

Inferencia en redes bayesianas

Francisco J. Palmero Moya

Enero 2022

Ejercicio 2.1 Sea la red bayesiana

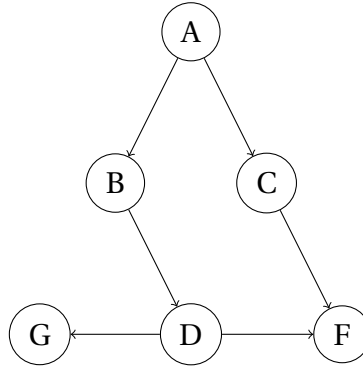


Figura 1: Red bayesiana

y las probabilidades asociadas del enunciado. Se cumple que

$$P(+b, +f, \neg g) = \sum_{a,c,d} P(a, +b, c, d, +f, \neg g) \quad (1)$$

$$P(+f, \neg g) = \sum_b \sum_{a,c,d} P(a, b, c, d, +f, \neg g) = \sum_b P(b, +f, \neg g) \quad (2)$$

$$P(+b|+f, \neg g) = \frac{P(+b, +f, \neg g)}{P(+f, \neg g)} \quad (3)$$

Se tiene que

$$P(+b, +f, \neg g) = \sum_{a,c,d} P(a, +b, c, d, +f, \neg g) \quad (4)$$

$$= \sum_a \sum_c \sum_d P(a)P(+b|a)P(c|a)P(d|+b)P(+f|c, d)P(\neg g|d) \quad (5)$$

Solo por aclarar la notación, las variables de la red son binarias y, por tanto, el sumatorio sería sobre $a, b, c \in \{+, \neg\}$. Utilizando el método de la fuerza bruta, esto daría lugar a $2^3 = 8$ sumandos, cada uno de ellos obtenido tras realizar 5 multiplicaciones.

Veamos como se puede realizar el cálculo anterior mediante el **método de eliminación de variables**. Tomamos la variable D y reunimos todos los potenciales que dependen de ella

$$\begin{aligned} \psi_1(c) &= \sum_d P(d|+b)P(+f|c, d)P(\neg g|d) \\ &= P(+d|+b)P(+f|c, +d)P(\neg g|+d) + P(\neg d|+b)P(+f|c, \neg d)P(\neg g|\neg d) \\ &= 0,6 * 0,6 * P(+f|c, +d) + 0,4 * 0,9 * P(+f|c, \neg d) \end{aligned}$$

Sustituyendo los valores que puede tomar C

$$\begin{aligned}\psi_1(+c) &= 0,6 * 0,8 * 0,6 + 0,4 * 0,6 * 0,9 = 0,504 \\ \psi_1(\neg c) &= 0,6 * 0,5 * 0,6 + 0,4 * 0 * 0,9 = 0,18\end{aligned}$$

Se obtiene así

$$P(+b, +f, \neg g) = \sum_a \sum_c P(a)P(+b|a)P(c|a)\psi_1(c)$$

Pasamos ahora a calcular el segundo potencial

$$\begin{aligned}\psi_2(a) &= \sum_c P(c|a)\psi_1(c) \\ &= P(+c|a)\psi_1(+c) + P(\neg c|a)\psi_1(\neg c) \\ &= 0,504 * P(+c|a) + 0,18 * P(\neg c|a)\end{aligned}$$

Sustituyendo los valores que puede tomar A

$$\begin{aligned}\psi_2(+a) &= 0,504 * 0,75 + 0,18 * 0,25 = 0,423 \\ \psi_2(\neg a) &= 0,504 * 0,25 + 0,18 * 0,75 = 0,261\end{aligned}$$

Hemos eliminado así la variable C , lo que da lugar a

$$P(+b, +f, \neg g) = \sum_a P(a)P(+b|a)\psi_2(a)$$

Finalmente, eliminamos la variable A

$$\begin{aligned}P(+b, +f, \neg g) &= P(+a)P(+b|+a)\psi_2(+a) + P(\neg a)P(+b|\neg a)\psi_2(\neg a) \\ &= 0,3 * 0,9 * 0,423 + 0,7 * 0,2 * 0,261 = 0,15075\end{aligned}$$

Hay que tener en cuenta que esta probabilidad, equivale a sustituir $b = +b$ en el potencial que estamos calculando $\psi_3(b) = P(b, +f, \neg g)$. Por otro lado,

$$\begin{aligned}P(+f, \neg g) &= \sum_b P(b, +f, \neg g) = P(+b, +f, \neg g) + P(\neg b, +f, \neg g) \\ &= \sum_b \psi_3(b) = 0,15075 + [0,3 * 0,1 * 0,423 + 0,7 * 0,8 * 0,261] \\ &= 0,15075 + 0,15885 = 0,3177\end{aligned}$$

La probabilidad buscada es, usando la ec. (3)

