INTELIGENCIA ARTIFICIAL (INF371)

RAZONAMIENTO PROBABILÍSTICO CON REDES BAYESIANAS (PARTE 2)

Dr. Edwin Villanueva Talavera

Contenido

- Redes Bayesianas
  - Representación de conocimiento
  - Semántica de Redes Bayesianas
  - Construyendo una Red Bayesiana
  - Inferencia en Redes Bayesianas

Bibliografía:

Capitulo 14.1, 14.2, 14.4 del libro:

Stuart Russell & Peter Norvig "Artificial Intelligence: A modern Approach", Prentice Hall, Third Edition, 2010

Redes Bayesianas

- Modelos para especificar de forma concisa la distribución de probabilidad conjunta de un conjunto de variables aleatorias explotando las relaciones de independencia condicional entre variables
- Especificación de una RB:
  - Un Grafo Acíclico Direccionado (DAG) con n nodos representando n variables aleatorias $[X_1, X_2, \ldots, X_n]$
  - Cada arista $X_i \rightarrow X_j$ en el DAG representa una dependencia directa entre las variables, donde X_i se llama padre de X_j (X_j es hijo de X_i)
  - Cada nodo X_i tiene asociado una distribución de probabilidad condicional $\mathbf{P}(X_i \mid \mathit{Parents}(X_i))$ que cuantifica el efecto de los padres en el nodo. Cuando las variables son discretas tales distribuciones corresponden a tablas de probabilidad condicional (CPTs)

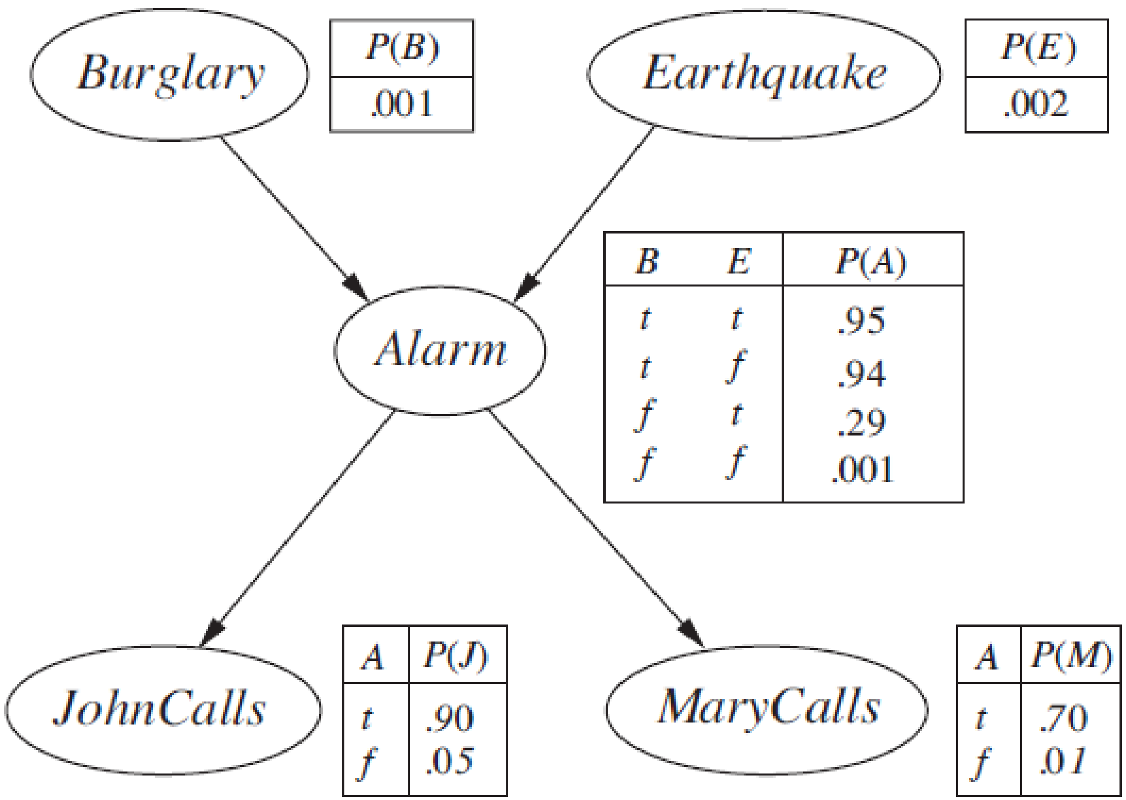
Redes Bayesianas

Ejemplo:

- "Tienes una alarma en casa que es confiable detectando intrusos, pero también puede activarse con temblores. Tienes dos vecinos: John y Mary que han prometido llamarte cuando escuchen la alarma. John confunde a veces la alarma con el timbre del teléfono. Mary le gusta escuchar música en alto volumen y a veces no escucha la alarma. ¿Cómo representamos este conocimiento en una RB)?"
- Variables: Robo (*Burglary*), Temblor (*Earthquake*), Alarma (*Alarm*), LlamadaJohn (*JohnCalls*), LlamadaMary (*MaryCalls*)
- Se tiene cierto conocimiento causal para construir la topología de la RB:
  - Un robo puede activar la alarma
  - Un temblor puede activar la alarma
  - La alarma provoca que Mary llame
  - La alarma provoca que John llame

