Inteligencia Artificial

Conocimientos y razonamientos bajo incertidumbre

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

Parte II

Razonamiento Probabilístico

En el que explicamos cómo construir modelos de redes para razonar bajo incertidumbre de acuerdo con las leyes de la teoría de la probabilidad.

Intelligent Robotics Lab

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

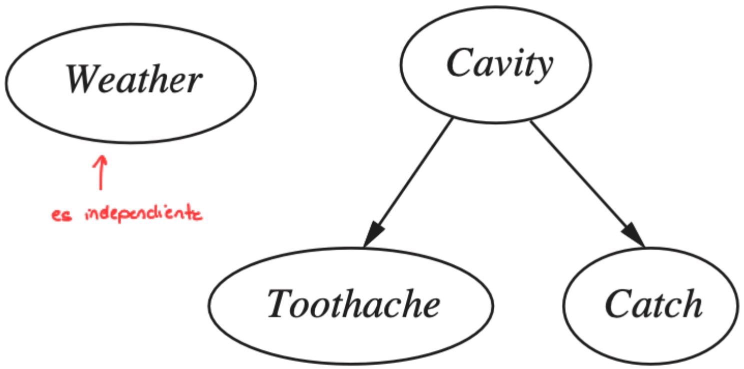
Representando incertidumbre en dominios inciertos

- La distribución de probabilidad conjunta completa puede responder a cualquier pregunta sobre el dominio, pero:
  - Puede volverse increíblemente grande a medida que aumenta el número de variables.
  - Especificar las probabilidades de los mundos posibles uno por uno es antinatural y tedioso.

- Las relaciones de independencia e independencia condicional entre variables pueden reducir en gran medida el número de probabilidades que deben especificarse para definir la distribución conjunta completa.

- Las **redes bayesianas** pueden representar esencialmente cualquier distribución de probabilidad conjunta completa de manera muy concisa.

- Grafo acíclico dirigido donde cada nodo está anotado con información de probabilidad cuantitativa
  - Cada nodo es una variable aleatoria
  - Si hay un arco de A →B, A es el padre de B
  - Cada nodo X_i tiene una distribución de probabilidad condicional $\mathbf{P}(X_i \mid parents(X_i))$ que cuantifica el efecto de los padres en el nodo. CPTs (Conditional Probability Table)



es independiente

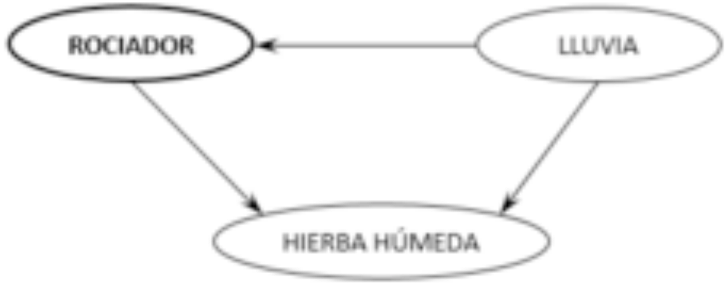
dolor de muelas depende de cavity

la probabilidad del nodo depende del padre.
y nos ahorramos saber de los "primos", "sobrinos"...

Ejemplo: Un sistema de riego

G = Hierba húmeda, *S = Rociador activado*, y *R = Lloviendo*

	ROCIADOR	
LLUVIA	T	F
F	0.4	0.6
T	0.01	0.99



LLUVIA	
T	F
0.2	0.8

Conditional Probability Table (CPT)

		HIERBA HÚMEDA	
ROCIADOR	LLUVIA	T	F
F	F	0	1
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Cada fila es un **caso condicional** (debe sumar 1)

Es una combinación de valores de los padres

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) \qquad P(x_1, \ldots, x_n) = \prod_{i=1}^{n} \theta(x_i \,|\, parents(X_i))$$

Ejemplo: SE ACOTA EL PROBLEMA: padre.

De abajo hacia arriba

$$p(r, s, \neg g) = P(\neg g \,|\, s \wedge r)P(s \,|\, r)P(r) = 0.01 \times 0.01 \times 0.2 = 0.00002$$

$$p(r, \neg s, g) = P(g \,|\, \neg s \wedge r)P(\neg s \,|\, r)P(r) = 0.8 \times 0.99 \times 0.2 = 0.15$$

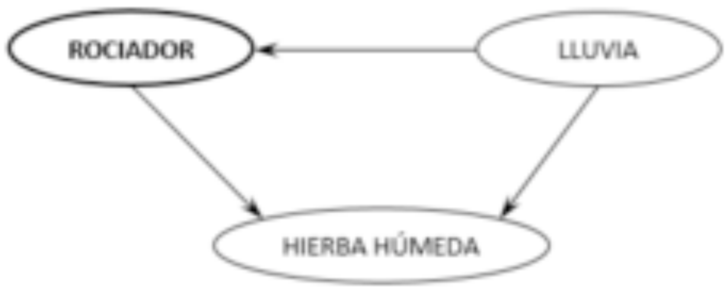
$$p(\neg r, \neg s, \neg g) = P(\neg g \,|\, \neg s \wedge \neg r)P(\neg s \,|\, \neg r)P(\neg r) = 1 \times 0.6 \times 0.8 = 0.48$$

Intelligent Robotics Lab

Ejemplo: Un sistema de riego

G = Hierba húmeda, S = Rociador activado, y R = Lloviendo

	ROCIADOR	
LLUVIA	T	F
F	0.4	0.6
T	0.01	0.99



LLUVIA	
T	F
0.2	0.8

		HIERBA HÚMEDA	
ROCIADOR	LLUVIA	T	F
F	F	0	1
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

- ¿Cuál es la probabilidad de que esté lloviendo dado que la hierba está húmeda?

