Incertidumbre y conocimiento

- `SI síntoma(P,dolor-de-muelas) ENTONCES enfermedad(P,caries)`
  - **¿Expresa esta regla un conocimiento correcto?**
  - **Quizás sería mejor un conocimiento más exhaustivo:**

```
SI sintoma(P,dolor-de-muelas)
ENTONCES enfermedad(P,caries) O
         enfermedad(P,sinusitis) O
         enfermedad(P,muela-del-juicio) O
         .....
```

- **¿Por qué a veces el conocimiento categórico no nos sirve?**
  - **Demasiado trabajo ser exhaustivo**
  - **Desconocimiento teórico**
  - **Desconocimiento práctico**

Incertidumbre y conocimiento

- **Otra forma de expresar el conocimiento: grado de creencia**
  - ***Creemos*, basándonos en nuestras *percepciones*, que un paciente que tenga dolor de muelas, tiene caries con una probabilidad del 80 %**
  - $P(Caries = true|Dolor = true) = 0{,}8$
  - **La probabilidad expresa el *grado de creencia*, no el *grado de verdad***
  - **Por tanto, la probabilidad puede cambiar a medida que se conocen nuevas *evidencias***
- **La teoría de la probabilidad servirá como medio de representación de conocimiento incierto**

Variables aleatorias

- **Variables aleatorias:** una "parte" del mundo cuyo estado podemos desconocer
  - Ejemplo: la variable aleatoria $Caries$ describe el hecho de que un paciente pueda o no tener caries
  - Nuestra descripción del mundo vendrá dada por un conjunto de variables aleatorias
- Una variable aleatoria puede tomar diferentes ***valores*** de su ***dominio***
  - Los posibles valores de $Caries$ son $true$ y $false$
  - Notación: variables aleatorias en mayúsculas y sus valores en minúsculas
- Tipos de variables aleatorias:
  - Booleanas (notación: $caries$ y $\neg caries$ son equivalentes a $Caries = true$ y $Caries = false$, respectivamente)
  - Discretas (incluyen a las booleanas)
  - Continuas
- En lo que sigue, nos centraremos en las variables discretas

Proposiciones

- **Usando las conectivas proposicionales y las variables aleatorias, podemos expresar *proposiciones***
- **Ejemplos:**
  - $caries \wedge \neg dolor$
  - $Caries = true \vee Tiempo = nublado$
  - $Tirada1 = 5 \wedge (Tirada2 = 4 \vee Tirada2 = 5 \vee Tirada2 = 6)$
- **Asignaremos probabilidades a las proposiciones para expresar nuestro grado de creencia en las mismas**

Eventos atómicos

- **Dado un conjunto de variables aleatorias que describen nuestro "mundo", un *evento atómico* es un tipo particular de proposición:**
  - **Conjunción de proposiciones elementales, que expresan un valor concreto para todas y cada una de las variables**
  - **Ejemplo:** $caries \wedge \neg dolor$, **siempre que** $Caries$ **y** $Dolor$ **sean todas las variables aleatorias en nuestra descripción del mundo**
- **Características:**
  - **Mútuamente excluyentes**
  - **Todos los eventos atómicos son exhaustivos (alguno debe ocurrir)**
  - **Un evento atómico implica la verdad o falsedad de toda proposición**
  - **Toda proposición es equivalente a la disyunción de un conjunto de eventos atómicos: por ejemplo,** $caries$ **es equivalente a** $(caries \wedge dolor) \vee (caries \wedge \neg dolor)$

Probabilidad incondicional

- **Probabilidad *incondicional* (o *a priori*) asociada a una proposición a:**
  - **Grado de creencia sobre a, *en ausencia de cualquier otra información o evidencia*, notada $P(a)$**
  - **Ejemplo:** $P(caries) = 0{,}1$, $P(caries \wedge \neg dolor) = 0{,}05$
  - **Notación:** $P(caries, \neg dolor)$ **es equivalente a** $P(caries \wedge \neg dolor)$
- ***Distribución de probabilidad* de una variable aleatoria**
  - **Ejemplo:** $Tiempo$ **es una v.a. con valores** $lluvia$, sol, $nubes$ **y** $nieve$; **una distribución de probabilidad viene dada por las probabilidades de que la variable pueda tomar los diferentes valores**
  - $P(Tiempo = sol) = 0{,}7, P(Tiempo = lluvia) = 0{,}2, P(Tiempo = nubes) = 0{,}08, P(Tiempo = nieve) = 0{,}02$

