

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS MATEMÁTICAS**  
**Departamento de Estadística e Investigación Operativa**



**ANÁLISIS DE SENSIBILIDAD EN REDES BAYESIANAS  
GAUSSIANAS**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR  
PRESENTADA POR**

**Rosario Susi García**

Bajo la dirección de los doctores:  
Miguel Ángel Gómez Villegas y Paloma Maín Yaque

**Madrid, 2007**

**ISBN 978-84-669-3133-5**

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE CIENCIAS MATEMÁTICAS**

**Departamento de Estadística e Investigación Operativa**



**ANÁLISIS DE SENSIBILIDAD EN  
REDES BAYESIANAS GAUSSIANAS**

**MEMORIA PRESENTADA PARA OPTAR AL GRADO DE  
DOCTOR POR**

Rosario Susi García

Bajo la dirección de los Doctores:

Miguel Ángel Gómez Villegas  
Paloma Maín Yaque

**Madrid, 2007**



# **ANÁLISIS DE SENSIBILIDAD EN REDES BAYESIANAS GAUSSIANAS**

Tesis Doctoral de Rosario Susi García

Realizada bajo la dirección de  
Dr. Miguel Ángel Gómez Villegas  
Dra. Paloma Maín Yaque

Madrid, 2007

Departamento de Estadística e Investigación Operativa  
FACULTAD DE CIENCIAS MATEMÁTICAS  
UNIVERSIDAD COMPLUTENSE DE MADRID

A Gelo.

# Índice

|   |          |
|---|----------|
| <b>Prólogo</b>  | <b>v</b> |
| <b>1 Definiciones Fundamentales</b>   | <b>1</b> |
| 1.1. Introducción . . . . .   | 1        |
| 1.2. Conceptos de la Teoría de Grafos . . . . .                                 | 1        |
| 1.2.1. Tipos de Grafos . . . . .  | 2        |
| 1.2.2. Conceptos básicos de Grafos no dirigidos . . . . .                       | 3        |
| 1.2.3. Conceptos básicos de Grafos dirigidos . . . . .                          | 6        |
| 1.2.4. Conceptos básicos de grafos mixtos . . . . .                             | 8        |
| 1.2.5. Estructuras gráficas de interés . . . . .                                | 10       |
| 1.3. Grafos y distribuciones de probabilidad . . . . .                          | 17       |
| 1.3.1. Independencia condicionada . . . . .                                     | 17       |
| 1.3.2. Factorización de una función . . . . .                                   | 19       |
| 1.3.3. Propiedades de Markov sobre Grafos . . . . .                             | 21       |
| 1.3.4. Separación gráfica . . . . .   | 30       |
| 1.4. Modelos gráficos probabilísticos . . . . .                                 | 36       |
| 1.4.1. Modelos gráficos probabilísticos no dirigidos. Redes de Markov . . . . . | 38       |

|          |  |            |
|----------|--|------------|
| 1.4.2.   | Modelos gráficos probabilísticos dirigidos. Redes Bayesianas                     | 40         |
| 1.4.3.   | Modelos gráficos probabilísticos mixtos. Redes Cadena . . .                      | 42         |
| <b>2</b> | <b>Redes Bayesianas</b>  | <b>45</b>  |
| 2.1.     | Introducción . . . . .   | 45         |
| 2.2.     | Redes Bayesianas . . . . .   | 46         |
| 2.2.1.   | Tipos de Redes Bayesianas . . . . .  | 51         |
| 2.2.2.   | Aplicaciones de las Redes Bayesianas . . . . .                                   | 60         |
| 2.3.     | Propagación de la evidencia en Redes Bayesianas . . . . .                        | 65         |
| 2.3.1.   | Evidencia y propagación . . . . .  | 67         |
| 2.3.2.   | Algoritmos de propagación de la evidencia . . . . .                              | 68         |
| 2.4.     | Redes Bayesianas Gaussianas . . . . .  | 94         |
| 2.4.1.   | Construcción de una Red Bayesiana Gaussiana . . . . .                            | 97         |
| 2.4.2.   | Estructura de dependencia . . . . .  | 101        |
| 2.4.3.   | Independencia condicionada . . . . .   | 103        |
| 2.4.4.   | Propagación de la evidencia en Redes Bayesianas Gaussianas . . . . .             | 106        |
| <b>3</b> | <b>Sensibilidad en Redes Bayesianas Gaussianas</b>                               | <b>113</b> |
| 3.1.     | Introducción . . . . .   | 113        |
| 3.2.     | Análisis de sensibilidad para Redes Bayesianas: aproximación histórica . . . . . | 115        |
| 3.3.     | Análisis de sensibilidad de una vía para Redes Bayesianas Gaussianas . . . . .   | 120        |
| 3.3.1.   | Medida de sensibilidad . . . . .   | 123        |
| 3.3.2.   | Análisis de sensibilidad de una vía . . . . .                                    | 124        |

|        |   |     |
|--------|---|-----|
| 3.3.3. | Casos extremos en la relación entre la variable de interés y la variable evidencial . . . . . | 143 |
| 3.3.4. | Algoritmo para el cálculo de la medida de sensibilidad . . . . .                              | 147 |
| 3.4.   | Sensibilidad para perturbaciones extremas . . . . .   | 150 |
| 3.4.1. | Casos extremos en la relación entre la variable de interés y la variable evidencial . . . . . | 158 |
| 3.5.   | Análisis de Sensibilidad de n vías para Redes Bayesianas Gaussianas . . . . .                 | 160 |
| 3.5.1. | Medida de sensibilidad . . . . .  | 163 |
| 3.5.2. | Análisis de sensibilidad de n vías . . . . .  | 165 |
| 3.6.   | Robustez en Redes Bayesianas Gaussianas . . . . .   | 178 |
| 3.7.   | Conclusiones y futuras líneas de investigación . . . . .                                      | 184 |

|                     |            |
|---------------------|------------|
| <b>Bibliografía</b> | <b>189</b> |
|---------------------|------------|





# Prólogo

El objetivo de esta Memoria se centra en desarrollar una metodología para efectuar un análisis de sensibilidad en Redes Bayesianas Gaussianas.

Las Redes Bayesianas son un tipo de modelos gráficos probabilísticos, que se caracterizan por modelizar dependencias de tipo causal. Dentro de las Redes Bayesianas, se describen las Redes Bayesianas Discretas y las Redes Bayesianas Gaussianas, siendo estas últimas el objeto fundamental de estudio en la presente Memoria.

Los modelos gráficos probabilísticos se componen de una parte cualitativa, dada por un grafo que representa la estructura de dependencia entre las variables del problema, y una parte cuantitativa, que hace referencia a la distribución condicionada o conjunta de las variables del mismo. Por tanto, en los modelos gráficos probabilísticos se aúnan conceptos de la Teoría de Grafos y de la Teoría de la Probabilidad con la finalidad de modelizar un conjunto de variables relacionadas entre sí.

