

# **Sistemas Inteligentes**

**Escuela Técnica Superior de Informática**

**Universitat Politècnica de València**

**Tema B2T1**

**Razonamiento probabilístico**

# Índice

- 1 Introducción: Incertidumbre y probabilidad ▷ 2
- 2 Teoría de la probabilidad ▷ 5
- 3 Razonamiento probabilístico: inferencia ▷ 15
- 4 Incertidumbre y decisiones óptimas ▷ 19
- 5 Bibliografía ▷ 23

# Índice

- 1 *Introducción: Incertidumbre y probabilidad* ▷ 2
- 2 Teoría de la probabilidad ▷ 5
- 3 Razonamiento probabilístico: inferencia ▷ 15
- 4 Incertidumbre y decisiones óptimas ▷ 19
- 5 Bibliografía ▷ 23

# Incertidumbre

Sea  $A_t$  el consecuente (acción) de alguna regla:

$A_t = \text{SALIR HACIA EL AEROPUERTO } t \text{ MINUTOS ANTES DEL VUELO}$

Pongamos  $t = 25$ . Pregunta clave: ¿Me permitirá  $A_{25}$  llegar a tiempo?

*Problemas:*

- observabilidad parcial (estado de la carretera, planes de otros conductores, etc.)
- información imprecisa sobre el estado del tráfico
- otras incertidumbres tales como posible pinchazo de neumáticos, etc.
- inmensa complejidad de modelar y predecir el tráfico

El razonamiento basado solo en reglas (lógica) tiene dos tipos de riesgo:

- posible falsedad: ( $A_{25}$  ME PERMITIRÁ LLEGAR A TIEMPO) ¿en todos los casos?
- conclusiones poco útiles para decidir: ( $A_{25}$  ME PERMITIRÁ LLEGAR A TIEMPO SI NO HAY ACCIDENTE EN EL PUENTE Y NO LLUEVE Y LOS NEUMÁTICOS NO SE PINCHAN Y ...).

Obviamente, una acción tal como  $A_{1440}$  garantizaría llegar razonablemente a tiempo, pero ¡a costa de tener que pernoctar en el aeropuerto! ;)

# Aproximaciones para modelar la incertidumbre

Muchas aproximaciones se han propuesto históricamente:

- Lógica no monotónica [1]
- Reglas con factores de certeza (ej. en el sistema MYCIN [2])
- Lógica difusa (*fuzzy sets*) [3]
- Métodos basados en la *Teoría de la Probabilidad* [4,5]

En 1931 de Finetti demostró la siguiente proposición [5, p. 489-490]:

Si un agente [de bolsa] *A* basa sus decisiones [inversiones] en un sistema de grados de certeza que viola los axiomas de la Teoría de la Probabilidad, entonces existe una combinación de decisiones con las que otro agente *B* hará que *A* pierda [dinero] sistemáticamente.

Actualmente prevalecen los *métodos probabilísticos* como marco general para representar la incertidumbre. Estos métodos permiten:

- Modelar y combinar de forma adecuada y consistente:
  - La *inexactitud o vaguedad del conocimiento a priori*
  - La *imprecisión de los hechos, observaciones o datos*
- *Aprendizaje automático* de los modelos de representación

# Índice

- 1 Introducción: Incertidumbre y probabilidad ▷ 2
- 2 *Teoría de la probabilidad* ▷ 5
- 3 Razonamiento probabilístico: inferencia ▷ 15
- 4 Incertidumbre y decisiones óptimas ▷ 19
- 5 Bibliografía ▷ 23

# Espacio muestral y de probabilidad

Sea  $\Omega$  un conjunto, denominado *espacio muestral*.

Ejemplo: los 6 posibles resultados de una tirada de dado,

$$\Omega = \{t \in \mathbb{N} : 1 \leq t \leq 6\}$$

Un elemento  $\omega \in \Omega$  se denomina *evento atómico*, *punto muestral*, o simplemente *punto* o *muestra*.

