

## Tema 8: Introducción a las Redes Bayesianas

José L. Ruiz Reina

Dpto. de Ciencias de la Computación e Inteligencia Artificial

UNIVERSIDAD DE SEVILLA

# Redes bayesianas

- Como vimos en el tema anterior, las relaciones de independencia (condicional) nos permiten reducir el tamaño de la información necesaria para especificar una DCC
  - Las *redes bayesianas* (o *redes de creencia*) constituyen una manera práctica y compacta de representar el conocimiento incierto, basada en esta idea
- Una red bayesiana es un grafo dirigido *acíclico* que consta de:
  - Un conjunto de nodos, uno por cada variable aleatoria del “mundo”
  - Un conjunto de arcos dirigidos que conectan los nodos; si hay un arco de  $X$  a  $Y$  decimos que  $X$  es un *padre* de  $Y$  ( $padres(X)$  denota el conjunto de v.a. que son padres de  $X$ )
  - Cada nodo  $X_i$  contiene la distribución de probabilidad condicional  $P(X_i|padres(X_i))$
- Intuitivamente, en una red bayesiana una arco entre  $X$  e  $Y$  significa *una influencia directa* de  $X$  sobre  $Y$ 
  - Es tarea del experto en el dominio el decidir las relaciones de dependencia directa (es decir, la *topología* de la red)

# Ejemplo de red bayesiana (Russell y Norvig)

$P(sol)$	$P(lluv)$	$P(nubl)$	$P(nieve)$
0.7	0.2	0.08	0.02

Tiempo

$P(caries)$
0.8

Caries

Caries	$P(dolor)$
caries	0.6
no caries	0.1

Dolor

Huecos

Caries	$P(hueco)$
caries	0.9
no caries	0.2

.

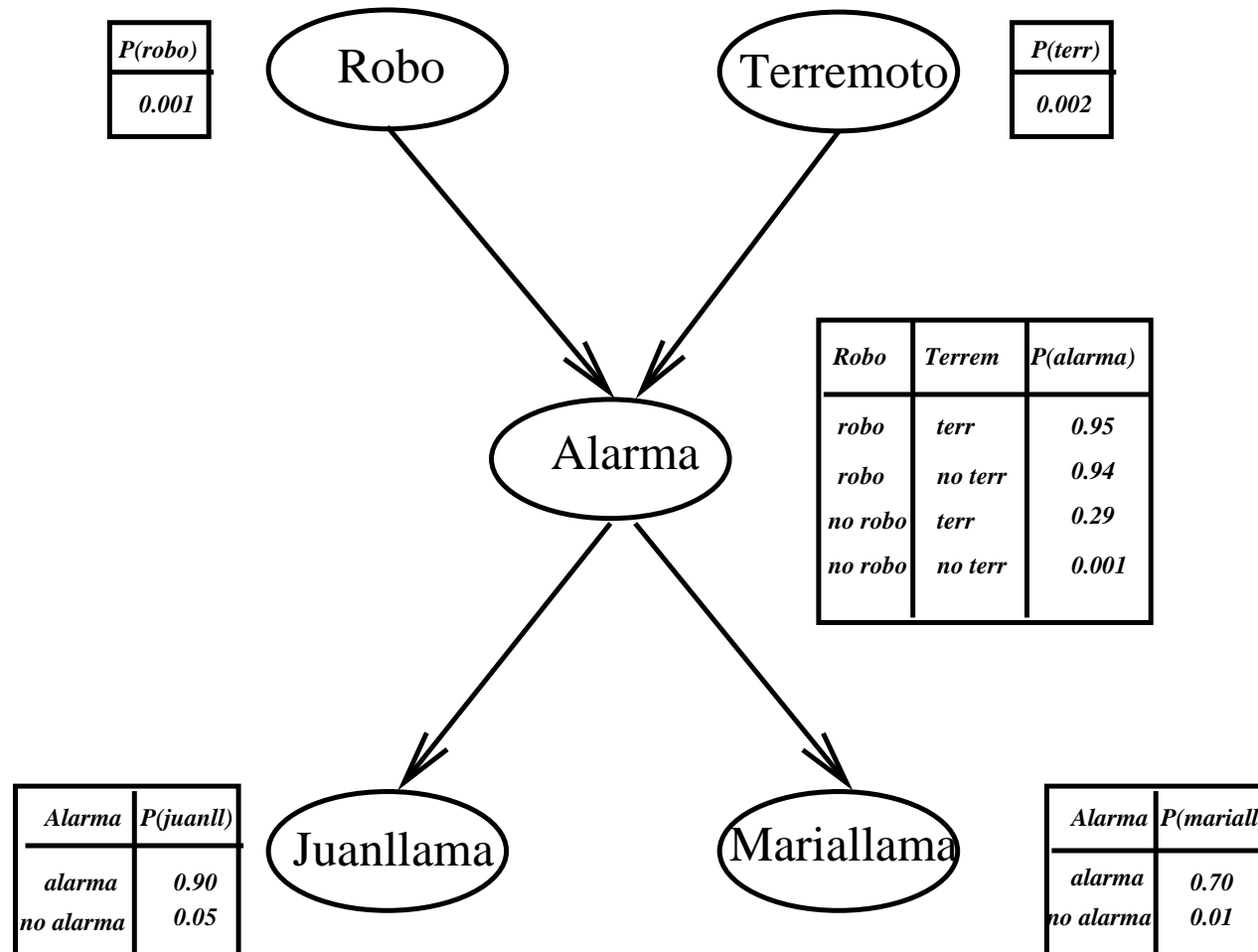
## Observaciones sobre el ejemplo

- La topología de la red anterior nos expresa que:
  - *Caries* es una *causa directa* de *Dolor* y *Huecos*
  - *Dolor* y *Huecos* son condicionalmente independientes dada *Caries*
  - *Tiempo* es independiente de las restantes variables
- No es necesario dar la probabilidad de las negaciones de *caries*, *dolor*, ...

