Modelos Gráficos Probabilísticos

Tema 4. Problemas de inferencia en modelos gráficos: Cálculo de probabilidades condicionadas de forma exacta

Andrés Cano Utrera

UNIVERSIDAD DE GRANADA

ETSIIT Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación

DECSAI

Curso 2024-2025

Referencias I

- E. Castillo, J.M. Gutiérrez, A.S. Hadi (1997) *Sistemas Expertos y Modelos de Redes Probabilísticas*. Monografías de la Academia de Ingeniería. Academia de Ingeniería, Madrid. (disponible libremente en http://personales.unican.es/gutierjm/papers/BookCGH.pdf)
- F.V. Jensen, T. D. Nielsen (2007) *Bayesian Networks and Decision Graphs*, 2nd Edi. Springer-Verlag, Nueva York. (disponible en pdf desde red de la UGR)
- D. Koller, N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques.* MIT Press.
- R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter (1999) *Probabilistic Networks and Expert Systems*. Springer-Verlag, Nueva York. (disponible en pdf desde red de la UGR)

Referencias II

- U.B. Kjaerulff, A.L. Madsen (2008) *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. Springer Verlag, Nueva York. (disponible en pdf desde red de la UGR)
- F.V. Jensen (1996) *An Introduction to Bayesian Networks*. UCL Press, Londres.
- J. Pearl (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

Índice

Contenido del tema

Objetivo Básico

Cálculo con Redes Bayesianas

- Cálculo de una probabilidad condicionada
- Cálculo de la configuración más probable
- Análisis de la sensibilidad
  - A los parámetros
  - A las observaciones

Aquí veremos cálculo de la probabilidad condicionada de forma exacta. Más adelante veremos algoritmos aproximados y el problema de la explicación más probable.

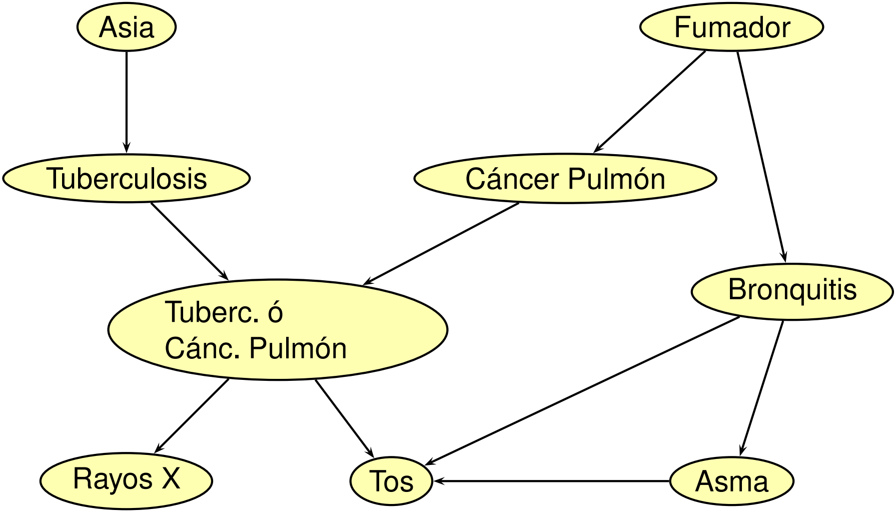
Objetivo Básico

Cálculo con Redes Bayesianas

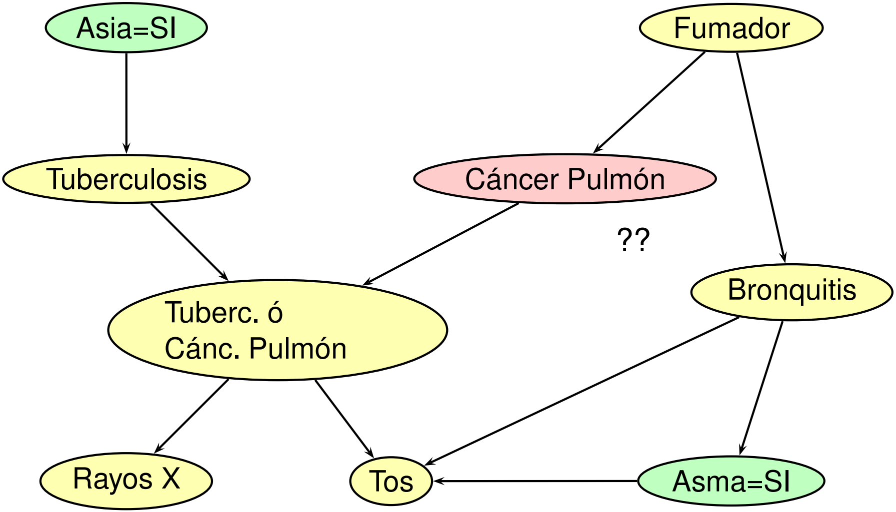
- Cálculo de una probabilidad condicionada
- Cálculo de la configuración más probable
- Análisis de la sensibilidad
  - A los parámetros
  - A las observaciones

Aquí veremos cálculo de la probabilidad condicionada de forma exacta. Más adelante veremos algoritmos aproximados y el problema de la explicación más probable.

El Problema de la Probabilidad Condicional



El Problema de la Probabilidad Condicional



El Problema de la Probabilidad Condicional

- Tenemos una red que representa un modelo de una situación.
- Para un caso concreto hemos realizado algunas observaciones, por ejemplo, Visita Asia, Asma.
- Estamos interesados en calcular la probabilidad de alguna variable, por ejemplo, Cáncer de Pulmón condicionado a las observaciones de las que disponemos.

Problemas

Aplicaciones de la probabilidad condicionada

- Diagnosis: Razonar desde los síntomas (efectos) hacia las causas. ¿Cuál es la prob. de padecer una enfermedad dados ciertos síntomas? Ver red carstart, Prostanet
- Predicción: Razonar desde las causas hacia los síntomas (efectos) ¿cuál es la probabilidad de que se manifieste un síntoma dado que se padece una enfermedad?. Ver redes Hailfinder, insurance.
- Intercausal: Razonar sobre las causas mutuas de un efecto común.

Reglas de Cálculo: Resumen

- Marginalización: Cálculo de $P(X)$ desde $P(X,Y)$:

$$\begin{aligned} P(X) &= P(X, Y=y_1) + P(X, Y=y_2) + \cdots + P(X, Y=y_m) \\ &= \sum_i^m P(X, Y=y_i) \end{aligned}$$

- Cálculo de la distribución condicional:

$$\begin{aligned} P(X|Y) &= P(X,Y)/P(Y) \\ P(Y|X) &= P(X,Y)/P(X) \end{aligned}$$

- Combinación marginal y condicional:

$$P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$$

- Regla de la cadena:

$$P(X_1, X_2, \ldots, X_n) = \prod_i P(X_i|X_1, \ldots, X_{i-1})$$

Reglas de Cálculo: Resumen

Ejemplo con dos variables (ver EjemploCalculosAPartirProbabilidadConjunta.xls)

Disponemos de la distribución conjunta para las variables Temperatura (alta, media y baja) y Ambiente (seco, nublado y lluvioso).
Podemos usarla para hacer varios cálculos empleando las anteriores reglas de cálculo.

Representación

Según vimos en el tema 3:

- La forma más directa de representar una distribución de probabilidad es mediante una tabla (*array*) con tantas dimensiones como variables. Denominaremos **potencial** a una función que asigna un valor de probabilidad a cada configuración de sus variables.
- Para cada dimensión correspondiente a una variable hay un índice con tantos valores como casos posibles tiene esta variable.

Ejemplo. Distribución de probabilidad condicionada

Una distribución de probabilidad condicionada sobre X *Cáncer de Pulmón* condicionada a las variables Y *Fumador* y Z *Sexo* se puede representar como un potencial con lista de variables: X, Y, Z y tabla:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.5	0.4	0.2	0.1
X=No	0.5	0.6	0.8	0.9

Ejemplo. Distribución Conjunta

Sean X *Cáncer de Pulmón*, Y *Fumador* y Z *Sexo*.
Un potencial puede representar también una distribución conjunta:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.024	0.018
X=No	0.14	0.252	0.096	0.162

Desde un punto de vista sintáctico, este potencial tiene la misma estructura que el anterior.

Ejemplo. Observaciones

Supongamos que tenemos una variable X con tres posibles valores a, b, c.
Si hemos observado $X = b$, esta observación se puede representar mediante el potencial:

$X = a$	$X = b$	$X = c$
0.0	1.0	0.0

Contenido del tema

Operaciones Básicas con Potenciales

Descripción intuitiva:

- Combinación: Mediante esta operación se unen las informaciones representadas por dos potenciales t_1 y t_2 en un único potencial definido para la unión de las variables de t_1 y t_2.
- Marginalización: Mediante esta operación se transforma la información representada por el potencial t en un potencial t' definido para un subconjunto de las variables de t.
- Selección: Dado un potencial t y un conjunto de observaciones $\mathbf{Y} = \mathbf{y}$, se obtiene un potencial representando la restricción de t a las observaciones $\mathbf{Y} = \mathbf{y}$. Es equivalente a la combinación con un potencial representando las observaciones y la marginalización del resultado sobre las variables no observadas de t.