Para poder describir el modelo, dado por una Red Bayesiana Gaussiana, es fundamental introducir algunos conceptos y definiciones de la Teoría de Grafos y de la Teoría de la Probabilidad, en los que se apoya el modelo de estudio. Con este objetivo, en el Capítulo 1 de esta Memoria, se presentan definiciones fundamentales para el desarrollo posterior de las Redes Bayesianas Gaussianas. Dichas definiciones hacen referencia a estructuras gráficas de interés, como los grafos no dirigidos, los grafos dirigidos y los grafos mixtos. También se introducen algunas manipulaciones sobre el grafo, que mantienen la estructura de dependencia y que facilitan el tratamiento local de las distribuciones de probabilidad que definen el problema.

Ligadas a las definiciones de la Teoría de Grafos presentadas, se introducen

conceptos de separación gráfica, que se reflejan en la distribución de probabilidad mediante relaciones de independencia condicionada entre variables, lo cual permite la descripción probabilística de la red, a través de una factorización de la distribución conjunta.

Una vez que se han introducido estos conceptos, básicos para el desarrollo de un modelo gráfico probabilístico, se definen los tres tipos de modelos gráficos probabilísticos existentes, sobre grafos no dirigidos, grafos dirigidos y grafos mixtos. Además, se caracterizan las Redes de Markov, las Redes Bayesianas y las Redes Cadena, como las clases más destacables dentro de los modelos gráficos probabilísticos introducidos.

En el Capítulo 2 se describen las Redes Bayesianas y sus principales características. Cabe destacar, entre los autores que han estudiado los fundamentos y aplicaciones de las Redes Bayesianas, nombres como Dawid (1979), Pearl (1988), Heckerman (1995), Lauritzen (1996) o Jensen (2001).

Una Red Bayesiana, además de describir un problema formado por un conjunto de variables relacionadas entre sí, puede actualizar la información que se tiene acerca de las variables del problema, cuando se conoce el valor que toma alguna de ellas para casos concretos. Este proceso se denomina actualización de la evidencia y se realiza para conocer y describir las variables de interés del problema cuando se tiene información o evidencia acerca de la ocurrencia de otra u otras variables del mismo.

Se han desarrollado diversos algoritmos para propagar la evidencia en Redes Bayesianas. En esta Memoria, se presentan dos algoritmos fundamentales para Redes Bayesianas Discretas y un algoritmo para Redes Bayesianas Gaussianas, que se utilizará posteriormente en el desarrollo del análisis de sensibilidad propuesto.

Además de introducir un algoritmo de propagación, se profundiza en los conceptos de las Redes Bayesianas Gaussianas. La característica principal de este modelo es la de ser una Red Bayesiana en la que todas las variables del problema tienen distribución normal. De manera que, se definen las Redes Bayesianas Gaussianas como aquellas cuya distribución conjunta es normal multivariante. A continuación, se enuncian y estudian propiedades fundamentales de los parámetros que caracterizan relaciones de independencia condicionada.

Cuando se modeliza un problema mediante una Red Bayesiana Gaussiana, y

en general mediante una Red Bayesiana, los expertos en el campo de aplicación de la red han de determinar las relaciones de dependencia que existen entre las variables, para reflejarlas posteriormente en el grafo acíclico dirigido (DAG) que representa la parte cualitativa de la red. Posteriormente, los expertos han de asignar valores a los parámetros de la distribución de probabilidad conjunta o condicionada que describen la parte cuantitativa de la misma.

En el caso de una Red Bayesiana Gaussiana, se deben determinar por ejemplo, los valores de los parámetros que describen la distribución conjunta normal multivariante. Esta tarea es compleja, porque a veces se trabaja con información parcial del problema, que puede llevar a detallar la red de forma inexacta. Por este motivo, surge la necesidad de estudiar la sensibilidad de la Red Bayesiana Gaussiana que modeliza el problema de interés.

En el Capítulo 3, se desarrollan dos análisis de sensibilidad para determinar la respuesta de la red frente a cambios en los parámetros. Se estudia el problema tras realizarse la propagación de la evidencia y se da una medida que permite cuantificar la sensibilidad de los resultados.

Se han propuesto diversas técnicas para determinar la sensibilidad de las Redes Bayesianas, la mayoría para Redes Discretas. Así, autores como Laskey (1995), Coupé, et al. (2002) o Chan, et al. (2004) describen análisis de sensibilidad para este tipo de redes. En Redes Bayesianas Gaussianas, cabe destacar el desarrollo propuesto por Castillo, et al. (2003), que generaliza la idea de Laskey (1995) estudiando así, pequeñas incertidumbres y pequeñas perturbaciones alrededor de los parámetros propuestos inicialmente al describir el problema, efectuando por tanto un análisis de sensibilidad local.

Con el análisis de sensibilidad de una vía que se presenta en el Capítulo 3, se diseña un análisis de sensibilidad global para Redes Bayesianas Gaussianas que cuantifica el efecto de la incertidumbre acerca de los parámetros que describen la distribución conjunta de las variables del problema. Posteriormente, se generalizan los resultados obtenidos y se desarrolla un análisis de sensibilidad de  $n$  vías para Redes Bayesianas Gaussianas.

Ambos análisis de sensibilidad consisten en utilizar la divergencia de Kullback-Leibler después de la propagación de la evidencia. Otros autores como Chan, et al. (2004) también apoyan sus análisis de sensibilidad en medidas de discrepancia.

La metodología seguida en ambos análisis consiste en, modificar los parámetros

que describen la red, cuantificando la incertidumbre mediante unas perturbaciones aditivas, de forma que se llega a un modelo perturbado. Tanto el modelo original, con los parámetros inicialmente asignados, como el modelo perturbado, describen la red inicialmente, es decir, antes de introducir evidencia acerca de las variables del problema. A continuación, se introduce la evidencia en la red y se propaga por la misma, obteniéndose dos salidas de interés, la del modelo original y la del modelo perturbado. Con la medida de sensibilidad propuesta, se comparan ambas salidas.

En el análisis de sensibilidad de una vía desarrollado, se introduce en cada paso una única perturbación asociada a un parámetro. De esta forma, se obtiene una medida de sensibilidad para cada tipo de parámetro perturbado, adoptando diferentes expresiones que permiten evaluar la importancia de la perturbación en los distintos casos.

En el análisis de sensibilidad de  $n$  vías, se consideran, en cada paso, un conjunto de parámetros perturbados, respondiendo a unas características concretas. De nuevo, en función del conjunto de parámetros perturbados, se obtienen diferentes medidas de sensibilidad que pueden ser comparadas y teniendo en cuenta sus valores concluirse resultados acerca de la sensibilidad de la red.

En el Capítulo 3, también se propone una aproximación para estudiar la robustez de una Red Bayesiana Gaussiana. La idea de la robustez surge de los resultados asociados al análisis de sensibilidad para distintas Redes Bayesianas Gaussianas.

Para obtener un resultado concreto acerca de la robustez de una Red Bayesiana Gaussiana, se desarrolla un análisis de robustez siguiendo la propuesta introducida para los análisis de sensibilidad. De esta forma, se calcula una única medida de robustez, para todas las posibles imprecisiones que se determinan al asignar los parámetros de la red, y en función de su valor se puede concluir que la Red Bayesiana Gaussiana de estudio, es más o menos robusta frente a las perturbaciones propuestas.

Finalmente, quiero agradecer la inestimable ayuda que los profesores Miguel Ángel Gómez Villegas y Paloma Maín Yaque me han brindado en todo momento, sobresaliendo su valía profesional y humana. También quiero darle las gracias a mi familia y a todos los que con su apoyo, ánimo y comprensión han hecho posible la realización de esta Memoria.