## Otro ejemplo (Pearl, 1990):

- Tenemos una alarma antirrobo instalada en una casa
  - La alarma salta normalmente con la presencia de ladrones
  - Pero también cuando ocurren pequeños temblores de tierra
- Tenemos dos vecinos en la casa, Juan y María, que han prometido llamar a la policía si oyen la alarma
  - Juan y María podrían no llamar aunque la alarma sonara: por tener música muy alta en su casa, por ejemplo
  - Incluso podrían llamar aunque no hubiera sonado: por confundirla con un teléfono, por ejemplo

# Red bayesiana para el ejemplo de la alarma



## Observaciones sobre el ejemplo

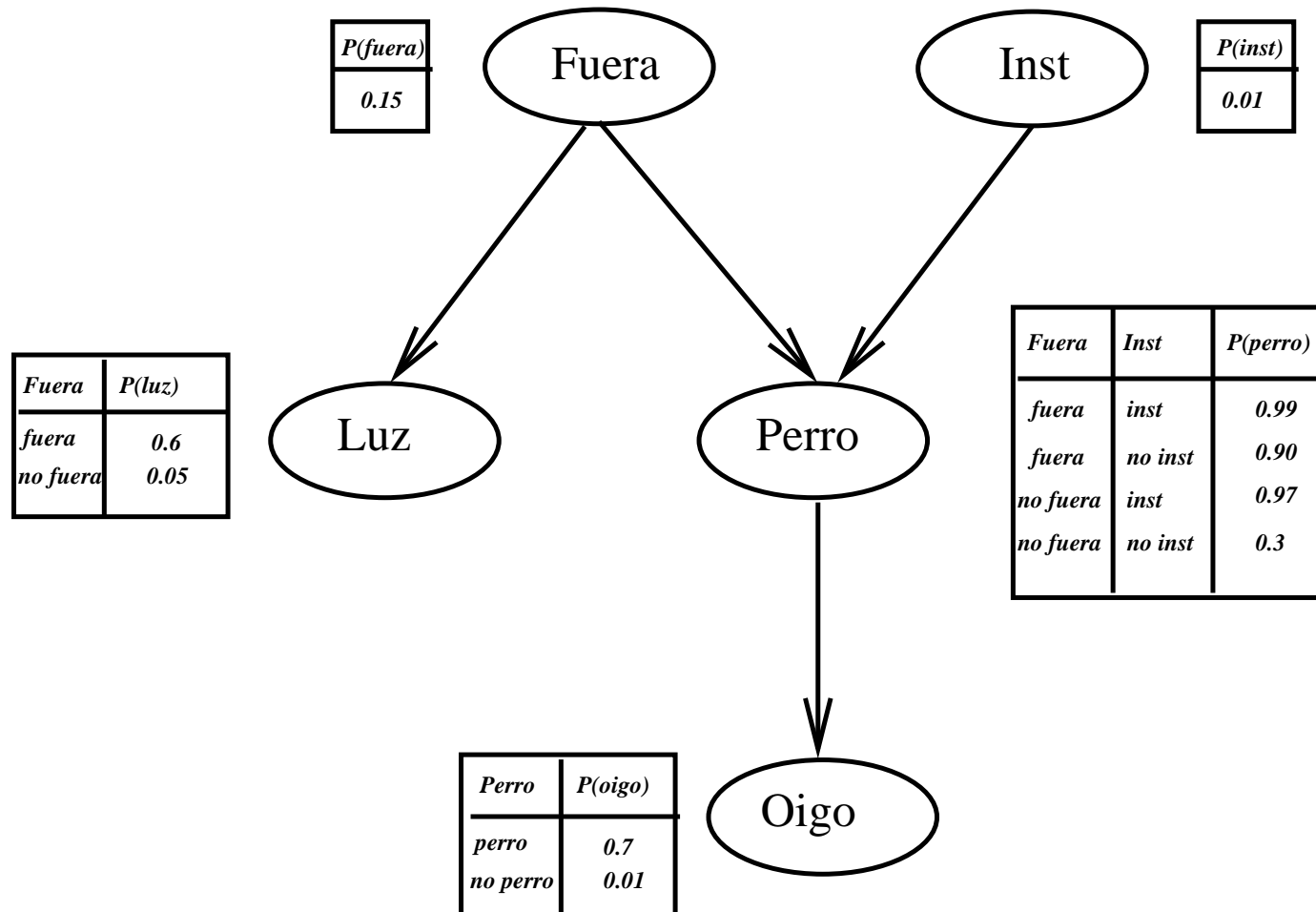
- La topología de la red nos expresa que:
  - *Robo* y *Terremoto* son causas directas para *Alarma*
  - También, *Robo* y *Terremoto* son causas para *Juanllama* y para *Mariallama*, pero esa influencia sólo se produce a través de *Alarma*: ni Juan ni María detectan directamente el robo ni los pequeños temblores de tierra
  - En la red no se hace referencia directa, por ejemplo, a las causas por las cuales María podría no oír la alarma: éstas están implícitas en la tabla de probabilidades  $P(\text{Mariallama}|\text{Alarma})$

## Un tercer ejemplo (Charniak, 1991):

- Supongamos que quiero saber si alguien de mi familia está en casa, basándome en la siguiente información
  - Si mi esposa sale de casa, usualmente (pero no siempre) enciende la luz de la entrada
  - Hay otras ocasiones en las que también enciende la luz de la entrada
  - Si no hay nadie en casa, el perro está fuera
  - Si el perro tiene problemas intestinales, también se deja fuera
  - Si el perro está fuera, oigo sus ladridos
  - Podría oír ladridos y pensar que son de mi perro aunque no fuera así
- Variables aleatorias (booleanas) en este problema:
  - *Fuera* (nadie en casa), *Luz* (luz en la entrada), *Perro* (perro fuera), *Inst* (problemas intestinales en el perro) y *Oigo* (oigo al perro ladrar)



# Red bayesiana para el ejemplo de la familia fuera de casa



## Las redes bayesianas representan DCCs

- Consideremos una red bayesiana con  $n$  variables aleatorias
  - Y un orden entre esas variables:  $X_1, \dots, X_n$
- En lo que sigue, supondremos que:
  - $\text{padres}(X_i) \subseteq \{X_{i-1}, \dots, X_1\}$  (para esto, basta que el orden escogido sea consistente con el orden parcial que induce el grafo)
  - $P(X_i | X_{i-1}, \dots, X_1) = P(X_i | \text{padres}(X_i))$  (es decir, cada variable es condicionalmente independiente de sus anteriores, dados sus padres en la red)
- Estas condiciones expresan formalmente nuestra intuición al representar nuestro “mundo” mediante la red bayesiana correspondiente
  - En el ejemplo de la alarma, la red expresa que creemos que  $P(\text{Mariallama} | \text{Juanllama}, \text{Alarma}, \text{Terremoto}, \text{Robo}) = P(\text{Mariallama} | \text{Alarma})$