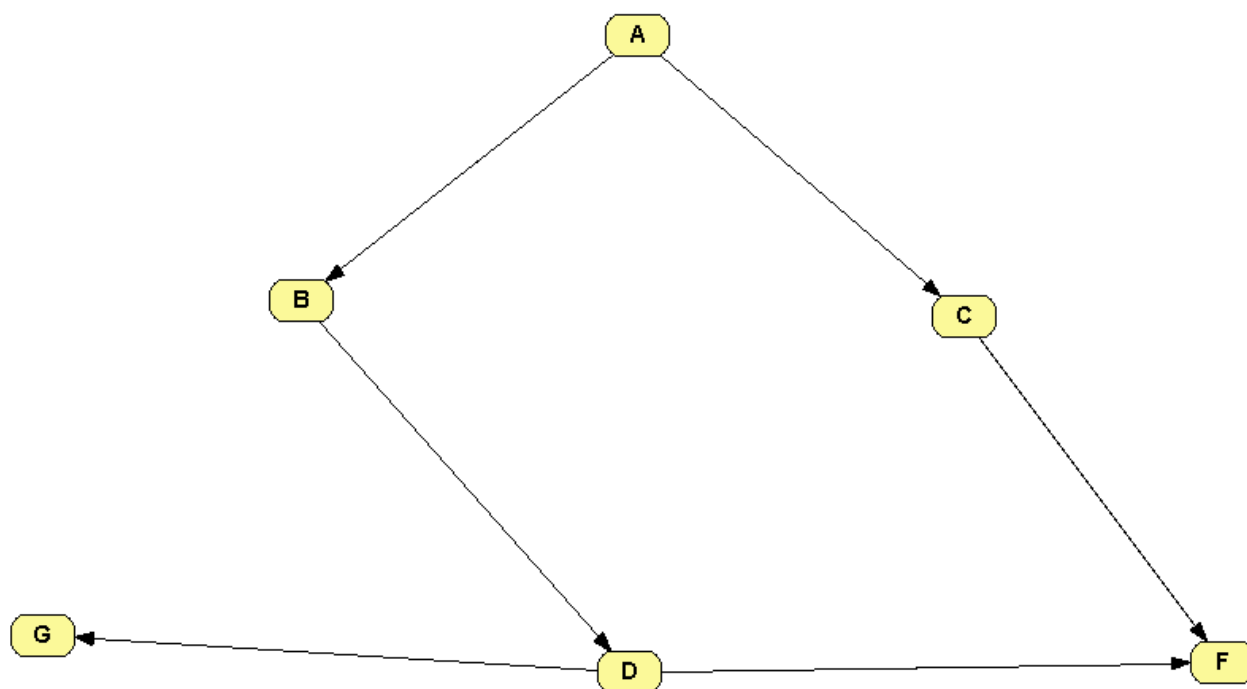
$$P(+b|+f, \neg g) = \frac{P(+b, +f, \neg g)}{P(+f, \neg g)} = \frac{0,15075}{0,3177} \simeq 0,5$$

Ejercicio 2.2 Sea la red bayesiana del ejercicio anterior y las mismas probabilidades asociadas, vamos a realizar los cálculos usando OpenMarkov.

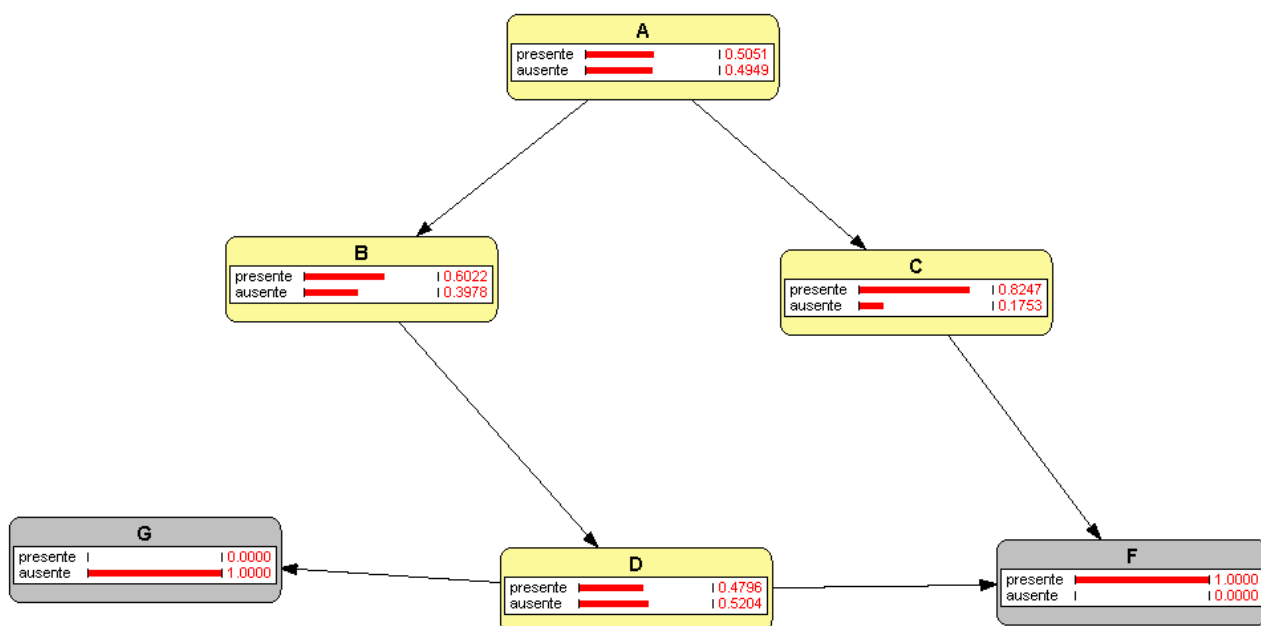
Los pasos seguidos son los siguientes:

1. Definir los nodos y sus enlaces. Véase Figura 2a.
2. Definir las tablas de probabilidad correspondientes a cada uno de los nodos.
3. Introducir los hallazgos. Véase Figura 2b.

Aunque los resultados difieren ligeramente, suponemos que es por errores de redondeo o porque el método de inferencia de OpenMarkov no es exacto.



(a) Red bayesiana



(b) Inferencia con hallazgos

Figura 2: Solución Ejercicio 2.2

Ejercicio 2.3 Los métodos estudiados con anterioridad son métodos exactos, i.e., calculan las probabilidades condicionales de los nodos de forma exacta. Se ha demostrado (véase Cooper (1990)) que este tipo de tarea de propagación exacta es NP-compleja.

Surgen así los métodos de propagación aproximada donde la idea básica consiste en generar una muestra de tamaño N , a partir de la función de probabilidad conjunta de las variables, y luego utilizar la muestra generada para calcular valores aproximados de las probabilidades de ciertos sucesos dada la evidencia. Estas probabilidades se aproximan mediante los cocientes de la frecuencia de aparición de los sucesos en la muestra y el tamaño de la muestra.

Vamos a hacer un estudio de los pasos a seguir en los siguientes métodos aproximados para calcular la probabilidad anterior:

Muestreo lógico Se trata de un método de simulación en el que se pretende calcular la distribución de probabilidad real a partir de una distribución simulada. Este método se conoce como método aceptación-rechazo y se puede fundamentar en el siguiente teorema.

Teorema 1 (El método de aceptación-rechazo). *Sea X una variable aleatoria con función de probabilidad $p(x)$. Supóngase que $p(x)$ puede ser expresada como*

$$p(x) = c g(x) h(x) \quad (6)$$

donde $c \geq 1$, $0 \leq g(x) \leq 1$ y $h(x)$ es una función de probabilidad. Sea U una variable aleatoria uniforme $U(0, 1)$ y sea Y una variable aleatoria con función de probabilidad $h(y)$ independiente de U . Entonces, la función de probabilidad condicional de Y dado que $u \leq g(y)$ coincide con la función de probabilidad de X . Por otra parte, la probabilidad de aceptar la muestra es $1/c$.

El método de aceptación-rechazo para nuestro caso, genera las variables una a una de forma que se muestra una variable sólo cuando ya han sido muestreados todos sus padres. Según este método se simulan todas las variables, incluyendo las evidenciales.

Sea la red bayesiana de la Figura 1. La función de probabilidad conjunta puede factorizarse como en (4). Los pasos a seguir son:

1. Muestreo de la variable A : Se consulta un generador de muestras aleatorias que, basándose en la distribución de probabilidad $P(a)$, de $+a$ con probabilidad 0.3 y $\neg a$ con probabilidad 0.7. Seguidamente se utiliza este valor \hat{a}_1 para calcular las probabilidades de los restantes nodos de extracción.
2. Muestreo de la variable B : dado $A = \hat{a}_1$, se consulta la probabilidad condicionada $P(b|\hat{a}_1)$ y se genera una muestra aleatoria con las correspondientes probabilidades. Supóngase que dicho valor para B resulta ser \hat{b}_1 .
3. Muestreo de la variable C : dado $A = \hat{a}_1$, se consulta la probabilidad condicionada $P(c|\hat{a}_1)$ y se genera una muestra aleatoria con las correspondientes probabilidades. Supóngase que dicho valor para C resulta ser \hat{c}_1 .
4. Muestreo de la variable D : dado $B = \hat{b}_1$, se consulta la probabilidad condicionada $P(d|\hat{b}_1)$ y se genera una muestra aleatoria con las correspondientes probabilidades. Supóngase que dicho valor para D resulta ser \hat{d}_1 .
5. Muestreo de la variable F : dado $D = \hat{d}_1$ y $C = \hat{c}_1$, se consulta la probabilidad condicionada $P(f|\hat{c}_1, \hat{d}_1)$ y se genera una muestra aleatoria con las correspondientes probabilidades. Supóngase que dicho valor para F resulta ser \hat{f}_1 . Si el valor resulta ser $\neg f$, entonces se rechaza esta extracción porque no coincide con la evidencia $F = +f$ y se comienza de nuevo desde 1. Si toma el valor de la evidencia, la simulación continúa.
6. Muestreo de las variables G : dado $D = \hat{d}_1$, se consulta la probabilidad condicionada $P(g|\hat{d}_1)$ y se genera una muestra aleatoria con las correspondientes probabilidades. Supóngase que dicho valor para G resulta ser \hat{g}_1 . Si el valor resulta ser $+g$, entonces se rechaza esta extracción porque no coincide con la evidencia $G = \neg g$ y se comienza de nuevo desde 1. Si toma el valor de la evidencia, la simulación concluye aquí por ser el último nodo.