# 1

## Definiciones Fundamentales

### 1.1. Introducción

En este Capítulo se recogen algunas definiciones fundamentales para la descripción del modelo de estudio en esta Memoria, las Redes Bayesianas Gaussianas.

Conceptos asociados a la Teoría de Grafos, a la separación gráfica o a la independencia condicionada, se introducen como base de los modelos gráficos probabilísticos; modelos que se definen en la última Sección del Capítulo y en los que se enmarcan las Redes Bayesianas.

### 1.2. Conceptos de la Teoría de Grafos

Con el fin de profundizar en los modelos gráficos probabilísticos, en este Apartado se introducen definiciones e ideas de la Teoría de Grafos sobre los que se apoyarán algunos resultados posteriormente introducidos.

En este Apartado, además de definir conjuntos y formaciones básicas de un grafo y de describir algunos tipos de grafos, se presentan las estructuras necesarias para poder desarrollar resultados computacionales localmente de forma sencilla, reduciéndose así la complejidad de los cálculos generales.

### 1.2.1. Tipos de Grafos

Cuando se representa gráficamente una colección de objetos  $V = \{V_1, \dots, V_n\}$  que se relacionan entre sí mediante aristas, siendo  $E_{ij}$  la arista que une los elementos  $V_i$  y  $V_j$  de  $V$ , estamos definiendo implícitamente un grafo, donde  $V = \{V_1, \dots, V_n\}$  son los *nodos* y  $E$  el conjunto de *aristas* que lo forman.

**Definición 1.1** (*Grafo*)

Un Grafo se presenta como un par  $\mathcal{G} = (V, E)$ , donde  $V = \{V_1, V_2, \dots, V_n\}$  es el conjunto finito de nodos o vértices y  $E$  es el conjunto de aristas, es decir, el conjunto de pares ordenados de los distintos elementos de  $V$  que se relacionan.

En la Figura 1.1 se muestra un grafo con dos tipos de aristas.

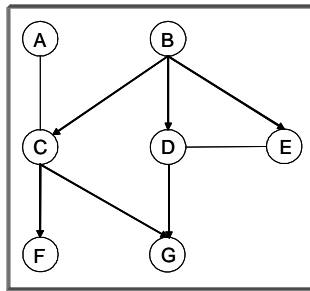


Figura 1.1. Grafo

Dependiendo de la relación y el orden que existe entre los nodos del grafo, se puede hablar de dos tipos de aristas: aristas dirigidas y aristas no dirigidas. De esta manera, se utilizan *aristas dirigidas* cuando  $E_{ij} \in E$  pero  $E_{ji} \notin E$ , y se denota como  $V_i \rightarrow V_j$ , de forma que  $V_i$  se conecta con  $V_j$  y no viceversa. Por el contrario, las *aristas no dirigidas*, cuya notación será  $V_i - V_j$ , se presentan cuando  $E_{ij} \in E$  y  $E_{ji} \in E$ , quedando ambos nodos  $V_i$  y  $V_j$  conectados.

El tipo de arista puede determinar el grafo, así si un grafo tiene todas sus aristas dirigidas, se define el grafo como *grafo dirigido*; si todas las aristas del mismo son no dirigidas, el grafo se denomina *grafo no dirigido* y cuando el grafo tiene aristas dirigidas y no dirigidas, el grafo se dice *grafo mixto*.

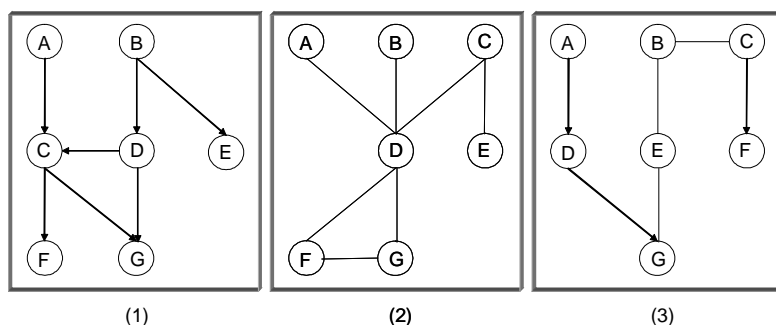


Figura 1.2. Grafo dirigido (1), grafo no dirigido (2) y grafo mixto (3)

En la Figura 1.2 aparece un grafo dirigido (1), un grafo no dirigido (2) y un grafo mixto (3).

Como se puede ver, en el grafo dirigido se observa un orden entre los distintos nodos mientras que no existe orden aparente entre los nodos del grafo no dirigido.

Se define un *camino entre nodos* como una sucesión de nodos conectados por una arista, de este modo si se busca un camino entre los nodos  $V_i$  y  $V_j$  se tendrá la sucesión de nodos  $(V_{k_1}, \dots, V_{k_r})$  donde  $V_i = V_{k_1}$  y  $V_j = V_{k_r}$ , de forma que existe una arista entre los nodos  $V_{k_l}$  y  $V_{k_{l+1}} \forall l = 1, \dots, r$ . Se dice que el *camino es cerrado* cuando el nodo inicial del camino coincide con el nodo final del mismo, es decir, si  $V_{k_1} = V_{k_r}$ .

### 1.2.2. Conceptos básicos de Grafos no dirigidos

En este Apartado se introducen definiciones y características particulares de los grafos no dirigidos  $\mathcal{G} = (V, E)$ , que van a ser utilizadas para reducir la estructura gráfica y solucionar aspectos computacionales.

Algunas definiciones surgen de las relaciones entre nodos y como se agrupan, de esta manera, se utiliza el concepto de *vecino de un nodo*  $V_i$  como el conjunto de nodos directamente alcanzables desde  $V_i$ , y se denota como  $vec(V_i)$ ; y se define la *frontera de un conjunto de nodos*  $C$  como la unión de los conjuntos de vecinos de los nodos que hay en  $C$ , excluyendo los propios elementos de  $C$ , es decir, la frontera denotada como  $frn(C)$ , es tal que  $frn(C) = \left( \bigcup_{V_i \in C} vec(V_i) \right) \setminus C$ . En la Figura 1.2 (2), los vecinos del nodo  $G$  son los nodos  $D$  y  $F$ , tal que



$vec(G) = \{D, F\}$ , y si se busca la frontera del conjunto  $S = \{D, G\}$  se tiene que  $frn(S) = \{A, B, C, F\}$ .

Cuando el grafo de estudio es un grafo no dirigido se define un *bucle* como un camino cerrado, esto es, como una sucesión de nodos conectados tales que el nodo inicial coincide con el nodo final, por tanto, en la Figura 1.2 (2) sólo existe un bucle que viene dado por los nodos  $\{D, F, G\}$ .

A continuación, se presentan algunos tipos de grafos no dirigidos.

**Definición 1.2** (*Subgrafo asociado a un conjunto C*)

Sea  $\mathcal{G} = (V, E)$  un grafo no dirigido, y  $C$  un conjunto de nodos del mismo. Se define el subgrafo  $\mathcal{C}$  asociado al conjunto  $C$  como el grafo no dirigido formado por los nodos de  $C$  y las aristas de  $E$  que unen a elementos de  $C$ .

