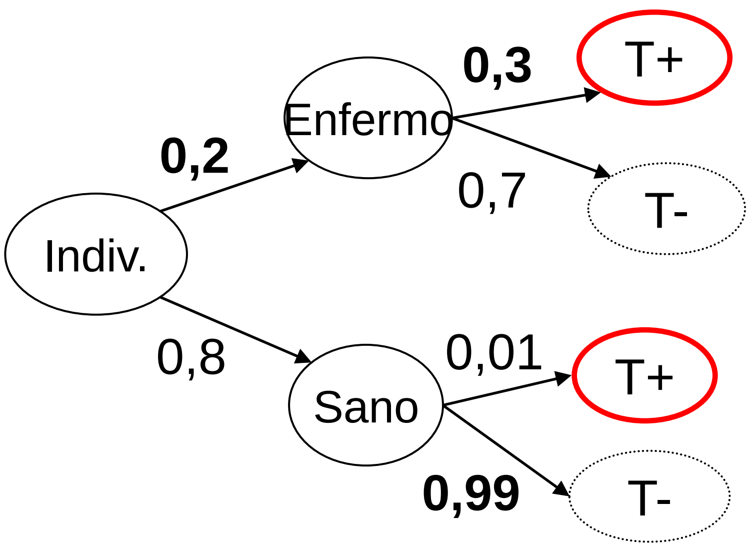
- P(Enfermo | Test +) = Índice Predictivo Positivo
- P(Sano | Test -) = Índice Predictivo Negativo

PROBLEMAS DE DIAGNÓSTICO



EJEMPLO

La diabetes afecta al 20% de los individuos de una población. La presencia de glucosuria se usa como indicador de diabetes. Su sensibilidad es de 0,3 y la especificidad de 0,99. Calcular los índices predictivos.



$$P(Enf\,|\,T+) = \frac{P(Enf,T+)}{P(Enf,T+) + P(Sano,T+)}$$

$$= \frac{0{,}2 \times 0{,}3}{0{,}2 \times 0{,}3 + 0{,}8 \times 0{,}01} = 0{,}88$$

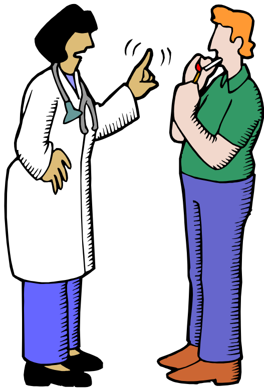
$$P(Sano\,|\,T-) = \frac{P(Sano,T-)}{P(Sano,T-) + P(Enf,T-)}$$

$$= \frac{0{,}8 \cdot 0{,}99}{0{,}8 \cdot 0{,}99 + 0{,}2 \cdot 0{,}7} = 0{,}85$$

En el ejemplo anterior, al llegar un individuo a la consulta tenemos un idea ***a priori*** sobre la probabilidad de que tenga una enfermedad.

-¿Qué probabilidad tengo de estar enfermo?

- En principio un 20%. Le haremos unas pruebas.



A continuación se le pasa una prueba diagnóstica que nos aportará nueva información: Presenta glucosuria o no.

En función del resultado tenemos una nueva idea (***a posteriori***) sobre la probabilidad de que esté enfermo.

- Nuestra opinión a priori ha sido modificada por el resultado de un experimento.

- Presenta glucosuria. La probabilidad ahora es del 88%.

USANDO LA REGLA DE BAYES

M= meningitis, C= cuello rígido

$P(C \mid M) = 0.5$

$P(M) = 1/50000$

$P(C) = 1/20$

$P(M \mid C) = P(C|M) \cdot P(M) \;/\; P(C) = 0.0002$

PROBABILIDAD RELATIVA (FORMA RACIONAL DE BAYES)

L = “latigazos”en el cuello

P(C | L)=0.8   P(L)=1/1000

$$\frac{P(M\,|\,C)}{P(L\,|\,C)} = \frac{P(C\,|\,M)}{P(C\,|\,L)} \cdot \frac{P(M)}{P(L)} = \frac{0.5}{0.8} \cdot \frac{1/50000}{1/1000} = \frac{1}{80}$$