Combinación

Combinación

Si $p(\mathbf{x},\mathbf{y})$ y $q(\mathbf{y},\mathbf{z})$ son dos potenciales sobre $(\mathbf{X},\mathbf{Y})$ y $(\mathbf{Y},\mathbf{Z})$ respectivamente, entonces su combinación es el potencial $p\cdot q(\mathbf{x},\mathbf{y},\mathbf{z})$ sobre $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ dado por

$$p\cdot q(\mathbf{x},\mathbf{y},\mathbf{z}) = p(\mathbf{x},\mathbf{y})\cdot q(\mathbf{y},\mathbf{z})$$

Ejemplo de combinación

p	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.5	0.4	0.2	0.1
X=No	0.5	0.6	0.8	0.9

q	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
	0.3	0.4	0.2	0.1

$p \cdot q$	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.15	0.16	0.04	0.01
X=No	0.15	0.24	0.16	0.09

Otro ejemplo de combinación

Variables bivaluadas A, B, C. Tenemos $p(A, B), q(A, C)$ y calculamos la combinación:

$$f(a, b, c) = p(a, b) \cdot q(a, c)$$

$$\begin{array}{|c|cc|}\hline & b & \overline{b} \\ \hline a & 0{,}5 & 0{,}8 \\ \overline{a} & 0{,}5 & 0{,}2 \\ \hline \end{array} \times \begin{array}{|c|cc|}\hline & c & \overline{c} \\ \hline a & 1{,}0 & 0{,}4 \\ \overline{a} & 0{,}0 & 0{,}6 \\ \hline \end{array}$$

Resultado:

$$= \begin{array}{|c|cccc|}\hline & b, c & b, \overline{c} & \overline{b}, c & \overline{b}, \overline{c} \\ \hline a & 0{,}50 & 0{,}20 & 0{,}80 & 0{,}32 \\ \overline{a} & 0{,}00 & 0{,}30 & 0{,}00 & 0{,}12 \\ \hline \end{array}$$

Otro ejemplo de combinación

Variables bivaluadas A, B, C. Tenemos $p(A, B), q(A, C)$ y calculamos la combinación:

$$f(a, b, c) = p(a, b) \cdot q(a, c)$$

$$\begin{array}{|c|cc|} \hline & b & \overline{b} \\ \hline a & 0{,}5 & 0{,}8 \\ \overline{a} & 0{,}5 & 0{,}2 \\ \hline \end{array} \times \begin{array}{|c|cc|} \hline & c & \overline{c} \\ \hline a & 1{,}0 & 0{,}4 \\ \overline{a} & 0{,}0 & 0{,}6 \\ \hline \end{array}$$

Resultado:

$$= \begin{array}{|c|cccc|} \hline & b, c & b, \overline{c} & \overline{b}, c & \overline{b}, \overline{c} \\ \hline a & 0{,}50 & 0{,}20 & 0{,}80 & 0{,}32 \\ \overline{a} & 0{,}00 & 0{,}30 & 0{,}00 & 0{,}12 \\ \hline \end{array}$$

Marginalización

Si tenemos un conjunto de variables $\mathbf{Y} = (\mathbf{X}, \mathbf{Z})$, entonces la marginalización permite obtener la distribución de probabilidad sobre $\mathbf{X}$ (distribución marginal) a partir de la de $\mathbf{Y}$.

Marginalización

Si $p(\mathbf{x}, \mathbf{z})$ es un potencial sobre $(\mathbf{X}, \mathbf{Z})$ entonces su marginalización sobre $\mathbf{X}$ es el potencial que se obtiene de la forma:

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z})$$

La marginalización sobre $\mathbf{X}$ se llama también borrado de las variables en $\mathbf{Z}$.

Por ejemplo, si tengo una distribución $p(x, y, z)$ sobre (X, Y, Z), la marginalización sobre (X, Y) se obtiene como $p(x, y) = \sum_z p(x, y, z)$

Ejemplo de marginalización

Sean X *Cáncer de Pulmón*, Y *Fumador* y Z *Sexo*. Supongamos que tenemos la siguiente distribución de probabilidad conjunta

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.024	0.018
X=No	0.14	0.252	0.096	0.162

Ejemplo de marginalización

Sean X *Cáncer de Pulmón*, Y *Fumador* y Z *Sexo*. Supongamos que tenemos la siguiente distribución de probabilidad conjunta

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.024	0.018
X=No	0.14	0.252	0.096	0.162

La marginalización sobre (Y, Z) viene dada por la distribución de probabilidad:

Y= Si	Y = Si	Y = No	Y = No
Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
0.28	0.42	0.12	0.18

Ejemplo de marginalización

La distribución sobre (Y, Z) la podemos marginalizar sobre cualquiera de sus variables.

Y= Si	Y = Si	Y = No	Y = No
Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
0.28	0.42	0.12	0.18

Sobre Y obtenemos

Y = Si	Y = No
0.7	0.3

Sobre Z obtenemos

Z = Hombre	Z = Mujer
0.4	0.6

El resultado de borrar dos variables consecutivas es el mismo que si dichas variables se borran en un solo paso.

Otro ejemplo de marginalización

Suponiendo un potencial f definido sobre tres variables binarias A; B; C la marginalización sobre A es el potencial:

$$f_1(A) = \sum_{B,C} f(A,B,C) = \sum_{B,C} \begin{array}{c} \\ \\ \end{array}\left(\begin{array}{ccccc} & b,c & b,\bar{c} & \bar{b},c & \bar{b},\bar{c} \\ a & 0{,}50 & 0{,}20 & 0{,}80 & 0{,}32 \\ \bar{a} & 0{,}00 & 0{,}30 & 0{,}00 & 0{,}12 \end{array} \right) =$$

$$\left(\begin{array}{cc} a & 1{,}82 \\ \bar{a} & 0{,}42 \end{array} \right)$$

Contenido del tema

Cálculo de Probabilidades

Problema Básico

Tenemos una red bayesiana asociada a un conjunto de probabilidades. El problema fundamental de las redes es: dado un conjunto $\mathbf{O}$ de variables observadas: $\mathbf{O} = \mathbf{o}$ y una variable objetivo Z, queremos calcular $p(z|\mathbf{o})$, para todos los valores de la variable Z.

Fuerza Bruta

Podríamos calcular la distribución conjunta, marginalizarla en las variables $\mathbf{O} \cup \{Z\}$ y entonces calcular la distribución de probabilidad condicionada deseada, pero esto tiene complejidad exponencial en el número de variables.

Gran reto: Calcular la probabilidad condicionada sin tener que calcular la conjunta (usando las distribuciones de cada variable condicionadas a sus padres).

Cálculo sin Observaciones

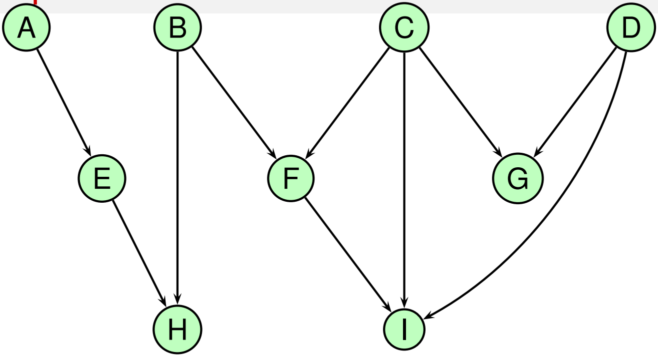
Tenemos un conjunto de variables $\mathbf{X}$ y queremos calcular las probabilidades sobre Z sin observaciones: $p(z)$.
Supongamos que $\mathbf{X} = \mathbf{Y} \cup \{Z\}$
Tenemos que:

$$p(z) = \sum_{\mathbf{y}} p(\mathbf{y}, z)$$

Donde $p(\mathbf{y}, z)$ es la distribución conjunta. Es decir, es la marginalización sobre Z de la distribución conjunta.
La forma más evidente de hacerlo es:

- Combinar todas las distribuciones condicionadas para calcular la conjunta
- Marginalizar sobre Z

Ejemplo



Objetivo: Calcular las probabilidades sobre H.

$$p(h) = \sum_{a,b,c,d,e,f,g,i} p(a,b,c,d,e,f,g,h,i) =$$

$$\sum_{a,b,c,d,e,f,g,i} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

Algoritmo de Borrado. Bases

$$\sum_{a,b,c,d,e,f,g,i} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

Algoritmo de Borrado. Bases

$$\sum_{a,b,c,d,e,f,g,i} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

$$\downarrow$$

$$\sum_{b,c,d,e,f,g,i} \sum_{a} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

Algoritmo de Borrado. Bases

$$\sum_{a,b,c,d,e,f,g,i} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

$$\downarrow$$

$$\sum_{b,c,d,e,f,g,i} \sum_{a} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

$$\downarrow$$

$$\sum_{b,c,d,e,f,g,i} p(b).p(c).p(d).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f) \sum_a p(a).p(e|a)$$

Algoritmo de Borrado. Bases

$$\sum_{a,b,c,d,e,f,g,i} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

$$\downarrow$$

$$\sum_{b,c,d,e,f,g,i} \sum_{a} p(a).p(b).p(c).p(d).p(e|a).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)$$

$$\downarrow$$

$$\sum_{b,c,d,e,f,g,i} p(b).p(c).p(d).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)\textstyle\sum_a p(a).p(e|a)$$

$$r(e) = \textstyle\sum_a p(a).p(e|a) \downarrow$$

$$\sum_{b,c,d,e,f,g,i} p(b).p(c).p(d).p(f|b,c).p(g|c,d).p(h|b,e).p(i|c,d,f)r(e)$$

Tenemos un problema similar, pero con una variable menos

Algoritmo de Borrado

T: Conjunto de potenciales (inicialmente probabilidades condicionadas)
$\mathbf{X}$: variables iniciales
H: variable objetivo
$\mathbf{Y}$: variables iniciales, excepto H

Algoritmo:

1. Para cada variable $Z \in \mathbf{Y}$
    2. Sea T_Z el conjunto de los potenciales en T que contienen la variable Z
    3. Sea q el potencial combinación de todos los potenciales en T_Z
    4. Sea r el resultado de borrar Z en q
    5. Hacer T igual a $(T - T_Z) \cup \{r\}$
6. $p(h)$ es la combinación de todos los potenciales en T

Algoritmo de Borrado: Ejemplo

$T = \{p(a), p(b), p(c), p(d), p(e|a), p(f|b,c), p(g|c,d), p(h|b,e), p(i|c,d,f)\}$

Elegimos variable: A.

Calculamos:

$$r(e) = \sum_a p(a).p(e|a)$$

Calculamos el nuevo conjunto:

$$T = \{p(b), p(c), p(d), p(f|b,c), p(g|c,d), p(h|b,e), p(i|c,d,f), r(e)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(b), p(c), p(d), p(f|b,c), p(g|c,d), p(h|b,e), p(i|c,d,f), r(e)\}$
Elegimos variable: I.

Calculamos:

$$s(c,d,f) = \sum_i p(i|c,d,f)$$

Calculamos el nuevo conjunto:

$$T = \{p(b), p(c), p(d), p(f|b,c), p(g|c,d), p(h|b,e), s(c,d,f), r(e)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(b), p(c), p(d), p(f|b,c), p(g|c,d), p(h|b,e), s(c,d,f), r(e)\}$
Elegimos variable: B.

Calculamos:

$$q(c,e,f,h) = \sum_b p(b).p(f|b,c).p(h|b,e)$$

Calculamos el nuevo conjunto:

$$T = \{p(c), p(d), p(g|c,d), s(c,d,f), r(e), q(c,e,f,h)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(c), p(d), p(g|c,d), s(c,d,f), r(e), q(c,e,f,h)\}$
Elegimos variable: D.

Calculamos:

$$t(c,f,g) = \sum_d p(d).s(c,d,f).p(g|c,d)$$

Calculamos el nuevo conjunto:

$$T = \{p(c), r(e), q(c,e,f,h), t(c,f,g)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(c), r(e), q(c,e,f,h), t(c,f,g)\}$
Elegimos variable: F.

Calculamos:

$$w(c,e,g,h) = \sum_f q(c,e,f,h).t(c,f,g)$$

Calculamos el nuevo conjunto:

$$T = \{p(c), r(e), w(c,e,g,h)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(c), r(e), w(c, e, g, h)\}$
Elegimos variable: E.