Redes Bayesianas

Ejemplo:



P(B)
.001

P(E)
.002

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

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Redes Bayesianas

- Las probabilidades resumen un conjunto potencialmente infinito de circunstancias:
  - María no oyó la alarma porque estaba escuchando música
  - John llama cuando timbra el teléfono
  - Un ratón activó la alarma
  - John y Mary no están en casa, etc.
- Es difícil lidiar con todas las situaciones por causa de: laxitud e ignorancia!

Redes Bayesianas

Tablas de probabilidad condicional (CPT):

- Cada fila de un CPT contiene la distribución de probabilidades para una combinación de valores de las variables condicionantes (padres del nodo).
- Si el nodo es booleano, cada fila de su CPT será un número *p* relativo a la probabilidad de X_i =*true*. La probabilidad de X_i =*false* no se indica, ya que se puede calcular como 1-*p*
- El CPT de un nodo sin padres es solo una fila con probabilidades "a priori"
- Un CPT de una variable booleana con *k* padres posee 2^k probabilidades
- Si una RB posee n nodos booleanos con *k* padres como máximo, para especificar la red completamente se necesita $O(n\ 2^k)$ probabilidades (parámetros de la RB)
  - Esto es, crece linealmente con n (la especificación de la probabilidad conjunta total requeriría $O(2^n)$ valores)

Redes Bayesianas

Semántica de las RBs: dos formas de entenderlas

- Semántica Global (o numérica): Se entiende las RBs como una forma de representar la distribución de probabilidad conjunta de un conjunto de variables aleatorias
  - Indica como se puede construir una RB
- Semántica Local (o topológica): Se entiende las RBs como una forma de codificar una serie de aseveraciones de independencia condicional
  - Indica como se puede hacer inferencia en una RB

Redes Bayesianas

Semántica Global

- La RB define la distribución de probabilidad conjunta total como el producto de distribuciones condicionales locales de cada nodo:

  Así, para n variables $[X_1, X_2, ..., X_n]$, una RB permite calcular la prob. de cualquier conjunción $(X_1=x_1 \wedge \ldots \wedge X_n=x_n)$ (denotado $P(x_1, \ldots, x_n)$):

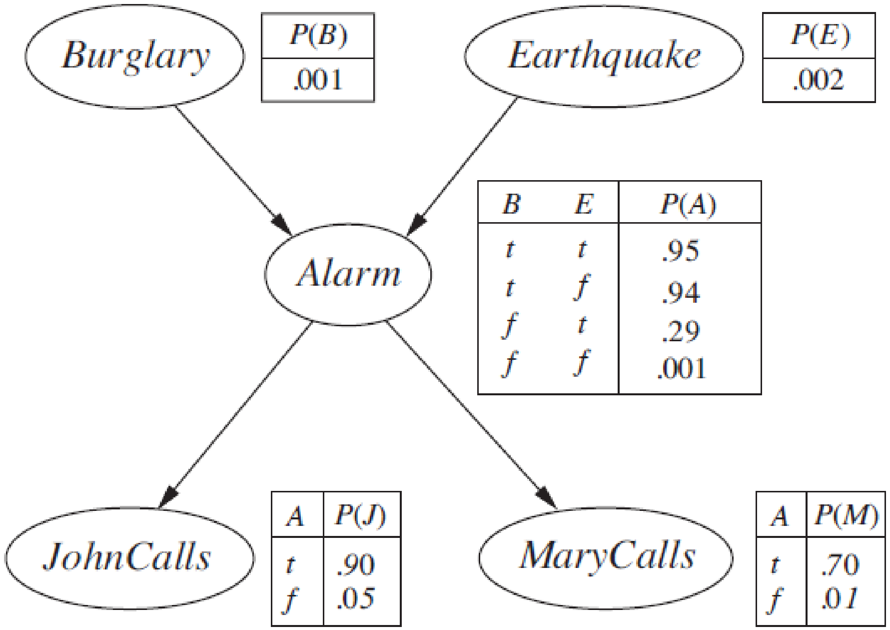
  $$P(x_1, \ldots, x_n) = \prod_{i=1 \ldots n} P(x_i \mid parents(X_i))$$

  donde $parents(X_i)$ denota los valores específicos de las variables padres de X_i, $Parents(X_i)$, que aparecen en la conjunción $x_1, \ldots, x_n$

- Al ser una RB especificación de la distribución de probabilidad conjunta de una serie de variables, esta puede ser usada para responder cualquier consulta acerca de las variables efectuándose el productorio de las probabilidades cond. locales

Redes Bayesianas

Semántica Global: Ejemplo



Cuánto es $P(j, m, a, \neg b, \neg e)$?