En forma de redes bayesianas

$$\mathrm{P}(R=T \mid G=T) = \frac{\mathrm{P}(G=T, R=T)}{\mathrm{P}(G=T)} = \frac{\sum_{S\in\{T,F\}} \mathrm{P}(G=T, S, R=T)}{\sum_{S,R\in\{T,F\}} \mathrm{P}(G=T, S, R)}$$

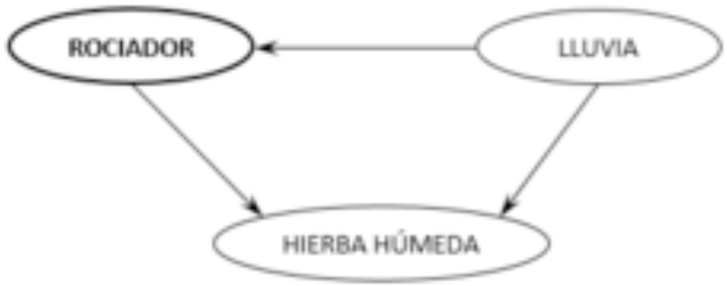
$$= \frac{(0.99 \times 0.01 \times 0.2 = 0.00198_{TTT}) + (0.8 \times 0.99 \times 0.2 = 0.1584_{TFT})}{0.00198_{TTT} + 0.288_{TTF} + 0.1584_{TFT} + 0_{TFF}} \approx 35.77\%.$$

Intelligent Robotics Lab

Ejemplo: Un sistema de riego

G = Hierba húmeda, S = Rociador activado, y R = Lloviendo

LLUVIA	ROCIADOR T	ROCIADOR F
F	0.4	0.6
T	0.01	0.99



LLUVIA T	LLUVIA F
0.2	0.8

ROCIADOR	LLUVIA	HIERBA HÚMEDA T	HIERBA HÚMEDA F
F	F	0	1
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

- ¿Cuál es la probabilidad de que esté lloviendo dado que la hierba está húmeda?

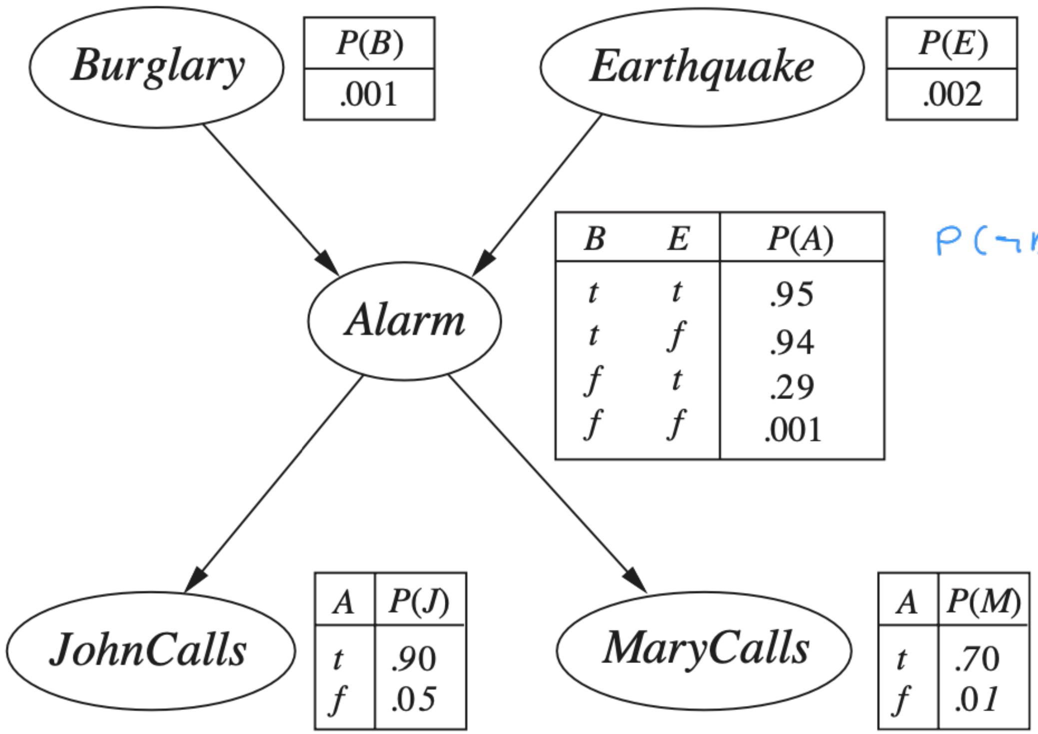
$$\mathrm{P}(R=T \mid G=T) = \frac{\mathrm{P}(G=T, R=T)}{\mathrm{P}(G=T)} = \frac{\sum_{S\in\{T,F\}} \mathrm{P}(G=T, S, R=T)}{\sum_{S,R\in\{T,F\}} \mathrm{P}(G=T, S, R)}$$

$$= \frac{(0.99 \times 0.01 \times 0.2 = 0.00198_{TTT}) + (0.8 \times 0.99 \times 0.2 = 0.1584_{TFT})}{0.00198_{TTT} + 0.288_{TTF} + 0.1584_{TFT} + 0_{TFF}} \approx 35.77\%.$$

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

Ejemplo: Un sistema de alarma

Red Bayesiana típica, mostrando la topología y las Tablas de Probabilidad Condicional (TPC).