La extracción concluye con la primera realización

$$x^1 = (a_1, b_1, c_1, d_1, f_1, g_1) = (\hat{a}_1, \hat{b}_1, \hat{c}_1, \hat{d}_1, \hat{f}_1, \hat{g}_1)$$

Se repite el proceso hasta que se obtienen N realizaciones. Entonces, la distribución de probabilidad de cualquier variable se puede estimar por el porcentaje de muestras en las que ocurre el suceso de interés. Para nuestro caso, nos fijamos en el número de veces que hemos obtenido $B = +b$ y dividimos entre el número de realizaciones N .

Ponderación por verosimilitud El método de la función de verosimilitud pensante trata de resolver los problemas de alto rechazo como en el método anterior. La distribución que se utiliza en la simulación es

$$h(x_i) = \begin{cases} P_e(x_i|\pi_i), & \text{si } X_i \notin E, \\ 1, & \text{si } X_i \in E \text{ y } x_i = e, \\ 0, & \text{en otro caso.} \end{cases}$$

siendo E el subconjunto de variables de la evidencia, π_i los antepasados de la variable X_i (sólo los padres en el caso de una red bayesiana por la propiedad de Markov) y

$$P_e(x_i|\pi_i) = \begin{cases} P(x_i|\pi_i), & \text{si } x_i \text{ y } \pi_i \text{ son consistentes con } e, \\ 0, & \text{en otro caso.} \end{cases}$$

Sea la red bayesiana de la Figura 1. La función de probabilidad conjunta puede factorizarse como en (4). Los pasos a seguir son:

1. Se comienza por asignar a los nodos evidenciales sus correspondientes valores observados $\hat{f}_1 = +f$ y $\hat{g}_1 = \neg g$.
2. Siguiendo la ordenación de los nodos no evidenciales (orden ancestral), se simula A . Esto es, usando la función $P(a)$, se simula de forma aleatoria A de forma que devuelva $+a$ con probabilidad 0.3 y $\neg a$ con probabilidad 0.7. Supóngase que el valor simulado para A es \hat{a}_1 .
3. Seguidamente se simula B usando $P(b|\hat{a}_1)$. Supóngase que se obtiene \hat{b}_1 .
4. Seguidamente se simulan el resto de variables no evidenciales siguiendo la misma metodología y se guarda su resultado.

El peso¹ de la muestra obtenida

$$x^1 = (\hat{a}_1, \hat{b}_1, \hat{c}_1, \hat{d}_1, +f, \neg g)$$

resulta

$$s(x^1) = P(+f|\hat{c}_1, \hat{d}_1) \times P(\neg g|\hat{d}_1)$$

Se repite el proceso hasta obtener el número deseado de realizaciones. Al final de las simulaciones se normalizan los pesos y se utilizan los pesos normalizados para estimar la probabilidad de nuestro suceso de interés $P(+b|+f, \neg g)$. Nótese que esto se hace para mejorar la eficiencia que supone utilizar una constante c del método de aceptación-rechazo. A la función $s(x)$ se la conoce como función de peso².

¹El peso asociado a una realización se define

$$s(x) = \frac{P_e(x)}{h(x)} = \prod_{X_i \in E} P(e_i|\pi_i),$$

donde la última igualdad es específica del método de ponderación por verosimilitud

²Véase [1]

Ejercicio 2.4 Sea la red bayesiana

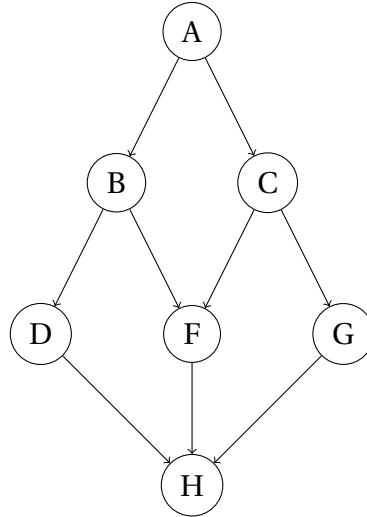


Figura 3: Red bayesiana

donde se supone que conocemos las tablas de probabilidades condicionadas que definen la red. Vamos a mostrar los pasos necesarios para calcular la probabilidad $P(a|h)$, para todo $a \in A$, $h \in H$.