$= P(j \mid a)\, P(m \mid a)\, P(a \mid \neg b, \neg e)\, P(\neg b)\, P(\neg e)$

$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$

$= 0.00063$

Redes Bayesianas

Semántica Local (Topológica)

- La topología de una RB nos indica ciertas independencias condicionales:
  - Un nodo X_i es condicionalmente independiente (CI) de cualquier subconjunto **Z** de sus no descendientes (**ND**(X_i)) dados sus padres **Parents**(X_i):

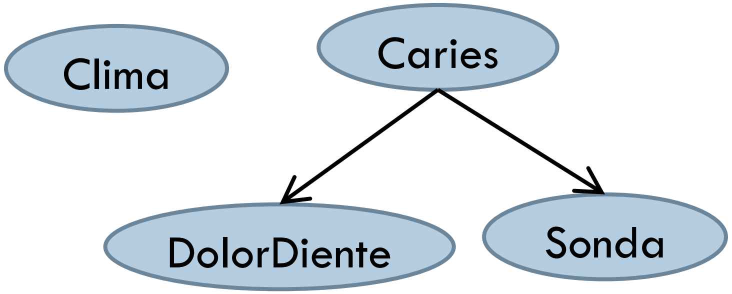
$$X_i \perp \mathbf{Z} \mid \textbf{Parents}(X_i)\ , \qquad \mathbf{Z} \subset \textbf{ND}(X_i)$$

  - Ausencia de arista entre dos nodos X_i, X_j indica que existe una independencia condicional entre ellos dado algún subconjunto Z de otras variables de la RB (puede ser vacío):

$$X_i \perp X_j \mid \mathbf{Z}\ , \qquad \mathbf{Z} \subset \mathbf{U} \backslash \{X_i, X_j\}$$

Redes Bayesianas

Semántica Local (Topológica): Ejemplo

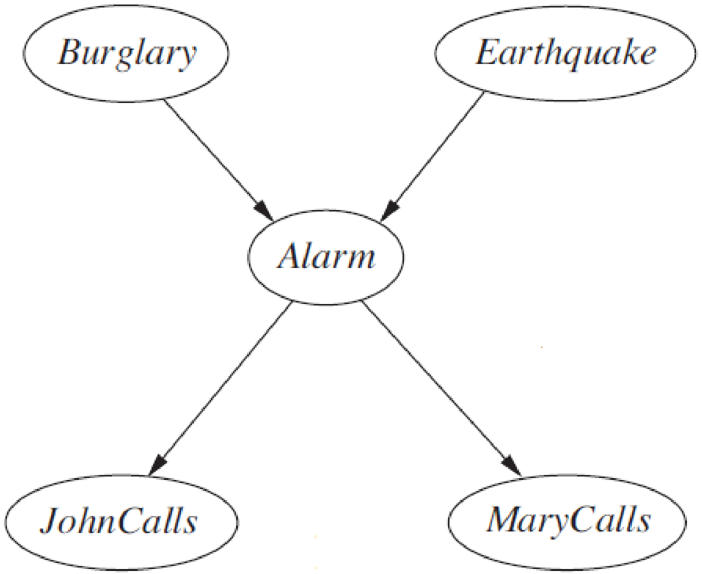


- Independencias condicionales implicadas:
  - *Clima* es marginalmente independiente de cualquier otra(s) variable(s):

    $Clima \perp \mathbf{Z}, \quad \mathbf{Z} \subset \{Caries, DolorDiente, Sonda\}$
  - $DolorDiente \perp Sonda \mid Caries$