Calculamos:

$$m(c, g, h) = \sum_{e} r(e).w(c, e, g, h)$$

Calculamos el nuevo conjunto:

$$T = \{p(c), m(c, g, h)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(c), m(c,g,h)\}$
Elegimos variable: G.

Calculamos:

$$n(c,h) = \sum_g m(c,g,h)$$

Calculamos el nuevo conjunto:

$$T = \{p(c), n(c,h)\}$$

Algoritmo de Borrado: Ejemplo

$T = \{p(c), n(c,h)\}$
Elegimos variable: C.

Calculamos:

$$v(h) = \sum_c p(c).n(c,h)$$

Calculamos el nuevo conjunto:

$$T = \{v(h)\}$$

La probabilidad buscada es: $p(h) = v(h)$

Notas

- Calculamos la probabilidad deseada sin calcular la probabilidad conjunta
- En nuestro caso el número máximo de variables en un potencial es 5
- Las variables se pueden elegir en cualquier orden. El resultado es siempre correcto
- Distintos órdenes pueden producir distinto número de operaciones
- Una buena heurística: elegir la variable más fácil de borrar en cada momento
- El problema es NP-difícil, pero dependiendo de los grafos se pueden resolver problemas con miles de variables

Variables Observadas

Hemos observado $\mathbf{O} = \mathbf{o}$ y queremos calcular $p(z|\mathbf{o})$ para una variable Z.
El algoritmo de borrado calcula: $p(z, \mathbf{o})$ para todos los valores de Z
Después, el valor deseado $p(z|\mathbf{o})$ se obtiene dividiendo cada valor $p(z, \mathbf{o})$ por $\sum_{z'} p(z', \mathbf{o})$ (normalizando).

Para calcular $p(z, \mathbf{o})$ se aplica el mismo algoritmo de antes, pero transformando los potenciales iniciales. El proceso consiste en hacer iguales a cero los valores correspondientes a los valores no observados de las variables $\mathbf{O}$.

Ejemplo. Variables Observadas

Supongamos el potencial p:

p	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.024	0.018
X=No	0.14	0.252	0.096	0.162

Y que hemos observado, $Y = Si$.

Antes del algoritmo tendríamos que transformar p en el potencial:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.0	0.0
X=No	0.14	0.252	0.0	0.0

Ejemplo 2. Variables Observadas

Supongamos el potencial p:

p	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.168	0.024	0.018
X=No	0.14	0.252	0.096	0.162

Y que hemos observado, $Y = Si, Z = Hombre$.

Antes del algoritmo tendríamos que transformar p en el potencial:

	Y= Si	Y = Si	Y = No	Y = No
	Z= Hombre	Z= Mujer	Z = Hombre	Z = Mujer
X=Si	0.14	0.0	0.0	0.0
X=No	0.14	0.0	0.0	0.0

Ejemplo: normalización

Si Z tiene tres valores $\{z_1, z_2, z_3\}$, y al final del algoritmo hemos obtenido el potencial:

q	$Z = z_1$	$Z = z_2$	$Z = z_3$
	0.2	0.2	0.1

Entonces, las probabilidades condicionadas se obtienen dividiendo estos valores por su suma:

	$p(z_1\|\mathbf{o})$	$p(z_2\|\mathbf{o})$	$p(z_3\|\mathbf{o})$
	0.4	0.4	0.2

Integración de Observaciones: Método Alternativo

Un método alternativo y que produce el mismo resultado, aunque no siempre lleve asociada la misma complejidad, consiste en añadir para cada observación su potencial asociado a la lista de potenciales, aplicando el algoritmo de borrado normalmente y normalizando al final. Recordemos que el potencial asociado a una observación $X = b$ sería:

$X = a$	$X = b$	$X = c$
0.0	1.0	0.0

Integración de Observaciones: Método Alternativo II

Otro método alternativo y que suele ser bastante mejor en complejidad consiste en añadir para cada observación su potencial y, al mismo tiempo, realizar una selección de los potenciales dada la evidencia (operación también conocida como absorción de la evidencia).

Selección - absorción de la evidencia

- Hacer 0 los valores de los potenciales incompatibles con las observaciones
- Marginalizar sobre las variables no observadas, borrando las observadas

Integración de Observaciones: Método Alternativo II, Ejemplo

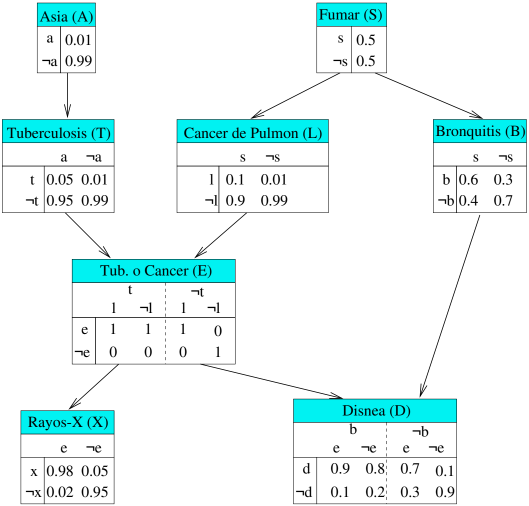
Ejemplo: Si observamos $(B = b)$

$$\begin{array}{c|cccc} & b,c & b,\bar{c} & \bar{b},c & \bar{b},\bar{c} \\ \hline a & 0{,}5 & 0{,}2 & 0{,}8 & 0{,}32 \\ \bar{a} & 0{,}0 & 0{,}3 & 0{,}0 & 0{,}12 \end{array} \longrightarrow \begin{array}{c|cccc} & b,c & b,\bar{c} & \bar{b},c & \bar{b},\bar{c} \\ \hline a & 0{,}5 & 0{,}2 & 0{,}0 & 0{,}0 \\ \bar{a} & 0{,}0 & 0{,}3 & 0{,}0 & 0{,}0 \end{array} \longrightarrow \begin{array}{c|cc} & c & \bar{c} \\ \hline a & 0{,}50 & 0{,}20 \\ \bar{a} & 0{,}00 & 0{,}30 \end{array}$$

Estos dos pasos se pueden dar de forma eficiente de una vez sin tener que sumar.

Ejemplos

Consideraremos la siguiente red (ASIA) para los ejemplos:



Ejemplos (cont.)

Ejemplo 1. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = T, S, E, A, L, B, X$.

1. $\mathcal{L} = \{f_A(A), \underbrace{f_T(T, A)}, f_S(S), f_L(L, S), f_B(B, S), \underbrace{f_E(E, T, L)}, f_X(X, E), f_D(D, E, B)\}$. **Eliminar T**.

$$g_1(A, E, L) = \sum_T (f_T(A, T) \times f_E(E, T, L))$$

tamaño = 16

2. $\mathcal{L} = \{f_A(A), \underbrace{f_S(S), f_L(L, S), f_B(B, S)}, f_X(X, E), f_D(D, E, B), g_1(A, E, L)\}$. **Eliminar S**.

$$g_2(L, B) = \sum_S (f_S(S) \times f_L(L, S) \times f_B(B, S))$$

tamaño = 8

3. $\mathcal{L} = \{f_A(A), \underbrace{f_X(X, E), f_D(D, E, B), g_1(A, E, L)}, g_2(L, B)\}$. **Elim. E**

$$g_3(X, D, B, A, L) = \sum_E (f_X(X, E) \times f_D(D, E, B) \times g_1(A, E, L))$$

tamaño = 64

Ejemplos (cont.)

Ejemplo 1. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = T, S, E, A, L, B, X$.

1. $\mathcal{L} = \{f_A(A), \underbrace{f_T(T, A)}, f_S(S), f_L(L, S), f_B(B, S), \underbrace{f_E(E, T, L)}, f_X(X, E), f_D(D, E, B)\}$. **Eliminar T**.

$$g_1(A, E, L) = \sum_T (f_T(A, T) \times f_E(E, T, L))$$

tamaño = 16

2. $\mathcal{L} = \{f_A(A), \underbrace{f_S(S), f_L(L, S), f_B(B, S)}, f_X(X, E), f_D(D, E, B), g_1(A, E, L)\}$. **Eliminar S**.

$$g_2(L, B) = \sum_S (f_S(S) \times f_L(L, S) \times f_B(B, S))$$

tamaño = 8

3. $\mathcal{L} = \{f_A(A), \underbrace{f_X(X, E), f_D(D, E, B), g_1(A, E, L)}, g_2(L, B)\}$. **Elim. E**

$$g_3(X, D, B, A, L) = \sum_E (f_X(X, E) \times f_D(D, E, B) \times g_1(A, E, L))$$

tamaño = 64

Ejemplos (cont.)

Ejemplo 1. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = T, S, E, A, L, B, X$.

1 $\mathcal{L} = \{f_A(A), \underbrace{f_T(T,A)}, f_S(S), f_L(L,S), f_B(B,S), \underbrace{f_E(E,T,L)}, f_X(X,E), f_D(D,E,B)\}$. **Eliminar T**.

$$g_1(A,E,L) = \sum_T (f_T(A,T) \times f_E(E,T,L))$$

tamaño = 16

2 $\mathcal{L} = \{f_A(A), \underbrace{f_S(S), f_L(L,S), f_B(B,S)}, f_X(X,E), f_D(D,E,B), g_1(A,E,L)\}$. **Eliminar S**.

$$g_2(L,B) = \sum_S (f_S(S) \times f_L(L,S) \times f_B(B,S))$$

tamaño = 8

3 $\mathcal{L} = \{f_A(A), \underbrace{f_X(X,E), f_D(D,E,B), g_1(A,E,L)}, g_2(L,B)\}$. **Elim. E**

$$g_3(X,D,B,A,L) = \sum_E (f_X(X,E) \times f_D(D,E,B) \times g_1(A,E,L))$$

tamaño = 64

Ejemplos (cont.)

4 $\mathcal{L} = \{\underbrace{f_A(A)}, g_2(L,B), \underbrace{g_3(X,D,B,A,L)}\}$. **Eliminar A** tamaño = 32

$$g_4(X,D,B,L) = \sum_A (f_A(A) \times g_3(X,D,B,A,L))$$

5 $\mathcal{L} = \{\underbrace{g_2(L,B), g_4(X,D,B,L)}\}$. **Eliminar L.** tamaño = 16

$$g_5(X,D,B) = \sum_L g_2(L,B) \times g_4(X,D,B,L)$$

6 $\mathcal{L} = \{\underbrace{g_5(X,D,B)}\}$. **Eliminar B.** tamaño = 8

$$g_6(X,D) = \sum_B g_5(X,D,B)$$

7 $\mathcal{L} = \{\underbrace{g_6(X,D)}\}$. **Eliminar X.** tamaño = 8

$$g_7(D) = \sum_X g_6(X,D)$$

8 Devolver (normalizando si es necesario) $g_7(D)$

Ejemplos (cont.)