*Espacio o modelo de probabilidad* es un espacio muestral junto con una función  $P : \Omega \rightarrow \mathbb{R}$  que asigna a cada  $\omega \in \Omega$  un número real tal que:

$$0 \leq P(\omega) \leq 1; \quad \sum_{\omega} P(\omega) = 1$$

Ejemplo:  $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$ ,

$$\sum_{t=1}^6 P(t) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

# Eventos, variables aleatorias y distribución de probabilidad

Un *evento*  $\mathcal{A}$  es cualquier subconjunto de  $\Omega$ ; su probabilidad es:

$$P(\mathcal{A}) = \sum_{\omega \in \mathcal{A}} P(\omega)$$

Ejemplo:  $P(1 < t < 4) = P(2) + P(3) = 1/6 + 1/6 = 1/3$

Una *variable aleatoria* es una función del espacio muestral en algún rango, como por ejemplo los booleanos ( $\mathbb{B}$ ).

Ejemplo: la función *impar* ( $I$ ).  $I : \Omega \rightarrow \mathbb{B}$ ;  $I(5) = \text{verdad}$ ,  $I(2) = \text{falso}$ .

Si  $X$  es una variable aleatoria, “ $(X = x)$ ” se usa para denotar el evento:

$$(X = x) \equiv \{\omega \in \Omega : X(\omega) = x\}$$

Dada una variable aleatoria  $X$ ,  $P$  induce una *distribución de probabilidad*:

$$P(X = x) \stackrel{\text{def}}{=} \sum_{\omega \in (X=x)} P(\omega)$$

Ejemplo:  $P(I = \text{verdad}) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$



# Eventos, variables aleatorias y proposiciones

Una proposición en lógica puede verse como un evento (subconjunto puntos muestrales) en el que la proposición es verdadera.

Dadas dos variables aleatorias booleanas  $A$  y  $B$ :

$$\text{evento } a \equiv \{\omega \in \Omega : A(\omega) = \text{verdad}\}$$

$$\text{evento } \neg a \equiv \{\omega \in \Omega : A(\omega) = \text{falso}\}$$

$$\text{evento } \neg a \wedge b \equiv \{\omega \in \Omega : A(\omega) = \text{falso} \wedge B(\omega) = \text{verdad}\}$$

Simplificaciones de notación (siempre que no se presten a confusión):

$$P(A = \text{verdad}) \rightarrow P(a), \quad P(A = \text{falso}) \rightarrow P(\neg a),$$

$$P(X = x) \rightarrow P(x)$$

## Axiomas de la Teoría de la Probabilidad

Se han propuesto varias formulaciones axiomáticas para la Teoría de la Probabilidad. Un ejemplo son los axiomas de Kolmogorov:

$$0 \leq P(\omega) \leq 1 \quad (1)$$

$$\sum_{\omega \in \Omega} P(\omega) = 1 \quad (2)$$

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \quad (3)$$

A partir de estos axiomas pueden derivarse todas las propiedades de la Teoría de la Probabilidad.

*Ejercicio:* demostrar que  $P(\neg a) = 1 - P(a)$

Como se ha comentado antes, cualquier agente cuyo sistema de valoración de la incertidumbre viole los axiomas (1-3) caerá en contradicciones y/o producirá resultados prácticos indeseables.

# Probabilidad incondicional, conjunta y condicional

*Probabilidad* (incondicional o “a priori”) de una variable aleatoria  $X$ :

$$P(X = x) \equiv P(x) : \sum_x P(x) = 1$$

*Probabilidad conjunta* de dos variables aleatorias  $X, Y$ :

$$P(X = x; Y = y) \equiv P(x, y) : \sum_x \sum_y P(x, y) = 1$$

*Probabilidad condicional:*

$$P(X = x \mid Y = y) \equiv P(x \mid y) : \sum_x P(x \mid y) = 1 \quad \forall y$$

Lo que está en la parte derecha de la barra '|' (se dice 'dado') son condiciones, y puede haber más de una.