$P(B)$
.001

$P(E)$
.002

B	E	$P(A)$
t	t	.95
t	f	.94
f	t	.29
f	f	.001

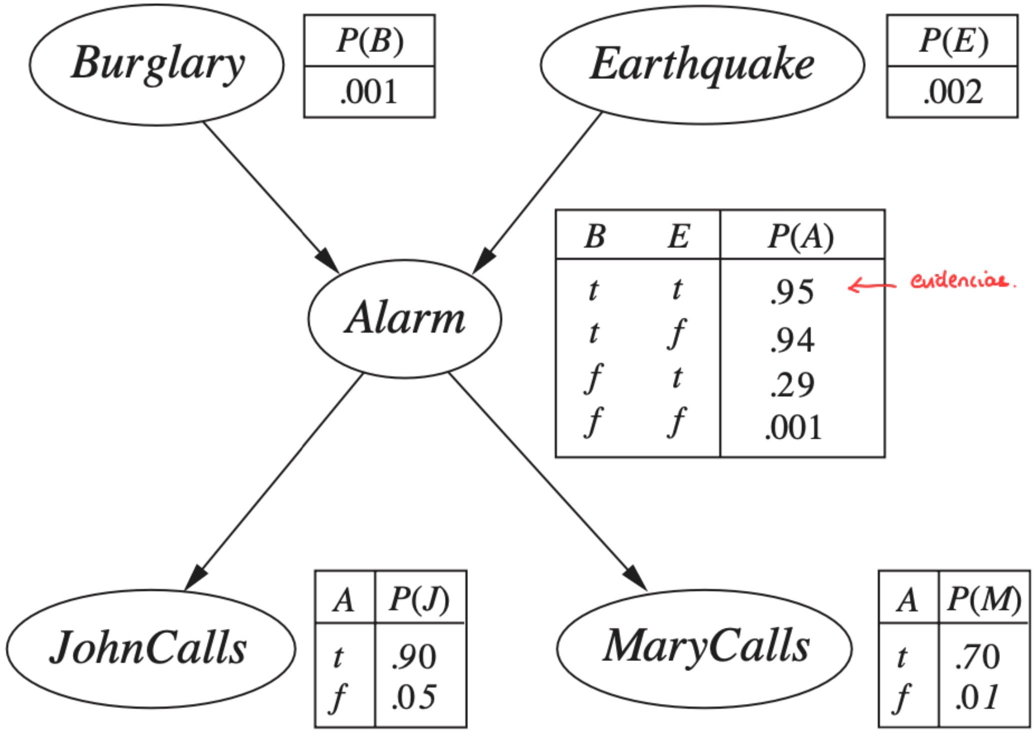
$P(\neg A) = 1 - P(A)$

A	$P(J)$
t	.90
f	.05

A	$P(M)$
t	.70
f	.01

$$\begin{aligned} P(j, m, a, \neg b, \neg e) &= P(j \mid a) P(m \mid a) P(a \mid \neg b \wedge \neg e) P(\neg b) P(\neg e) \\ &= 0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 = 0.000628 \end{aligned}$$

Ejemplo: Un sistema de alarma



- ¿Cuál es la probabilidad de que John llame? $P(j|a) + P(j|\neg a)$
- ¿Cuál es la probabilidad de que la alarma suene y Mary y John llamen sin que haya habido un robo o un terremoto?

$P(a, m, j, \neg r, \neg e)$

Intelligent Robotics Lab

Método para construir de Redes Bayesianas

1. Reescribe las entradas en la distribución conjunta en términos de probabilidad condicional, usando la **regla del producto**

$$\boxed{P(x_1, \ldots, x_n) = P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1}, \ldots, x_1)}$$

2. Repite el proceso, reduciendo cada probabilidad conjuntiva a una probabilidad condicional y una conjunción más pequeña

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n \mid x_{n-1}, \ldots, x_1) P(x_{n-1} \mid x_{n-2}, \ldots, x_1) \cdots P(x_2 \mid x_1) P(x_1) \\ &= \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \ldots, x_1) . \end{aligned}$$

Regla de la Cadena

var. aleatoria
La prob nodo depende de la probabilidad de los padres.

$$\boxed{\mathbf{P}(X_i \mid X_{i-1}, \ldots, X_1) = \mathbf{P}(X_i \mid Parents(X_i))}$$

La red bayesiana es una representación correcta del dominio solo si cada nodo es condicionalmente independiente de sus otros predecesores en el orden de los nodos, dados sus padres