Redes Bayesianas

Semántica Local (Topológica): Ejemplo

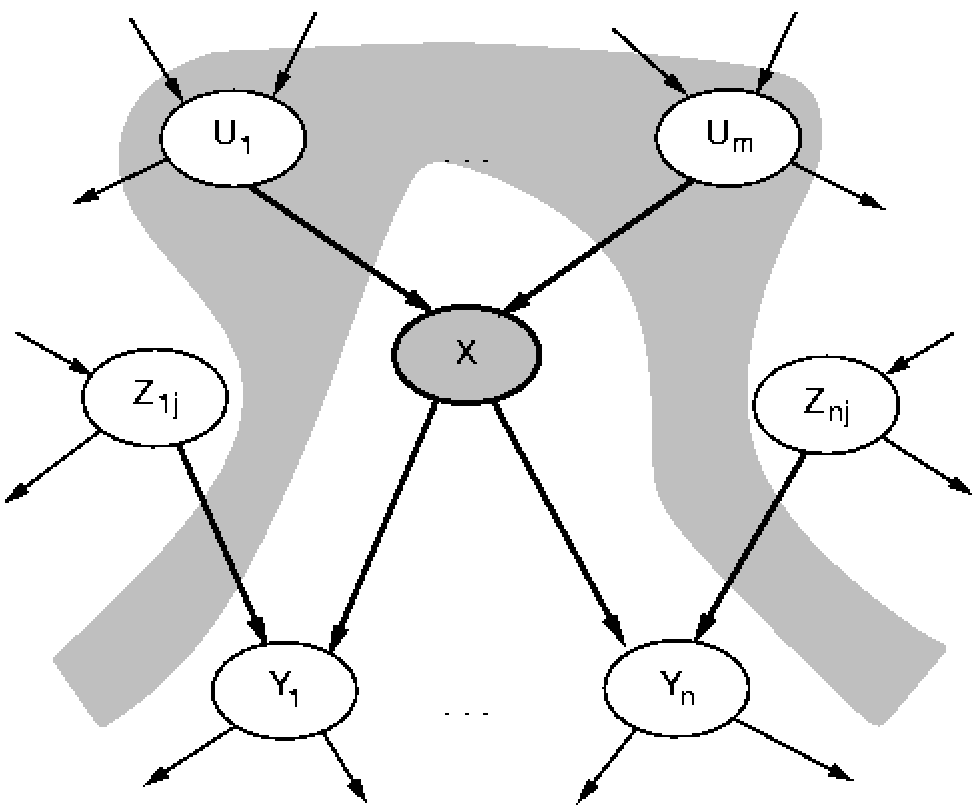


- Independencias condicionales implicadas:
  - *JohnCalls* $\perp$ *Burglary* | *Alarm*; *MaryCalls* $\perp$ *Burglary* | *Alarm*
  - *JohnCalls* $\perp$ *Earthquake* | *Alarm*; *MaryCalls* $\perp$ *Earthquake* | *Alarm*
  - *JohnCalls* $\perp$ *MaryCalls* | *Alarm*; *Burglary* $\perp$ *Earthquake*

Redes Bayesianas

Semántica Local (Topológica)

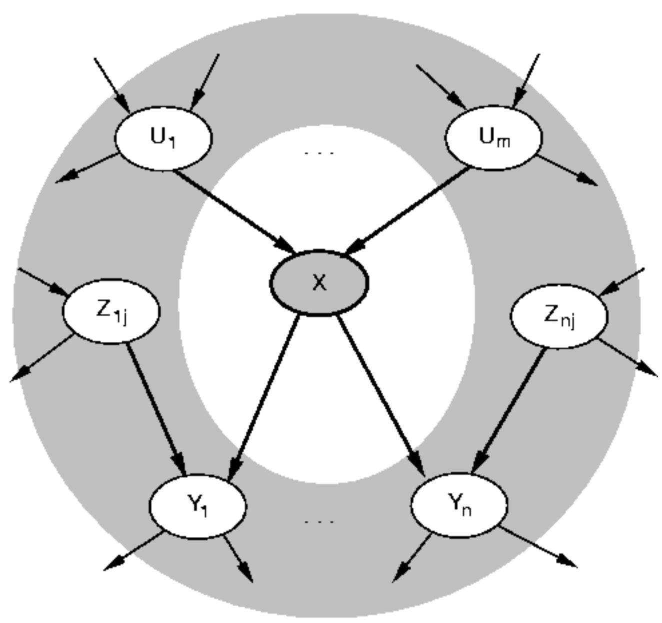
- Interpretación gráfica de: $X \perp \mathbf{Z} \subset \mathbf{ND}(X) \mid \mathbf{Parents}(X)$



Redes Bayesianas

Semántica Local (Topológica)

- **Manto de Markov** de un nodo X (**MB**(X)): Todos los padres, los hijos y los padres de los hijos (esposos) de X
- Un nodo X es condicionalmente independiente del resto de nodos de la RB dado el **MB**(X) : $X \perp \mathbf{Z} \mid \mathbf{MB}(X)$, $\quad \mathbf{Z} \subset \mathbf{U} \setminus \{X \cup \mathbf{MB}(X)\}$
  - Toda la información para inferir acerca de X esta en su Manto de Markov



Redes Bayesianas

Construyendo una RB

1. Escoger un orden de las variables (orden ancestral): $X_1, \ldots, X_n$
   Una buena opción es colocar las variables causa precediendo sus efectos en el orden

2. Para $i = 1 \ldots n$

   - Del conjunto $X_1, \ldots, X_{i-1}$ escoger un conjunto mínimo de padres (**Parents**(X_i)) para X_i tal que se cumpla lo siguiente:

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

     Intuitivamente los padres deben ser aquellos que directamente influyen en X_i

   - Para cada padre escogido colocar una arista hasta X_i

3. Escribir las tablas de probabilidades para cada nodo X_i:
   $\mathbf{P}(X_i \mid \mathbf{Parents}(X_i))$

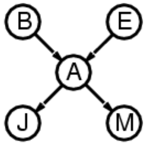
Redes Bayesianas

Construyendo una RB

- El orden ancestral correcto es aquel donde las variables causas de raíz aparecen primero, luego las variables que ellas influencian y así sucesivamente hasta llegar a las variables hojas que no tienen influencia causal directa sobre las otras variables
- Qué pasa si no se escoge un orden ancestral correcta?