# Variable aleatoria y probabilidad incondicional: ejemplos

Espacio muestral: Desplazamientos por carretera ( $\Omega$ ). Factores a considerar:

- *Climatología (C)*: despejado (DES), nublado (NUB), lluvia (LLU)
- *Luminosidad (L)*: día (DIA), noche (NOC)
- *Seguridad (S)*: desplazamiento seguro (SEG), o con accidente (ACC)

# Variable aleatoria y probabilidad incondicional: ejemplos

Espacio muestral: Desplazamientos por carretera ( $\Omega$ ). Factores a considerar:

- *Climatología* ( $C$ ): *despejado* (DES), *nublado* (NUB), *lluvia* (LLU)
- *Luminosidad* ( $L$ ): *día* (DIA), *noche* (NOC)
- *Seguridad* ( $S$ ): desplazamiento *seguro* (SEG), o con *accidente* (ACC)

*Variables aleatorias:*

$$C: \Omega \rightarrow \{\text{DES}, \text{NUB}, \text{LLU}\}, \quad L: \Omega \rightarrow \{\text{DIA}, \text{NOC}\}, \quad S: \Omega \rightarrow \{\text{SEG}, \text{ACC}\}.$$

Ejemplo:  $(L = \text{DIA}) \equiv \{\omega \in \Omega : L(\omega) = \text{DIA}\} \rightarrow$  desplazamientos diurnos ...

# Variable aleatoria y probabilidad incondicional: ejemplos

Espacio muestral: Desplazamientos por carretera ( $\Omega$ ). Factores a considerar:

- *Climatología* ( $C$ ): *despejado* (DES), *nublado* (NUB), *lluvia* (LLU)
- *Luminosidad* ( $L$ ): *día* (DIA), *noche* (NOC)
- *Seguridad* ( $S$ ): desplazamiento *seguro* (SEG), o con *accidente* (ACC)

*Variables aleatorias:*

$$C: \Omega \rightarrow \{\text{DES}, \text{NUB}, \text{LLU}\}, \quad L: \Omega \rightarrow \{\text{DIA}, \text{NOC}\}, \quad S: \Omega \rightarrow \{\text{SEG}, \text{ACC}\}.$$

Ejemplo:  $(L = \text{DIA}) \equiv \{\omega \in \Omega : L(\omega) = \text{DIA}\} \rightarrow$  desplazamientos diurnos ...

*Probabilidades incondicionales o a priori.*

Ejemplos:  $P(L = \text{DIA}) = 0.62$ ,  $P(L = \text{NOC}) = 0.38$

# Variable aleatoria y probabilidad incondicional: ejemplos

Espacio muestral: Desplazamientos por carretera ( $\Omega$ ). Factores a considerar:

- *Climatología* ( $C$ ): despejado (DES), nublado (NUB), lluvia (LLU)
- *Luminosidad* ( $L$ ): día (DIA), noche (NOC)
- *Seguridad* ( $S$ ): desplazamiento seguro (SEG), o con accidente (ACC)

*Variables aleatorias:*

$$C: \Omega \rightarrow \{\text{DES}, \text{NUB}, \text{LLU}\}, \quad L: \Omega \rightarrow \{\text{DIA}, \text{NOC}\}, \quad S: \Omega \rightarrow \{\text{SEG}, \text{ACC}\}.$$

Ejemplo:  $(L = \text{DIA}) \equiv \{\omega \in \Omega : L(\omega) = \text{DIA}\} \rightarrow$  desplazamientos diurnos ...

*Probabilidades incondicionales o a priori.*

Ejemplos:  $P(L = \text{DIA}) = 0.62$ ,  $P(L = \text{NOC}) = 0.38$

Más ejemplos:

$c$	DES	NUB	LLU	$\Sigma$	$l$	DIA	NOC	$\Sigma$	$s$	SEG	ACC	$\Sigma$
$P(C = c)$	0.46	0.33	0.21	1.00	$P(L = l)$	0.62	0.38	1.00	$P(S = s)$	0.86	0.14	1.00

## Probabilidad conjunta: ejemplos

*Probabilidades conjuntas.* Ejemplo:

Probabilidad de desplazamiento lloviendo y de noche  $\rightarrow P(C = \text{LLU}, L = \text{NOC}) = 0.11$

Probab. de desplaz. en día despejado y sin accidente  $\rightarrow P(C = \text{DES}, S = \text{SEG}) = 0.43$



## Probabilidad conjunta: ejemplos

*Probabilidades conjuntas. Ejemplo:*

Probabilidad de desplazamiento lloviendo y de noche  $\rightarrow P(C=\text{LLU}, L=\text{NOC}) = 0.11$

Probab. de desplaz. en día despejado y sin accidente  $\rightarrow P(C=\text{DES}, S=\text{SEG}) = 0.43$