- Metodología para cumplirlo
  1. Nodos: Determina las variables requeridas y ordénalas de tal manera que las causas preceden a los efectos $\{X_1, \ldots, X_n\}$
  2. Arcos. for *i*=0 to n: (cuando identificamos los padres)
     a. Elige de $X_1, \ldots, X_{i-1}$ el conjunto mínimo de padres que realmente tengan una influencia directa
     b. Haz un arco de cada padre a X_i
     c. CPTs: Haz la tabla $\boxed{\mathbf{P}(X_i \mid Parents(X_i))}$

$$\mathbf{P}(MaryCalls \mid JohnCalls, Alarm, Earthquake, Burglary) = \mathbf{P}(MaryCalls \mid Alarm)$$

Francisco Martín Rico francisco.rico@urjc.es

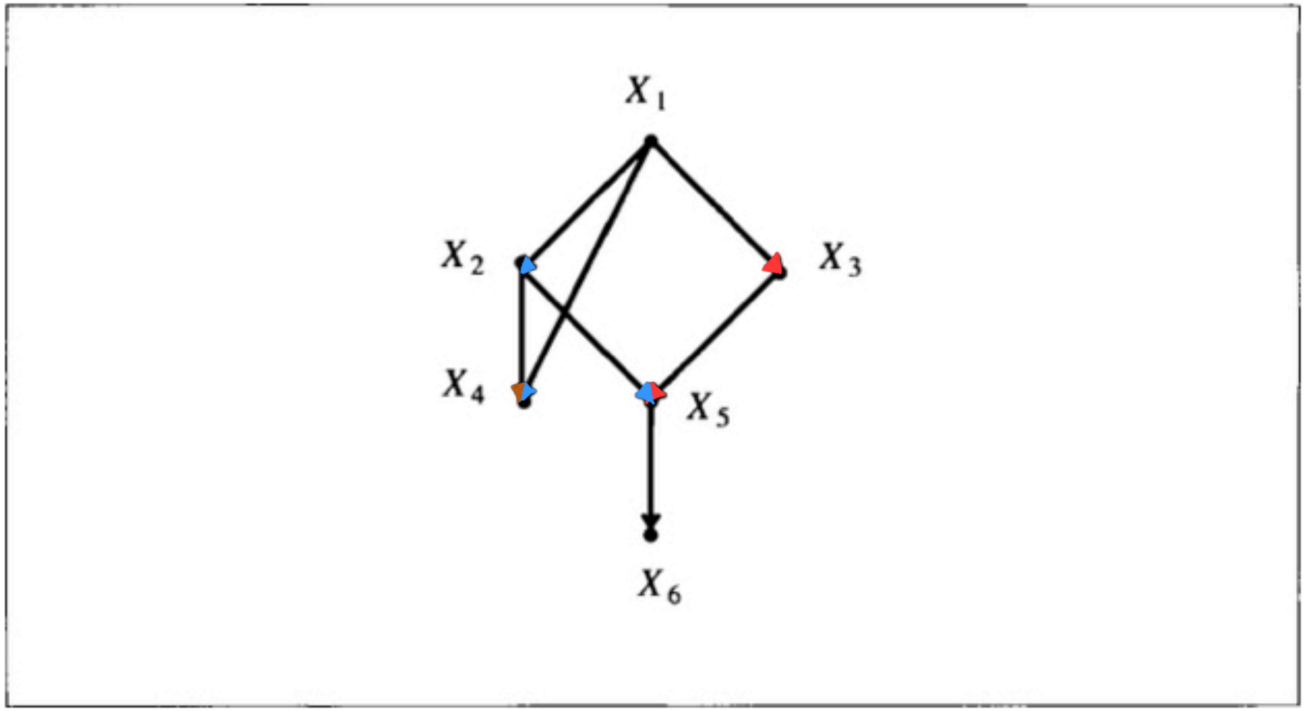


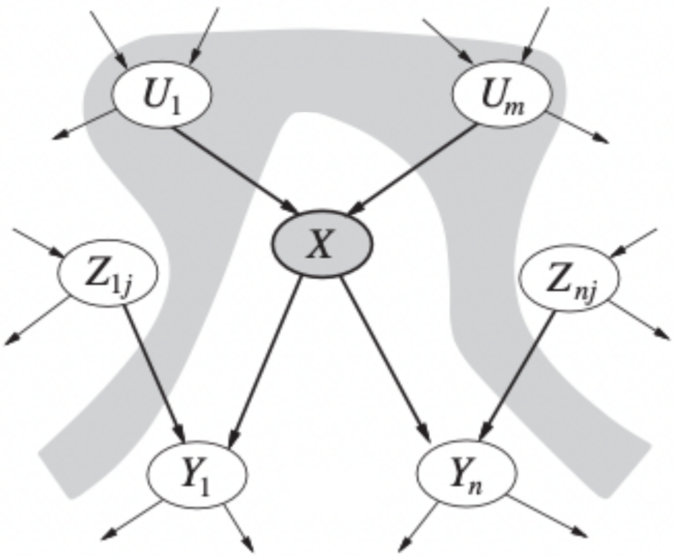
FIGURE 3.11 A Bayesian network representing the distribution
$P(x_6 \mid x_5)P(x_5 \mid x_2, x_3)P(x_4 \mid x_1, x_2)P(x_3 \mid x_1)P(x_2 \mid x_1)P(x_1)$

$$P(x_1, x_2, x_3, x_4, x_5, x_6)$$
$$= P(x_6 \mid x_5)P(x_5 \mid x_2, x_3)P(x_4 \mid x_1, x_2) \,.\, P(x_3 \mid x_1)P(x_2 \mid x_1)P(x_1) \,.$$