La probabilidad conjunta se factoriza como

$$\begin{aligned} P(a, h) &= \sum_{b, c, d, f, g} P(a, b, c, d, f, g, h) \\ &= \sum_{b, c, d, f, g} P(a)P(b|a)P(c|a)P(d|b)P(f|b, c)P(g|c)P(h|d, f, g) \end{aligned}$$

Marginalizando, obtenemos la probabilidad de la evidencia

$$P(h) = \sum_a P(a, h)$$

Finalmente, obtenemos la probabilidad deseada

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

Eliminación de variables Tomamos la variable G y calculamos el primer potencial

$$\psi_1(c, d, f, h) = \sum_g P(g|c)P(h|d, f, g)$$

Por tanto,

$$P(a, h) = \sum_{b, c, d, f} P(a)P(b|a)P(c|a)P(d|b)P(f|b, c)\psi_1(c, d, f, h)$$

Seguidamente vamos a eliminar la variable F

$$\psi_2(b, c, d, h) = \sum_f P(f|b, c)\psi_1(c, d, f, h)$$

De modo que,

$$P(a, h) = \sum_{b, c, d} P(a)P(b|a)P(c|a)P(d|b)\psi_2(b, c, d, h)$$

Ahora eliminamos la variable D

$$\psi_3(b, c, h) = \sum_d P(d|b) \psi_2(b, c, d, h)$$

Lo que da lugar a

$$P(a, h) = \sum_{b,c} P(a) P(b|a) P(c|a) \psi_3(b, c, h)$$

por último, eliminamos la variable B

$$\psi_4(a, c, h) = \sum_b P(b|a) \psi_3(b, c)$$

Finalmente, sumando sobre $c \in C$ obtenemos

$$P(a, h) = \sum_c P(a) P(c|a) \psi_4(a, c, h)$$

La probabilidad $P(h)$ se calcula como hemos visto anteriormente.

Agrupamiento El método de agrupamiento se divide en dos fases

1. Construcción del árbol de grupos: construimos el grafo moral casando los padres y eliminando la dirección de los enlaces Eliminación de variables: intentamos eliminar

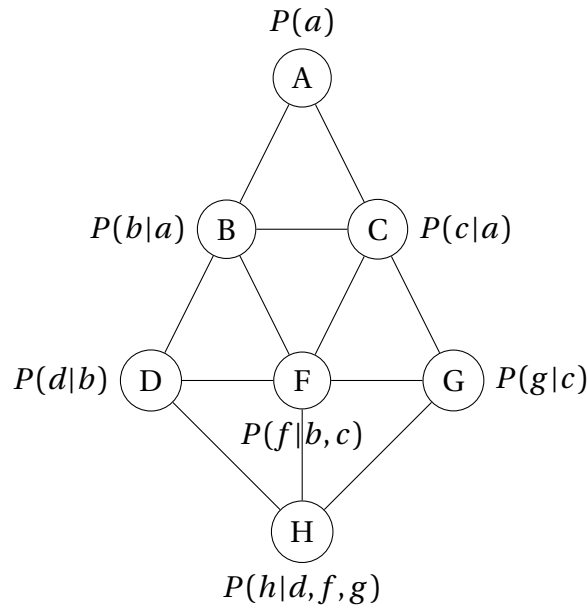


Figura 4: Grafo moral. Paso 1

primero los nodos que tienen todos sus vecinos casados. Tomamos la variable A y eliminamos su correspondiente nodo.

Eliminamos A

- Vecinos: $\{B, C\}$
- Grupo: $\{A, B, C\}$
- Separador: $\{B, C\}$

Al eliminar A el grafo quedaría de la forma

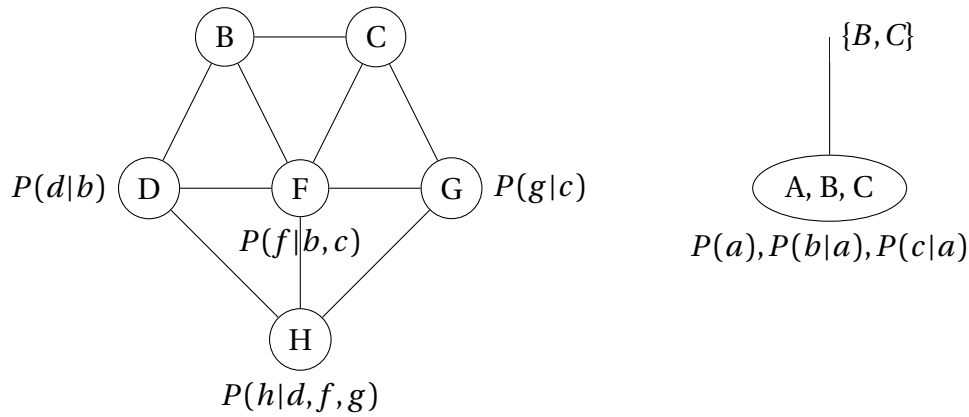


Figura 5: Árbol de grupos. Paso 2

Ahora nos encontramos ante un grafo en el que no hay ningún nodo que tenga todos sus vecinos casados. Tenemos que casar los vecinos de la variable que queremos eliminar para poder formar un grupo.