## Las redes bayesianas representan DCCs

- En las anteriores condiciones, y aplicando repetidamente la regla del producto:

$$\begin{aligned} P(X_1, \dots, X_n) &= P(X_n | X_{n-1} \dots, X_1) P(X_{n-1} \dots, X_1) = \dots \\ &= \prod_{i=1}^n P(X_i | X_{i-1}, \dots, X_1) = \prod_{i=1}^n P(X_i | \text{padres}(X_i)) \end{aligned}$$

- Es decir, una red bayesiana *representa una DCC* obtenida mediante la expresión  $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{padres}(X_i))$ 
  - Por ejemplo, en el ejemplo de la alarma, la probabilidad de que la alarma suene, Juan y María llamen a la policía, pero no haya ocurrido nada es (usamos iniciales, por simplificar) :

$$\begin{aligned} P(j, m, a, \neg r, \neg t) &= P(j|a)P(m|a)P(a|\neg r, \neg t)P(\neg r)P(\neg t) = \\ &= 0,9 \times 0,7 \times 0,001 \times 0,999 \times 0,998 = 0,00062 \end{aligned}$$

# Representaciones compactas

- Dominios *localmente estructurados*:
  - Las relaciones de independencia que existen entre las variables de un dominio hacen que las redes bayesianas sean una representación mucho más compacta y eficiente de una DCC que la tabla con todas las posibles combinaciones de valores
  - Además, para un experto en un dominio de conocimiento suele ser más natural dar probabilidades condicionales que directamente las probabilidades de la DCC
  - Con  $n$  variables, si cada variable está directamente influenciada por  $k$  variables a lo sumo, entonces una red bayesiana necesitaría  $n2^k$  números, frente a los  $2^n$  números de la DCC
  - Por ejemplo, Para  $n = 30$  y  $k = 5$ , esto supone 960 números frente a  $2^{30}$  (billones)
- Hay veces que una variable influye directamente sobre otra, pero esta dependencia es muy tenue
  - En ese caso, puede compensar no considerar esa dependencia, perdiendo algo de precisión en la representación, pero ganado manejabilidad

# Algoritmo de construcción de una red bayesiana

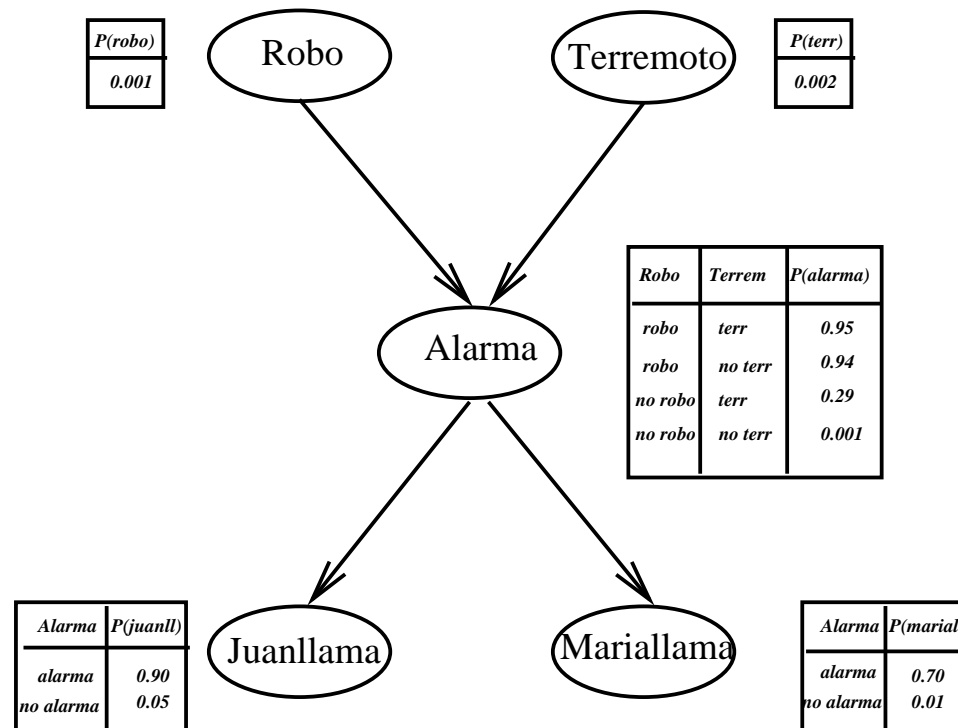
- Supongamos dado un conjunto de variables aleatorias VARIABLES que representan un dominio de conocimiento (con incertidumbre)

FUNCION CONSTRUYE\_RED(VARIABLES)

1. Sea  $(X_1, \dots, X_n)$  una ordenación de las variables de VARIABLES
2. Sea RED una red bayesiana ‘‘vacía’’
3. PARA  $i=1, \dots, n$  HACER
  - 3.1 Añadir un nodo etiquetado con  $X_i$  a RED
  - 3.2 Sea  $\text{padres}(X_i)$  un subconjunto minimal de  $\{X_{i-1}, \dots, X_1\}$  tal que existe una independencia condicional entre  $X_i$  y cada elemento de  $\{X_{i-1}, \dots, X_1\}$  dado  $\text{padres}(X_i)$
  - 3.3 Añadir en RED un arco dirigido entre cada elemento de  $\text{padres}(X_i)$  y  $X_i$
  - 3.4 Asignar al nodo  $X_i$  la tabla de probabilidad  $P(X_i | \text{padres}(X_i))$
4. Devolver RED