Judea Pearl, Probabilistic Reasoning in Intelligent Systems Published by Morgan Kaufmann

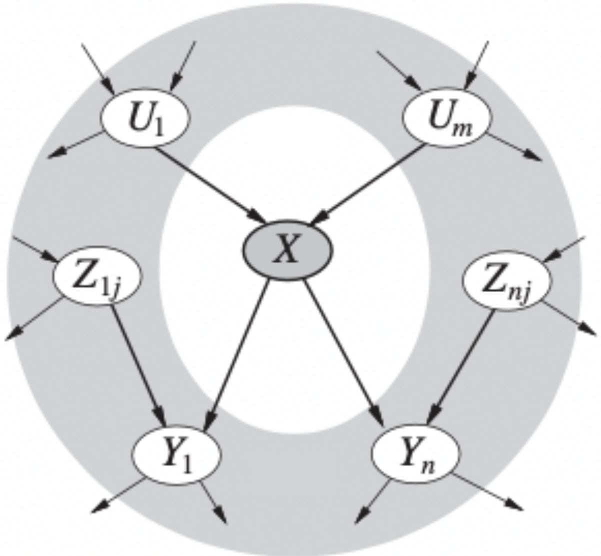
Relaciones condicionales de independencia

Markov blanket



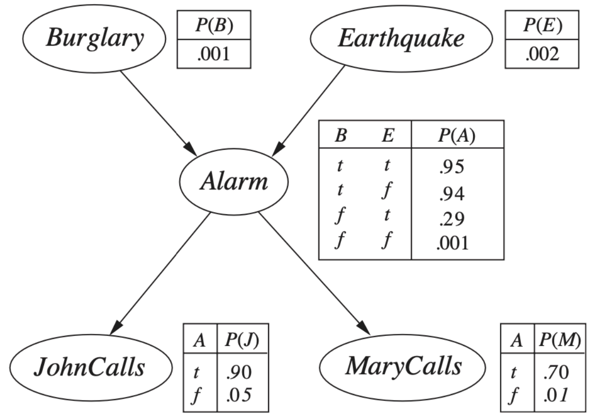
Un nodo X es condicionalmente independiente de sus no descendientes (los Z_{is}) dados sus padres (los U_i mostrados en el área gris).

JohnCalls es independiente de *Burglary, EarthQuake* y *MaryCalls* dado el valor de *Alarm*.



Un nodo X está relacionado con los nodos dentro de su **manto de Markov** (el área gris).

Burglary es independiente de *JohnCalls,* y *MaryCalls* dado el valor de *Alarm* y *EarthQuake*.



Inferencia Exacta en Redes Bayesianas

- La tarea básica de cualquier sistema de inferencia probabilístico es calcular la distribución de probabilidad posterior para un conjunto de **variables de consulta**, dado algún **evento observado**, es decir, alguna asignación de valores a un conjunto de variables de evidencia.

evidencia

- Notación
  - X es la variable de consulta
  - $\mathbf{E}$ es el conjunto de variables de evidencias $E_1, \ldots, E_m$ y $\mathbf{e}$ es un evento particular observado
  - $\mathbf{Y}$ es el conjunto de variables de no evidencias, ni de consulta $Y_1, \ldots, Y_l$ (**variables ocultas**)
  - La consulta es $\mathbf{P}(X \mid \mathbf{e})$

preguntamos algo en base a las evidencias.

$$\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$$

$$\mathbf{P}(Burglary \mid JohnCalls = true, MaryCalls = true) = \langle 0.284, 0.716 \rangle$$

Intelligent Robotics Lab

Inferencia por enumeración

- Como vimos en la Parte I

normalización.

no tiene nada asignado.

$$\mathbf{P}(B\,|\,j,m) = \alpha \mathbf{P}(B,j,m) = \alpha \sum_e \sum_a \mathbf{P}(B,j,m,e,a)$$

variable aleatoria (todos los posibles casos).

- Usando Redes Bayesianas, en función de las CPTs

$$P(b\,|\,j,m) = \alpha \sum_e \sum_a P(b)p(e)P(a\,|\,b,e)P(j\,|\,a)P(m\,|\,a) \qquad O(n2^n)$$

