

# Herramientas de probabilidad

Verónica E. Arriola-Rios

Inteligencia Artificial

31 de mayo de 2020

# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades
- 4 Factores
- 5 Distribuciones de probabilidad con factores
- 6 Cómo programar factores

# Experimento de azar

## Definición

Un *experimento causal, indeterminista o de azar* es aquel para el cual no necesariamente podemos predecir con certeza lo que va a ocurrir al realizarlo. Cuando se repite el mismo experimento bajo las mismas condiciones se puede obtener un resultado diferente.

*Ej: “predecir el resultado de lanzar dos monedas.”*



# Eventos

En un experimento aleatorio:

- *Evento simple*. Cualquier resultado elemental.

*Ej: a: "obtener el 2 al lanzar un dado."*

- *Evento*. Cualquier conjunto de resultados posibles.

*Ej: A: "obtener un par al lanzar un dado."*

$A = \{2, 4, 6\} = \{ \text{"obtengo el dos", "o... cuatro", "o... seis"} \} = \{x|x = 2, 4, 6\}$

- *Espacio de eventos, muestras o Universo S*. Conjunto de todos los eventos simples posibles.

*Ej: "lanzar un dado."  $S = \{1, 2, 3, 4, 5, 6\}$*

- *Tamaño de un evento*. Número de eventos simples que satisfacen la definición del evento (cardinalidad del conjunto).

*Ej: tamaño de A:  $|A| = 3$ .*

# Definición de probabilidad

## Definición

Definimos a la probabilidad de un evento  $A$  como:

$$P(A) = \frac{|A|}{|S|}. \quad (1)$$

Destaca también lo que se conoce como la *definición frecuentista* de probabilidad, en la cual esta cantidad  $P(A)$  se define con respecto a una secuencia potencialmente infinita de repeticiones del experimento aleatorio  $A$ .

# Variable aleatoria

## Definición

Una variable aleatoria es una función que va del espacio de muestras a los reales  $X : S \rightarrow \mathbb{R}^a$ .

<sup>a</sup>Por convención, escribiremos las variables aleatorias con mayúsculas y sus valores concretos con minúsculas.

*Ej: Para el evento  $E$ : “obtener un 7 al lanzar dos dados”.  
Sea la variable aleatoria  $X$ : “La suma del resultado en cada uno de dos dados”.*

$$X(\text{dado}_1, \text{dado}_2) = \text{dado}_1 + \text{dado}_2 \quad (2)$$

*Entonces el evento  $E$  se expresa como*

$$E : X(\text{dado}_1, \text{dado}_2) = 7 \quad (3)$$

# Casos particulares de variable aleatoria

- Función identidad  $X: \mathbb{R} \rightarrow \mathbb{R}$ .
- Un mapeo uno a uno del espacio de eventos hacia los reales, cuando los valores del dominio son:
  - 1 Mutuamente exclusivos.
  - 2 Exhaustivos.

*Ej: La variable aleatoria  $C(\text{Clima})$  tendría el rango:*

$C(\text{Clima})$

$= C(< \text{despejado, lluvioso, nublado, granizado, nevado} >)$

$= < 1, 2, 3, 4, 5 >$

*Esto también se puede escribir  $C \in \{c^1, c^2, c^3, c^4, c^5\}$ , que abrevia los casos  $C = 1, C = 2, C = 3, C = 4, C = 5$ .*

- Por comodidad y legibilidad, frecuentemente se hará referencia a la variable del espacio de eventos como una *variable aleatoria*.

# Extensión de la lógica proposicional

- Es posible utilizar la teoría de probabilidades para extender a la lógica proposicional.
- Se utilizan proposiciones que afirman que una variable aleatoria tiene un valor particular tomado de su dominio:

*Ej:* "Temperatura = 310".

- Se asocia una probabilidad al evento de que la proposición dada sea verdadera o falsa.