Más ejemplos \* :

$P(c, s)$	DES	NUB	LLU		$P(l, s)$	DIA	NOC		$P(c, l)$	DES	NUB	LLU		
SEG	0.43	0.30	0.13		SEG	0.57	0.29		DIA	0.31	0.21	0.10		
ACC	0.03	0.03	0.08		ACC	0.05	0.09		NOC	0.15	0.12	0.11		
				$\Sigma=1$					$\Sigma=1$					$\Sigma=1$

	DIA			NOC			
$P(s, c, l)$	DES	NUB	LLU	DES	NUB	LLU	
SEG	0.30	0.20	0.07	0.13	0.10	0.06	
ACC	0.01	0.01	0.03	0.02	0.02	0.05	
							$\Sigma=1$

\* Todos estas probabilidades son inventadas, pero *no* arbitrarias. Los valores de estas tablas y las de las páginas anterior y siguiente están relacionados. Ver ejercicio de la página [18](#).

## Probabilidad condicional: ejemplos

Probab. de accidente *dado* desplazam. nocturno:  $P(S = \text{ACC} \mid L = \text{NOC}) = 0.24$

Probab. de *no* accidente *dado* desplazam. nocturno:  $P(S = \text{SEG} \mid L = \text{NOC}) = 0.76$

Probab. de desplazam. diurno *dado* día lluvioso:  $P(L = \text{DIA} \mid C = \text{LLU}) = 0.48$

Probab. de desplazam. nocturno *dado* día lluvioso:  $P(L = \text{NOC} \mid C = \text{LLU}) = 0.52$

## Probabilidad condicional: ejemplos

Probab. de accidente *dado* desplazam. nocturno:  $P(S = \text{ACC} \mid L = \text{NOC}) = 0.24$

Probab. de *no* accidente *dado* desplazam. nocturno:  $P(S = \text{SEG} \mid L = \text{NOC}) = 0.76$

Probab. de desplazam. diurno *dado* día lluvioso:  $P(L = \text{DIA} \mid C = \text{LLU}) = 0.48$

Probab. de desplazam. nocturno *dado* día lluvioso:  $P(L = \text{NOC} \mid C = \text{LLU}) = 0.52$

Más ejemplos:

$P(s \mid c)$	DES	NUB	LLU
SEG	0.93	0.91	0.62
ACC	0.07	0.09	0.38
$\Sigma$	1.00	1.00	1.00

$P(s \mid l)$	DIA	NOC
SEG	0.92	0.76
ACC	0.08	0.24
$\Sigma$	1.00	1.00

$P(l \mid c)$	DES	NUB	LLU
DIA	0.67	0.64	0.48
NOC	0.33	0.36	0.52
$\Sigma$	1.00	1.00	1.00

## Probabilidad condicional: ejemplos

Probab. de accidente *dado* desplazam. nocturno:  $P(S = \text{ACC} \mid L = \text{NOC}) = 0.24$

Probab. de *no* accidente *dado* desplazam. nocturno:  $P(S = \text{SEG} \mid L = \text{NOC}) = 0.76$

Probab. de desplazam. diurno *dado* día lluvioso:  $P(L = \text{DIA} \mid C = \text{LLU}) = 0.48$

Probab. de desplazam. nocturno *dado* día lluvioso:  $P(L = \text{NOC} \mid C = \text{LLU}) = 0.52$

Más ejemplos:

$P(s \mid c)$	DES	NUB	LLU
SEG	0.93	0.91	0.62
ACC	0.07	0.09	0.38
$\Sigma$	1.00	1.00	1.00

$P(s \mid l)$	DIA	NOC
SEG	0.92	0.76
ACC	0.08	0.24
$\Sigma$	1.00	1.00

$P(l \mid c)$	DES	NUB	LLU
DIA	0.67	0.64	0.48
NOC	0.33	0.36	0.52
$\Sigma$	1.00	1.00	1.00

Más ejemplos:

$P(c \mid s)$	DES	NUB	LLU	$\Sigma$
SEG	0.50	0.35	0.15	1.0
ACC	0.21	0.21	0.58	1.0

$P(l \mid s)$	DIA	NOC	$\Sigma$
SEG	0.66	0.34	1.0
ACC	0.36	0.64	1.0

$P(c \mid l)$	DES	NUB	LLU	$\Sigma$
DIA	0.50	0.34	0.16	1.0
NOC	0.39	0.32	0.29	1.0

# Variables aleatorias continuas: Densidad de Probabilidad

Si  $X$  es una variable aleatoria definida en  $\mathbb{R}$ , entonces  $P(X = x) \equiv 0 \ \forall x \in \mathbb{R}$

*Densidad de Probabilidad:* 
$$p(x) \stackrel{\text{def}}{=} \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}$$

En general,  $p(x) \in [0, \infty[$ , aunque: 
$$\int_{-\infty}^{+\infty} p(x) dx = 1$$

Todas las ecuaciones vistas anteriormente para variables discretas pueden aplicarse a variables continuas cambiando los sumatorios por integrales.

*Ejemplo:* para la *probabilidad conjunta*  $P(x, y)$ :

si  $Y$  es continua:  $\sum_x \int_y p(x, y) dy = 1$ ; si  $X, Y$  continuas:  $\int_x \int_y p(x, y) dx dy = 1$

*Ejemplo:* Para la *probabilidad condicional*  $P(x | y)$ , si  $X$  es continua:

$$\int_x p(x | y) dx = 1 \quad \forall y$$

# Índice

- 1 Introducción: Incertidumbre y probabilidad ▷ 2
- 2 Teoría de la probabilidad ▷ 5
- 3 *Razonamiento probabilístico: inferencia* ▷ 15
- 4 Incertidumbre y decisiones óptimas ▷ 19
- 5 Bibliografía ▷ 23

# Marginal, regla de la cadena y regla de Bayes

Las probabilidades incondicionales  $P(x)$ ,  $P(y)$  son las *marginales* de la probabilidad conjunta  $P(x, y)$ :

$$P(x) = \sum_y P(x, y), \quad P(y) = \sum_x P(x, y)$$

La probabilidad conjunta está relacionada con las probabilidades condicionales e incondicionales según:

$$P(x, y) = P(x) P(y | x) = P(y) P(x | y)$$

*Regla de la cadena:*

$$P(x_1, x_2, \dots, x_n) = P(x_1) \prod_{i=2}^n P(x_i | x_1, \dots, x_{i-1})$$

*Regla de Bayes:*

$$P(y | x) = \frac{P(x, y)}{P(x)} = \frac{P(y) P(x | y)}{P(x)} = \frac{P(y) P(x | y)}{\sum_{y'} P(y') P(x | y')}$$

## Inferencia: ejemplos

En el ejemplo de la página 11, podemos inferir las incondicionales de esa misma página mediante marginalización de probabilidades conjuntas de la página 12. Por ejemplo:

$$P(C = \text{DES}) = \sum_{s \in \{\text{SEG}, \text{ACC}\}} P(C = \text{DES}, S = s) = \sum_{l \in \{\text{DIA}, \text{NOC}\}} P(C = \text{DES}, L = l) = 0.46$$

$$P(S = \text{ACC}) = \sum_{c \in \{\text{DES}, \text{NUB}, \text{LLU}\}} P(C = c, S = \text{ACC}) = \sum_{l \in \{\text{DIA}, \text{NOC}\}} P(L = l, S = \text{ACC}) = 0.14$$

Y, mediante la regla de Bayes, podemos inferir las probabilidades condicionales de la página 13. Por ejemplo:

$$P(S = \text{ACC} \mid L = \text{NOC}) = \frac{P(S = \text{ACC}, L = \text{NOC})}{P(L = \text{NOC})} = 0.24$$

$$P(L = \text{DIA} \mid C = \text{LLU}) = \frac{P(C = \text{LLU}, L = \text{DIA})}{P(C = \text{LLU})} = 0.48$$

Siendo la coma equivalente a una 'y', da igual invertir el orden de los operandos en la parte de arriba, pero siempre hay que dividirlo entre la condición



## Ejercicio

La última tabla de la página 12 contiene todos los valores de la probabilidad conjunta  $P(s, c, l)$ , referente al ejemplo de la página 11.