4 $\mathcal{L} = \{\underbrace{f_A(A)}, g_2(L,B), \underbrace{g_3(X,D,B,A,L)}\}$. **Eliminar A** tamaño = 32

$$g_4(X,D,B,L) = \sum_A (f_A(A) \times g_3(X,D,B,A,L))$$

5 $\mathcal{L} = \{\underbrace{g_2(L,B), g_4(X,D,B,L)}\}$. **Eliminar L.** tamaño = 16

$$g_5(X,D,B) = \sum_L g_2(L,B) \times g_4(X,D,B,L)$$

6 $\mathcal{L} = \{\underbrace{g_5(X,D,B)}\}$. **Eliminar B.** tamaño = 8

$$g_6(X,D) = \sum_B g_5(X,D,B)$$

7 $\mathcal{L} = \{\underbrace{g_6(X,D)}\}$. **Eliminar X.** tamaño = 8

$$g_7(D) = \sum_X g_6(X,D)$$

8 Devolver (normalizando si es necesario) $g_7(D)$

Ejemplos (cont.)

4 $\mathcal{L} = \{\underbrace{f_A(A)}, g_2(L,B), \underbrace{g_3(X,D,B,A,L)}\}$. **Eliminar A** tamaño = 32

$$g_4(X,D,B,L) = \sum_A (f_A(A) \times g_3(X,D,B,A,L))$$

5 $\mathcal{L} = \{\underbrace{g_2(L,B), g_4(X,D,B,L)}\}$. **Eliminar L**. tamaño = 16

$$g_5(X,D,B) = \sum_L g_2(L,B) \times g_4(X,D,B,L)$$

6 $\mathcal{L} = \{\underbrace{g_5(X,D,B)}\}$. **Eliminar B**. tamaño = 8

$$g_6(X,D) = \sum_B g_5(X,D,B)$$

7 $\mathcal{L} = \{\underbrace{g_6(X,D)}\}$. **Eliminar X**. tamaño = 8

$$g_7(D) = \sum_X g_6(X,D)$$

8 Devolver (normalizando si es necesario) $g_7(D)$

Ejemplos (cont.)

4 $\mathcal{L} = \{\underbrace{f_A(A)}, g_2(L,B), \underbrace{g_3(X,D,B,A,L)}\}$. **Eliminar A** tamaño = 32

$$g_4(X,D,B,L) = \sum_A (f_A(A) \times g_3(X,D,B,A,L))$$

5 $\mathcal{L} = \{\underbrace{g_2(L,B), g_4(X,D,B,L)}\}$. **Eliminar L**. tamaño = 16

$$g_5(X,D,B) = \sum_L g_2(L,B) \times g_4(X,D,B,L)$$

6 $\mathcal{L} = \{\underbrace{g_5(X,D,B)}\}$. **Eliminar B**. tamaño = 8

$$g_6(X,D) = \sum_B g_5(X,D,B)$$

7 $\mathcal{L} = \{\underbrace{g_6(X,D)}\}$. **Eliminar X**. tamaño = 8

$$g_7(D) = \sum_X g_6(X,D)$$

8 Devolver (normalizando si es necesario) $g_7(D)$

Ejemplos (cont.)

4. $\mathcal{L} = \{\underbrace{f_A(A)}, g_2(L,B), \underbrace{g_3(X,D,B,A,L)}\}$. **Eliminar A** tamaño = 32

$$g_4(X,D,B,L) = \sum_A (f_A(A) \times g_3(X,D,B,A,L))$$

5. $\mathcal{L} = \{\underbrace{g_2(L,B), g_4(X,D,B,L)}\}$. **Eliminar L**. tamaño = 16

$$g_5(X,D,B) = \sum_L g_2(L,B) \times g_4(X,D,B,L)$$

6. $\mathcal{L} = \{\underbrace{g_5(X,D,B)}\}$. **Eliminar B**. tamaño = 8

$$g_6(X,D) = \sum_B g_5(X,D,B)$$

7. $\mathcal{L} = \{\underbrace{g_6(X,D)}\}$. **Eliminar X**. tamaño = 8

$$g_7(D) = \sum_X g_6(X,D)$$

8. Devolver (normalizando si es necesario) $g_7(D)$

Ejemplos (cont.)

Ejemplo 2. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = A, X, T, S, L, E, B$.

$$f_A(A), f_T(T, A), f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B)$$

Eliminar A Usar: $f_A(A), f_T(A, T)$ Nuevo: $g_1(T)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B), g_1(T)$$

Eliminar X Usar: $f_X(X, E)$ Nuevo: $g_2(E)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_D(D, E, B), g_1(T), g_2(E)$$

Eliminar T Usar: $f_E(E, T, L), g_1(T)$ Nuevo: $g_3(E, L)$ 8

$$f_S(S), f_L(L, S), f_B(B, S), f_D(D, E, B), g_2(E), g_3(E, L)$$

Ejemplos (cont.)

Ejemplo 2. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = A, X, T, S, L, E, B$.

$$f_A(A), f_T(T, A), f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B)$$

Eliminar A   Usar: $f_A(A), f_T(A, T)$ Nuevo: $g_1(T)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B), g_1(T)$$

Eliminar X   Usar: $f_X(X, E)$ Nuevo: $g_2(E)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_D(D, E, B), g_1(T), g_2(E)$$

Eliminar T   Usar: $f_E(E, T, L), g_1(T)$ Nuevo: $g_3(E, L)$ 8

$$f_S(S), f_L(L, S), f_B(B, S), f_D(D, E, B), g_2(E), g_3(E, L)$$

Ejemplos (cont.)

Ejemplo 2. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = A, X, T, S, L, E, B$.

$$f_A(A), f_T(T, A), f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B)$$

Eliminar A Usar: $f_A(A), f_T(A, T)$ Nuevo: $g_1(T)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B), g_1(T)$$

Eliminar X Usar: $f_X(X, E)$ Nuevo: $g_2(E)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_D(D, E, B), g_1(T), g_2(E)$$

Eliminar T Usar: $f_E(E, T, L), g_1(T)$ Nuevo: $g_3(E, L)$ 8

$$f_S(S), f_L(L, S), f_B(B, S), f_D(D, E, B), g_2(E), g_3(E, L)$$

Ejemplos (cont.)

Ejemplo 2. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = A, X, T, S, L, E, B$.

$$f_A(A), f_T(T, A), f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B)$$

Eliminar A Usar: $f_A(A), f_T(A, T)$ Nuevo: $g_1(T)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B), g_1(T)$$

Eliminar X Usar: $f_X(X, E)$ Nuevo: $g_2(E)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_D(D, E, B), g_1(T), g_2(E)$$

Eliminar T Usar: $f_E(E, T, L), g_1(T)$ Nuevo: $g_3(E, L)$ 8

$$f_S(S), f_L(L, S), f_B(B, S), f_D(D, E, B), g_2(E), g_3(E, L)$$

Ejemplos (cont.)

Ejemplo 2. Obtener la probabilidad de padecer Disnea (D), usando el orden de eliminación $\sigma_1 = A, X, T, S, L, E, B$.

$$f_A(A), f_T(T, A), f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B)$$

Eliminar A Usar: $f_A(A), f_T(A, T)$ Nuevo: $g_1(T)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_X(X, E), f_D(D, E, B), g_1(T)$$

Eliminar X Usar: $f_X(X, E)$ Nuevo: $g_2(E)$ 4

$$f_S(S), f_L(L, S), f_B(B, S), f_E(E, T, L), f_D(D, E, B), g_1(T), g_2(E)$$

Eliminar T Usar: $f_E(E, T, L), g_1(T)$ Nuevo: $g_3(E, L)$ 8

$$f_S(S), f_L(L, S), f_B(B, S), f_D(D, E, B), g_2(E), g_3(E, L)$$

Ejemplos (cont.)

$$f_S(S), f_L(L,S), f_B(B,S), f_D(D,E,B), g_2(E), g_3(E,L)$$

Eliminar S Usar: $f_S(S), f_L(L,S), f_B(B,S)$ Nuevo: $g_4(L,B)$ 8

$$f_D(D,E,B), g_2(E), g_3(E,L), g_4(L,B)$$

Eliminar L Usar: $g_3(E,L), g_4(L,B)$ Nuevo: $g_5(E,B)$ 8

$$f_D(D,E,B), g_2(E), g_5(E,B)$$

Eliminar E Usar: $f_D(D,E,B), g_2(E), g_5(E,B)$ Nuevo: $g_6(D,B)$ 8

$$g_6(D,B)$$

Eliminar B Usar: $g_6(D,B)$ Nuevo: $g_7(D)$ 4

$$g_7(D)$$

Ejemplos (cont.)

$$f_S(S), f_L(L,S), f_B(B,S), f_D(D,E,B), g_2(E), g_3(E,L)$$

Eliminar S Usar: $f_S(S), f_L(L,S), f_B(B,S)$ Nuevo: $g_4(L,B)$ 8

$$f_D(D,E,B), g_2(E), g_3(E,L), g_4(L,B)$$

Eliminar L Usar: $g_3(E,L), g_4(L,B)$ Nuevo: $g_5(E,B)$ 8

$$f_D(D,E,B), g_2(E), g_5(E,B)$$

Eliminar E Usar: $f_D(D,E,B), g_2(E), g_5(E,B)$ Nuevo: $g_6(D,B)$ 8

$$g_6(D,B)$$

Eliminar B Usar: $g_6(D,B)$ Nuevo: $g_7(D)$ 4

$$g_7(D)$$

Ejemplos (cont.)

$$f_S(S), f_L(L,S), f_B(B,S), f_D(D,E,B), g_2(E), g_3(E,L)$$

Eliminar S Usar: $f_S(S), f_L(L,S), f_B(B,S)$ Nuevo: $g_4(L,B)$ 8

$$f_D(D,E,B), g_2(E), g_3(E,L), g_4(L,B)$$

Eliminar L Usar: $g_3(E,L), g_4(L,B)$ Nuevo: $g_5(E,B)$ 8

$$f_D(D,E,B), g_2(E), g_5(E,B)$$

Eliminar E Usar: $f_D(D,E,B), g_2(E), g_5(E,B)$ Nuevo: $g_6(D,B)$ 8

$$g_6(D,B)$$

Eliminar B Usar: $g_6(D,B)$ Nuevo: $g_7(D)$ 4

$$g_7(D)$$

Ejemplos (cont.)