← los posibles casos

suma o no la alarma

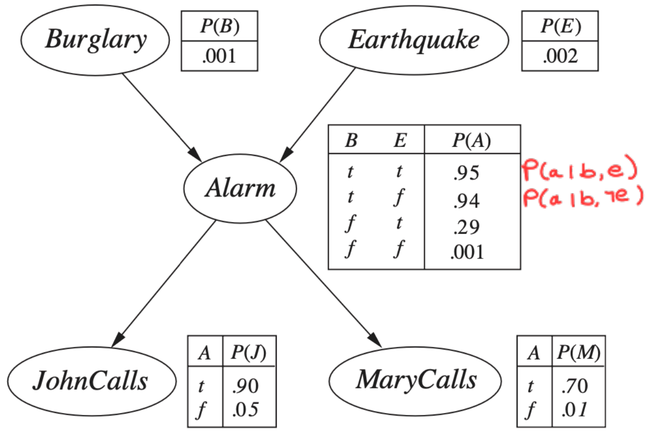
$$P(b\,|\,j,m) = \alpha P(b) \sum_e p(e) \sum_a P(a\,|\,b,e)P(j\,|\,a)P(m\,|\,a)$$

$$P(b\,|\,j,m) = \alpha \times 0.00059224$$

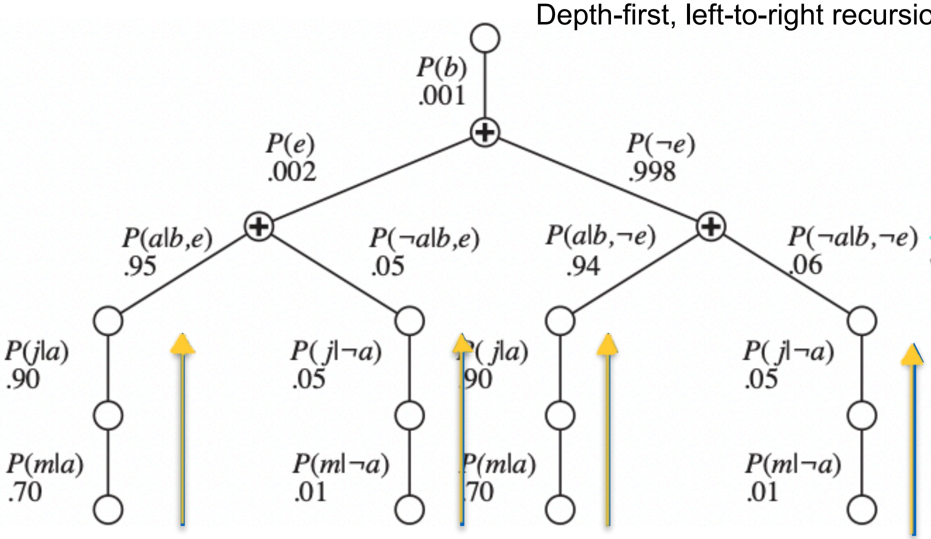
$$\mathbf{P}(B\,|\,j,m) = \alpha \langle 0.00059224, 0.0014919 \rangle \approx \langle 0.284, 0.716 \rangle$$

distribución de las probabilidades.

La posibilidad de un robo, dadas las llamadas de ambos vecinos, es de alrededor del 28%.



Depth-first, left-to-right recursion




```
function ENUMERATION-ASK(X, e, bn) returns a distribution over X
  inputs: X, the query variable
          e, observed values for some set of variables E
          bn, a Bayes net
  Q ← a distribution over X, where Q(x_i) is P(X=x_i)
  for each value x_i that X can have do
      Q(x_i) ← ENUMERATE-ALL(bn.VARS, e_xi), where e_xi is the evidence e plus the assignment X=x_i
  return NORMALIZE(Q)

function ENUMERATE-ALL(vars, e) returns a probability (a real number in [0,1])
  inputs: vars, a list of all the variables
          e, observed values for some set of variables E
  if EMPTY(vars) then return 1.0
  Y ← FIRST(vars)
  if Y is assigned a value (call it y) in e then
      return P(Y=y | values assigned to Y's parents in e) × ENUMERATE-ALL(REST(vars), e)
  else
      return Σ_yi [P(Y=y_i | values assigned to Y's parents in e) × ENUMERATE-ALL(REST(vars), e_yi)],
          where e_yi is the evidence e plus the assignment Y=y_i
```

te da un valor de prob distinta

variables de la red bayesiana.

vector evidencias + (asignación X = x_i)

la variable aleatoria toma un valor en concreto.

tienen que tener un orden topológico: los padres vienen primero.

variables red bayesiana.

pq vamos a multiplicar por 1

vars red bayesiana - las que has usado.

B no estaría asignado.

si no está asignado, hay que iterar.

http://courses.csail.mit.edu/6.034s/handouts/spring12/bayesnets-pseudocode.pdf

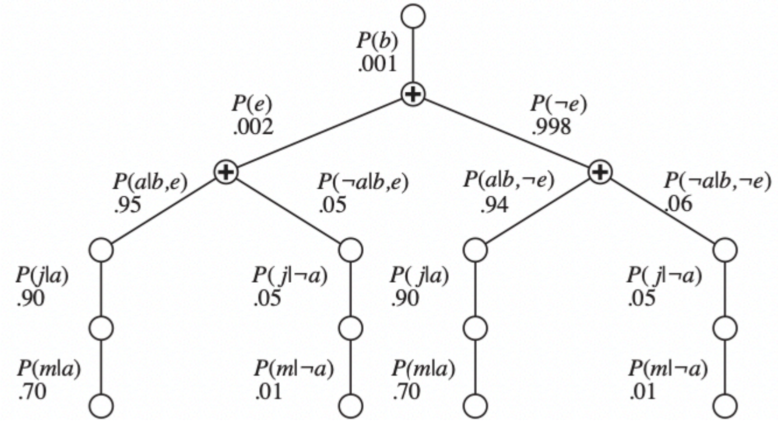
Caso base (⊛) recursividad. (✝)