Las definiciones que se muestran en este Apartado, son fundamentales para tratar el grafo localmente en función de los conjuntos que presentan las características que se muestran a continuación.

**Definición 1.3** (*Grafo completo*)

Un grafo no dirigido  $\mathcal{G} = (V, E)$  es un grafo completo cuando existe una arista entre cada par de nodos.

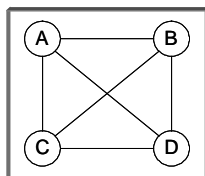


Figura 1.3. Grafo completo

Cuando un grafo no dirigido tampoco es completo, se podrá estudiar si existe algún conjunto completo dentro del propio grafo, a continuación se define dicho concepto.

**Definición 1.4** (*Conjunto completo*)

Sea  $\mathcal{C}$  un subgrafo asociado a un subconjunto  $C$  de un grafo no dirigido, se dice que  $C$  es un subconjunto completo cuando existe una arista entre cada par de nodos de  $C$ .

Por tanto, cada par de nodos unidos por una arista en un grafo no dirigido

formará un conjunto completo.

Una estructura básica en los modelos gráficos probabilísticos que se estudiarán en posteriores capítulos, es el *ciclado* que se obtiene cuando un subconjunto completo del grafo es maximal. Formalmente

**Definición 1.5** (*Ciclado*)

Sea  $C$  un subconjunto completo de un grafo. Se dice que  $C$  es un *ciclado* cuando además  $C$  no es subconjunto propio de otro subconjunto completo, es decir, cuando  $C$  es maximal.

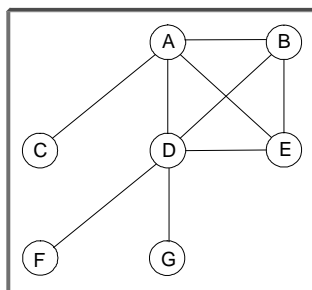


Figura 1.4. Ciclado  $C = \{A, B, D, E\}$  asociado a un grafo no dirigido

Teniendo en cuenta los caminos que aparecen en un grafo, se enuncian dos tipos de grafos: los *grafos conexos no dirigidos* que son aquellos en los que existe al menos un camino entre cada par de nodos, véase la Figura 1.4 como ejemplo también de grafo conexo no dirigido, y los *árboles* que son grafos conexos no dirigidos en los que existe un único camino entre cada par de nodos. El grafo de la Figura 1.5 muestra un árbol obtenido a partir de la Figura 1.4 tras eliminar las aristas  $E_{AE}$ ,  $E_{BD}$ ,  $E_{DE}$ .

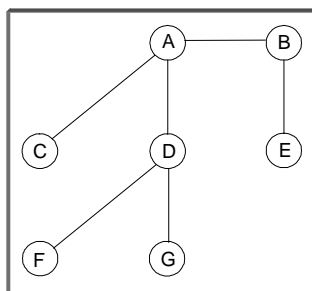


Figura 1.5. Árbol

### 1.2.3. Conceptos básicos de Grafos dirigidos

Al igual que en los grafos no dirigidos, las relaciones y agrupaciones entre los nodos definen conjuntos de nodos específicos. En los grafos dirigidos las relaciones familiares marcan algunas definiciones de igual carácter, así, si  $V_i \rightarrow V_j$  se dice que  $V_i$  es *padre* de  $V_j$ , y se denota como  $pa(V_j)$ , y que  $V_j$  es *hijo* de  $V_i$ .

Al conjunto formado por un nodo  $V_i$  y sus padres  $pa(V_i)$  se le llama *familia del nodo*  $V_i$ , de forma que  $fa(V_i) = V_i \cup pa(V_i)$ .

En el grafo dirigido (1) de la Figura 1.2 se tiene que  $pa(G) = \{C, D\}$  y los hijos de  $C$  son los nodos  $\{F, G\}$ , además  $fa(G) = \{G, C, D\}$ .

Dependiendo del número de padres de un nodo se presentan distintos tipos de grafos dirigidos, así, si cada nodo tiene como máximo un padre, el grafo dirigido se denomina *grafo o árbol simple* y en caso contrario *poliárbol*.

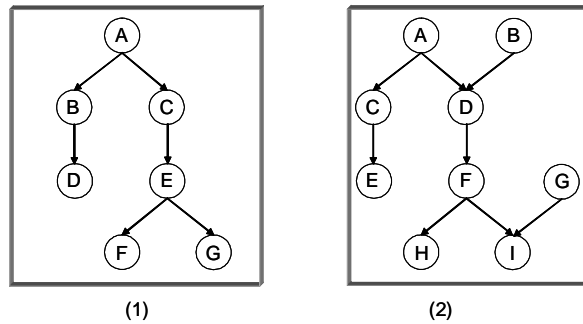


Figura 1.6. Árbol simple (1) y poliárbol (2)

Continuando con las relaciones entre nodos, se definen los *ascendientes de un nodo*  $V_i$ , y se denota por  $as(V_i)$ , al conjunto de nodos que tienen un camino hasta  $V_i$  y se denominan *descendientes del nodo*  $V_i$ ,  $de(V_i)$ , al conjunto de nodos a los que se puede ir desde  $V_i$ , así, en el poliárbol (2) de la Figura 1.6, el nodo  $D$  tiene de ascendientes a los nodos  $\{A, B\}$  y de descendientes a  $\{F, H, I\}$ . Igualmente, se define el conjunto de *no ascendientes de un nodo*  $V_i$ ,  $na(V_i)$ , como el conjunto de nodos  $V$  menos los ascendientes de  $V_i$  y el propio  $V_i$ , de forma que  $na(V_i) = V \setminus (as(V_i) \cup V_i)$  y el conjunto de *no descendientes de un nodo*  $V_i$ ,  $nd(V_i)$ , como el conjunto de nodos dado por  $V \setminus (de(V_i) \cup V_i)$ .

Además, se dice que un conjunto  $C$  es un *conjunto ancestral*, y se denota como

$an(C)$ , cuando  $C$  contiene todos los ascendientes de los nodos que conforman dicho conjunto. En la Figura 1.6 (2) el conjunto  $C = \{A, B, C, D\}$  es un conjunto ancestral.

Como se ha comentado anteriormente, un grafo dirigido, cuya notación vendrá dada por  $\mathcal{D} = (V, E)$ , refleja una ordenación entre sus nodos. Si se le asigna un número a cada uno de los nodos, se dice que se tiene una *numeración ancestral* cuando el número de cada nodo es menor que el correspondiente a sus hijos. En la Figura 1.7 se presenta un numeración ancestral de un grafo dirigido.

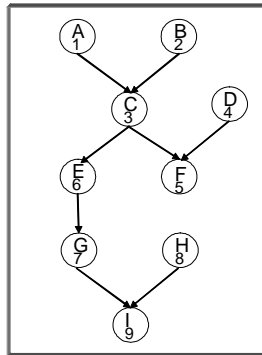


Figura 1.7. Numeración ancestral de un grafo dirigido

El camino cerrado en un grafo dirigido  $\mathcal{D}$  se llama *ciclo* (recuérdese que bucle es cuando el camino cerrado se encuentra en un grafo no dirigido). En función de la existencia o no de ciclos en el grafo se tiene la siguiente definición.