$$f_S(S), f_L(L,S), f_B(B,S), f_D(D,E,B), g_2(E), g_3(E,L)$$

Eliminar S   Usar: $f_S(S), f_L(L,S), f_B(B,S)$ Nuevo: $g_4(L,B)$ 8

$$f_D(D,E,B), g_2(E), g_3(E,L), g_4(L,B)$$

Eliminar L   Usar: $g_3(E,L), g_4(L,B)$ Nuevo: $g_5(E,B)$ 8

$$f_D(D,E,B), g_2(E), g_5(E,B)$$

Eliminar E   Usar: $f_D(D,E,B), g_2(E), g_5(E,B)$ Nuevo: $g_6(D,B)$ 8

$$g_6(D,B)$$

Eliminar B   Usar: $g_6(D,B)$ Nuevo: $g_7(D)$ 4

$$g_7(D)$$

Ejemplos (cont.)

$$f_S(S), f_L(L,S), f_B(B,S), f_D(D,E,B), g_2(E), g_3(E,L)$$

Eliminar S   Usar: $f_S(S), f_L(L,S), f_B(B,S)$ Nuevo: $g_4(L,B)$ 8

$$f_D(D,E,B), g_2(E), g_3(E,L), g_4(L,B)$$

Eliminar L   Usar: $g_3(E,L), g_4(L,B)$ Nuevo: $g_5(E,B)$ 8

$$f_D(D,E,B), g_2(E), g_5(E,B)$$

Eliminar E   Usar: $f_D(D,E,B), g_2(E), g_5(E,B)$ Nuevo: $g_6(D,B)$ 8

$$g_6(D,B)$$

Eliminar B   Usar: $g_6(D,B)$ Nuevo: $g_7(D)$ 4

$$g_7(D)$$

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés* (**I**), *variables observadas o evidencia* ($\mathbf{E} = \mathbf{e}$) y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en* **R***?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés* (**I**), *variables observadas o evidencia* ($\mathbf{E} = \mathbf{e}$) y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en* **R***?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés (**I**), variables observadas o evidencia (**E** = **e**)* y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en **R**?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés* (**I**), *variables observadas o evidencia* ($\mathbf{E} = \mathbf{e}$) y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en* **R***?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés* (**I**), *variables observadas o evidencia* ($\mathbf{E} = \mathbf{e}$) y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en* **R***?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto

- En un escenario concreto las variables de la red pueden dividirse en *variables de interés* (**I**), *variables observadas o evidencia* ($\mathbf{E} = \mathbf{e}$) y el resto de variables (**R**).
- La pregunta es: *¿es necesario considerar todas las variables en* **R***?*
- Respuesta: normalmente no
- Como herramienta para descartar variables de **R** usaremos los conceptos de nodos *barren* y *d-separación*
- D-separación: Podemos descartar todas las variables **K** tal que $I(\mathbf{I}; \mathbf{K}|\mathbf{E})$. Existen algoritmos eficientes para detectar **K** como BayesBall (Schacther, 1998)

Sacando partido del escenario concreto (II)

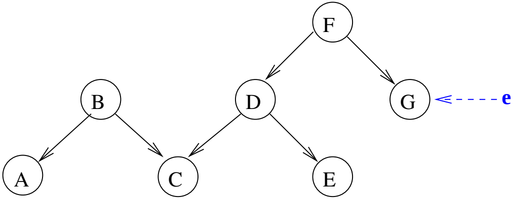
- Nodo barren: ni él ni sus descendientes han recibido evidencia

Sacando partido del escenario concreto (II)

- Nodo barren: ni él ni sus descendientes han recibido evidencia
- La idea es que si un nodo (X) sin descendientes no ha sido observado, al eliminarlo la operación a realizar es $\sum_X P(X|padres(X)) = \mathbf{1}$, obteniéndose así un potencial unitario (y así recursivamente ....)

Sacando partido del escenario concreto (II)

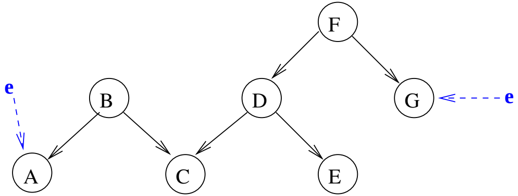
- Nodo barren: ni él ni sus descendientes han recibido evidencia
- La idea es que si un nodo (X) sin descendientes no ha sido observado, al eliminarlo la operación a realizar es $\sum_X P(X|padres(X)) = \mathbf{1}$, obteniéndose así un potencial unitario (y así recursivamente ....)



- Ejemplo: Para calcular $P(F|G = g)$ sólo las probabilidades $P(F)$ y $P(G|F)$ son necesarias puesto que A, B, C, D, E son nodos barren.

Sacando partido del escenario concreto (II)

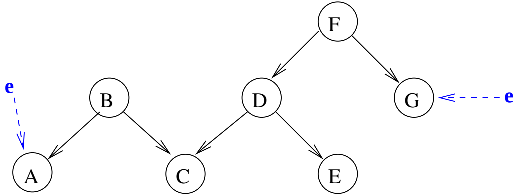
- Nodo barren: ni él ni sus descendientes han recibido evidencia
- La idea es que si un nodo (X) sin descendientes no ha sido observado, al eliminarlo la operación a realizar es $\sum_X P(X|padres(X)) = \mathbf{1}$, obteniéndose así un potencial unitario (y así recursivamente ....)



- Ejemplo: Para calcular $P(F|G = g)$ sólo las probabilidades $P(F)$ y $P(G|F)$ son necesarias puesto que A, B, C, D, E son nodos barren.
- Ejemplo 2: Si ahora la evidencia es $A = a, G = g$ de nuevo sólo las tablas anteriores son necesarias para obtener la prob. a posteriori en F puesto que F y A están d-separadas y, por tanto, la observación $A = a$ no tiene influencia.

Sacando partido del escenario concreto (II)

- Nodo barren: ni él ni sus descendientes han recibido evidencia
- La idea es que si un nodo (X) sin descendientes no ha sido observado, al eliminarlo la operación a realizar es $\sum_X P(X|padres(X)) = \mathbf{1}$, obteniéndose así un potencial unitario (y así recursivamente ....)



- Ejemplo: Para calcular $P(F|G=g)$ sólo las probabilidades $P(F)$ y $P(G|F)$ son necesarias puesto que A, B, C, D, E son nodos barren.
- Ejemplo 2: Si ahora la evidencia es $A = a, G = g$ de nuevo sólo las tablas anteriores son necesarias para obtener la prob. a posteriori en F puesto que F y A están d-separadas y, por tanto, la observación $A = a$ no tiene influencia.
- Regla nodos barren: Sea ϕ un cjto de potenciales y queremos obtener $\phi^{\downarrow \mathbf{V}}$. Si $A \notin V$, y sólo hay un potencial en ϕ conteniendo A, y es de la forma $P(A|\ldots)$, entonces podemos marginalizar A simplemente descartando $P(A|\ldots.)$.

Sacando Partido del Escenario Concreto III

EL algoritmo Bayes-Ball

- Cada variable tiene una variable booleana *VALIDO* que es falsa para todos los nodos excepto para el nodo de interés que es cierta.
- Comienza un recorrido de la red en la variable de interés *F* recorriendo los arcos en ambas direcciones.
- Cada vez que llega a un nodo:
  - Si VALIDO es cierto no hace nada, en otro caso:
    - Si llega a un nodo procedente de un hijo, se pone la variable VALIDO a cierta. Si ese nodo no está observado pasa a visitar todos los vecinos y si está observado no se visita ninguno.
    - Si llega a un nodo precedente de un padre, si no está observado se visitan todos los hijos. Si está observado se pone la variable VALIDO a cierta y se visitan los padres.

Al final del algoritmo podemos prescindir de todas las variables NO VALIDAS.

Comentarios sobre Eliminación de Variables

- Importancia del orden σ: Como hemos visto en el ejemplo 1 la mayor tabla construida es de tamaño **64** mientras que en el ejemplo 2 es **8**, lo que muestra claramente que el orden elegido juega un papel fundamental en la eficiencia del algoritmo.
  - La determinación de la secuencia óptima es un problema NP-difícil, pero se pueden seguir heurísticas que producen buenos resultados. *En cada paso eliminar la variable que produzca la tabla de menor tamaño. Romper los empates arbitra/aleato(riamente)*
- Explota la idea de trabajar con cálculos locales mediante el desplazamiento de factores fuera de las sumas
- Permite descartar partes de la red usando el criterio de d-separación
- El principal problema es que el proceso debe repetirse para cada variable objetivo

Contenido del tema

Árboles de cliques o grupos (*Join Trees*)

- El algoritmo de borrado es muy bueno si sólo queremos calcular la probabilidad condicionada en una variable de interés y es el algoritmo que habría que utilizar en ese caso.
- Si queremos calcular la probabilidad condicionada para más de una variable, se pueden organizar los cálculos de manera que con el equivalente al doble de esfuerzo para una variable se calcule para todas las variables.
- Para ello, la inferencia no se realiza sobre la red original, sino sobre una estructura gráfica de orden superior. La idea es agrupar de forma adecuada las variables de la red y, estructurarlas en forma de árbol.
- Sobre esta estructura secundaria se organiza un método equivalente al algoritmo de borrado para todas las variables (compartiendo cálculos entre variables).
- Es el método de propagación incorporado en la mayoría de las herramientas actuales.

Árboles de cliques o grupos (*Join Trees*)

- El algoritmo de borrado es muy bueno si sólo queremos calcular la probabilidad condicionada en una variable de interés y es el algoritmo que habría que utilizar en ese caso.
- Si queremos calcular la probabilidad condicionada para más de una variable, se pueden organizar los cálculos de manera que con el equivalente al doble de esfuerzo para una variable se calcule para todas las variables.
- Para ello, la inferencia no se realiza sobre la red original, sino sobre una estructura gráfica de orden superior. La idea es agrupar de forma adecuada las variables de la red y, estructurarlas en forma de árbol.
- Sobre esta estructura secundaria se organiza un método equivalente al algoritmo de borrado para todas las variables (compartiendo cálculos entre variables).
- Es el método de propagación incorporado en la mayoría de las herramientas actuales.

Árboles de cliques o grupos (*Join Trees*)

- El algoritmo de borrado es muy bueno si sólo queremos calcular la probabilidad condicionada en una variable de interés y es el algoritmo que habría que utilizar en ese caso.
- Si queremos calcular la probabilidad condicionada para más de una variable, se pueden organizar los cálculos de manera que con el equivalente al doble de esfuerzo para una variable se calcule para todas las variables.
- Para ello, la inferencia no se realiza sobre la red original, sino sobre una estructura gráfica de orden superior. La idea es agrupar de forma adecuada las variables de la red y, estructurarlas en forma de árbol.
- Sobre esta estructura secundaria se organiza un método equivalente al algoritmo de borrado para todas las variables (compartiendo cálculos entre variables).
- Es el método de propagación incorporado en la mayoría de las herramientas actuales.