Eliminamos B , casando D y C

- Vecinos: $\{C, D, F\}$
- Grupo: $\{B, C, D, F\}$
- Separador: $\{C, D, F\}$

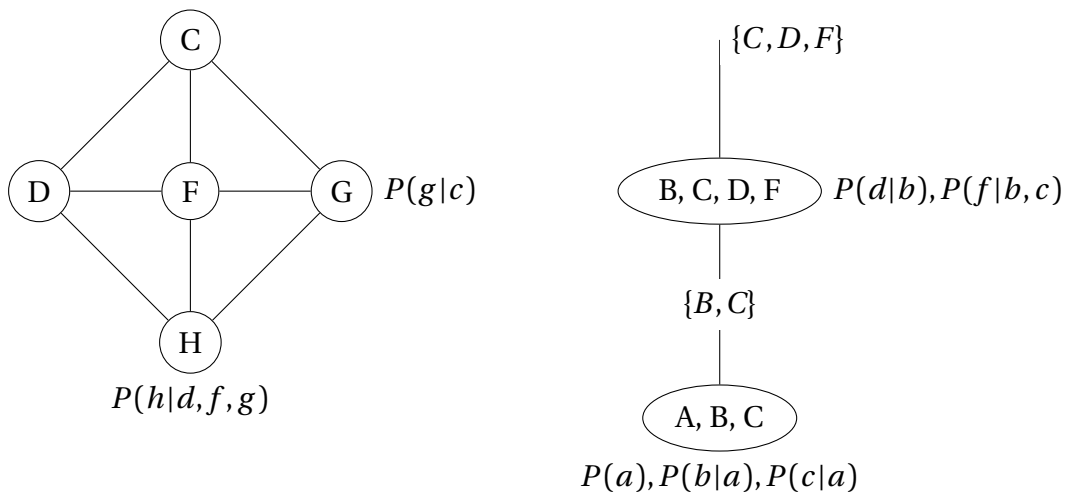


Figura 6: Árbol de grupos. Paso 3

Dado que el separador del grupo $\{A, B, C\}$, esto es $\{B, C\}$, está contenido en el grupo $\{B, C, D, F\}$, este puede adoptar al mismo.

Volvemos a encontrarnos ante un grafo en el que ninguno de los nodos tiene todos sus vecinos casados. Tenemos, por tanto, que casar, por ejemplo, D y G , para eliminar C .

Eliminamos C , casando D y G

- Vecinos: $\{D, F, G\}$
- Grupo: $\{C, D, F, G\}$
- Separador: $\{D, F, G\}$

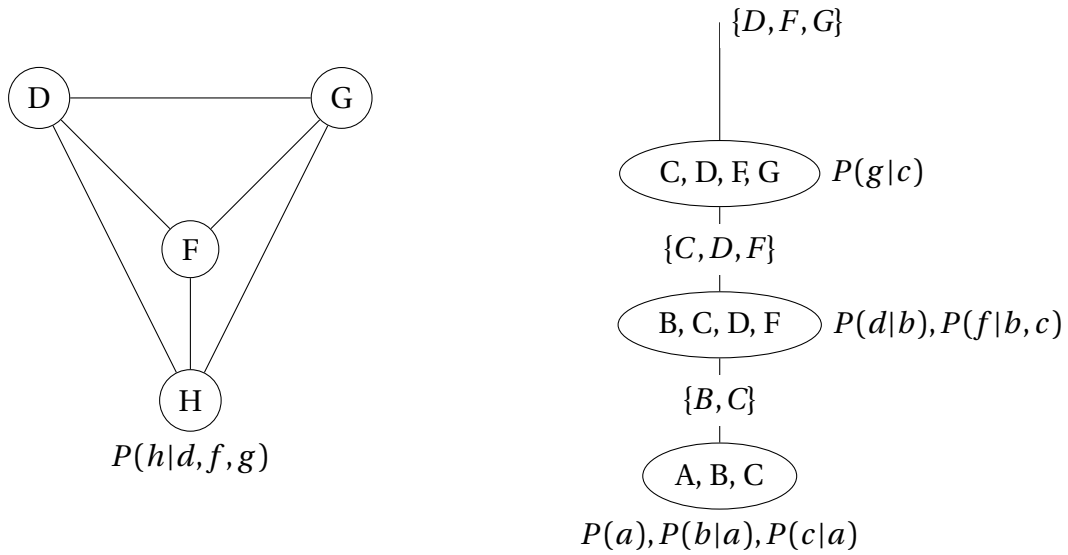


Figura 7: Árbol de grupos. Paso 4

Vemos que ahora todos los nodos tienen todos sus vecinos casados (grafo completamente conexo), por lo que podemos eliminar cualquier nodo. Una alternativa más eficiente sería eliminar todos los nodos a la vez, de forma que quedaría un solo grupo. Esto es posible por una propiedad de los grafos completamente conexos.