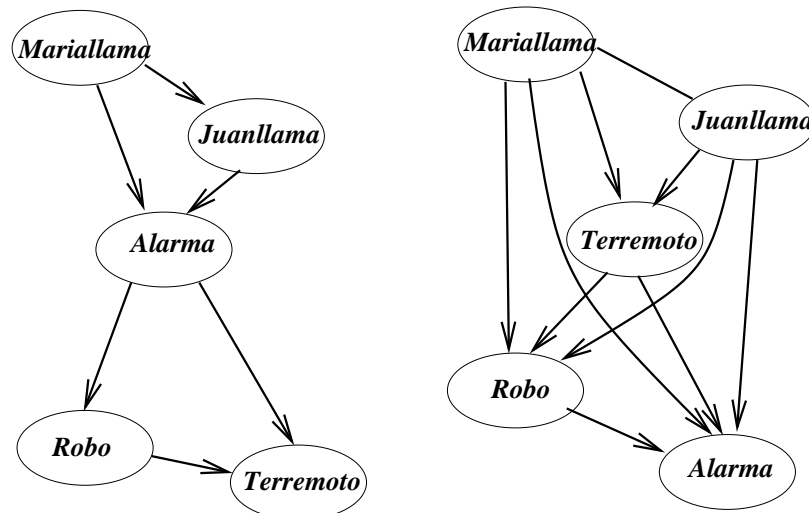
## Ejemplo de construcción de red bayesiana (alarma)

- Partiendo del orden *Robo*, *Terremoto*, *Alarma*, *Juanllama*, *Mariallama*, y aplicando el algoritmo anterior obtenemos la red del ejemplo:



# Construcción de redes bayesianas

- Problema: elección del orden entre variables
  - En general, deberíamos empezar por las “causas originales”, siguiendo con aquellas a las que influyen directamente, etc..., hasta llegar a las que no influyen directamente sobre ninguna (modelo *causal*)
  - Esto hará que las tablas reflejen probabilidades “causales” más que “diagnósticos”, lo cual suele ser preferible por los expertos
- Un orden malo puede llevar a representaciones poco eficientes
  - Red izquierda (*Mariallama*, *Juanllama*, *Alarma*, *Robo* y *Terremoto*); red derecha (*Mariallama*, *Juanllama*, *Terremoto*, *Robo* y *Alarma*)



# Inferencia probabilística en una red bayesiana

- El problema de la inferencia en una red bayesiana
  - Calcular la probabilidad a posteriori para un conjunto de *variables de consulta*, dado que se han observado algunos valores para las *variables de evidencia*
  - Por ejemplo, podríamos querer saber qué probabilidad hay de que realmente se haya producido un robo, sabiendo que tanto Juan como María han llamado a la policía
  - Es decir, calcular  $P(\text{Robo} | \text{juanllama}, \text{mariallama})$
- Notación:
  - $X$  denotará la variable de consulta (sin pérdida de generalidad supondremos sólo una variable)
  - $E$  denota un conjunto de *variables de evidencia*  $E_1, E_2, \dots, E_n$  y  $e$  una observación concreta para esas variables
  - $Y$  denota al conjunto de las restantes variables de la red (variables *ocultas*) e  $y$  representa un conjunto cualquiera de valores para esas variables



## Inferencia por enumeración

- Recordar la fórmula para la inferencia probabilística a partir de una DCC:

$$P(X|e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$$

- Esta fórmula será la base para la inferencia probabilística:
  - Puesto que una red bayesiana es una representación de una DCC, nos permite calcular cualquier probabilidad a posteriori a partir de la información de la red bayesiana
  - Esencialmente, se trata de una suma de productos de los elementos de las distribuciones condicionales

## Un ejemplo de inferencia probabilística

- Ejemplo de la alarma (usamos iniciales por simplificar)

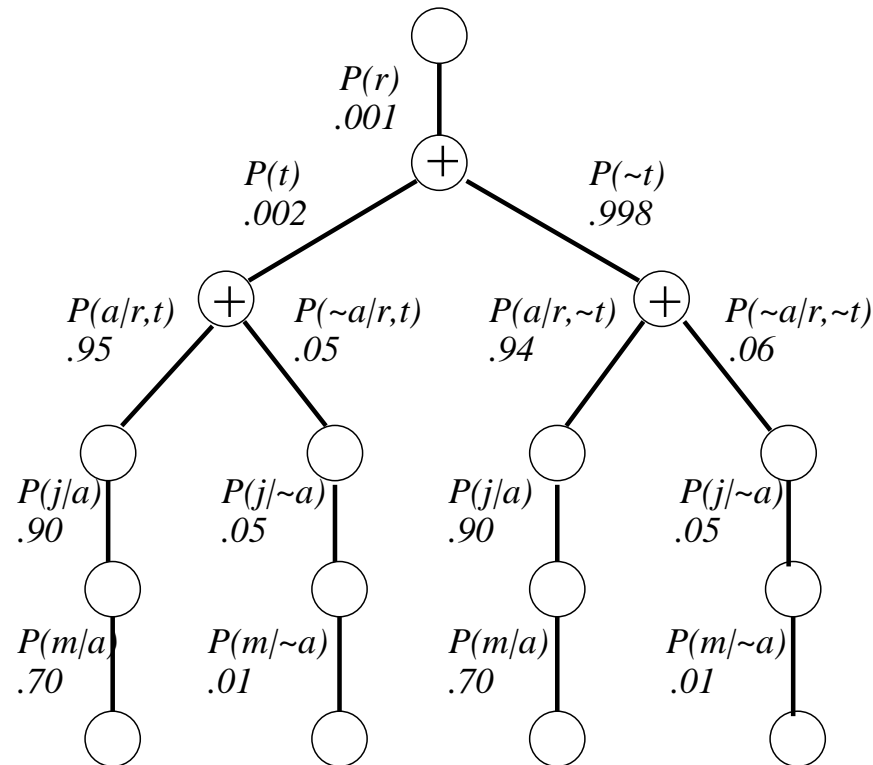
$$\begin{aligned} P(R|j, m) &= \alpha \langle P(r|j, m), P(\neg r|j, m) \rangle = \alpha \langle \sum_t \sum_a P(r, t, a, j, m), \sum_t \sum_a P(\neg r, t, a, j, m) \rangle = \\ &= \alpha \langle \sum_t \sum_a P(r)P(t)P(a|r, t)P(j|a)P(m|a), \sum_t \sum_a P(\neg r)P(t)P(a|\neg r, t)P(j|a)P(m|a) \rangle \end{aligned}$$