Razón de verosimilitud

Razón de probabilidad

Teniendo cuello rígido, la presencia de latigazos es 80 veces más probable que la meningitis

NORMALIZACIÓN

$$P(M|C) = \frac{P(C|M)P(M)}{P(C)}$$

$P(C) = P(C|M) \cdot P(M) + P(C|\neg M) \cdot P(\neg M)$

$1 / P(C) = \alpha$ es un <u>factor de normalización</u>

(hace que los términos condicionales sumen 1)

En general $P(H \mid E) = \alpha P(E \mid H)P(H)$

NORMALIZACIÓN

P(e|h)	h1	h2	h3
e1	0.30	0.25	0.20
e2	0.40	0.35	0.50
e3	0.25	0.30	0.10
e4	0.05	0.10	0.20

$P(h1)=0.46$
$P(h2)=0.36$
$P(h3)=0.18$

$$P(h1|e4)= \alpha P(e4|h1)P(h1)= \alpha \times 0.05 \times 0.46 = 0.023\alpha$$
$$P(h2|e4)= \alpha P(e4|h2)P(h2)= \alpha \times 0.10 \times 0.36 = 0.036\alpha$$
$$P(h3|e4)= \alpha P(e4|h3)P(h3)= \alpha \times 0.20 \times 0.18 = 0.036\alpha$$

$$0.023\alpha + 0.036\alpha + 0.036\alpha = 1 \;\Downarrow\; \alpha=10.526$$

BAYES CLÁSICO

C =Caries, D = dolor de muelas, E = Enganchar

$P(C \mid D) = 0.8$, $P(C \mid E) = 0.95$

$$P(C \mid D, E) = \frac{P(D, E \mid C)P(C)}{P(D, E)}$$

$$\underbrace{P(C}_{\text{Diagnóstico}} \mid \underbrace{D, E}_{\text{Hallazgos}}) = \frac{P(D, E \mid C)P(C)}{P(DE \mid C)\,P(C) + P(DE \mid \neg C)P(\neg C)}$$

BAYES CLÁSICO

¿Y si hay varias posibles enfermedades H1,H2,...,Hm ?

- Se asume que no pueden coexistir en el paciente (exclusividad mutua) y que el paciente tiene una de las enfermedades (completitud).
- En estas condiciones se puede aplicar la forma general del Teorema de Bayes.
- Se necesita:
  - La prevalencia de cada enfermedad: P(hi)
  - La sensibilidad y especificidad del test para cada enfermedad: P(e|hi), P(¬ e| ¬ hi).
- Se toma la enfermedad con mayor probabilidad: max P(H|E)

BAYES CLÁSICO

¿Y si hay varias enfermedades y varios tests E1,E2,...,En?

- Enfermedad más probable dadas las evidencias: necesitamos calcular el máximo de

  $P(H|E1,E2,...,En)=P(E1,E2,...,En|\ H)P(H)/P(E1,E2,...,En)$
- 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 $P(E1,E2,...,En\ |hi)$ se necesitan guardar unos 2^{30} números reales (unos 10 Terabytes por paciente)
  - ¿Cómo estimamos los números a partir de casos (en la Tierra hay 2^{32} personas aproximadamente)?
  - ¿Cómo hacemos los cálculos computacionalmente eficientes?

BAYES CLÁSICO

Hipótesis 1: los diagnósticos son exclusivos (no puede haber diagnósticos simultáneos) y exhaustivos (no hay otros diagnósticos posibles)

$$P(h_i \mid e_1,..,e_m) = \frac{P(h_i)\cdot P(e_1,..,e_m \mid h_i)}{\sum_j P(h_j)\cdot P(e_1,..,e_m \mid h_j)}$$

Requiere *n* probabilidades a priori y $2^m \cdot n$ condicionadas → $(2^m-1)n+n-1$ parámetros independientes

- Ejemplo: para 3 diagnósticos y 10 hallazgos requiere 3.071 parámetros, con 5 diagnósticos y 20 hallazgos se requieren 5.242.879 parámetros

BAYES SIMPLIFICADO

Hipótesis 2: que los hallazgos son condicionalmente independientes entre sí para cada diagnóstico.