Probabilidad incondicional

- **Notación: usaremos $\boldsymbol{P}$ (en negrita), para expresar de manera compacta una distribución de probabilidad (fijado un orden entre sus valores)**
  - **Ejemplo:** $\boldsymbol{P}(Tiempo) = \langle 0{,}7, 0{,}2, 0{,}08, 0{,}02 \rangle$

- **Distribución de probabilidad *conjunta*: probabilidad de cada combinación de valores de dos o más variables aleatorias**
  - **Notación** $\boldsymbol{P}(X, Y)$**: manera compacta de denotar a una tabla con esas probabilidades**
  - $\boldsymbol{P}(Tiempo, Caries)$ **denota una tabla con** 4×2 **entradas**
  - **Distribución de probabilidad *conjunta y completa* (DCC): probabilidad de cada evento atómico**
  - **Una DCC es una *especificación completa* de la incertidumbre que se tiene sobre el "mundo" descrito**

Probabilidad condicional

- **Probabilidad *condicional* (o *a posteriori*) asociada a a dada b (a y b proposiciones):**
  - **Grado de creencia sobre a, dado que *todo lo que sabemos* es que b ocurre, notada** $P(a|b)$
  - **Ejemplo:** $P(caries|dolor) = 0{,}8$ **significa que una vez sabido que un paciente tiene dolor de muelas (y sólamente sabemos eso), nuestra creencia es que el paciente tendrá caries con probabilidad** $0{,}8$
- **Relación entre probabilidad condicional e incondicional:**
  - $P(a|b) = \frac{P(a \wedge b)}{P(b)}$**, o también**
  - $P(a \wedge b) = P(a|b)P(b)$ **(regla del producto)**
- **Notación $\boldsymbol{P}(X|Y)$ para expresar la tabla de probabilidades condicionales**
  - **Forma compacta de la regla del producto:** $\boldsymbol{P}(X, Y) = \boldsymbol{P}(X|Y)\boldsymbol{P}(Y)$

Probabilidad condicional *vs* implicación lógica

- **La probabilidad condicional formaliza el hecho de que los grados de creencias se actualizan a medida que se van conociendo nuevas evidencias en el mundo incierto**

- **La probabilidad condicional *no* es lo mismo que una implicación lógica con incertidumbre**

    - $P(a|b) = 0{,}8$ **no es lo mismo que decir que "siempre que** b **sea verdad, entonces** $P(a) = 0{,}8$**"**

    - **Ya que** $P(a|b)$ **refleja que** b **es *la única* evidencia conocida**

Axiomas de probabilidad y cálculo

- **Interpretación de la función P**
  - **Frecuentista: casos posibles entre casos totales**
  - **Subjetiva: grado de creencia basado en nuestras percepciones**
- **Axiomas sobre la función de probabilidad**
  - $0 \leq P(a) \leq 1$
  - $P(true) = 1$ **y** $P(false) = 0$ **donde** $true$ **y** $false$ **representan a cualquier proposición tautológica o insatisfactible, respectivamente**
  - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- **El resto del cálculo de probabilidades se construye sobre estos tres axiomas. Por ejemplo:**
  - $P(\neg a) = 1 - P(a)$
  - $\sum_{i=1}^{n} P(D = d_i) = 1$**, siendo** D **una v.a. y** $d_i, i = 1, \ldots, n$ **sus posibles valores**
  - $P(a) = \sum_{e_i \in \boldsymbol{e}(a)} P(e_i)$**, siendo** a **una proposición y** $\boldsymbol{e}(a)$ **el conjunto de eventos atómicos cuya disyunción es equivalente a** a

Inferencia probabilística

- Por ***inferencia probabilística*** entendemos el cálculo de la probabilidad de una proposición dada ***condicionada*** por la observación de determinadas evidencias
  - Es decir, cálculos del tipo $P(a|b)$ donde a es la proposición que se ***consulta*** y b es la proposición que se ha ***observado***
  - El ***conocimiento base*** vendrá dado por una DCC (representada de alguna manera ***eficiente***, como ya veremos)
- **Ejemplo de DCC:**

	$dolor$	$dolor$	$\neg dolor$	$\neg dolor$
	$hueco$	$\neg hueco$	$hueco$	$\neg hueco$
$caries$	0.108	0.012	0.072	0.008
$\neg caries$	0.016	0.064	0.144	0.576