# Axiomas de la probabilidad

Andrei Kolmogorov demostró cómo desarrollar el resto de la teoría probabilista a partir de los tres axiomas que llevan su nombre<sup>[1]</sup>:

- 1 Todas las probabilidades están entre 0 y 1. Para cualquier proposición  $\alpha$ ,

$$0 \leq P(\alpha) \leq 1 \quad (4)$$

- 2 Las proposiciones necesariamente ciertas (es decir, válidas) tienen probabilidad 1, y las proposiciones necesariamente falsas (es decir, insatisfacibles) tienen probabilidad 0.

$$P(\text{cierto}) = 1 \quad P(\text{falso}) = 0 \quad (5)$$

- 3 La probabilidad de una disyunción viene dada por

$$P(\alpha \vee b) = P(\alpha) + P(b) - P(\alpha \wedge b) \quad (6)$$

<sup>[1]</sup>Russell y Norving 2004

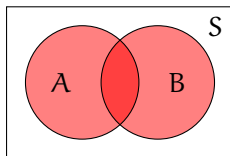
## Tercer axioma con conjuntos

- Se puede interpretar el axioma:

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b) \quad (7)$$

- utilizando teoría de conjuntos:

$$\begin{aligned} P(a \vee b) &= \frac{|a \cup b|}{|S|} = \frac{|a| + |b| - |a \cap b|}{|S|} \\ &= P(a) + P(b) - P(a \wedge b). \end{aligned}$$

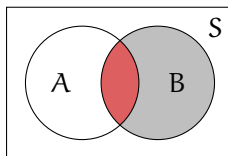


**Figura:** La probabilidad  $P(a \vee b)$  corresponde a la probabilidad de la unión de ambos eventos.

# Probabilidad condicional

## Definición

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \quad (8)$$



**Figura:** La probabilidad  $P(a|b)$  corresponde a la probabilidad de la intersección de ambos eventos, como si B fuera el nuevo universo.

# Teorema de Bayes

- Tomando la definición de probabilidad condicional:

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

- Despejando:

$$P(a \wedge b) = P(b)P(a|b) = P(a)P(b|a)$$

- Se obtiene el teorema de bayes.

## Definición

$$P(b|a) = \frac{P(a|b)P(b)}{P(a)} \qquad (9)$$

# A priori, a posteriori y verosimilitud

- Reescribiendo el teorema de Bayes como:

$$p(\theta|e) = \frac{p(e|\theta)p(\theta)}{p(e)} \quad (10)$$

- Leemos cada variable como sigue:

- $p(\theta)$ : *Creencia a priori (prior)*. Probabilidad de que ocurra el evento  $\theta$  en ausencia de evidencia.
- $p(e)$ : probabilidad de encontrar la evidencia por sí misma.
- $p(e|\theta)$ : *Verosimilitud (likelihood)*. Probabilidad de observar la evidencia dado un evento que la causa.
- $p(\theta|e)$ : *Creencia a posteriori (posterior)*. Probabilidad de que ocurra el evento  $\theta$  dado que se posee información sobre su evidencia.

Esta es la cantidad que se desea calcular cuando se quiere realizar un *diagnóstico*.

# Ejemplo

- Supongamos que un médico quiere saber si un niño tiene varicela, dado que observa manchas en la piel. Sin embargo, recordemos que también pudiera se sarampión.
- Se busca entonces:

$$p(\text{varicela}|\text{machas}) = \frac{p(\text{manchas}|\text{varicela})p(\text{varicela})}{p(\text{manchas})} \quad (11)$$

- $p(\text{varicela})$ : Probabilidad de que cualquier persona tenga varicela.
- $p(\text{manchas})$ : probabilidad que una persona tenga manchas en la piel por cualquier causa.
- $p(\text{manchas}|\text{varicela})$ : probabilidad de que una persona enferma de varicela tenga manchas en la piel.
- $p(\text{varicela}|\text{manchas})$ : probabilidad de que la persona tenga varicela, dado que tiene manchas en la piel.

# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades
- 4 Factores
- 5 Distribuciones de probabilidad con factores
- 6 Cómo programar factores

# Distribución de probabilidad

Una *distribución de probabilidad* asocia a cada valor posible de una variable aleatoria la probabilidad de que se obtenga ese valor al realizar un experimento. Por ejemplo:

Lluvia	P(Lluvia)
0	0.6375
1	0.3625
$\sum = 1$	



# Distribución de probabilidad conjunta

Una *distribución de probabilidad conjunta* asocia a cada combinación posible de los valores de varias variables que se revisan simultáneamente, la probabilidad de que se obtenga esa combinación al realizar un experimento. Por ejemplo:

Estación	Lluvia	$P(\text{Lluvia} \wedge \text{Estación})$
Primavera	0	0.1875
Primavera	1	0.0625
Verano	0	0.075
Verano	1	0.175
Otoño	0	0.175
Otoño	1	0.075
Invierno	0	0.2
Invierno	1	0.05
		$\Sigma = 1$

# Distribución de probabilidad conjunta completa

Si las variables siendo consideradas, son todas las variables del sistema, entonces se dice que se tiene la *distribución de probabilidad conjunta completa*.

# Distribución de probabilidad condicional

La *distribución de probabilidad condicional* asocia una probabilidad a cada combinación de valores de variables, dado un valor determinado para un conjunto de variables evidencia.

*Por ejemplo: “¿Cuál es la probabilidad de que llueva, si es primavera?” y la respuesta es 0.25.*

Lluvia / Estación	Primavera	Verano	Otoño	Invierno
0	0.75	0.30	0.70	0.80
1	0.25	0.70	0.30	0.20
	$\sum = 1$	$\sum = 1$	$\sum = 1$	$\sum = 1$

# La regla de la cadena

- Definición de probabilidad condicional:

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

$$\Rightarrow P(a, b) = P(a|b)P(b)$$

- Definición de probabilidad condicional para más variables:

$$P(X_1, \dots, X_{n_1} | Y_1, \dots, Y_{n_2}) = \frac{P(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})}{P(Y_1, \dots, Y_{n_2})}$$

$$\Rightarrow P(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}) = P(X_1, \dots, X_{n_1} | Y_1, \dots, Y_{n_2})P(Y_1, \dots, Y_{n_2})$$

- Aplicando lo anterior recursivamente podemos derivar:

$$P(X_1, X_2, \dots, X_n) = P(X_1 | X_2, \dots, X_n) P(X_2 | X_3, \dots, X_n) \dots P(X_n) \quad (12)$$

# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades**
- 4 Factores
- 5 Distribuciones de probabilidad con factores
- 6 Cómo programar factores

# Motivación

*Si podemos escribir las funciones lógicas  $\wedge$ ,  $\vee$  y  $\neg$ , podemos escribir cualquier función booleana.*

# Usando distribuciones de probabilidad

- *No*

$\neg$		
x	y	$P(y x)$
0	0	0
0	1	1
1	0	1
1	1	0

- *Y*

$\wedge$			
$x_1$	$x_2$	y	$P(y x_1, x_2)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

# Compuertas lógicas con ruido

Es posible modelar compuertas ruidosas, que no siempre devuelven la respuesta correcta, asociando una pequeña probabilidad  $\varepsilon \neq 0$  a las respuestas incorrectas y ajustando acordemente las correctas  
Koller y Friedman 2009:

- No

$\neg$		
x	y	$P(y x)$
0	0	$\varepsilon$
0	1	$1 - \varepsilon$
1	0	$1 - \varepsilon$
1	1	$\varepsilon$

- Y

$\wedge$			
$x_1$	$x_2$	y	$P(y x_1, x_2)$
0	0	0	$1 - \varepsilon$
0	0	1	$\varepsilon$
0	1	0	$1 - \varepsilon$
0	1	1	$\varepsilon$
1	0	0	$1 - \varepsilon$
1	0	1	$\varepsilon$
1	1	0	$\varepsilon$
1	1	1	$1 - \varepsilon$



# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades
- 4 Factores**
- 5 Distribuciones de probabilidad con factores
- 6 Cómo programar factores