- En este ejemplo hay que hacer  $2 \times 4$  sumas, cada una de ellas con un producto de cinco números tomados de la red bayesiana
  - En el peor de los casos, con  $n$  variables booleanas, este cálculo toma  $O(n2^n)$
- Una primera mejora consiste en sacar factor común de aquellas probabilidades que sólo involucran variables que no aparecen en el sumatorio:

$$\begin{aligned} P(R|j, m) &= \alpha \langle P(r) \sum_t P(t) \sum_a P(a|r, t)P(j|a)P(m|a), P(\neg r) \sum_t P(t) \sum_a P(a|\neg r, t)P(j|a)P(m|a) \rangle = \\ &= \alpha \langle 0,00059224, 0,0014919 \rangle = \langle 0,284, 0,716 \rangle \end{aligned}$$

## Inferencia por enumeración

- Las operaciones realizadas en la fórmula anterior se pueden simbolizar con el siguiente árbol



# Algoritmo de inferencia por enumeración

- Entrada: una v.a.  $X$  de consulta, un conjunto de valores observados  $e$  para la variables de evidencia y una red bayesiana. Salida:  $P(X|e)$

FUNCION INFERENCIA\_ENUMERACION( $X, e, RED$ )

1. Sea  $Q(X)$  una distribución de probabilidad sobre  $X$ , inicialmente vacía
2. PARA cada valor  $x_i$  de  $X$  HACER
  - 2.1 Extender  $e$  con el valor  $x_i$  para  $X$
  - 2.2 Hacer  $Q(x_i)$  el resultado de  $ENUM\_AUX(VARIABLES(RED), e, RED)$
3. Devolver  $NORMALIZA(Q(X))$

FUNCION ENUM\_AUX( $VARS, e, RED$ )

1. Si  $VARS$  es vacío devolver 1
2. Si no,
  - 2.1 Hacer  $Y$  igual a PRIMERO( $VARS$ )
  - 2.2 Si  $Y$  tiene un valor  $y$  en  $e$  devolver  $P(y|padres(Y, e)) \cdot ENUM\_AUX(RESTO(VARS), e)$   
Si no, devolver  $SUMATORIO(y, P(y|padres(Y, e)) \cdot ENUM\_AUX(RESTO(VARS), e_y))$   
(donde:  $padres(Y, e)$  es el conjunto de valores que toman en  $e$   
los padres de  $Y$  en la  $RED$ , y  $e_y$  extiende  $e$  con el valor  $y$  para  $Y$ )

# Algoritmo de inferencia por enumeración

- Observación:
  - Para que el algoritmo funcione,  $\text{VARIABLES}(\text{RED})$  debe devolver las variables en un orden consistente con el orden implícito en el grafo de la red (de arriba hacia abajo)
- Recorrido en profundidad:
  - El algoritmo genera el árbol de operaciones anterior de arriba hacia abajo, en profundidad
  - Por tanto, tiene un coste lineal en espacio
- Puede realizar cálculos repetidos
  - En el ejemplo,  $P(j|a)P(m|a)$  y  $P(j|\neg a)P(m|\neg a)$  se calculan dos veces
  - Cuando hay muchas variables, estos cálculos redundantes son inaceptables en la práctica

## Evitando cálculos redundantes

- Idea para evitar el cálculo redundante:
  - Realizar las operaciones *de derecha a izquierda* (o, equivalentemente, de abajo a arriba en el árbol de operaciones)
  - Almacenar los cálculos realizados para posteriores usos
- En lugar de multiplicar *números*, multiplicaremos *tablas de probabilidades*
  - Denominaremos *factores* a estas tablas
  - Por ejemplo, la operación  $P(R|j, m) = \alpha P(R) \sum_t P(t) \sum_a P(a|R, t) P(j|a) P(m|a)$  puede verse como la *multiplicación* de cinco tablas o *factores* en los que hay intercaladas dos operaciones de suma o *agrupamiento*
  - Se trata de hacer esas operaciones entre factores de derecha a izquierda
  - Es el denominado algoritmo de *eliminación de variables*
  - Veamos con más detalle cómo actuaría este algoritmo para calcular  $P(R|j, m)$

## El algoritmo de eliminación de variables: un ejemplo

- En primer lugar, tomamos un orden fijado entre las variables de la red
  - Un orden adecuado será esencial para la eficiencia del algoritmo (más adelante comentaremos algo más sobre el orden)
  - En nuestro caso, tomaremos el inverso de un orden consistente con la topología de la red:  $M, J, A, T, R$
- El factor correspondiente a  $M$  se obtiene a partir de la distribución condicional  $P(M|A)$ 
  - Como  $M$  es una variable de evidencia y su valor está fijado a *true*, el factor correspondiente a  $P(m|A)$ , que notamos  $f(A)$ , es la tabla con componentes  $P(m|a)$  y  $P(m|\neg a)$ :

A		$f_1(A)$
-----		
T		0.70
F		0.01

## El algoritmo de eliminación de variables: un ejemplo

- La siguiente variable en el orden es  $J$ 
  - De manera análoga  $f_2(A)$  (tomado de  $P(j|A)$ ) es el factor correspondiente
- Siguiendo variable:  $A$ 
  - El factor correspondiente, notado  $f_3(A, R, T)$  se obtiene a partir de  $P(A|R, T)$
  - Ninguna de esas variables es de evidencia, y por tanto no están fijados sus valores
  - Es una tabla con  $2 \times 2 \times 2$ , una por cada combinación de valores de  $R$  (variable de consulta)  $A$  y  $T$  (variables ocultas)
  - En este caso, esta tabla está directamente en la propia red