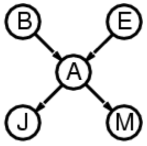
Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M



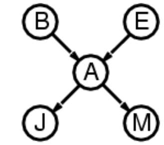
Redes Bayesianas

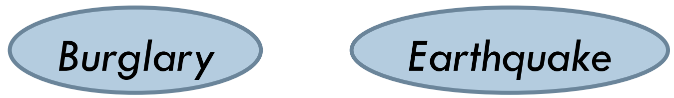
Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M



Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M

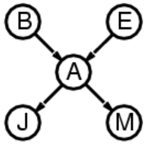


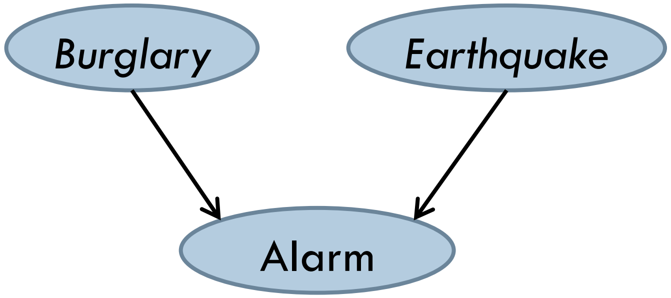


$P(E \mid B) = P(E)$? **Si**

Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M





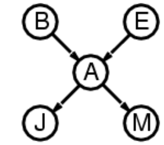
$P(E \mid B) = P(E)$? **Si**

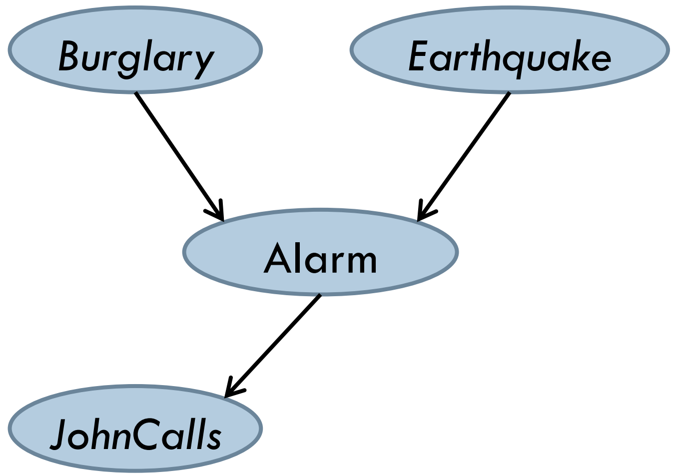
$P(A \mid B, E) = P(A)$? **No,** $P(A \mid B) = P(A)$? **No,** $P(A \mid E) = P(A)$? **No,**

⇒ *B, E* son padres de A

Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M





$P(E \mid B) = P(E)$? **Si**

$P(A \mid B, E) = P(A)$? **No** $P(A \mid B) = P(A)$? **No,** $P(A \mid E) = P(A)$? **No,**

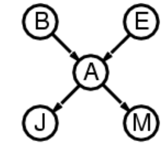
⇒ *B, E* son padres de A

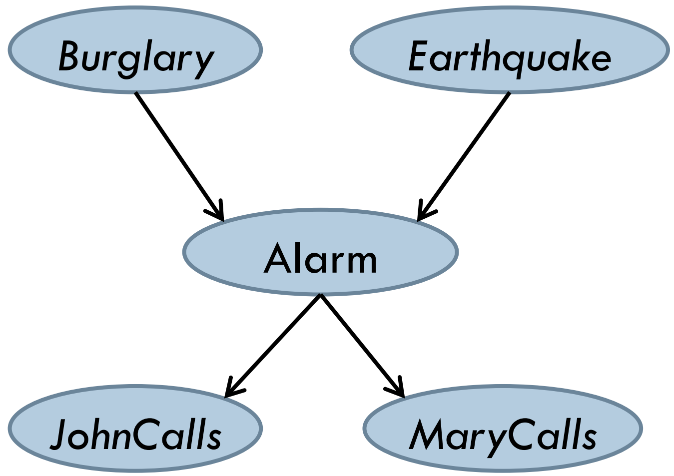
$P(J \mid B, E, A) = P(J)$? **No,** $P(J \mid B, E, A) = P(J \mid A)$? **SI**

⇒ *A* es padre de J

Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden B, E, A, J, M





$P(E \mid B) = P(E)$? **Si**

$P(A \mid B, E) = P(A)$? **No** $P(A \mid B) = P(A)$? **No,** $P(A \mid E) = P(A)$? **No,**