Inferencia probabilística a partir de una DCC

- **Cálculo de probabilidades incondicionales basado en la ecuación** $P(a) = \sum_{e_i \in \boldsymbol{e}(a)} P(e_i)$
  - $P(caries \vee dolor) = 0{,}108 + 0{,}012 + 0{,}072 + 0{,}008 + 0{,}016 + 0{,}064 = 0{,}28$
  - $P(caries) = 0{,}108 + 0{,}012 + 0{,}072 + 0{,}008 = 0{,}2$
- **En general:**
  - $\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} \boldsymbol{P}(\boldsymbol{Y}, \boldsymbol{z})$ **(regla de marginalización)**
- **Notación:**
  - $\boldsymbol{Y}$ **es un vector de variables aleatorias, simbolizando cualquier combinación de valores de esas variables**
  - $\boldsymbol{z}$ **representa una combinación de valores concretos para un conjunto** $\boldsymbol{Z}$ **de variables aleatorias (las restantes)**
- **Variante:** $\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\boldsymbol{z}} \boldsymbol{P}(\boldsymbol{Y}|\boldsymbol{z}) \cdot P(\boldsymbol{z})$ ***(regla de condicionantes)***

Inferencia probabilística a partir de una DCC

- **Cálculo de probabilidades condicionales usando la DCC (un ejemplo)**
  - **Probabilidad de tener caries, observado que hay dolor:**

$$P(caries|dolor) = \frac{P(caries \wedge dolor)}{P(dolor)} = \frac{0{,}108 + 0{,}012}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}6$$

  - **Como comprobación podemos calcular el opuesto:**

$$P(\neg caries|dolor) = \frac{P(\neg caries \wedge dolor)}{P(dolor)} = \frac{0{,}016 + 0{,}064}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}4$$

  - $P(dolor)$ **puede verse como una constante que *normaliza* la distribución $\boldsymbol{P(Caries|dolor)}$ haciendo que sume 1:**

$$\boldsymbol{P}(Caries|dolor) = \alpha \boldsymbol{P}(Caries, dolor) =$$
$$= \alpha[\boldsymbol{P}(Caries, dolor, hueco) + \boldsymbol{P}(Caries, dolor, \neg hueco)] =$$
$$= \alpha[\langle 0{,}108, 0{,}016\rangle + \langle 0{,}012, 0{,}064\rangle] = \alpha\langle 0{,}12, 0{,}08\rangle = \langle 0{,}6, 0{,}4\rangle$$

Inferencia probabilística a partir de una DCC

- En general, dada una variable aleatoria X, un conjunto de variables observadas $\boldsymbol{E}$ (con valor concreto $\boldsymbol{e}$), se tiene:

$$\boldsymbol{P}(X|\boldsymbol{e}) = \alpha \boldsymbol{P}(X, \boldsymbol{e}) = \alpha \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

  - un sumando para cada combinación $\boldsymbol{y}$ de valores de las variables restantes $\boldsymbol{Y}$ no observadas
  - α es una constante de normalización, que hace que la distribución de probabilidades sume 1
- Dada una DCC, la fórmula anterior nos da un método para realizar inferencia probabilística
- Problema en la práctica: exponencialidad
  - Con n variables, procesar la DCC necesita un tiempo $O(2^n)$
  - En un problema real, podría haber cientos o miles de variables

Independencia probabilística

- **En muchos casos prácticos, muchas de las variables de un problema son *independientes* entre sí**
  - **Ejemplo:** $P(Tiempo = nublado|dolor, caries, hueco) = P(Tiempo = nublado)$
  - **Si la variable** $Tiempo$ **(con 4 posibles valores) formara parte de una descripción en la que están** $Caries$, $Hueco$ **y** $Dolor$, **no necesitaríamos una tabla con 32 entradas para describir la DCC, sino dos tablas independientes (8+4 entradas)**
- **Dos variables aleatorias** X **e** Y **son independientes si** $\boldsymbol{P}(X|Y) = \boldsymbol{P}(X)$ **(equivalentemente,** $\boldsymbol{P}(Y|X) = \boldsymbol{P}(Y)$ **ó** $\boldsymbol{P}(X,Y) = \boldsymbol{P}(X) \cdot \boldsymbol{P}(Y)$**)**
  - **En general, dos proposiciones** a **y** b **son independientes si** $P(a|b) = P(a)$
- **Asumir independencia entre ciertas variables ayuda a que la representación del mundo sea más manejable**
  - **Reduce la exponencialidad (*factorización* del problema)**
  - **El asumir que dos variables son independientes está basado normalmente el en conocimiento previo del dominio que se modela**

La regla de Bayes

- **De $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$ podemos deducir la siguiente fórmula, conocida como *regla de Bayes*:**

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

- **Regla de Bayes para variables aleatorias:**

$$\boldsymbol{P}(Y|X) = \frac{\boldsymbol{P}(X|Y)\boldsymbol{P}(Y)}{\boldsymbol{P}(X)}$$

  - **recuérdese que esta notación representa un conjunto de ecuaciones, una para cada valor específico de las variables**