## El algoritmo de eliminación de variables: un ejemplo

- Hasta ahora, no hemos realizado ninguna operación
  - Sólomente hemos construido los factores
- Pero  $A$  es una variable oculta, así que hemos de realizar el sumatorio sobre sus valores
  - Por tanto, multiplicamos ahora los tres factores y sumamos sobre  $A$
  - La multiplicación de  $f_1$ ,  $f_2$  y  $f_3$ , notada  $f_4(A, R, T)$  se obtiene multiplicando las entradas correspondientes a los mismos valores de  $A$ ,  $R$  y  $T$
  - Es decir, para cada valor  $v_1$  de  $A$ ,  $v_2$  de  $R$  y  $v_3$  de  $T$  se tiene  $f_4(v_1, v_2, v_3) = f_1(v_1)f_2(v_1)f_3(v_1, v_2, v_3)$
  - Por ejemplo:  $f_4(true, false, true) = f_1(true)f_2(true)f_3(true, false, true) = 0,70 \times 0,90 \times 0,29 = 0,1827$
  - Almacenamos  $f_4$  y nos olvidamos de  $f_1$ ,  $f_2$  y  $f_3$
  - Nota: aunque no es éste el caso, si alguno de los factores no tuviera a la variable  $A$  como argumento , lo conservaríamos y no participaría en la multiplicación, ni en el agrupamiento posterior

## El algoritmo de eliminación de variables: un ejemplo

- Ahora hay que *agrupar* el valor de  $A$  en  $f_4$  (realizar el sumatorio  $\sum_a$ )
  - Así, obtenemos una tabla  $f_5(R, T)$  haciendo  $f_5(v_1, v_2) = \sum_a f_4(a, v_1, v_2)$  para cada valor  $v_1$  de  $R$  y  $v_2$  de  $T$ , y variando  $a$  en los posibles valores de  $A$
  - Llamaremos a esta operación *agrupamiento*
  - Hemos *eliminado* la variable  $A$
  - Una vez realizada la agrupación, guardamos  $f_5$  y nos olvidamos de  $f_4$
- Continuamos con la siguiente variable  $T$ :
  - El factor correspondiente a esta variable, que notaremos  $f_6(T)$ , es la tabla  $P(T)$
- $T$  es una variable oculta
  - Por tanto, debemos multiplicar y agrupar, eliminando la variable  $T$
  - Notemos por  $f_7(R)$  al resultado de multiplicar  $f_6$  por  $f_5$  y agrupar por  $T$
  - Podemos olvidarnos de  $f_5$  y  $f_6$

## El algoritmo de eliminación de variables: un ejemplo

- Última variable:  $R$ 
  - El factor correspondiente a esta variable, que notaremos  $f_8(R)$ , es la tabla  $P(R)$
- Para finalizar:
  - Multiplicamos los factores que nos quedan ( $f_7$  y  $f_8$ ) para obtener  $f_9(R)$  y normalizamos para que sus dos entradas sumen 1
  - La tabla finalmente devuelta es justamente la distribución  $P(R|j.m)$

## Observaciones sobre el algoritmo de eliminación de variables

- En cada momento tenemos un conjunto de factores, que va cambiando:
  - Al añadir un nuevo factor, uno por cada variable que se considera
  - Al multiplicar factores
  - Al agrupar por una variable oculta
- Multiplicación de tablas:
  - Si  $f_1(X, Y)$  y  $f_2(Y, Z)$  son dos tablas cuyas variables en común son las de  $Y$ , se define su producto  $f(X, Y, Z)$  como la tabla cuyas entradas son  $f(x, y, z) = f_1(x, y)f_2(y, z)$
  - Similar a una operación *join* en bases de datos, multiplicando los valores correspondientes
- En el algoritmo de eliminación de variables, sólo multiplicaremos tablas:
  - Previo a realizar cada operación de agrupamiento
  - Y en el paso final

# Observaciones sobre el algoritmo de eliminación de variables

- Agrupamiento de tablas:
  - Dado *un conjunto de factores*, la operación de agrupar (respecto de los valores de una v.a.  $X$ ) consiste en obtener *otro conjunto de factores*
  - Se dejan igual aquellos que no tienen a la variable  $X$  entre sus argumentos
  - Y el resto de factores se multiplican y *se sustituyen* por el resultado multiplicarlos y sumar en la tabla por cada posible valor de  $X$
  - Por ejemplo, si en el conjunto de factores  $f_1(T)$ ,  $f_2(A, R, T)$ ,  $f_3(A)$ ,  $f_4(A)$  tuvierámos que agrupar por la variable  $T$ , dejamos  $f_3$  y  $f_4$  y sustituimos  $f_1$  y  $f_2$  por el resultado de agregar (por cada valor de  $T$ ) la multiplicación de  $f_1$  y  $f_2$
  - La operación de sumar por un valor es similar a la *agregación* de una columna en bases de datos

## Un paso previo de optimización: variables irrelevantes

- En el algoritmo de eliminación de variables, se suele realizar un paso previo de eliminación de variables irrelevantes para la consulta
- Ejemplo:
  - Si la consulta a la red del ejemplo es  $P(J|r)$ , hay que calcular  $\alpha P(r) \sum_t P(t) \sum_a P(a|r, t) P(J|a) \sum_m P(m|a)$
  - Pero  $\sum_m P(m|a) = 1$ , así que la variable  $M$  es irrelevante para la consulta
  - Siempre podemos eliminar cualquier variable que sea una hoja de la red y que no sea de consulta ni de evidencia
  - Y en general, se puede demostrar que toda variable que no sea antecesor (en la red) de alguna de las variables de consulta o de evidencia, es irrelevante para la consulta y por tanto *puede ser eliminada*

## El algoritmo de eliminación de variables

- Entrada: una v.a.  $X$  de consulta, un conjunto de valores observados  $e$  para la variables de evidencia y una red bayesiana. Salida:  $P(X|e)$

FUNCION INFERENCIA\_ELIMINACION\_VARIABLEES( $X, e, RED$ )