⇒ *B, E* son padres de A

$P(J \mid B, E, A) = P(J)$? **No,** $P(J \mid B, E, A) = P(J \mid A)$? **SI**

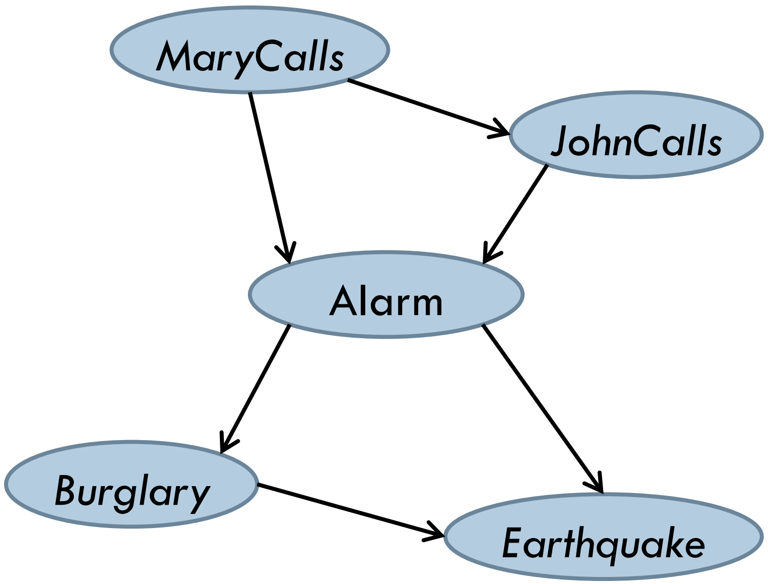
⇒ *A* es padre de J

$P(M \mid B, E, A, J) = P(M)$? **No,** $P(M \mid B, E, A, J) = P(M \mid A)$? **SI**

⇒ *M* es padre de J

Redes Bayesianas

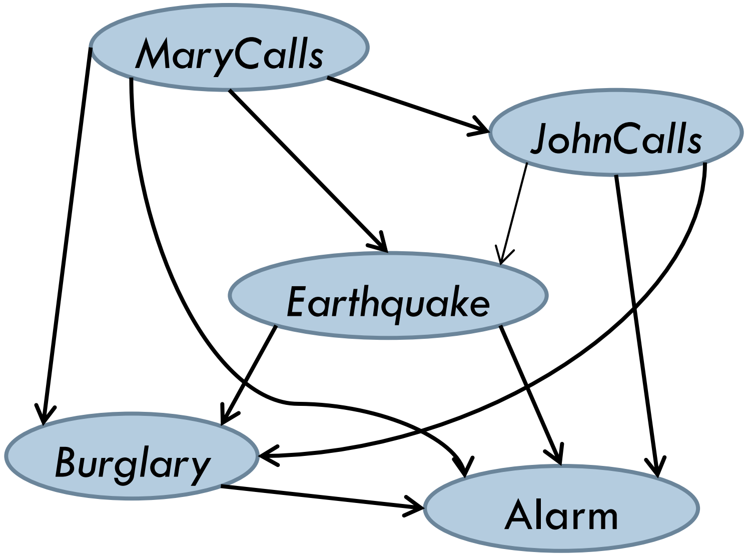
Construyendo una RB: Ejemplo asumiendo orden *M, J, A, B, E*



- La RB resultante tendrá dos aristas mas que la red original y requerirá mas parámetros (probabilidades condicionales) para ser especificada
- Algunas aristas pueden representar relaciones tenues que implican probabilidades difíciles o antinaturales de determinar
  - Ej. $P(Earthquake \mid Alarm, Burglary)$
- En general es preferible tratar de construir la RB de causas para efectos (modelo causal) y no al contrario (modelo de diagnóstico)

Redes Bayesianas

Construyendo una RB: Ejemplo asumiendo orden *M, J, E, B, A*



- El orden *M, J, E, B, A* no es recomendable, ya que genera una RB con 9 aristas (5 mas que la RB original) y requerirá especificar 31 probabilidades distintas, lo mismo que se necesita para especificar la distribución de probabilidad conjunta
- Sin embargo, las tres RB codifican la misma distribucion de prob. conjunta, solo que las dos ultimas no representan todas las independencias condicionales que existen

Inferencia en Redes Bayesianas

- Inferencia en una rede bayesiana es la tarea de calcular la distribución de probabilidad posterior de una variable de consulta dados los valores de un conjunto de variables (evidencia)