Intelligent Robotics Lab

El algoritmo de eliminación de variables

- La idea es simple: **hacer el cálculo una vez y guardar los resultados para su uso posterior**
- La eliminación de variables funciona:
  1. Evaluando expresiones de derecha a izquierda, o en la la figura de abajo hacia arriba

$$P(b \mid j,m) = \alpha P(b) \sum_e p(e) \sum_a P(a \mid b,e) P(j \mid a) P(m \mid a)$$

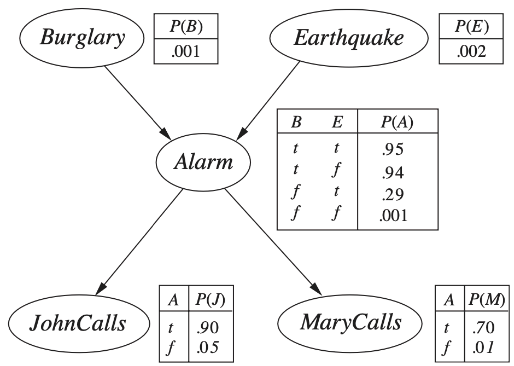


  2. Los resultados intermedios se almacenan y las sumas de cada variable se realizan solo para aquellas partes de la expresión que dependen de la variable.

$$\mathbf{P}(B \mid j,m) = \alpha \underbrace{\mathbf{P}(B)}_{\mathbf{f}_1(B)} \sum_e \underbrace{P(e)}_{\mathbf{f}_2(E)} \sum_a \underbrace{\mathbf{P}(a \mid B,e)}_{\mathbf{f}_3(A,B,E)} \underbrace{P(j \mid a)}_{\mathbf{f}_4(A)} \underbrace{P(m \mid a)}_{\mathbf{f}_5(A)}$$

$$\mathbf{P}(B \mid j,m) = \alpha \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

cada **factor** es una **matriz indexada por los valores de sus variables de argumento**



$$\mathbf{f}_4(A) = \begin{pmatrix} P(j \mid a) \\ P(j \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.90 \\ 0.05 \end{pmatrix} \qquad \mathbf{f}_5(A) = \begin{pmatrix} P(m \mid a) \\ P(m \mid \neg a) \end{pmatrix} = \begin{pmatrix} 0.70 \\ 0.01 \end{pmatrix}$$

$\mathbf{f}_3(A,B,E)$ es una matriz de 2 x 2 x 2 $\quad P(a \mid b,e) = 0.95 \rightarrow P(\neg a \mid \neg b, \neg e) = 0.999$

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

$$\mathbf{P}(B \mid j, m) = \alpha \, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

- × es un operación llamada **pointwise product:**
  - El pointwise product de dos factores f1 y f2 produce un nuevo factor f cuyas variables **son la unión** de las variables en f1 y f2 y cuyos elementos están dados por el producto de los elementos correspondientes en los dos factores.

$$\boxed{\mathbf{f}_1(A, B) \times \mathbf{f}_2(B, C) = \mathbf{f}_3(A, B, C).}$$

A	B	$\mathbf{f}_1(A,B)$	B	C	$\mathbf{f}_2(B,C)$	A	B	C	$\mathbf{f}_3(A,B,C)$
T	T	.3	T	T	.2	T	T	T	$.3 \times .2 = .06$
T	F	.7	T	F	.8	T	T	F	$.3 \times .8 = .24$
F	T	.9	F	T	.6	T	F	T	$.7 \times .6 = .42$
F	F	.1	F	F	.4	T	F	F	$.7 \times .4 = .28$
						F	T	T	$.9 \times .2 = .18$
						F	T	F	$.9 \times .8 = .72$
						F	F	T	$.1 \times .6 = .06$
						F	F	F	$.1 \times .4 = .04$

Intelligent Robotics Lab

$$\boxed{\mathbf{P}(B\mid j,m)=\alpha\,\overset{2\times1}{\mathbf{f}_1(B)}\times\sum_e \overset{2\times1}{\mathbf{f}_2(E)}\times\sum_a \overset{2\times2\times2}{\mathbf{f}_3(A,B,E)}\times\overset{2\times1}{\mathbf{f}_4(A)}\times\overset{2\times1}{\mathbf{f}_5(A)}}$$

- El proceso de evaluación es un proceso de **sumar variables (de derecha a izquierda)** a partir de productos puntuales de factores para producir nuevos factores, lo que finalmente produce un factor que es la solución, es decir, la distribución posterior sobre la variable de consulta:
  1. Primero, sumamos A del producto de f3, f4 y f5 para obtener un nuevo factor f6(B|E)

$$\begin{aligned}\mathbf{f}_6(B,E) &= \sum_a \mathbf{f}_3(A,B,E)\times\mathbf{f}_4(A)\times\mathbf{f}_5(A)\\ &= (\underset{2\times2}{\mathbf{f}_3(a,B,E)}\times\underset{1}{\mathbf{f}_4(a)}\times\underset{1}{\mathbf{f}_5(a)})+(\underset{2\times2}{\mathbf{f}_3(\neg a,B,E)}\times\underset{1}{\mathbf{f}_4(\neg a)}\times\underset{1}{\mathbf{f}_5(\neg a)})\,.\end{aligned}$$

eliminamos variables

$$\boxed{\mathbf{P}(B\mid j,m)=\alpha\,\overset{2\times1}{\mathbf{f}_1(B)}\times\sum_e \overset{2\times1}{\mathbf{f}_2(E)}\times\overset{2\times2}{\mathbf{f}_6(B,E)}}$$