Calcular:

- $P(S = \text{ACC}), P(c) \forall c \in \{\text{DES}, \text{LLU}\} \rightarrow 0.14$  (suma de la fila ACC)
- $P(c, s) \forall c \in \{\text{DES}, \text{NUB}, \text{LLU}\}, s \in \{\text{SEG}, \text{ACC}\}; \rightarrow 1$
- $P(s | c) \forall s \in \{\text{SEG}, \text{ACC}\}, c = \text{NUB} \rightarrow 0.33$

Comprobar que los resultados coinciden con las probabilidades que se muestran en las páginas 11, 12 y 13

$$P(S = s, C = \text{llu}, L = \text{noc}) / (P(C = \text{llu}, L = \text{noc}) = 0.06 / 0.11 = 0.545, 0.05 / 0.11 = 0.45$$

Calcular  $\forall s \in \{\text{SEG}, \text{ACC}\}$ :

- $P(S = s | C = \text{LLU}, L = \text{NOC})$
- $P(S = s | C = \text{DES}, L = \text{DIA})$

$P(s, c, l)$	DIA			NOC		
	DES	NUB	LLU	DES	NUB	LLU
SEG	0.30	0.20	0.07	0.13	0.10	0.06
ACC	0.01	0.01	0.03	0.02	0.02	0.05
$\Sigma = 1$						

*Soluciones:* 0.545, 0.455, 0.968, 0.032, respectivamente

$$\rightarrow P(S = s, C = \text{DES}, L = \text{DIA}) / P(C = \text{DES}, L = \text{DIA}) = 0.3 / 0.31 = 0.968, 0.01 / 0.31 = 0.032$$

# Índice

- 1 Introducción: Incertidumbre y probabilidad ▷ 2
- 2 Teoría de la probabilidad ▷ 5
- 3 Razonamiento probabilístico: inferencia ▷ 15
- 4 *Incertidumbre y decisiones óptimas* ▷ 19
- 5 Bibliografía ▷ 23

# Incertidumbre y decisiones óptimas: Teoría de la decisión

En el ejemplo del viaje al aeropuerto de la pág. 3, supongamos lo siguiente:

$$P(\neg \text{PERDER-EL-AVIÓN} \mid A_{25}) = 0.04$$

$$P(\neg \text{PERDER-EL-AVIÓN} \mid A_{90}) = 0.70$$

$$P(\neg \text{PERDER-EL-AVIÓN} \mid A_{120}) = 0.95$$

$$P(\neg \text{PERDER-EL-AVIÓN} \mid A_{1440}) = 0.999$$

¿Por qué acción nos decidimos?

Depende de nuestras *preferencias* sobre la posibilidad de perder el avión con respecto a disfrutar de las tiendas o la cocina del aeropuerto, etc.

La *Teoría de la Utilidad* puede usarse para representar e inferir preferencias o costes de los efectos no deseables de las decisiones.

*Teoría de la Probabilidad + Teoría de la Utilidad*

=

*Teoría de la Decisión Estadística*

# Teoría de la decisión: minimizar el riesgo de error

*Simplificación:* las decisiones solo pueden ser “acertadas” o “erróneas” y sus costes son 0 y 1, respectivamente.

Sea  $y \in \mathcal{Y}$  un *hecho* o *dato* y sea  $d \in \mathcal{D}$  una *decisión* que se toma para  $y$ .

*Probabilidad de error si se toma la decisión  $d$ :*

# Teoría de la decisión: minimizar el riesgo de error

*Simplificación:* las decisiones solo pueden ser “acertadas” o “erróneas” y sus costes son 0 y 1, respectivamente.

Sea  $y \in \mathcal{Y}$  un *hecho* o *dato* y sea  $d \in \mathcal{D}$  una *decisión* que se toma para  $y$ .

*Probabilidad de error si se toma la decisión  $d$ :*

$$P_d(\text{error} \mid y) = 1 - P(d \mid y)$$

*Mínima probabilidad de error:*

$$\forall y \in \mathcal{Y} : P_{\star}(\text{error} \mid y) = \min_{d \in \mathcal{D}} P_d(\text{error} \mid y) = 1 - \max_{d \in \mathcal{D}} P(d \mid y)$$

Es decir, para cada  $y$  la probabilidad de error se minimiza si se toma la decisión con mayor probabilidad a posteriori.

# Teoría de la decisión: minimizar el riesgo de error

*Simplificación:* las decisiones solo pueden ser “acertadas” o “erróneas” y sus costes son 0 y 1, respectivamente.