$$\mathbf{P}(X \mid \mathbf{E} = \mathbf{e})$$

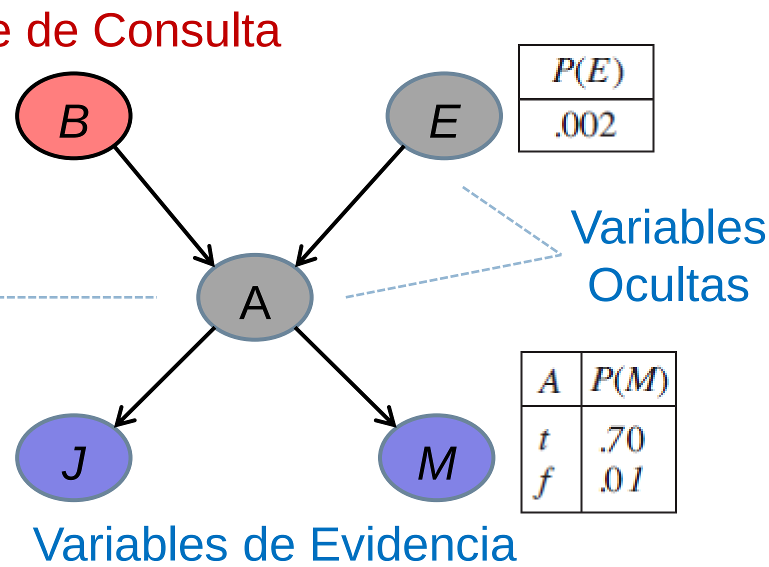
Notación

- X: variable de consulta
- $\mathbf{E}$: variables de evidencia
- $\mathbf{e}$: evidencia (atribución particular de las variables $\mathbf{E}$)
- $\mathbf{Y}$: variables no evidenciadas (ocultas)

El universo de variables de la RB es $\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$

Inferencia en Redes Bayesianas

Ejemplo:



$P(B)$
.001

$P(E)$
.002

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

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

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

Se quiere encontrar $\mathbf{P}(B \mid j, m)$?

- variable de consulta = B
- variables de evidencia = $\{J, M\}$
- evidencia = $\{j, m\}$
- variables ocultas = $\{E, A\}$

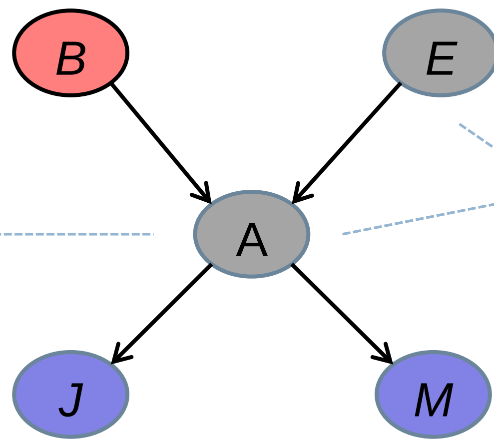
Inferencia en Redes Bayesianas

Ejemplo: **P** $(B \mid j, m)$?

Variable de Consulta

$P(B)$
.001

$P(E)$
.002



Variables Ocultas

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

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

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

De forma general: $\mathbf{P}(X \mid \mathbf{e}) = \alpha \, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$

Usando la semantica de la RB

En el ejemplo: $\mathbf{P}(B \mid j, m) = \alpha \sum_e \sum_a \mathbf{P}(B, j, m, e, a)$

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

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

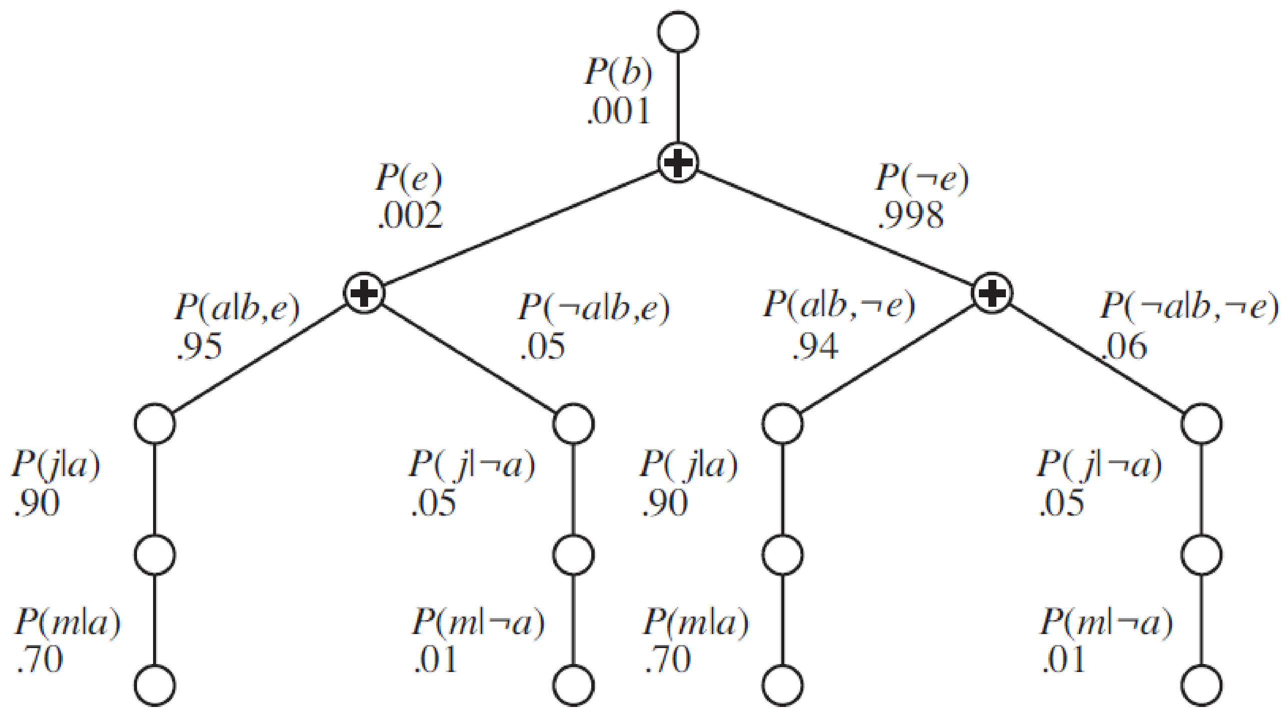
$P(b \mid j, m) = 0.0005922\ \alpha$; $P(\neg b \mid j, m) = 0.00149\ \alpha$; $\mathbf{P}(B \mid j, m) = [\mathbf{0.28, 0.72}]$