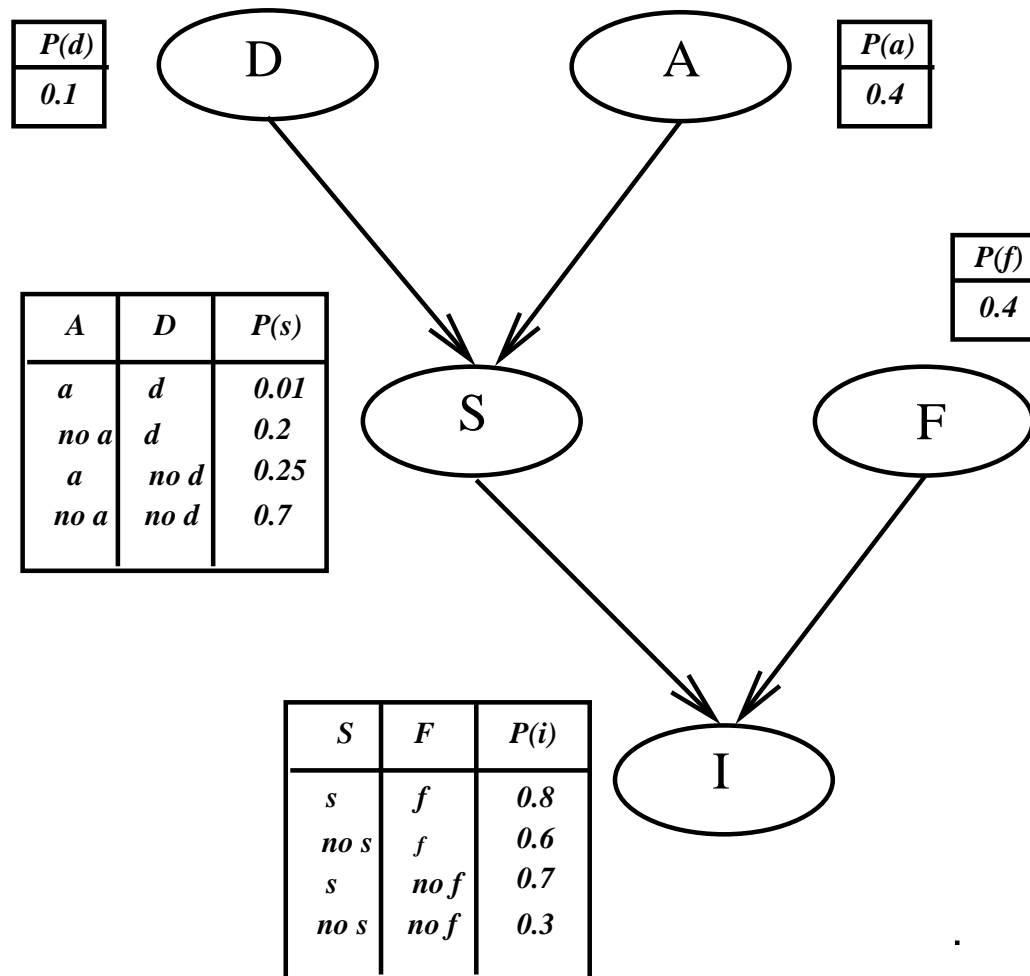
1. Sea  $RED\_E$  el resultado de eliminar de  $RED$  las variables irrelevantes para la consulta realizada
2. Sea  $FACTORES$  igual a vacío
3. Sea  $VARIABLES$  el conjunto de variables de  $RED\_E$
4. Sea  $VAR\_ORD$  el conjunto de  $VARIABLES$  ordenado según un orden de eliminación
5. PARA cada  $V$  en  $VAR\_ORD$  HACER
  - 5.1 Sea  $FACTOR$  el factor correspondiente a  $VAR$  (respecto de  $e$ )
  - 5.2 Añadir  $FACTOR$  a  $FACTORES$
  - 5.3 Si  $VAR$  es una variable oculta hacer  $FACTORES$  igual a  $AGRUPA(VAR, FACTORES)$
6. Devolver  $NORMALIZA(MULTIPLICA(FACTORES))$

## Otro ejemplo (Béjar)

- Consideremos las siguientes variables aleatorias:
  - $D$ : práctica deportiva habitual
  - $A$ : alimentación equilibrada
  - $S$ : presión sanguínea alta
  - $F$ : fumador
  - $I$ : ha sufrido un infarto de miocardio
- Las relaciones causales y el conocimiento probabilístico asociado están reflejadas en la siguiente red bayesiana



## Otro ejemplo: red bayesiana



## Ejemplo de inferencia probabilística

- Podemos usar la red bayesiana para calcular la probabilidad de ser fumador si se ha sufrido un infarto y no se hace deporte,  $P(F|i, \neg d)$
- Directamente:
  - Aplicamos la fórmula:  $P(F|i, \neg d) = \alpha P(F, i, \neg d) = \alpha \sum_{S,A} P(F, i, \neg d, A, S)$
  - Factorizamos según la red:  $P(F|i, \neg d) = \alpha \sum_{S,A} P(\neg d)P(A)P(S|\neg d, A)P(F)P(i|S, F)$
  - Sacamos factor común:  $P(F|i, \neg d) = \alpha P(\neg d)P(F) \sum_A P(A) \sum_S P(S|\neg d, A)P(i|S, F)$
- Calculemos:
  - Para  $F = \text{true}$ :

$$\begin{aligned} P(f|i, \neg d) &= \alpha \cdot P(\neg d) \cdot P(f) [P(a) \cdot (P(s|\neg d, a) \cdot P(i|s, f) + P(\neg s|\neg d, a) \cdot P(i|\neg s, f)) + \\ &\quad + P(\neg a) \cdot (P(s|\neg d, \neg a) \cdot P(i|s, f) + P(\neg s|\neg d, \neg a) \cdot P(i|\neg s, f))] = \\ &= \alpha \cdot 0,9 \cdot 0,4 \cdot [0,4 \cdot (0,25 \cdot 0,8 + 0,75 \cdot 0,6) + 0,6 \cdot (0,7 \cdot 0,8 + 0,3 \cdot 0,6)] = \alpha \cdot 0,253 \end{aligned}$$

- Análogamente, para  $F = \text{false}$ ,  $P(\neg f|i, \neg d) = \alpha \cdot 0,274$
- Normalizando,  $P(F|i, \neg d) = \langle 0,48, 0,52 \rangle$

## Aplicando eliminación de variables

- Seguiremos el siguiente orden de variables, inspirado en la topología de la red (de abajo a arriba):  $I, F, S, A, D$ 
  - Aunque igual otro orden sería más eficiente, pero eso no lo sabemos a priori
- Variable  $I$ :
  - El factor  $f_I(S, F)$  es  $P(i|S, F)$  (no depende de  $I$  ya que su valor está determinado a  $i$ ):

S	F		$f_{\{I\}}(S, F) = P(i S, F)$
-----			
s	f		0.8
no s	f		0.6
s	no f		0.7
no s	no f		0.3

- $FACTORES = \{f_I(S, F)\}$

## Aplicando eliminación de variables

- Variable  $F$ :