Sea  $y \in \mathcal{Y}$  un *hecho* o *dato* y sea  $d \in \mathcal{D}$  una *decisión* que se toma para  $y$ .

*Probabilidad de error si se toma la decisión  $d$ :*

$$P_d(\text{error} \mid y) = 1 - P(d \mid y)$$

*Mínima probabilidad de error:*

$$\forall y \in \mathcal{Y} : P_{\star}(\text{error} \mid y) = \min_{d \in \mathcal{D}} P_d(\text{error} \mid y) = 1 - \max_{d \in \mathcal{D}} P(d \mid y)$$

Es decir, para cada  $y$  la probabilidad de error se minimiza si se toma la decisión con mayor probabilidad a posteriori.

*Mínimo riesgo global o mínima esperanza de error de decisión:*

$$P_{\star}(\text{error}) = \sum_{y \in \mathcal{Y}} P_{\star}(\text{error} \mid y) P(y)$$

*Regla de decisión de mínimo riesgo de error o de Bayes:*

$$\forall y \in \mathcal{Y} : d^{\star}(y) = \operatorname{argmax}_{d \in \mathcal{D}} P(d \mid y)$$

## Ejercicio (para hacer en clase)

Un problema clásico de decisión consiste en clasificar flores de la familia *Iris* en tres clases; *setosa*, *versicolor* y *virginica*, en base a los tamaños de sus pétalos y sépalos ( $y$ ).

Para ello se han calculado sendos histogramas de las superficies de los pétalos de una muestra de 50 flores de cada clase. Normalizando estos histogramas, se ha estimado la siguiente distribución de tamaños de pétalos para cada clase ( $c$ ):

$P(y   c)$	tamaño de los pétalos en $\text{cm}^2$											
	<1	1	2	3	4	5	6	7	8	9	10	>10
SETO	0.90	0.10	0	0	0	0	0	0	0	0	0	0
VERS	0	0	0	0.20	0.30	0.32	0.12	0.06	0	0	0	0
VIRG	0	0	0	0	0	0	0.08	0.12	0.24	0.14	0.20	0.22

Asumiendo que las clases son equiprobables, calcular:

- Las probabilidades a posteriori  $P(c | y)$ ,  $c \in \{\text{SETO}, \text{VERS}, \text{VIRG}\}$ , para una flor cuyo tamaño de pétalos es  $y = 7 \text{ cm}^2$
- La decisión óptima de clasificación de esta flor y la probabilidad de que dicha decisión sea errónea
- La mejor decisión y la correspondiente probab. de error para tamaños de pétalos  $1, 2, \dots, 10 \text{ cm}^2$
- La mínima probabilidad de error de decisión esperada para cualquier flor Iris; es decir,  $P_*(\text{error})$
- Repetir los calculos anteriores, asumiendo que las probabilidades a priori son:

$$P(\text{SETO}) = 0.3, P(\text{VERS}) = 0.5, P(\text{VIRG}) = 0.2$$

Algunas soluciones: a) 0.0, 0.33, 0.67; b) VIRG, 0.33; d) 0.05 (5 %) e.a) 0.0, 0.55, 0.44; e.b) VERS, 0.44; e.d) 0.04 (4 %)

# Índice

- 1 Introducción: Incertidumbre y probabilidad ▷ 2
- 2 Teoría de la probabilidad ▷ 5
- 3 Razonamiento probabilístico: inferencia ▷ 15
- 4 Incertidumbre y decisiones óptimas ▷ 19
- 5 *Bibliografía* ▷ 23



## Bibliografía

- [1] A.N. Abdallah. The Logic of Partial Information. Springer Verlag, 1995.
- [2] B.G. Buchanan, E.H. Shortliffe (editores): Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison Wesley, 1984. (También en <http://aitopics.net/RuleBasedExpertSystems>).
- [3] J.F. Baldwin. Fuzzy sets and expert systems. Wiley, 1985.
- [4] R.O. Duda, D.G. Stork, P.E. Hart. Pattern Classification. Wiley, 2001.
- [5] S. Russell, P. Norvig. Artificial Intelligence: A Modern Approach. Pearson, third edition, 2010.

El material de este tema se basa principalmente en [5].