- **Generalización, en presencia de un conjunto $\boldsymbol{e}$ de observaciones:**

$$\boldsymbol{P}(Y|X, \boldsymbol{e}) = \frac{\boldsymbol{P}(X|Y, \boldsymbol{e})\boldsymbol{P}(Y|\boldsymbol{e})}{\boldsymbol{P}(X|\boldsymbol{e})}$$

Uso de la regla de Bayes

- **La regla de Bayes nos permite *diagnosticar* en función de nuestro conocimiento de relaciones *causales***
- **Ejemplo**
  - **Sabemos que la probabilidad de que un paciente de meningitis tenga el cuello hinchado es 0.5 (relación causal)**
  - **También sabemos la probabilidad (incondicional) de tener meningitis ($\frac{1}{50000}$) y de tener el cuello hinchado (0.05)**
  - **Estas probabilidades provienen del conocimiento y la experiencia**
  - **La regla de Bayes nos permite diagnosticar la probabilidad de tener meningitis una vez que se ha observado que el paciente tiene el cuello hinchado:**

$$P(m|h) = \frac{P(h|m)P(m)}{P(h)} = \frac{0{,}5 \times \frac{1}{50000}}{0{,}05} = 0{,}0002$$

Uso de la regla de Bayes

- **Una variante del ejemplo anterior:**
  - **Si M es una variable aleatoria booleana, podríamos calcular con la siguiente fórmula, donde α es el factor de normalización:**

$$\boldsymbol{P}(M|h) = \alpha\langle P(h|m)P(m), P(h|\neg m)P(\neg m)\rangle$$

- **Versión para variables aleatorias:** $\boldsymbol{P}(Y|X) = \alpha\boldsymbol{P}(X|Y)\boldsymbol{P}(Y)$
- **¿Por qué calcular el diagnóstico en función del conocimiento causal y no al revés?**
  - **Porque es más fácil y robusto disponer de probabilidades causales que de probabilidades de diagnóstico**
  - **La información probabilística está generalmente disponible en la forma** $P(efecto|causa)$
  - **Y usamos la regla de Bayes para calcular** $P(causa|efecto)$

La regla de Bayes: combinando evidencias

- **Evidencias múltiples y exponencialidad**
  - **Cuando manejamos varias variables para representar distintas evidencias (y es lo habitual), el uso de la regla de Bayes puede necesitar una cantidad exponencial de probabilidades de tipo** $P(efecto|causa)$
  - **Supongamos, por ejemplo, que tenemos evidencias sobre oquedades en los dientes (**$Hueco$**) y sobre dolor en el paciente (**$Dolor$**), y queremos diagnosticar si tiene caries (**$Caries$**)**
  - **Por la regla de Bayes:** $\boldsymbol{P}(Caries|Dolor, Hueco) = \alpha \boldsymbol{P}(Dolor, Hueco|Caries)\boldsymbol{P}(Caries)$
  - **En general, si se tienen** n **variables booleanas de evidencia, deberiamos tener** 2^n **probabilidades condicionales (por cada valor de una variable de diagnóstico) en nuestra base de conocimiento**
  - **Esta exponencialidad no se maneja bien desde el punto de vista práctico**
- **Nuevamente, asumir cierta noción de *independencia* entre variables simplificará la cuestión**

Independencia condicional

- **Sin embargo, en nuestro ejemplo $Dolor$ y $Hueco$ no son independientes**
  - **Ambas dependen de $Caries$**
- **Pero son independientes *una vez conocido el valor de* $Caries$**
  - **Es decir: $\boldsymbol{P}(Dolor|Hueco, Caries) = \boldsymbol{P}(Dolor|Caries)$ o equivalentemente $\boldsymbol{P}(Hueco|Dolor, Caries) = \boldsymbol{P}(Hueco|Caries)$**
  - **También equivalente: $\boldsymbol{P}(Hueco, Dolor|Caries) = \boldsymbol{P}(Hueco|Caries)\boldsymbol{P}(Dolor|Caries)$**
- **En general, dos v.a. X e Y son independientes dada una v.a. Z si $\boldsymbol{P}(X, Y|Z) = \boldsymbol{P}(X|Z)\boldsymbol{P}(Y|Z)$**
  - **O equivalentemente $\boldsymbol{P}(X|Y, Z) = \boldsymbol{P}(X|Z)$ ó $\boldsymbol{P}(Y|X, Z) = \boldsymbol{P}(Y|Z)$**
- **Esto simplifica la inferencia probabilística:**

$$\boldsymbol{P}(Caries|Dolor, Hueco) = \alpha \boldsymbol{P}(Dolor|Caries)\boldsymbol{P}(Hueco|Caries)\boldsymbol{P}(Caries)$$