El árbol de grupos finalmente sería

- Vecinos: $\{\emptyset\}$
- Grupo: $\{C, D, F, G\}$
- Separador: $\{\emptyset\}$

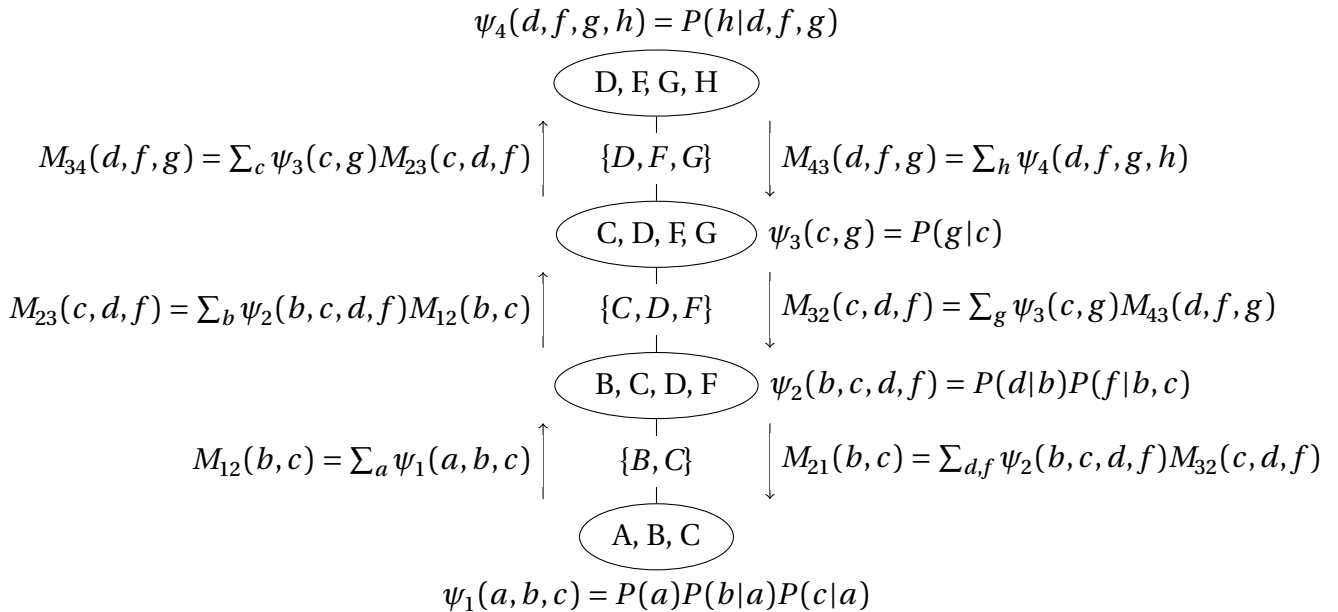


Figura 8: Árbol de grupos. Propagación de la evidencia

2. Propagación de la evidencia: Hay distintas heurísticas para calcular la propagación de la evidencia. Nosotros elegimos la propagación en dos fases por su versatilidad, ya que nos permitiría calcular cualquier otra probabilidad de una forma más sencilla.

Al tratarse de un árbol de grupos en el que no hay bifurcaciones, i.e., un padre con más de un hijo, la propagación de la evidencia es análoga a la que hicimos en la eliminación de variables.

La probabilidad deseada se calcularía

$$P(a|h) = \sum_{b,c} \psi_1(a, b, c) M_{21}(b, c)$$

siempre y cuando hayamos sustituido previamente el hallazgo h . De esta forma, la probabilidad ya será una probabilidad condicionada por haber incluido la evidencia desde el comienzo. Por ejemplo, tendríamos que

$$\begin{aligned} \psi_4(d, f, g, h) &\longrightarrow \psi_4(d, f, g) \\ M_{43}(d, f, g) &\longrightarrow \psi_4(d, f, g) \end{aligned}$$

La ventaja de utilizar la propagación en dos fases es que podríamos calcular cualquier otra probabilidad de nuestra red bayesiana sin necesidad de hacer muchos más cálculos como ocurriría en la eliminación de variables.

Inversión de arcos El método de inversión de arcos fue inicialmente propuesto para diagramas de influencia, sin embargo, también es aplicable a redes bayesianas como la de nuestro ejemplo.

La principal idea de este método consiste en invertir algunos enlaces de la red bayesiana, de forma que la probabilidad que queremos calcular no se vea afectada. Así, por cada inversión de arco, se deben añadir nuevos arcos que compensen los cambios realizados en el grafo de forma que las probabilidades no cambien su valor.

Una vez consigamos variables sumidero, i.e., variables que no son de interés y que no tiene ningún hijo, podemos ir eliminándolas del grafo hasta que obtengamos una red bayesiana que nos permita calcular la probabilidad deseada de forma sencilla.

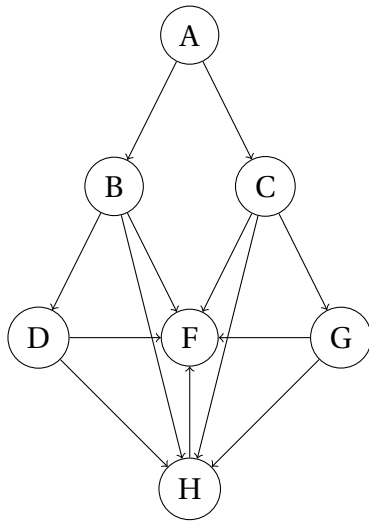
Los pasos a seguir en nuestro caso son:

1. Comenzamos invirtiendo el enlace $F \rightarrow H$ de forma que F se convierte en variable sumidero.
2. Ahora invertimos el enlace $G \rightarrow H$ de forma que G se convierte en variable sumidero.
3. Invertimos el arco $D \rightarrow H$ de forma que D se convierte en variable sumidero.
4. Finalmente invertimos el enlace $C \rightarrow H$ de forma que C se convierte en variable sumidero.