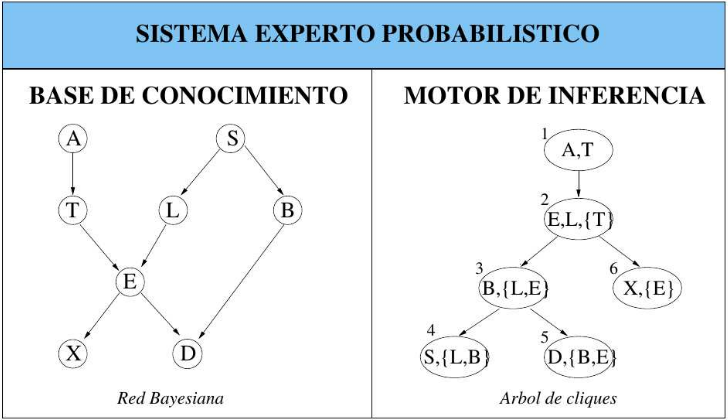
Árboles de cliques o grupos (*Join Trees*)

- El algoritmo de borrado es muy bueno si sólo queremos calcular la probabilidad condicionada en una variable de interés y es el algoritmo que habría que utilizar en ese caso.
- Si queremos calcular la probabilidad condicionada para más de una variable, se pueden organizar los cálculos de manera que con el equivalente al doble de esfuerzo para una variable se calcule para todas las variables.
- Para ello, la inferencia no se realiza sobre la red original, sino sobre una estructura gráfica de orden superior. La idea es agrupar de forma adecuada las variables de la red y, estructurarlas en forma de árbol.
- Sobre esta estructura secundaria se organiza un método equivalente al algoritmo de borrado para todas las variables (compartiendo cálculos entre variables).
- Es el método de propagación incorporado en la mayoría de las herramientas actuales.

Árboles de cliques o grupos (*Join Trees*)

- El algoritmo de borrado es muy bueno si sólo queremos calcular la probabilidad condicionada en una variable de interés y es el algoritmo que habría que utilizar en ese caso.
- Si queremos calcular la probabilidad condicionada para más de una variable, se pueden organizar los cálculos de manera que con el equivalente al doble de esfuerzo para una variable se calcule para todas las variables.
- Para ello, la inferencia no se realiza sobre la red original, sino sobre una estructura gráfica de orden superior. La idea es agrupar de forma adecuada las variables de la red y, estructurarlas en forma de árbol.
- Sobre esta estructura secundaria se organiza un método equivalente al algoritmo de borrado para todas las variables (compartiendo cálculos entre variables).
- Es el método de propagación incorporado en la mayoría de las herramientas actuales.

Ejemplo de *Join Tree*



Construcción de un árbol de cliques

- **Objetivos**:
  - El árbol obtenido debe ser un árbol en el que los nodos son grupos de variables verificando la propiedad de la intersección consecutiva:.
    Un árbol de grupos es un árbol que verifica la propiedad de intersección consecutiva, es decir, si para cada par de nodos G_1, G_2 *con intersección no vacía* $S = G_1 \cap G_2 \neq \emptyset$, *se verifica que* S *está contenido en todos los nodos que hay en el camino entre* G_1 *y* G_2.
  - Cada uno de los potenciales iniciales de la red caben en uno de los nodos del árbol. Un potencial cabe en un nodo cuando el conjunto de variables del potencial está incluido en el grupo variables del nodo.
- **Procedimiento:** El propio algoritmo de borrado nos proporciona un método de construcción.

Cliques o Grupos

- Un árbol de grupos se dice que es un árbol de cliques cuando todos los grupos son maximales.
- Siempre un árbol de grupos se puede transformar en un árbol de cliques eliminado los grupos no maximales y reconectando los vecinos de forma apropiada.
- Siempre podemos eliminar los grupos no maximales antes de construir el árbol de grupos, aunque no siempre es una buena idea.

Cliques o Grupos

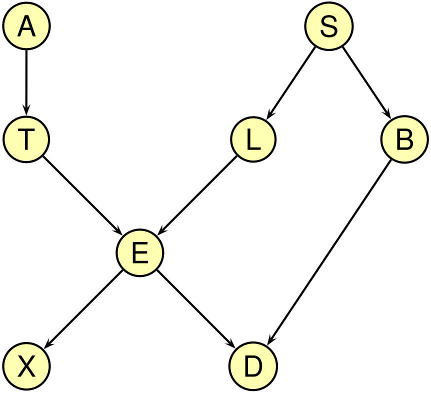
- Un árbol de grupos se dice que es un árbol de cliques cuando todos los grupos son maximales.
- Siempre un árbol de grupos se puede transformar en un árbol de cliques eliminado los grupos no maximales y reconectando los vecinos de forma apropiada.
- Siempre podemos eliminar los grupos no maximales antes de construir el árbol de grupos, aunque no siempre es una buena idea.

Cliques o Grupos

- Un árbol de grupos se dice que es un árbol de cliques cuando todos los grupos son maximales.
- Siempre un árbol de grupos se puede transformar en un árbol de cliques eliminado los grupos no maximales y reconectando los vecinos de forma apropiada.
- Siempre podemos eliminar los grupos no maximales antes de construir el árbol de grupos, aunque no siempre es una buena idea.

Obtención del Árbol de Grupos

Supongamos la red Asia:



Lista de potenciales:

$f_1(A), f_2(S), f_3(A,T), f_4(S,L), f_5(S,B), f_6(T,L,E), f_7(E,X), f_8(E,B,D)$

Obtención del Árbol de Grupos

Aplicamos el algoritmo de borrado hasta que quede sólo una variable (cualquiera).
Debemos de guardar una lista ordenada de todos los conjuntos de variables correspondientes a las combinaciones de los potenciales que vamos obteniendo en el proceso de borrado
Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: vacía
  Lista de potenciales:
  $f_1(A), f_2(S), f_3(A,T), f_4(S,L), f_5(S,B), f_6(T,L,E), f_7(E,X), f_8(E,B,D)$
  Borramos X, Obtenemos $f_9(E) = \sum_X f_7(E,X)$
- Lista de Grupos: $\{E, X\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_3(A,T), f_4(S,L), f_5(S,B), f_6(T,L,E), f_8(E,B,D), f_9(E)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_3(A, T), f_4(S, L), f_5(S, B), f_6(T, L, E), f_8(E, B, D), f_9(E)$
  Borramos T y obtenemos
  $f_{10}(A, L, E) = \sum_T f_3(A, T) f_6(T, L, E)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_8(E, B, D), f_9(E), f_{10}(A, L, E)$
  Borramos D y obtenemos $f_{11}(E, B) = \sum_D f_8(E, B, D)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_3(A, T), f_4(S, L), f_5(S, B), f_6(T, L, E), f_8(E, B, D), f_9(E)$
  Borramos T y obtenemos
  $f_{10}(A, L, E) = \sum_T f_3(A, T) f_6(T, L, E)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_8(E, B, D), f_9(E), f_{10}(A, L, E)$
  Borramos D y obtenemos $f_{11}(E, B) = \sum_D f_8(E, B, D)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_3(A, T), f_4(S, L), f_5(S, B), f_6(T, L, E), f_8(E, B, D), f_9(E)$
  Borramos T y obtenemos
  $f_{10}(A, L, E) = \sum_T f_3(A, T)f_6(T, L, E)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_8(E, B, D), f_9(E), f_{10}(A, L, E)$
  Borramos D y obtenemos $f_{11}(E, B) = \sum_D f_8(E, B, D)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$
  Borramos E y obtenemos:
  $f_{12}(A, L, B) = \sum_E f_9(E) f_{10}(A, L, E) f_{11}(E, B)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_{12}(A, L, B)$
  Borramos L y obtenemos:
  $f_{13}(A, S, B) = \sum_L f_4(S, L) f_{12}(A, L, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$
  Borramos E y obtenemos:
  $f_{12}(A, L, B) = \sum_E f_9(E) f_{10}(A, L, E) f_{11}(E, B)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_{12}(A, L, B)$
  Borramos L y obtenemos:
  $f_{13}(A, S, B) = \sum_L f_4(S, L) f_{12}(A, L, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_9(E), f_{10}(A, L, E), f_{11}(E, B)$
  Borramos E y obtenemos:
  $f_{12}(A, L, B) = \sum_E f_9(E) f_{10}(A, L, E) f_{11}(E, B)$
- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_4(S, L), f_5(S, B), f_{12}(A, L, B)$
  Borramos L y obtenemos:
  $f_{13}(A, S, B) = \sum_L f_4(S, L) f_{12}(A, L, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales:
  $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales: $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$
  Borramos S y obtenemos:
  $f_{14}(A, B) = \sum_S f_2(S) f_5(S, B) f_{13}(A, S, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}$
  Lista de potenciales: $f_1(A), f_{14}(A, B)$
  Borramos B y obtenemos $f_{15}(A) = \sum_B f_{14}(A, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}$
  Lista de potenciales: $f_1(A), f_{15}(A)$
- Añadinos la última variable no borrada y obtenemos la lista final:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}, \{A\}$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales: $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$
  Borramos S y obtenemos:
  $f_{14}(A, B) = \sum_S f_2(S) f_5(S, B) f_{13}(A, S, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}$
  Lista de potenciales: $f_1(A), f_{14}(A, B)$
  Borramos B y obtenemos $f_{15}(A) = \sum_B f_{14}(A, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}$
  Lista de potenciales: $f_1(A), f_{15}(A)$
- Añadinos la última variable no borrada y obtenemos la lista final:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}, \{A\}$

Obtención del Árbol de Grupos

Orden de borrado: X, T, D, E, L, S, B

- Lista de Grupos: $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$
  Lista de potenciales: $f_1(A), f_2(S), f_5(S, B), f_{13}(A, S, B)$
  Borramos S y obtenemos:
  $f_{14}(A, B) = \sum_S f_2(S) f_5(S, B) f_{13}(A, S, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}$
  Lista de potenciales: $f_1(A), f_{14}(A, B)$
  Borramos B y obtenemos $f_{15}(A) = \sum_B f_{14}(A, B)$
- Lista de Grupos:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}$
  Lista de potenciales: $f_1(A), f_{15}(A)$
- Añadinos la última variable no borrada y obtenemos la lista final:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}, \{A\}$

Construcción del árbol de grupos

- Lista final:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}, \{A, B\}, \{A\}$
- Ahora podemos eliminar grupos no maximales, pero entonces el proceso que viene a continuación no siempre puede ser aplicado. Hay que realizar un procedimiento un poco más sofisticado.
- Siempre podemos eliminar los últimos grupos no maximales de la lista y eso no cambia nada en nuestro procedimiento.
  Lista reducida:
  $\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}$
- No es necesario realizar las combinaciones y marginalizaciones. Sólo ver de qué variables dependen las variables que se van obteniendo.
- El árbol tendrá estos grupos como nodos y se irán introduciendo en orden inverso a como se han producido.
- Al introducir un grupo en el árbol, se enlaza con aquel que contenga todas las variables incluidas en la intersección de las suyas y las de los grupos que ya estén en el árbol. Si hay varias opciones, elegir cualquiera.
- Distintos órdenes de borrado producirán distintas listas y, por tanto, diferentes árboles de grupos. Todos ellos serán válidos.