**Definición 1.6** (*Grafo acíclico y cíclico*)

Un grafo dirigido  $\mathcal{D} = (V, E)$  es *acíclico* (DAG) cuando no contiene ningún ciclo; en caso de contener al menos un ciclo el grafo es un *grafo cíclico*.

En la Figura 1.7 se muestra un grafo acíclico dirigido (DAG). Esta estructura gráfica es básica para, posteriormente, poder especificar mediante un grafo un problema con incertidumbre en el que se relacionan un conjunto de variables.

Para finalizar se tiene, que asociado a un grafo dirigido siempre existe un grafo no dirigido que se obtiene cambiando las aristas dirigidas del grafo por aristas no dirigidas, siendo el nuevo grafo el *grafo no dirigido asociado*.

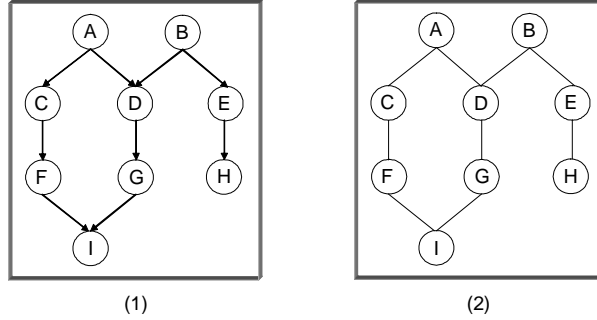


Figura 1.8. DAG (1) y su grafo no dirigido asociado (2)

#### 1.2.4. Conceptos básicos de grafos mixtos

Como se ha visto anteriormente, los grafos mixtos tienen una parte de grafo dirigido y otra de grafo no dirigido. Así, algunas de las definiciones introducidas en los apartados anteriores se ven ahora ampliadas debido al carácter del grafo mixto.

De esta forma, las definiciones de padre e hijo de un nodo presentadas para grafos dirigidos, se pueden aplicar a grafos mixtos. Además, se pueden definir tanto para grafos dirigidos como para grafos mixtos los *padres de un conjunto de nodos*  $C$  que viene dado por  $pa(C) = \left( \bigcup_{V_i \in C} pa(V_i) \right) \setminus C$ , y a su vez, los *hijos de un conjunto de nodos*  $C$  dados por  $hi(C) = \left( \bigcup_{V_i \in C} hi(V_i) \right) \setminus C$ .

En grafos mixtos también se utilizan conceptos asociados a los grafos no dirigidos, así, la definición de vecino de un nodo es válida tanto para grafos no dirigidos como para grafos mixtos. También, en ambos tipos de grafos, se presenta la definición de *vecinos de un conjunto de nodos*  $C$ , que viene dada por  $vec(C) = \left( \bigcup_{V_i \in C} vec(V_i) \right) \setminus C$ .

En los grafos mixtos se amplía la definición de *frontera de un conjunto de nodos*  $C$ , que viene dada por el conjunto de padres y vecinos de  $C$ , es decir  $frn(C) = \left( \left( \bigcup_{V_i \in C} pa(V_i) \right) \cup \left( \bigcup_{V_i \in C} vec(V_i) \right) \right) \setminus C = pa(C) \cup vec(C)$ .

Una nueva definición que aparece en los grafos mixtos es la de *cierre de un*

*conjunto de nodos*  $C$  que se obtiene uniendo a la frontera del conjunto, el propio conjunto  $C$ , es decir  $ci(C) = C \cup frn(C)$ . Tanto el concepto de frontera como el de cierre de un conjunto de nodos  $C$ , se aplica también en grafos dirigidos y en grafos no dirigidos.

Un conjunto ancestral en grafos dirigidos es el conjunto que contiene a todos los ascendientes de los nodos que conforman dicho conjunto, ahora, en los grafos mixtos hay conexiones dirigidas y no dirigidas, por tanto, el *conjunto ancestral* cambia y pasa a definirse como el conjunto más pequeño que contiene a la frontera de todos los elementos que conforman dicho conjunto, es decir, el conjunto  $C$  tal que  $\forall V_i \in C, frn(V_i) \subseteq C$ .

En un grafo mixto se entiende por *ciclo dirigido* al camino cerrado formado por aristas dirigidas.

El grafo mixto que se utiliza para representar un problema con incertidumbre, ha de cumplir que no tenga ciclos dirigidos y se define como

**Definición 1.7** (*Grafo cadena*)

Un grafo mixto  $\mathcal{K} = (V, E)$  es un grafo cadena cuando no contiene ningún ciclo dirigido.

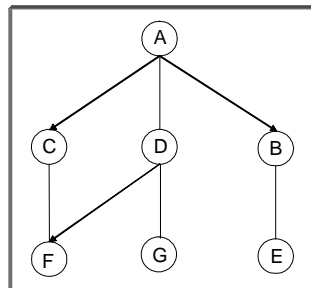


Figura 1.9. Grafo cadena

Esta estructura es básica para el desarrollo posterior de modelos gráficos probabilísticos con relaciones mixtas. Véase la Figura 1.9.

Al igual que sucede en los grafos dirigidos, es posible obtener el grafo no dirigido asociado a un grafo mixto. El *grafo no dirigido asociado* existe siempre y se obtiene tras cambiar las aristas dirigidas del grafo mixto por aristas no dirigidas.

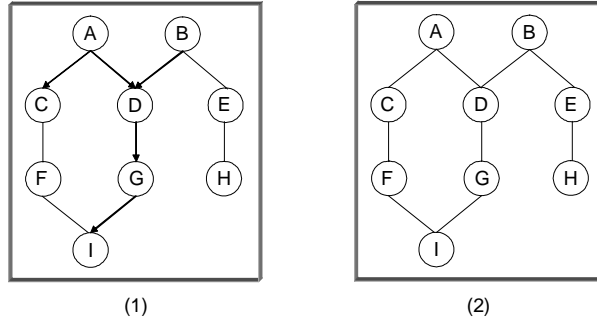


Figura 1.10. Grafo cadena (1) y su grafo no dirigido asociado (2)

### 1.2.5. Estructuras gráficas de interés

En este Apartado se presentan algunas estructuras gráficas que se obtienen tras manipular un grafo, para facilitar el tratamiento del mismo. De esta manera, tras modificar el grafo se obtiene finalmente una estructura simple que mantiene algunas propiedades del grafo original. Las aplicaciones de estos nuevos grafos son diversas y su conocimiento es fundamental para poder tratar los elementos del grafo localmente, sin dificultades computacionales, en procesos que se verán más adelante.

#### Grafo moral

##### Definición 1.8 (Grafo moral)

Sea un grafo no dirigido  $\mathcal{G} = (V, E)$ , que puede venir asociado a un grafo dirigido o un grafo mixto. Se define el grafo moral asociado a  $\mathcal{G}$ , y se denota con  $\mathcal{G}^m$ , como aquel que se obtiene tras añadir una arista entre cada par de nodos con algún hijo común.

La estructura que aquí se presenta se puede obtener igualmente a partir de un grafo dirigido como de un grafo mixto, ya que asociado a ambos tipos de grafos aparece un grafo no dirigido y partiendo de ese nuevo grafo  $\mathcal{G}$  se obtiene el grafo moral asociado  $\mathcal{G}^m$ .