$$P(e_1,..,e_m \mid h_i) = P(e_1 \mid h_i)\cdot..\cdot P(e_m \mid h_i), \forall i$$

$$P(h_i \mid e_1,..,e_m) = \frac{P(h_i)\cdot P(e_1 \mid h_i)\cdot..\cdot P(e_m \mid h_i)}{\sum_j P(h_j)\cdot P(e_1 \mid h_j)\cdot..\cdot P(e_m \mid h_j)}$$

Requiere *n* probabilidades a priori y mn condicionales → nm+n-1 parámetros independientes

- Ejemplo: 3 diagnósticos, 10 hallazgos requiere 32 parámetros, 5 diagnósticos y 20 hallazgos requiere 104 parámetros

BAYES SIMPLIFICADO

Si consideramos que X e Y son variables independientes, considerando la presencia de H, entonces

- $\mathbf{P}(X \mid Y,H) = \mathbf{P}(X \mid H)$
- $\mathbf{P}(H \mid X,Y) = \alpha \mathbf{P}(H)\ \mathbf{P}(X \mid H)\ \mathbf{P}(Y \mid H)$,
  donde α es la constante de normalización que hace que todos los términos $\mathbf{P}(h \mid x,y)$ sumen 1.

RESOLUCIÓN DEL CASO 1

Sensores independientes. Cada sensor (sonar, temperatura y cámara) queda definido por dos etiquetas, “humano” y “robot”.

Las probabilidades de las evidencias a priori se han calculado con una muestra de 1000 objetos:

- En el sensor **sonar** de 1000 objetos, han pasado por el sonar 500 objetos robot y 500 objetos humano. De los 500 objetos robot el sonar ha dicho 350 veces que eran robots cuando sí eran robots y 150 veces que eran humanos cuando realmente eran robots. De los 500 objetos humano, el sonar ha dicho que 400 eran humanos cuando realmente eran humanos pero 100 objetos los ha catalogado como robots cuando eran humanos
- En el sensor **temperatura** de 1000 objetos, han pasado 500 objetos robot y 500 objetos humanos. De los 500 objetos robot este sensor ha dicho 350 veces que eran robots cuando sí eran robots y 150 veces que eran humanos cuando realmente eran robots. De los 500 objetos humano, el sonar ha dicho que 450 eran humanos cuando realmente eran humanos pero 50 objetos los ha catalogado como robots cuando eran humanos
- En el sensor **cámara** de 1000 objetos, han pasado por el sensor 500 objetos robot y 500 objetos humanos. De los 500 objetos robot el sensor ha dicho que 300 eran robot y 200 que eran humanos. De los 500 objetos humanos, el sensor ha dicho que 200 eran robots y 300 humanos.

RESOLUCIÓN DEL CASO 1

Hipótesis:

- Humano (H) -> p(H) = 0.55
- Robot (R) -> p(R) = 0.45

Evidencias:

- Sensor Sonar:
  - p(R1/R) = 07
  - p(H1/R) = 0.3
  - p(R1/H) = 0.2
  - p(H1/H) = 0.8
- Sensor Temperatura:
  - p(R2/R) = 0.7
  - p(H2/R) = 0.3
  - p(R2/H) = 0.1
  - p(H2/H) = 0.9
- Sensor Cámara:
  - p(R3/R) = 0.60
  - p(H3/R) = 0.40
  - p(R3/H) = 0.40
  - p(H3/H) = 0.60

RESOLUCIÓN DEL CASO 1

¿Probabilidad que sea humano cuando los tres sensores dicen que es humano?

¿Probabilidad que sea humano cuando los sensores sonar y temperatura dicen que es humano y el sensor cámara dice que es robot?

¿Probabilidad que sea humano cuando el sensor sonar dice que es humano y los sensores temperatura y cámara dicen que es robot?

¿Probabilidad que sea humano cuando los tres sensores, sonar, temperatura y cámara, dicen que es robot?

¿Probabilidad que sea robot cuando los tres sensores dicen que es humano?

¿Probabilidad que sea robot cuando los tres sensores dicen que es robot?