# Factor

## Definición (Factor)

Un factor  $\phi(X_1, \dots, X_k)$ , sobre variables  $X_i$ , donde cada  $X_i$  puede tomar valores de un dominio  $D_{X_i}$  es una función:

$$\phi : \text{Val}(X_1, \dots, X_k) \rightarrow \mathbb{R} \quad (13)$$

que asocia un número real a posibles asignaciones de valores a las variables  $X_1, \dots, X_k$ . El *alcance*  $\mathbb{A}$  de un factor son las variables cuyos posibles valores están siendo considerados.

$$\mathbb{A}(\phi(X_1, \dots, X_k)) = \{X_1, \dots, X_k\} \quad (14)$$

# Ejemplo

$\mathbb{A} = \{\text{Estación, Lluvia}\}$

Estación	Lluvia	Frecuencia (días/mes)
Primavera	0	18
Primavera	1	6
Verano	0	7
Verano	1	17
Otoño	0	17
Otoño	1	7
Invierno	0	20
Invierno	1	5

# Marginalización

## Definición (Marginalización)

Dada una variable  $X_i$ , en el alcance  $\mathbb{A}$  de un factor  $\phi = \phi(X_1, \dots, X_k)$ , que se desea marginalizar, se define a la operación como:

$$\text{marginalización}(\phi(X_1, \dots, X_k), X_i) = \phi'(\mathbb{A}')$$
(15)

$$\mathbb{A}' = \mathbb{A}(\phi) - X_i \text{ con } i \in [1, k]$$
(16)

$$\phi'(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) = \sum_{\text{Val}\{X_i\}} \phi(x_1, \dots, x_k)$$
(17)

donde  $x_i$  es un valor particular de  $X_i$ ,  $\phi'(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$  es el renglón del factor  $\phi'$  con  $\{X_1 = x_1, \dots, X_k = x_k\}$  y  $\text{Val}\{X_i\}$  es el conjunto de valores posibles asignables a  $X_i$ .

# Ejemplo: Marginalizar Estación

Estación	Lluvia	P(Lluvia, Estación)
Primavera	0	0.1875
Primavera	1	0.0625
Verano	0	0.075
Verano	1	0.175
Otoño	0	0.175
Otoño	1	0.075
Invierno	0	0.2
Invierno	1	0.05

$$\sum = 1$$

	Lluvia	P(Lluvia)
⇒	0	0.63750
	1	0.36250

$$\sum = 1$$

# Reducción

## Definición (Reducción)

Dado un valor  $x_i = \alpha$  para una de las variables  $X_i$  en el alcance  $\mathbb{A}$  del factor  $\phi$ , se reduce el factor eliminando todas aquellas entradas en las cuales no se cumple que  $X_i = \alpha$ . La operación se define como:

$$\text{reducción}(\phi(X_1, \dots, X_k), X_i, \alpha) = \phi'(\mathbb{A}') \quad (18)$$

$$\mathbb{A}' = \mathbb{A}(\phi) - X_i \text{ con } i \in [1, k] \quad (19)$$

$$\phi'(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k) = \phi(x_1, \dots, x_k) \text{ con } X_i = \alpha \quad (20)$$

A diferencia de las distribuciones de probabilidad, la reducción en un factor no requiere renormalizar sus valores asociados, pues por definición, no es necesario que éstos sumen uno.

# Ejemplo: Reducir Estación = Primavera

Estación	Lluvia	P(Lluvia, Estación)
Primavera	0	0.1875
Primavera	1	0.0625
Verano	0	0.075
Verano	1	0.175
Otoño	0	0.175
Otoño	1	0.075
Invierno	0	0.2
Invierno	1	0.05
		$\Sigma = 1$

Lluvia	P(Lluvia, Estación = primavera)
0	0.1875
1	0.0625
	$\sum = 0.25$

⇒

# Normalización

## Definición (Normalización)

Dado un factor  $\phi$  con  $n$  renglones sea

$$s = \sum_{i=1}^n \phi_i \quad (21)$$

con  $\phi_i$  el valor asociado al renglón  $i$ ,  $s$  es la suma de los valores de todos los renglones. Entonces:

$$\text{normalización}(\phi) = \phi' \quad (22)$$

$$\phi'_i = \frac{\phi_i}{s} \quad (23)$$

donde el valor en cada renglón de  $\phi$  ha sido dividido entre  $s$ .