  - **Sólo es necesario saber las probabilidades causales de cada variable *por separado***
  - **“Reduciendo” la exponencialidad**

Independencia condicional

- **La independencia condicional entre algunas variables es esencial para un almacenamiento eficiente de las DCCs. Por ejemplo:**

$$\boldsymbol{P}(Dolor, Hueco, Caries) = \boldsymbol{P}(Dolor, Hueco|Caries)\boldsymbol{P}(Caries) =$$

$$= \boldsymbol{P}(Dolor|Caries)\boldsymbol{P}(Hueco|Caries)\boldsymbol{P}(Caries)$$

  - **En lugar de tener una tabla con 7 números independientes sólo necesitamos 5 números independientes (en tres tablas)**
  - **Si $Caries$ tuviera n síntomas independientes entre sí (dado $Caries$), el tamaño de la representación de la DCC crece $O(n)$ en lugar de $O(2^n)$**

- **Con una $Causa$ con n efectos E_i independientes entre sí dado $Causa$, se tiene $\boldsymbol{P}(Causa, E_1, \ldots, E_n) = \boldsymbol{P}(Causa)\prod_i \boldsymbol{P}(E_i|Causa)$**
  - **No siempre se dan estas condiciones de independencia tan fuertes, aunque a veces compensa asumirlas**

Ejemplo

- **Problema**
  - **Un 1 % de las mujeres de más de 40 años que se hacen un chequeo tienen cáncer de mama. Un 80 % de las que tienen cáncer de mama se detectan con una mamografía y el 9.6 % de las que no tienen cáncer de mama, al realizarse una mamografía se le diagnostica cáncer erróneamente**
- **Variables aleatorias: C (tener cáncer de mama) y M (mamografía positiva)**
- **Pregunta 1: ¿cuál es la probabilidad de tener cáncer si la mamografía así lo diagnostica?**
  - $\boldsymbol{P}(C|m) = \alpha\boldsymbol{P}(C,m) = \alpha\boldsymbol{P}(m|C)\boldsymbol{P}(C) = \alpha\langle P(m|c)P(c), P(m|\neg c)P(\neg c)\rangle = \alpha\langle 0{,}8 \cdot 0{,}01, 0{,}096 \cdot 0{,}99\rangle = \alpha\langle 0{,}008, 0{,}09504\rangle = \langle 0{,}0776, 0{,}9223\rangle$
  - **Luego el 7.8 % de las mujeres diagnosticadas positivamente con mamografía tendrán realmente cáncer de mama**

Ejemplo

- **Pregunta 2: ¿cuál es la probabilidad de tener cáncer si tras dos mamografías consecutivas en ambas se diagnostica cáncer?**
  - **Variables aleatorias:** M_1 **(primera mamografía positiva) y** M_2 **(segunda mamografía positiva)**
  - **Obviamente, no podemos asumir independencia incondicional entre** M_1 **y** M_2
  - **Pero es plausible asumir independencia condicional de** M_1 **y** M_2 ***dada*** C
  - **Por tanto,** $\boldsymbol{P}(C|m_1, m_2) = \alpha\boldsymbol{P}(C, m_1, m_2) = \alpha\boldsymbol{P}(m_1, m_2|C)\boldsymbol{P}(C) = \alpha\boldsymbol{P}(m_1|C)\boldsymbol{P}(m_2|C)\boldsymbol{P}(C) = \alpha\langle P(m_1|c)P(m_2|c)P(c), P(m_2|\neg c)P(m_2|\neg c)P(\neg c)\rangle = \alpha\langle 0{,}8 \cdot 0{,}8 \cdot 0{,}01, 0{,}096 \cdot 0{,}096 \cdot 0{,}99\rangle = \langle 0{,}412, 0{,}588\rangle$
  - **Luego aproximadamente el 41 % de las mujeres doblemente diagnosticadas positivamente con mamografía tendrán realmente cáncer de mama**

Redes Bayesianas

- **En general, las relaciones de independencia condicional permiten simplificar drásticamente las DCCs, haciendo que se puedan usar en la práctica**
- **Las Redes Bayesianas constituyen un método de representación de DCCs que explota las relaciones de independencia condicional**
  - **Permitiendo un tratamiento manejable de la inferencia probabilística**
  - **Es lo que veremos en el próximo tema**

Bibliografía

- Russell, S. y Norvig, P. *Inteligencia artificial (Un enfoque moderno)* (Prentice–Hall Hispanoamericana, 1996)
  - Cap. 13: "Incertidumbre"