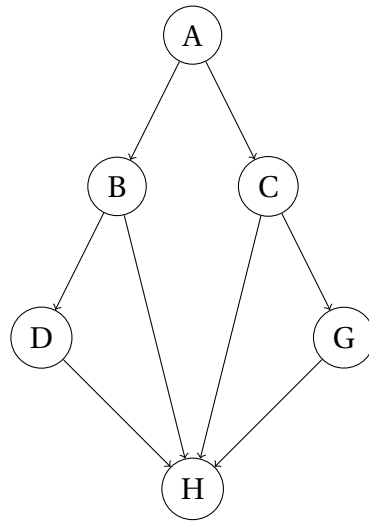
Una vez alcanzado un grafo en el que solo hay tres potenciales, ya podemos calcular la probabilidad deseada tal y como se explica en los apuntes.

Otra opción para calcular la probabilidad $P(a|h)$ sería invertir el enlace $B \rightarrow H$ de forma que B se convierte en variable sumidero, eliminar B y posteriormente invertir el enlace $A \rightarrow H$. Así, obtenemos un grafo simple de dos nodos en el que la única probabilidad condicionada que existe es la que estamos buscando.

Hemos buscado eliminar primero aquellas variables que requieran un menor número de inversiones de arcos. Esto disminuye el coste computacional, el cual es el principal objetivo de estos métodos.

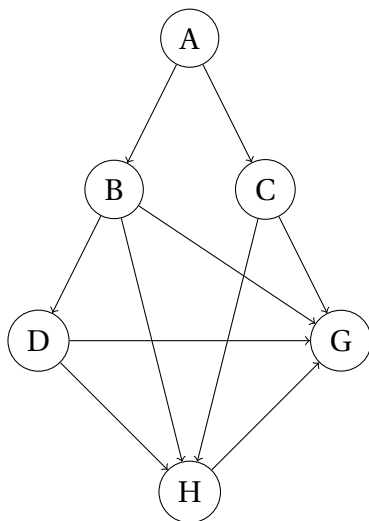


(a) Inversión de arco $F \rightarrow H$

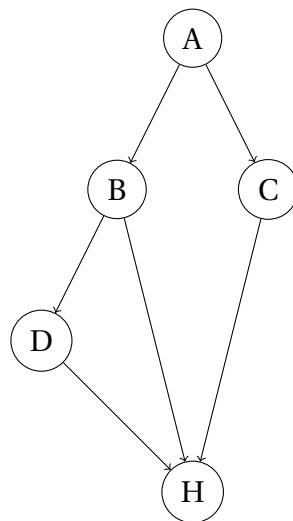


(b) Eliminación de variable F

Figura 9: Inversión de arcos. Paso 1

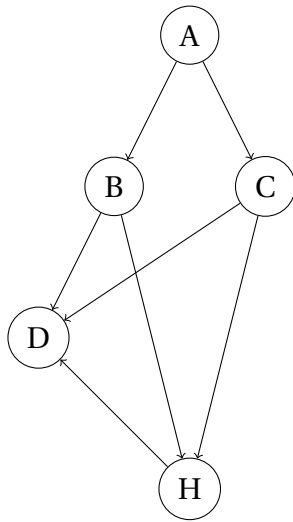


(a) Inversión de arco $G \rightarrow H$

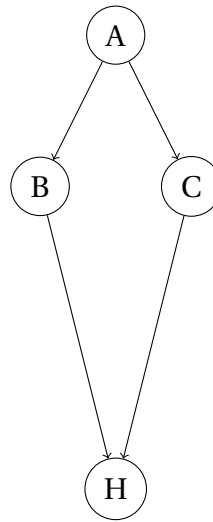


(b) Eliminación de variable G

Figura 10: Inversión de arcos. Paso 2

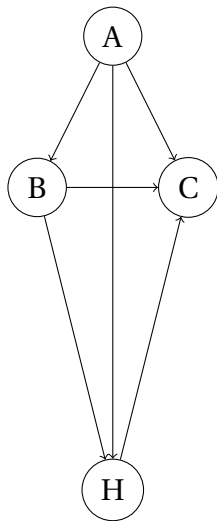


(a) Inversión de arco $D \rightarrow H$

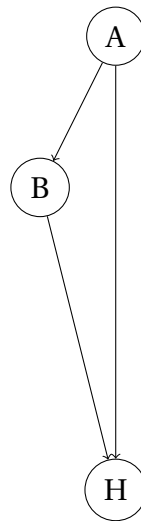


(b) Eliminación de variable D

Figura 11: Inversión de arcos. Paso 3



(a) Inversión de arco $C \rightarrow H$



(b) Eliminación de variable C

Figura 12: Inversión de arcos. Paso 4