# Ejemplo: Normalizar

Lluvia	P(Lluvia, primavera)
0	0.1875 / 0.25
1	0.0625 / 0.25
$\sum = 0.25$	

⇒

Lluvia	P(Lluvia primavera)
0	0.75
1	0.25
$\sum = 1$	

# Multiplicación

## Definición (Multiplicación)

Se define el producto de factores de tal modo que:

$$\phi_1 = \phi_1(X_1, \dots, X_k) \quad \phi_2 = \phi_2(Y_1, \dots, Y_l) \quad (24)$$

$$\phi = \phi_1 \phi_2 \quad (25)$$

$$\mathbb{A}(\phi) = \mathbb{A}(\phi_1) \cup \mathbb{A}(\phi_2)$$

$$\phi(z_1, \dots, z_m) = \phi_1(x_1, \dots, x_k) * \phi_2(y_1, \dots, y_l)$$

donde  $Z_k \in \mathbb{A}(\phi)$  por lo que:

$$\text{si } Z_k \in \mathbb{A}(\phi_1) \Rightarrow Z_k = X_i \wedge z_k = x_i \quad (26)$$

$$\text{si } Z_k \in \mathbb{A}(\phi_2) \Rightarrow Z_k = Y_j \wedge z_k = y_j$$

obsérvese que si  $Z_k \in \mathbb{A}(\phi_1)$  y  $Z_k \in \mathbb{A}(\phi_2)$  entonces  $Z_k = X_i = Y_j$  y en cada renglón deberá cumplirse  $z_k = x_i = y_j$ .

# Ejemplo 1: Multiplicar $P(\text{Lluvia}|\text{Estación})P(\text{Estación})$

Estación	Lluvia	$P(\text{Lluvia} \text{Estación})$
Primavera	0	0.75
Primavera	1	0.25
Verano	0	0.30
Verano	1	0.70
Otoño	0	0.70
Otoño	1	0.30
Invierno	0	0.80
Invierno	1	0.20

×

Estación	$P(\text{Estación})$
Primavera	0.25
Verano	0.25
Otoño	0.25
Invierno	0.25

 $\Sigma = 1$ 

Estación	Lluvia	$P(\text{Lluvia}, \text{Estación})$
Primavera	0	0.1875
Primavera	1	0.0625
Verano	0	0.075
Verano	1	0.175
Otoño	0	0.175
Otoño	1	0.075
Invierno	0	0.2
Invierno	1	0.05

 $\Sigma = 1$

## Ejemplo 2: Multiplicar $P(\text{Lluvia})P(\text{Estación})$

Lluvia	$P(\text{Lluvia})$	×	Estación	$P(\text{Estación})$	=
0	0.63750		Primavera	0.25	
1	0.36250		Verano	0.25	
	$\sum = 1$		Otoño	0.25	
			Invierno	0.25	
				$\sum = 1$	

Sea  $R = P(\text{Lluvia})P(\text{Estación})$

Estación	Lluvia	R	Estación	Lluvia	R
Primavera	0	0.159375	Primavera	0	0.159375
Primavera	1	0.090625	Verano	0	0.159375
Verano	0	0.159375	Otoño	0	0.159375
Verano	1	0.090625	Invierno	0	0.159375
Otoño	0	0.159375	Primavera	1	0.090625
Otoño	1	0.090625	Verano	1	0.090625
Invierno	0	0.159375	Otoño	1	0.090625
Invierno	1	0.090625	Invierno	1	0.090625
	$\sum = 1$			$\sum = 1$	

# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades
- 4 Factores
- 5 Distribuciones de probabilidad con factores**
- 6 Cómo programar factores

# Distribuciones de probabilidad con factores

- Claramente, los factores pueden contener, en particular, tablas de distribuciones de probabilidad sobre variables discretas.
- Las variables aleatorias del sistema aparecerán en el alcance del factor,
- las combinaciones posibles de asignaciones a estas variables quedarán registradas en los renglones del factor y
- la probabilidad de que ocurran será el número real asociado a esa entrada.