Inferencia en Redes Bayesianas

Inferencia por Enumeración

- Evalúa en profundidad el árbol de expresión

$$P(b \mid j, m) = \alpha\, P(b)\, \Sigma_e P(e)\, \Sigma_a P(a \mid b,e)\, P(j \mid a)\, P(m \mid a)$$



- Complejidad espacial es lineal en el numero de variables, pero la complejidad de tiempo es exponencial. Para n variables booleanas $O(2^n)$. Hay cálculos repetidos, Ej. $P(j \mid a)\, P(m \mid a)$

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables

- Elimina cálculos repetidos de la inferencia por enumeración
- Idea: hacer cálculos solo una vez y guardar los resultados
- Se convierte el problema en operaciones con factores.

Ej. RB Alarm:

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

  - Factor es una matriz definida en un conjunto de variables (no evidenciadas) que se indexa con sus valores. Ejemplos:

$f_4(A) =$

0.9	$P(j \mid a)$
0.05	$P(j \mid \neg a)$

$f_5(A) =$

0.7	$P(m \mid a)$
0.01	$P(m \mid \neg a)$

Em términos de factores:

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

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables

- Operaciones con factores
  - **Producto** (*pointwise product*): El producto de dos factores $\mathbf{f}_1$ y $\mathbf{f}_2$ genera un nuevo factor **f** cuyas variables son la unión de las variables y cuyos elementos son los productos de los elementos correspondientes

A	B	$\mathbf{f}_1(A,B)$
T	T	.3
T	F	.7
F	T	.9
F	F	.1

x

B	C	$\mathbf{f}_2(B,C)$
T	T	.2
T	F	.8
F	T	.6
F	F	.4

=

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

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables

- Operaciones con factores
  - **Suming out una variable**: Esta operación se realiza sumando las submatrices formadas de fijar la variable en cada uno de sus valores

Ej.

$$\sum_a \mathbf{f}_3(A,B,C) = \mathbf{f}_3(a,B,C) + \mathbf{f}_3(\neg a,B,C) \;=\; \mathbf{f}(B,C)$$

B	C	$\mathbf{f}_3(A,B,C)$
T	T	.06
T	F	.24
F	T	.42
F	F	.28

+

B	C	$\mathbf{f}_3(A,B,C)$
T	T	.18
T	F	.72
F	T	.06
F	F	.04

=

B	C	$\mathbf{f}_3(A,B,C)$
T	T	.24
T	F	.96
F	T	.48
F	F	.32

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables

- Factores que no dependen de la variable que se esta sumando dentro del sumatorio puede ser movidos fuera. Por ejemplo

$$\sum_e \mathbf{f}_2(E) \times \mathbf{f}_3(A,B,E) \times \mathbf{f}_4(A) \times \mathbf{f}_5(A) \quad = \mathbf{f}_4(A) \times \mathbf{f}_5(A) \times \sum_e \mathbf{f}_2(E) \times \mathbf{f}_3(A,B,E)$$

- Cualquier orden de eliminación de variables puede dar resultados válidos. Sin embargo, se pueden generar diferentes factores. El tamaño del mayor factor domina la complejidad de tiempo y espacio
- Encontrar una orden de eliminación optima es NP Completo. Una heurística importante es: considerar solamente en el proceso de eliminación las variables que son ancestros de la variable consulta y evidencia (las demás son irrelevantes).

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables: Ejemplo

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

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

$$\mathbf{P}(B \mid j, m) = \alpha \, \mathbf{f}_1(B) \times \underbrace{\sum_e \mathbf{f}_2(E) \times \mathbf{f}_6(B, E)}$$

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

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

Inferencia en Redes Bayesianas

Algoritmo de Eliminación de Variables

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 $\mathbf{P}(X_1, \ldots, X_n)$

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

Librerias de Redes Bayesianas

- Pgmpy: http://pgmpy.org/ (recomendado)
- Libpgm: https://pythonhosted.org/libpgm/
- http://www.bayespy.org (solo inferencia)
- Bnlearn (muy completa, em R)

Preguntas?