En la Figura 1.11 se presenta el grafo moral asociado al grafo no dirigido (2) de la Figura 1.8.

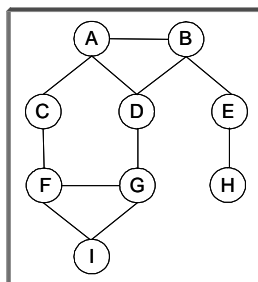


Figura 1.11. Grafo moral  $\mathcal{G}^m$  del grafo no dirigido (2) de la Figura 1.8

### Grafo triangulado y grafo descomponible

A continuación, se introducen los conceptos de grafo triangulado y grafo descomponible que definen estructuras gráficas obtenidas tras modificar un grafo no dirigido, aunque primero se han de introducir nuevos conceptos.

Si en un grafo no dirigido hay un bucle, a la arista que une dos nodos del bucle que no pertenece al bucle, se le denomina *cuerda del bucle*.

En la definición que se presenta a continuación es fundamental este concepto ya que se obtiene un grafo triangulado cuando a cada bucle de longitud mayor o igual que 4 se le añade al menos una cuerda.

#### **Definición 1.9** (*Grafo triangulado*)

*Un grafo no dirigido  $\mathcal{G} = (V, E)$  se define como grafo triangulado, cuando todos los bucles de longitud mayor o igual que cuatro contienen al menos una cuerda.*



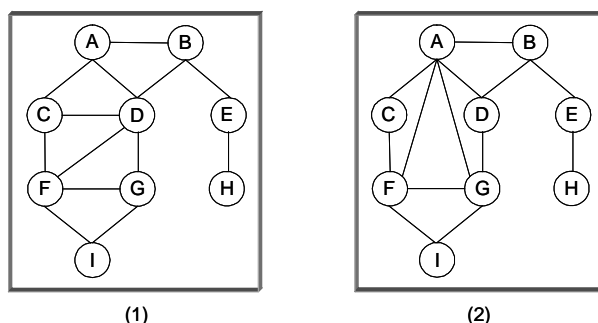


Figura 1.12. Dos grafos triangulados asociados al grafo moral de la Figura 1.11

En la Figura 1.12 se muestran dos grafos triangulados obtenidos a partir del grafo no dirigido (2) de la Figuras 1.8. Como se puede observar en el ejemplo, es posible convertir un grafo en triangulado añadiendo cuerdas que dividan los bucles, aunque este proceso no es trivial ya que se ha de mantener en lo posible la estructura original del grafo, buscando que la triangulación contenga el número mínimo de cuerdas, siendo así una triangulación minimal. Pese a que el problema de obtener la triangulación minimal de un grafo es NP-completo (Yannakakis, 1981), se han desarrollado varios algoritmos para triangular el grafo en tiempo lineal, aunque ninguno de ellos garantiza que la triangulación obtenida sea minimal<sup>1</sup>.

Lauritzen (1996) demuestra un resultado en el que se asocia la obtención de un grafo triangulado con la numeración perfecta de los nodos. Antes de ver este resultado se ha de introducir la definición de numeración perfecta.

**Definición 1.10** (*Numeración perfecta*)

Se dice que una numeración de los nodos de un grafo,  $\alpha$ , es perfecta, si el subconjunto de nodos  $Frn(\alpha(i)) \cap (\alpha(1), \dots, \alpha(i-1))$  es completo para  $i = 2, \dots, n$ , es decir, cuando para cualquier vértice sus vecinos con menor numeración forman un subgrafo completo.

En la Figura 1.13 se observa una numeración perfecta para el grafo triangulado (1) de la Figura 1.12. Aplicando la definición anterior se tiene que  $Frn(\alpha(2)) = Frn(B) = \{A, D, E\}$  y  $Frn(\alpha(2)) \cap \alpha(1)$  es  $\{A, D, E\} \cap \{A\} = \{A\}$  que es un conjunto completo trivial. Para  $Frn(\alpha(3)) = Frn(D) = \{A, B, C, F, G\}$  se tiene que  $Frn(\alpha(3)) \cap (\alpha(1), \alpha(2))$  es  $\{A, B, C, F, G\} \cap \{A, B\} = \{A, B\}$  conjunto

<sup>1</sup>Algunos de estos algoritmos se pueden ver en Rose, et al. (1976) o Tarjan, et al. (1984)

completo, y así para  $i = 4, \dots, 9$ .

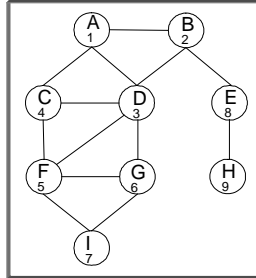


Figura 1.13. Numeración perfecta del grafo triangulado (1) de la Figura 1.12

Nótese que la numeración perfecta de un grafo no tiene por que ser única.

### Teorema 1.1

Un grafo no dirigido  $\mathcal{G} = (V, E)$  es triangulado si y solo si admite una numeración perfecta.

### Demostración 1.1

La demostración puede verse en Lauritzen (1996). ■

Una característica importante de los grafos triangulados, es la propiedad de intersección dinámica, que se define a continuación y que favorece la ordenación de ciclados, de forma que los nodos comunes a un ciclado específico y a todos los anteriores estén contenidos en algunos de los ciclados previos, es decir

### Definición 1.11 (Propiedad de intersección dinámica)

Una numeración de los ciclados de un grafo no dirigido  $\{C_1, \dots, C_k\}$  satisface la propiedad de intersección dinámica cuando  $C_i \cap (C_1 \cup \dots \cup C_{i-1})$  está contenido en, al menos, uno de los ciclados  $\{C_1, \dots, C_{i-1}\}$  para todo  $i = 1, \dots, k$ .

Esta propiedad es fundamental en el desarrollo de algunos modelos gráficos probabilísticos que se detallarán en secciones posteriores.

Continuando con la propiedad presentada y teniendo en cuenta que, en algunos casos, se pueden ordenar los ciclados, podrá aparecer una cadena de ciclados asociada a un grafo no dirigido. Esto sólo se da cuando el grafo es triangulado. En el grafo triangulado de la Figura 1.13 una cadena de ciclados puede venir dada

por los siguientes ciclados  $C_1 = \{A, B, D\}$ ,  $C_2 = \{A, C, D\}$ ,  $C_3 = \{C, D, F\}$ ,  $C_4 = \{D, F, G\}$ ,  $C_5 = \{F, G, I\}$ ,  $C_6 = \{B, E\}$  y  $C_7 = \{E, H\}$ .

Cuando el grafo con el que se trabaja es no dirigido, aparecen una serie de resultados enunciados por Berge (1973) y Golumbic (1980), que se muestran en el Teorema 1.2. Estos resultados relacionan el grafo triangularizado con un nuevo tipo de grafo denominado *grafo descomponible*, aunque antes de presentar su definición formal se ha de introducir el concepto de *descomposición de un grafo*.

**Definición 1.12** (*Descomposición de un grafo*)

Sea un grafo no dirigido  $\mathcal{G} = (V, E)$ , se dice que los subconjuntos disjuntos de nodos de  $V$  no vacíos  $(A, B, C)$  forman una *descomposición de  $\mathcal{G}$*  o *descomponen  $\mathcal{G}$*  si  $V = A \cup B \cup C$  tal que  $C$  es un subconjunto completo de  $V$  tal que cualquier camino de  $A$  a  $B$ , pasa por  $C$ .