- El factor  $f_F(F)$  es  $P(F)$ :

F		$f_{\{F\}}(F)=P(F)$
-----		
f		0.4
no f		0.6

- $FACTORES = \{f_I(S, F), f_F(F)\}$

## Aplicando eliminación de variables

- Variable  $S$ :

- El factor  $f_S(S, A)$  es  $P(S, \neg d, A)$  (no depende de  $D$  ya que su valor está determinado a  $\neg d$ ):

S	A		$f_{\{S\}}(S, A) = P(S   \text{no } d, A)$
-----			
s	a		0.25
s	no a		0.7
no s	a		0.75
no s	no a		0.3

- $FACTORES = \{f_I(S, F), f_F(F), f_S(S, A)\}$

## Aplicando eliminación de variables

- Como  $S$  es una variable oculta, agrupamos por  $S$
- Para ello, primero multiplicamos  $f_I(S, F)$  y  $f_S(S, A)$  obteniendo  $g(S, A, F)$

S	A	F		$g(F, A) = P(S   \text{no d}, A) \times P(i   S, F)$
-----				
s	a	f		0.25 x 0.8
s	no a	f		0.7 x 0.8
s	a	no f		0.25 x 0.7
s	no a	no f		0.7 x 0.7
no s	a	f		0.75 x 0.6
no s	no a	f		0.3 x 0.6
no s	a	no f		0.75 x 0.3
no s	no a	no f		0.3 x 0.3

## Aplicando eliminación de variables

- Y ahora, sumamos  $g(S, A, F)$  por la variable  $S$ , obteniendo  $h(A, F)$

A	F		$h(A, F) = \text{SUM}_{\{S\}} g(S, A, F)$
-----			
a	f		$0.25 \times 0.8 + 0.75 \times 0.6 = 0.65$
no a	f		$0.7 \times 0.8 + 0.3 \times 0.6 = 0.74$
a	no f		$0.25 \times 0.7 + 0.75 \times 0.3 = 0.4$
no a	no f		$0.7 \times 0.7 + 0.3 \times 0.3 = 0.58$

- Acabamos de eliminar la variable  $S$
- $FACTORES = \{h(A, F), f_F(F)\}$  (nótese que  $f_F(F)$  no se ha usado)

# Aplicando eliminación de variables

- Variable  $A$ :

- El factor  $f_A(A)$  es  $P(A)$ :

A		$f_{\{A\}}(A)=P(A)$
-----		
a		0.4
no a		0.6

- $FACTORES = \{f_A(A), h(A, F), f_F(F)\}$



## Aplicando eliminación de variables

- Como  $A$  es una variable oculta, agrupamos por  $A$ 
  - Para ello, primero multiplicamos  $f_A(A)$  y  $h(A, F)$  obteniendo  $k(A, F)$

A	F		$k(F, A) = P(A) \times h(A, F)$
a	f		$0.4 \times 0.65 = 0.26$
no a	f		$0.6 \times 0.74 = 0.444$
a	no f		$0.4 \times 0.4 = 0.16$
no a	no f		$0.6 \times 0.58 = 0.348$

- Y ahora, sumamos  $k(A, F)$  por la variable  $A$ , obteniendo  $l(F)$  (y eliminando, por tanto, la variable  $S$ )

F		$l(F) = \text{SUM}_{\{A\}} k(A, F)$
f		$0.26 + 0.444 = 0.704$
no f		$0.16 + 0.348 = 0.508$

- $FACTORES = \{l(F), f_F(F)\}$

## Aplicando eliminación de variables

- Variable  $D$ :
  - Factor  $f_D()$  (no depende de  $D$ , ya que su valor está fijado a  $\neg d$ , por tanto se trata de una tabla con una única entrada): 0,9
  - $FACTORES = \{f_D(), l(F), f_F(F)\}$
- Ultimo paso: multiplicamos y normalizamos
  - Obsérvese que sólo hasta este paso hacemos uso del factor correspondiente a  $F$
  - Multiplicación

F		$m(F)=f_d() \times l(F) \times f_F(F)$
-----		
f		$0.9 \times 0.704 \times 0.4 = 0.253$
no f		$0.9 \times 0.508 \times 0.6 = 0.274$

- Normalizando obtenemos finalmente:  $P(F|i, \neg d) = \langle 0,48, 0,52 \rangle$
- Por tanto, la probabilidad de ser fumador, dado que se ha tenido un infarto y no se hace deporte, es del 48 %

## Complejidad del algoritmo de eliminación de variables

- La complejidad del algoritmo (tanto en tiempo como en espacio) está dominada por el tamaño del mayor factor obtenido durante el proceso
- Y en eso influye el orden en el que se consideren las variables (*orden de eliminación*)
  - Podríamos usar un criterio heurístico para elegir el orden de eliminación
  - En general, es conveniente moverse “desde las hojas hacia arriba” (consistentemente con la topología de la red)
- Si la red está *simplemente conectada* (*poliárbol*) se puede probar que la complejidad del algoritmo (en tiempo y espacio) es *lineal* en el tamaño de la red (el número de entradas en sus tablas)
  - Una red está simplemente conectada si hay a lo sumo un camino (no dirigido) entre cada dos nodos

## Complejidad de la inferencia exacta

- Pero en general, el algoritmo tiene complejidad exponencial (en tiempo y espacio) en el peor de los casos
- Cuando la inferencia exacta se hace inviable, es esencial usar métodos aproximados de inferencia
- Métodos estocásticos, basados en muestreos que simulan las distribuciones de probabilidad de la red

# Aplicaciones de las redes bayesianas

- Aplicaciones en empresas
  - Microsoft: *Answer Wizard (Office)*, diagnóstico de problemas de impresora,...
  - Intel: Diagnóstico de fallos de procesadores
  - HP: Diagnóstico de problemas de impresora
  - Nasa: Ayuda a la decisión de misiones espaciales
- Otras aplicaciones: diagnóstico médico, *e-learning*,...

## Bibliografía

- Russell, S. y Norvig, P. *Inteligencia artificial (Un enfoque moderno)*, segunda edición (Prentice–Hall Hispanoamericana, 2004)
  - Cap. 14: “Razonamiento Probabilístico”