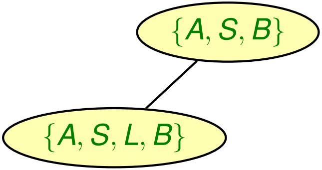
Construcción del árbol de grupos

$\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}, \{A, S, B\}$

$\{A, S, B\}$

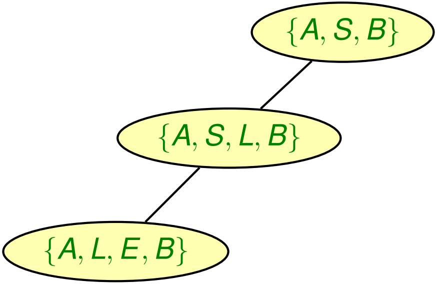
Construcción del árbol de grupos

$\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}, \{A, S, L, B\}$



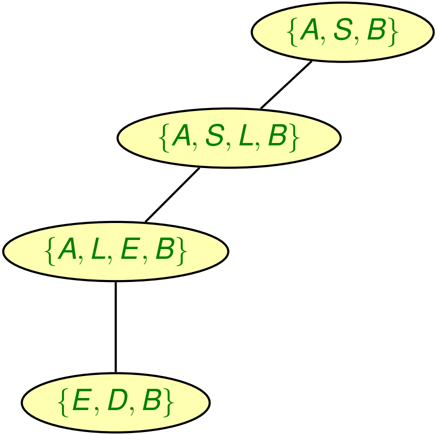
Construcción del árbol de grupos

$\{E, X\}, \{A, T, L, E\}, \{E, B, D\}, \{A, L, E, B\}$



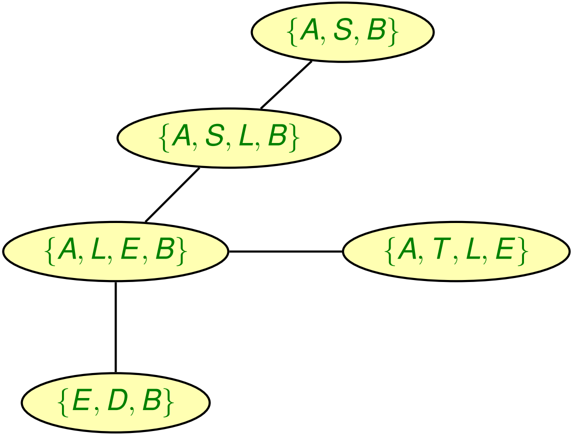
Construcción del árbol de grupos

$\{E, X\}, \{A, T, L, E\}, \{E, B, D\}$



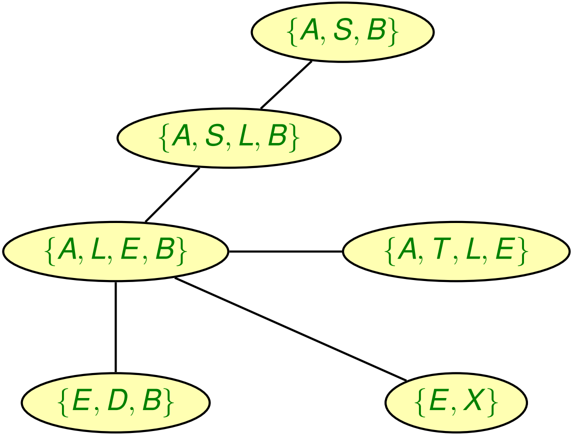
Construcción del árbol de grupos

$\{E, X\}, \{A, T, L, E\}$



Construcción del árbol de grupos

$\{E, X\}$



Construcción de la representación potencial

- A cada grupo C_i se le asigna un potencial $\psi(C_i)$ obtenido como:
  1. Inicializar todos los potenciales a 1.
  2. Para cada variable X_i de la RB identificar un grupo C_j que contenga a $F(X_i)$ (familia de X_i o variables en la CPD de X_i) (si hay varios elegir uno arbitrariamente). Multiplicar ψ_{C_j} por la familia de probabilidad f_{X_i}.
- Después de este proceso, la representación potencial es una factorización de la DPC:

$$P(\mathbf{x}) = \prod_{i=1}^{t} \psi_{C_i}(\mathbf{x}^{\downarrow C_i}), \quad \forall x \in \Omega_{\mathbf{X}}$$

Construcción de la representación potencial

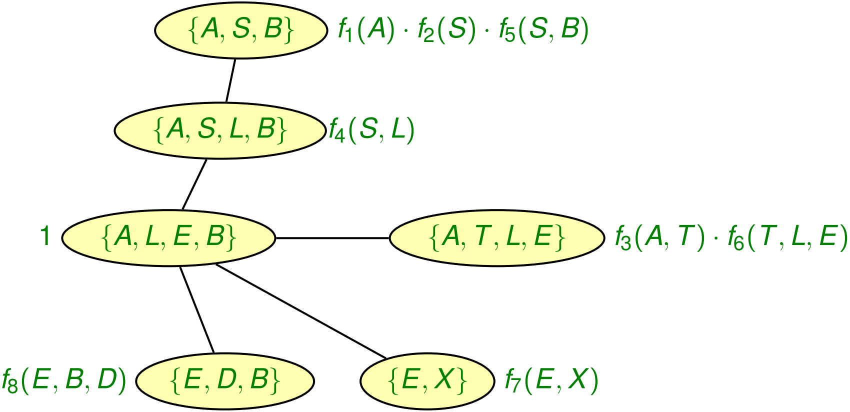
- A cada grupo C_i se le asigna un potencial $\psi(C_i)$ obtenido como:
  1. Inicializar todos los potenciales a 1.
  2. Para cada variable X_i de la RB identificar un grupo C_j que contenga a $F(X_i)$ (familia de X_i o variables en la CPD de X_i) (si hay varios elegir uno arbitrariamente). Multiplicar ψ_{C_j} por la familia de probabilidad f_{X_i}.
- Después de este proceso, la representación potencial es una factorización de la DPC:

$$P(\mathbf{x}) = \prod_{i=1}^{t} \psi_{C_i}(\mathbf{x}^{\downarrow C_i}), \quad \forall x \in \Omega_{\mathbf{X}}$$

Construcción de la representación potencial

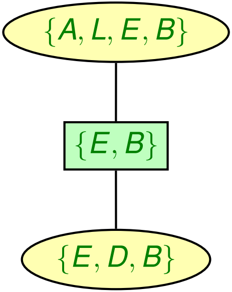
Lista de potenciales iniciales en la red bayesiana:

$f_1(A), f_2(S), f_3(A,T), f_4(S,L), f_5(S,B), f_6(T,L,E), f_7(E,X), f_8(E,B,D)$

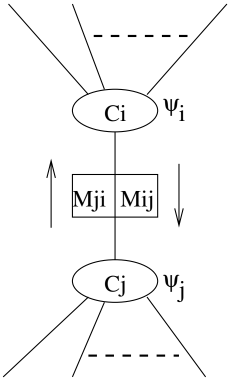


Construcción de la representación potencial: Separadores

El separador de dos grupos conectados de un árbol de grupos es la intersección de los mismos. El separador es necesario para los algoritmos de propagación y es donde se coloca la información que se traspasa entre los grupos. A veces se representan de forma explícita:



Propagación en árboles de grupos: Shenoy-Shafer



- Una vez inicializados los potenciales ψ_{C_i} de cada grupo C_i, integramos la evidencia disponible. Para cada variable $X_i \in \mathbf{E}$ se elige un grupo C_i que la contenga y se hacen igual a 0 aquellas configuraciones de ψ_{C_i} no consistentes con la evidencia.
- Se necesita espacio para almacenar dos mensajes en cada separador (uno en cada dirección).
- Durante la propagación los potenciales asociados a los grupos o cliques ya no se modifican más.
- Cálculo de los mensajes:

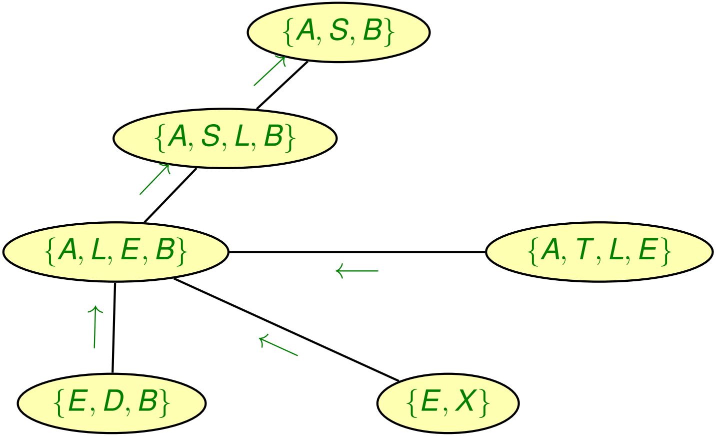
$$M^{i\to j} = \sum_{C_i \setminus S_{ij}} \psi(C_i) \prod_{k\neq j} M^{k\to i} \qquad (1)$$

Propagación en árboles de grupos: Shenoy-Shafer

- Un mensaje se puede calcular cuando todos los mensajes necesarios para su cálculo están disponibles.
- Existe un órden para calcularlos que garantiza que esto siempre ocurre. Se realiza en dos fases:
  1. Hacia arriba - 'upward' Se comienza por las hojas calculando los mensajes que los hijos mandan a los padres. Un nodo manda un mensaje cuando ha recibido todos los mensajes de los hijos.
  2. Hacia abajo - 'downward' Se comienza por el nodo raiz calculando los mensajes que los padres mandan a los hijos. Un nodo manda un mensaje cuando ha recibido todos los mensajes de los padres.