# Distribuciones de probabilidad condicional con factores

Estación	Lluvia	P(Lluvia Estación)	
Primavera	0	0.75	$\sum = 1$
Primavera	1	0.25	
Verano	0	0.30	$\sum = 1$
Verano	1	0.60	
Otoño	0	0.70	$\sum = 1$
Otoño	1	0.30	
Invierno	0	0.8	$\sum = 1$
Invierno	1	0.1	

# Temas

- 1 Antecedentes
- 2 Distribuciones de probabilidad
- 3 Compuertas lógicas con probabilidades
- 4 Factores
- 5 Distribuciones de probabilidad con factores
- 6 Cómo programar factores**



# Estructura

- Crear el clase *Variable*, con:
  - Nombre
  - Lista de valores posibles.  
Esta lista define un orden.
- Crear clase *Factor*, con:
  - Alcance. Una tupla de variables, esta tupla define un orden.
  - Valores  $\phi$ . Una lista con los valores correspondientes a cada renglón.

$$\mathbb{A} = \{\text{Estación, Lluvia}\}$$

Estación	Lluvia	Frecuencia (días/mes)
Primavera	0	18
Primavera	1	6
Verano	0	7
Verano	1	17
Otoño	0	17
Otoño	1	7
Invierno	0	20
Invierno	1	5

# Ejemplo

## Variables

- $E =$   
`Variable('Estación',  
 ['Primavera',  
 'Verano', 'Otoño',  
 'Invierno'])`
- $LL =$   
`Variable('Lluvia',  
 [0,1])`

## Factores

- $f = \text{Factor}((E, LL),$   
`[18,6,7,17,`  
`17,7,20,5])`

$$\mathbb{A} = \{\text{Estación}, \text{Lluvia}\}$$

Estación	Lluvia	Frecuencia (días/mes)
Primavera	0	18
Primavera	1	6
Verano	0	7
Verano	1	17
Otoño	0	17
Otoño	1	7
Invierno	0	20
Invierno	1	5

# Cómo encontrar un renglón

- Se utiliza un polinomio de direccionamiento:

$$i = (((i_0) * t_1 + i_1) * t_2 + i_2) * \dots * t_{n-1} + i_{n-1} \quad (27)$$

- Para determinar los valores de  $i_k$  y  $t_k$  usamos:
  - Los índices  $k$  son la posición de la variable en  $\mathbb{A}$ .
  - $t_k$  es el tamaño del dominio de la  $k$ -ésima variable.
  - $i_k$  es el índice del valor de la  $k$ -variable en su lista-dominio.

	Letra	Estación	Lluvia	Frecuencia (días/mes)
0	a	Primavera	0	18
1	a	Primavera	1	6
2	a	Verano	0	7
3	a	Verano	1	17
4	a	Otoño	0	17
...				
	b			
...				
	c			

$$i(\text{Letra} = b, \text{Estación} = \text{Otoño}, \text{Lluvia} = 0) = ((1) * 4 + 2) * 2 + 0$$

# Tips para implementar las operaciones

- ❶ Definir el dominio del factor resultante.
- ❷ Utilizar el polinomio de direccionamiento para encontrar el renglón del factor  $\phi$  donde las variables tienen los valores indicados.
  - Se recomienda tener una función auxiliar, que reciba un diccionario, y sólo lea los valores relevantes a las variables en su alcance.
  - A partir de esos valores, calcular sus índices según las listas con que se definieron los objetos.
  - Usar los índices en el polinomio para obter el índice del valor del factor.
- ❸ Iterar sobre los renglones del factor resultado, para calcular, según la definición de la operación correspondiente, el valor nuevo correspondiente.

# Referencias I



Koller, Daphne y Nir Friedman (2009). *Probabilistic Graphical Models, Principles and Techniques*. MIT Press Cambridge.



Russell, Stuart y Peter Norving (2004). *Inteligencia Artificial, Un Enfoque Moderno*. 2a. Pearson Prentice Hall.