  2. A continuación, sumamos E del producto de f2 y f6

$$\begin{aligned}\mathbf{f}_7(B) &= \sum_e \overset{2\times1}{\mathbf{f}_2(E)}\times\overset{2\times2}{\mathbf{f}_6(B,E)}\\ &= \underset{1}{\mathbf{f}_2(e)}\times\underset{2\times1}{\mathbf{f}_6(B,e)}+\underset{1}{\mathbf{f}_2(\neg e)}\times\underset{2\times1}{\mathbf{f}_6(B,\neg e)}\end{aligned}$$

$$\boxed{\mathbf{P}(B\mid j,m)=\alpha\,\overset{2\times1}{\mathbf{f}_1(B)}\times\overset{2\times1}{\mathbf{f}_7(B)}}$$

que se puede evaluar tomando el pointwise product y normalizando el resultado.

Algoritmo de derecha a izquierda.

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{f}_1(B) \times \sum_e \mathbf{f}_2(E) \times \sum_a \mathbf{f}_3(A, B, E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A)$$

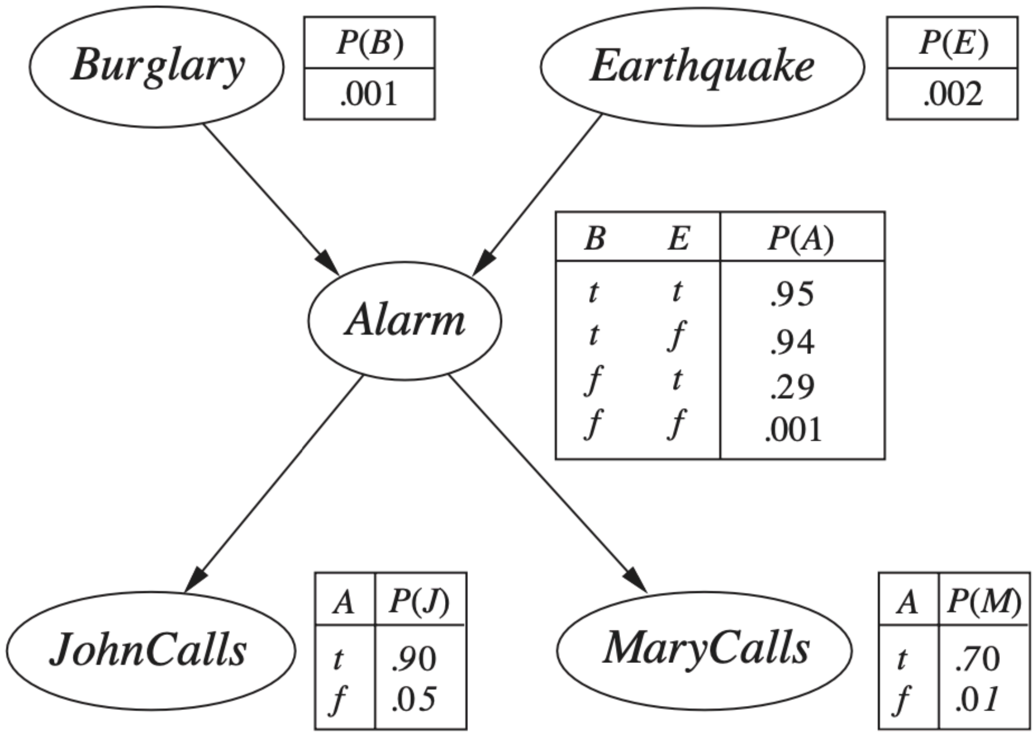
```
function ELIMINATION-ASK(X, e, bn) returns a distribution over X
  inputs: X, the query variable
          e, observed values for variables E
          bn, a Bayesian network specifying joint distribution P(X1, ..., Xn)

  factors ← [ ]
  for each var in ORDER(bn.VARS) do
      factors ← [MAKE-FACTOR(var, e)|factors]
      if var is a hidden variable then factors ← SUM-OUT(var, factors)
  return NORMALIZE(POINTWISE-PRODUCT(factors))
```

↑ definir una lista ...

$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{f}_1(B) \times \mathbf{f}_7(B)$$

Intelligent Robotics Lab

Ejercicio: Un sistema de alarma



- ¿Cuál es la probabilidad de que haya un ladrón dado que John y Mary llaman?

Alejandro Bascuñana Muñoz alejandro.bascunana@urjc.es

Referencias

Capítulo 13. Russell S., Norving P. “Artificial intelligence, A Modern Approach”. 2022, Pearson.

https://es.wikipedia.org/wiki/Red_bayesiana

https://www.youtube.com/watch?v=hEZjPZ-Ze0A

http://courses.csail.mit.edu/6.034s/handouts/spring12/bayesnets-pseudocode.pdf

Intelligent Robotics Lab