La idea introducida al exigir que cualquier camino entre  $A$  y  $B$  pase por  $C$ , es un nuevo concepto que se definirá en posteriores secciones y hace referencia a que  $C$  separa  $A$  de  $B$ .

La definición de *grafo descomponible* viene dada de forma recursiva y presenta un tipo de grafo básico para posteriores análisis.

**Definición 1.13** (*Grafo descomponible*)

Dado un grafo no dirigido  $\mathcal{G} = (V, E)$ , se dice que  $\mathcal{G}$  es un *grafo descomponible* si:

- (i) *o es completo*
- (ii) *o se obtiene una descomposición apropiada  $(A, B, C)$  del grafo, de forma que los subgrafos  $\mathcal{G}_{AUC}$  y  $\mathcal{G}_{BUC}$  son grafos descomponibles.*

**Teorema 1.2**

Sea un grafo no dirigido  $\mathcal{G} = (V, E)$ . Las siguientes condiciones son equivalentes:

- (i)  $\mathcal{G}$  es un grafo descomponible
- (ii)  $\mathcal{G}$  es un grafo triangulado
- (iii) *Cualquier separador mínimo existente entre dos vértices es completo.*

**Demostración 1.2**

Véase Lauritzen (1996). ■

**Árbol de unión**

Para finalizar con las estructuras gráficas de interés, se presenta el árbol de unión que agrupa en cada nodo un conjunto de nodos que forman conjuntos completos maximales, reduciéndose así la estructura inicial del grafo y facilitando la realización de cálculos locales que se presentarán en posteriores capítulos.

El árbol de unión que se recoge en este Punto, es el paso final en la transformación de un grafo dirigido o no dirigido en una nueva estructura que relaciona ciclados y simplifica la estructura topológica, manteniendo ciertas propiedades del grafo original. Para introducir este concepto, se necesita antes la definición siguiente.

**Definición 1.14** (*Grafo de ciclados*)

Sea  $\mathcal{G} = (V, E)$  un grafo no dirigido y sea  $C = \{C_1, \dots, C_k\}$  un conjunto de ciclados tal que  $V = C_1 \cup \dots \cup C_k$ . Se dice que el grafo  $\mathcal{G}' = (C, E')$  es un grafo de ciclados de  $\mathcal{G}$  si las aristas contenidas en  $E'$  solo unen ciclados con algún nodo común, es decir, si  $(C_i, C_j) \in E' \Rightarrow C_i \cap C_j \neq \emptyset$ .

**Definición 1.15** (*Árbol de unión*)

Un grafo de ciclados se dice que es árbol de unión de ciclados si es un árbol y todo nodo que pertenece a dos ciclados también pertenece a todos los ciclados contenidos en el camino que los une.

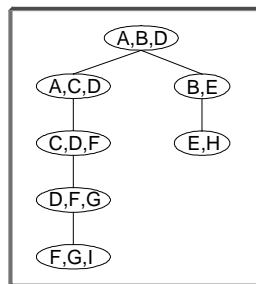


Figura 1.14. Árbol de unión del grafo moral de la Figura 1.11

En la Figura 1.14 se incluye el árbol de unión de ciclados obtenido a partir de la cadena de ciclados del grafo triangulado de la Figura 1.13.

Jensen (1988) presenta el resultado que se muestra en el siguiente teorema, asociado a la obtención de un árbol de unión a partir de un grafo no dirigido.

**Teorema 1.3**

*Un grafo no dirigido  $\mathcal{G} = (V, E)$  tiene un árbol de unión si y solo si es triangulado.*

**Demostración 1.3**

*La demostración puede verse en Jensen (1988). ■*

Existen algunos métodos fundamentales para actualizar la información que se tiene acerca de una red basados en el uso del árbol de unión de ciclados para simplificar los cálculos, tratando localmente la información entre grupos de variables. Por tanto, dada la importancia de la obtención del árbol de unión, se han desarrollado diversos algoritmos para obtener un árbol de unión de ciclados, o varios, partiendo de distintas estructuras asociadas al grafo inicial<sup>2</sup>.

En este Apartado se ha visto, cómo, a partir de un grafo no dirigido, de un grafo dirigido o de un grafo mixto se pueden obtener diversos grafos más simples en los que se agrupan nodos que conforman un mismo ciclado, pero con una estructura que mantiene parte de la topología inicial del grafo; así, se obtiene el grafo no dirigido asociado, para los grafos dirigidos (Figura 1.8) y los grafos mixtos y se determina el grafo moral  $\mathcal{G}^m$  (Figura 1.11) del mismo y de ahí el grafo triangulado (Figuras 1.12). Con los ciclados que se muestran en el grafo triangulado se obtiene una cadena de ciclados y posteriormente el árbol de unión de ciclados, como se presenta en la Figura 1.14.

De esta manera, se observa como se puede modificar la estructura de un grafo original para obtener un grafo en el que se agrupan los nodos, formándose así estructuras locales sobre las que se trabajará posteriormente. En el Capítulo 2 se definen de forma extensa los conceptos relativos a la actualización de la información en una red, que parte de un grafo modificado y simplifica los cálculos que se han de realizar, a cálculos sobre las estructuras locales.

---

<sup>2</sup>Se pueden encontrar algoritmos para la obtención del árbol de unión en Cowell, et al. (1999)

### 1.3. Grafos y distribuciones de probabilidad

Hasta aquí, se han introducido definiciones básicas asociadas a la Teoría de Grafos. En esta Sección se busca relacionar un grafo, con sus nodos y aristas, con un conjunto de variables aleatorias y su estructura de dependencia. Para ello, los nodos representan a las distintas variables y, debido a la potencia de los grafos, las aristas representan las relaciones de dependencia e independencia que se dan entre el conjunto de variables del problema.

Así, al trabajar con un grafo que representa el problema en estudio, se introducen de forma implícita determinadas propiedades. Lo que se busca en esta Sección es conocer el comportamiento de la distribución de probabilidad conjunta, a partir de dichas propiedades del grafo; con esta finalidad se analiza la posibilidad de factorizar una distribución de probabilidad, de manera que se obtenga mediante el producto de otras funciones más sencillas definidas para subconjuntos de variables. La posibilidad de factorizar la distribución de probabilidad permite simplificar los cálculos en posteriores desarrollos, así como, analizar los efectos de las interdependencias de las variables.

Estas propiedades del grafo se denominan *propiedades de Markov sobre grafos* y se enuncian teniendo en cuenta el tipo de grafo en estudio.

Conceptos como el de independencia condicionada o la factorización de una distribución y propiedades de los grafos ligadas a las relaciones de dependencia entre las variables del problema, como las propiedades de Markov fundamentales para obtener criterios de separación gráfica, son introducidos en esta Sección con el objetivo de describir mediante un grafo un conjunto de variables y sus dependencias. Además, las propiedades presentadas van a servir para justificar cálculos locales que se desarrollarán en el Capítulo 2.

#### 1.3.1. Independencia condicionada

Esta definición básica dentro de la Teoría de la Probabilidad, muestra la no relación entre dos conjuntos de variables dado un tercero.