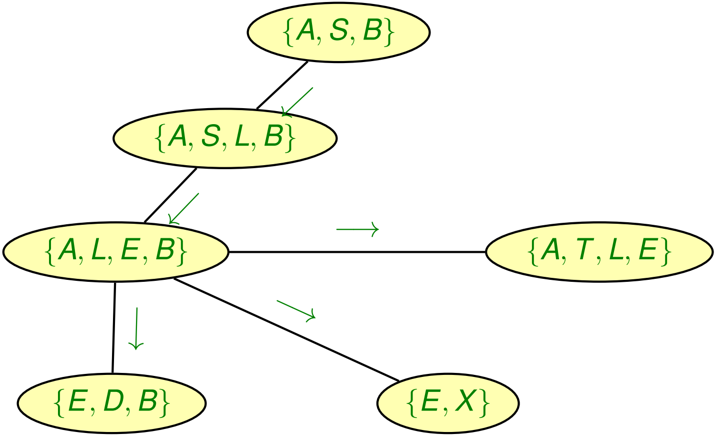
Propagación en árboles de grupos: Shenoy-Shafer

Propagación hacia arriba:



Propagación en árboles de grupos: Shenoy-Shafer

Propagación hacia abajo:



Propagación en árboles de grupos: Shenoy-Shafer

Tras una propagación completa (upward + downward) todos los separadores están llenos y se cumple:

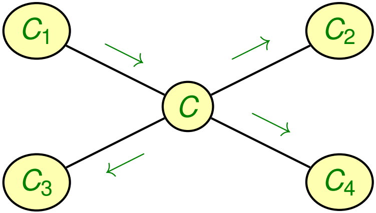
$$P(C_i, \mathbf{e}) = \psi(C_i) \prod_k M^{k \to i} \tag{2}$$

Cálculo de la probabilidad condicionada para una variable A

- Seleccionamos un grupo C_i que contenga esta variable.
- Calculamos $P(C_i, \mathbf{e})$ de acuerdo con la ecuación de arriba
- Marginalizamos el resultado sobre A
- Normalizamos el resultado

Contenido del tema

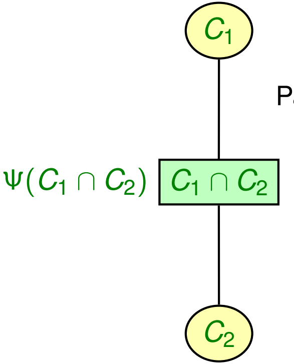
Propagación Hugin



En el algoritmo de Shenoy, el mensaje de C_1 hay que multiplicarlo por el potencial de C tres veces: una por cada uno de los mensajes salientes: a C_2, C_3 y C_4.
Para resolver esta ineficiencia, HUGIN sigue una estrategia distinta: En C acumula todos los mensajes que recibe: cada vez que recibe un mensaje, por ejemplo de C_1 a C, lo acumula al potencial de C:
$\Psi(C) \leftarrow \Psi(C).\Psi^{C_1 \to C}$
Cuando va a mandar de C a C_1, en $\Psi(C)$ está multiplicado $\Psi^{C_1 \to C}$ y si se marginaliza $\Psi(C)$ en $C \cap C_1$, estamos incluyendo de forma innecesaria $\Psi^{C_1 \to C}$. Esto se resuelve dividiendo: $\Psi(C)^{\downarrow C \cap C_1} / \Psi^{C_1 \to C}$.

Propagación Hugin

Hugin tiene otra diferencia respecto a Shafer y Shenoy: sólo almacena un potencial en cada separador. Este separador inicialmente es 1, y a medida que progresa el algoritmo contiene el producto de los dos mensajes de Shafer y Shenoy.

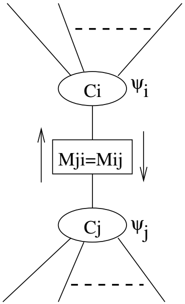


Para mandar un mensaje de C_1 a C_2:

- $\Psi(C_2) \leftarrow \Psi(C_2).(\Psi(C_1)^{\downarrow C_1 \cap C_2} / \Psi(C_1 \cap C_2))$
- $\Psi(C_1 \cap C_2) \leftarrow \Psi(C_1)^{\downarrow C_1 \cap C_2}$

Propagación Hugin

Consistencia.-



- En la arquitectura Hugin el objetivo es mantener la consistencia del árbol:

$$P(\mathbf{X}) = \frac{\prod_{i=1}^{n} \psi(C_i)}{\prod_{j}^{n-1} \psi(S_j)}$$

- Tras mandarse los mensajes $M^{i \to j}$ y $M^{j \to i}$ el enlace (i, j) es consistente, es decir,

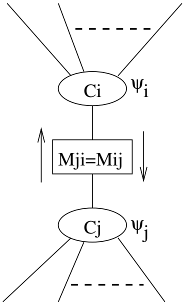
$$\sum_{C_i \setminus S_{ij}} \psi(C_i) = \psi(S_{ij}) = \sum_{C_j \setminus S_{ij}} \psi(C_j)$$

- Tras una propagación completa (upward + downward) se verifica que:

$$\psi(C_i) = P(C_i, \mathbf{e})$$

Propagación Hugin

Consistencia.-



- En la arquitectura Hugin el objetivo es mantener la consistencia del árbol:

$$P(\mathbf{X}) = \frac{\prod_{i=1}^{n} \psi(C_i)}{\prod_{j}^{n-1} \psi(S_j)}$$

- Tras mandarse los mensajes $M^{i\to j}$ y $M^{j\to i}$ el enlace (i,j) es consistente, es decir,

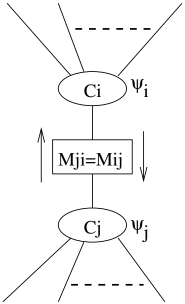
$$\sum_{C_i \setminus S_{ij}} \psi(C_i) = \psi(S_{ij}) = \sum_{C_j \setminus S_{ij}} \psi(C_j)$$

- Tras una propagación completa (upward + downward) se verifica que:

$$\psi(C_i) = P(C_i, \mathbf{e})$$

Propagación Hugin

Consistencia.-



- En la arquitectura Hugin el objetivo es mantener la consistencia del árbol:

$$P(\mathbf{X}) = \frac{\prod_{i=1}^{n} \psi(C_i)}{\prod_{j}^{n-1} \psi(S_j)}$$

- Tras mandarse los mensajes $M^{i \to j}$ y $M^{j \to i}$ el enlace (i, j) es consistente, es decir,

$$\sum_{C_i \setminus S_{ij}} \psi(C_i) = \psi(S_{ij}) = \sum_{C_j \setminus S_{ij}} \psi(C_j)$$

- Tras una propagación completa (upward + downward) se verifica que:

$$\psi(C_i) = P(C_i, \mathbf{e})$$

Contenido del tema

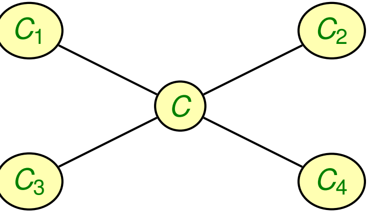
Arquitectura de Shafer y Shenoy en árboles binarios

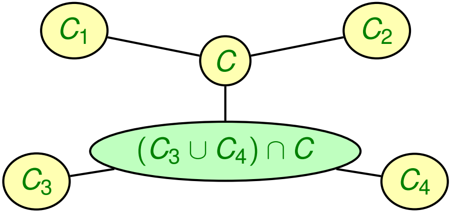
Árboles de Grupos Binarios

Son árboles de grupos en los que cada nodo tiene, a lo más, 3 vecinos.

Sobre árboles binarios, no hay ventaja en la arquitectura de Hugin sobre la de Shafer y Shenoy: en las dos se hace el mismo número de operaciones.

Transformación de un árbol cualquiera en un árbol binario





Se unen las conexiones de dos vecinos, por ejemplo C_3 y C_4, en una sóla conexión, introduciendo un nodo nuevo con variables: $(C_3 \cup C_4) \cap C$. El resultado sigue siendo un árbol de grupos válido y el número de vecinos de C se ha reducido en uno.

Este proceso se repite hasta que obtengamos un árbol binario.

Los mensajes en la conexión del nuevo nodo con C sirven de caché (automático) para almacenar algunos cálculos de forma que no haya que repetirlos.

Contenido del tema

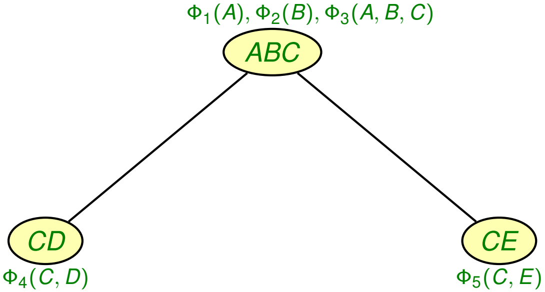
Propagación perezosa (Lazy)

- La idea de la propagación perezosa es mantener los potenciales en forma factorizada, es decir, se guardan listas de potenciales en cada clique o separador en lugar de su producto.
- El término perezoso responde a que los cálculos se retrasan lo máximo posible.
- Así, si tenemos en un clique C_i los potenciales $\psi(A,B), \psi(A,C), \psi(C)$ y tenemos que mandar un mensaje a C_j definido en $\{A,B\}$, haríamos:

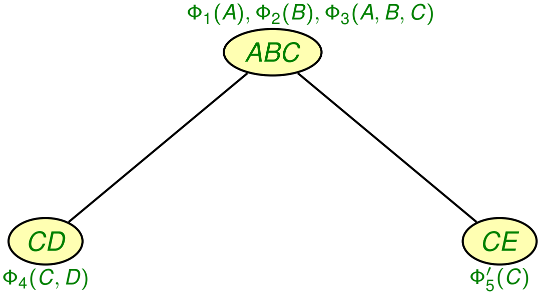
$$M^{i \to j} = \{\psi(A,B), \sum_{C} (\psi(A,C) \times \psi(C))\ \}$$

  evitándose así la necesidad de construir el potencial $\psi(A,B,C)$.
- La propagación perezosa permite explotar los conceptos de nodos *barren* y d-separación.

Propagación Perezosa

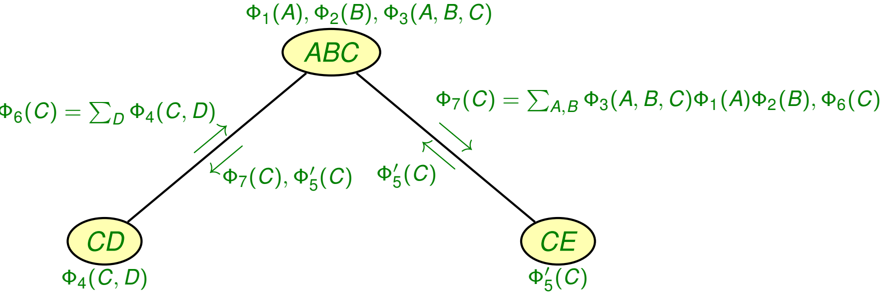


Propagación Perezosa



Observamos E: $\Phi_5(C, E) \rightarrow \Phi'_5(C)$

Propagación Perezosa



Observamos E: $\Phi_5(C,E) \rightarrow \Phi'_5(C)$