**Definición 1.16** (*Independencia condicionada*)

*Sean  $X$ ,  $Y$  y  $Z$  tres conjuntos disjuntos de variables aleatorias. Se dice que  $X$*

$X$  e  $Y$  son independientes condicionalmente dado  $Z$ , y se denota como  $X \perp\!\!\!\perp Y|Z$ , si y solo si

$$p(x|y, z) = p(x|z).$$

De forma equivalente se puede definir la independencia condicionada entre  $X$  e  $Y$  dado  $Z$  si y solo si se cumple cualquiera de las siguientes condiciones:

- (i)  $p(x, y|z) = p(x|z)p(y|z)$  con  $p(z) > 0$ .
- (ii)  $p(x, y, z) = p(x|z)p(y|z)p(z)$  con  $p(z) > 0$ .
- (iii)  $p(x, y, z) = \frac{p(x,z)p(y,z)}{p(z)}$  con  $p(z) > 0$ .

Como caso particular se puede enunciar la *independencia entre dos conjuntos disjuntos de variables*, de forma que dos conjuntos disjuntos de variables  $X$  e  $Y$  son independientes, y se denota como  $X \perp\!\!\!\perp Y|\emptyset$ , donde  $\emptyset$  es el conjunto vacío, cuando  $p(x|y) = p(x)$ .

Cabe resaltar el resultado que muestra que dos conjuntos disjuntos de variables pueden ser independientes y dejar de serlo cuando se condiciona la ocurrencia a otro conjunto de variables, de forma que puede darse el caso de  $X \perp\!\!\!\perp Y|\emptyset$  y ser condicionalmente dependientes dado  $Z$ . A su vez, dos conjuntos de variables  $X$  e  $Y$  pueden ser condicionalmente independientes dado  $Z$ , tal que  $X \perp\!\!\!\perp Y|Z$  y cuando no se condiciona a la ocurrencia de  $Z$ , ser variables dependientes, dando lugar a ciertas paradojas muy conocidas.

### Propiedades de la Independencia condicionada

Las propiedades que se muestran en este Punto facilitan la obtención de nuevas relaciones de independencia entre conjuntos de variables, de forma que se pueda tratar localmente la información entre subconjuntos de variables.

- (i) *Conmutativa*:  $X \perp\!\!\!\perp Y|Z \Leftrightarrow Y \perp\!\!\!\perp X|Z$
- (ii) *Descomposición*:  $X \perp\!\!\!\perp (Y \cup W)|Z \Rightarrow X \perp\!\!\!\perp Y|Z$  y  $X \perp\!\!\!\perp W|Z$
- (iii) *Unión Débil*:  $X \perp\!\!\!\perp (Y \cup W)|Z \Rightarrow X \perp\!\!\!\perp Y|(Z \cup W)$  y  $X \perp\!\!\!\perp W|(Z \cup Y)$

(iv) *Contracción*:  $X \perp\!\!\!\perp W | (Z \cup Y)$  y  $X \perp\!\!\!\perp Y | Z \Rightarrow X \perp\!\!\!\perp (Y \cup W) | Z$

(v) *Intersección*:  $X \perp\!\!\!\perp W | (Z \cup Y)$  y  $X \perp\!\!\!\perp Y | (Z \cup W) \Rightarrow X \perp\!\!\!\perp (Y \cup W) | Z$

(vi) *Unión fuerte*:  $X \perp\!\!\!\perp Y | Z \Rightarrow X \perp\!\!\!\perp Y | (Z \cup W)$

### 1.3.2. Factorización de una función

La posibilidad de factorizar la distribución de probabilidad asociada al conjunto de variables  $\mathbf{X} = \{X_1, \dots, X_n\}$ , permite trabajar con subconjuntos de variables y con funciones que sólo dependen de dichos subconjuntos, lo que simplificará posteriores cálculos.

A continuación, se introduce el concepto de *distribución jerárquica* que muestra como una distribución de probabilidad puede venir dada mediante la factorización de un conjunto de funciones, siendo estas funciones los *factores potenciales*.

**Definición 1.17** (*Distribución jerárquica*)

Sean  $C_1, \dots, C_k$  una colección de subconjuntos de  $\mathbf{X} = \{X_1, \dots, X_n\}$  y sea  $\psi_1(c_1), \dots, \psi_k(c_k)$  una colección de funciones no negativas tales que  $\psi_i(c_i)$  solo depende de  $c_i$ , donde  $c_i$  es una realización de  $C_i$ .

Se dice que una distribución de probabilidad conjunta es una distribución jerárquica si su distribución conjunta se puede obtener mediante el producto de las funciones  $\psi_i(c_i)$  de forma que la distribución se factoriza tal que

$$p(x_1, \dots, x_n) = \prod_{i=1}^k \psi_i(c_i).$$

A las funciones  $\psi_i(c_i)$  se las denomina factores potenciales de la distribución de probabilidad.

En consecuencia, se dice que una distribución de probabilidad conjunta factoriza si y solo si dicha distribución es jerárquica.

Además, al trabajar conjuntamente con el grafo y con una distribución jerárquica asociada al conjunto de variables del problema, se pueden conocer las propiedades de independencia condicionada que se presentan asociadas a dicha



distribución jerárquica. Para ello, se ha de tratar el grafo como un grafo no dirigido y a la colección de subconjuntos  $C_1, \dots, C_k$  de  $\mathbf{X} = \{X_1, \dots, X_n\}$  como los ciclados que se presentan en el grafo. Esto permite trabajar y realizar cálculos localmente con las variables que conforman cada uno de los ciclados, para posteriormente obtener la distribución conjunta jerárquica de las variables del grafo como el producto de los factores potenciales definidos para cada uno de los ciclados.

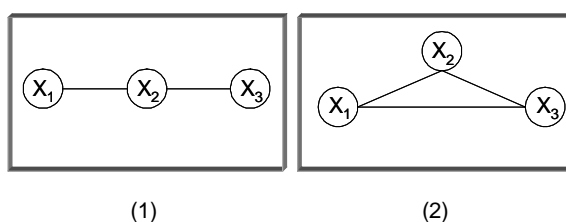


Figura 1.15. Grafos en los que a partir de la factorización de la densidad conjunta se obtienen relaciones de independencia condicionada. En (1)

$X_1 \perp\!\!\!\perp X_3 | X_2$  mientras que en (2) no se muestra ninguna relación de independencia condicionada

Por ejemplo, si se considera el grafo presentado en la Figura 1.15 (1) los ciclados del grafo son  $C_1 = \{X_1, X_2\}$  y  $C_2 = \{X_2, X_3\}$  de forma que la distribución conjunta factoriza tal que  $p(x_1, x_2, x_3) = \psi_1(x_1, x_2)\psi_2(x_2, x_3)$ , donde  $\psi_1(x_1, x_2)$  y  $\psi_2(x_2, x_3)$  son los factores potenciales asociados a la distribución jerárquica. En este caso, la factorización que se muestra para la distribución conjunta es equivalente a tener  $X_1 \perp\!\!\!\perp X_3 | X_2$ .

Si ahora se considera el grafo de la Figura 1.15 (2) con el mismo conjunto de nodos pero con los ciclados  $C_1 = \{X_1, X_2\}$ ,  $C_2 = \{X_2, X_3\}$  y  $C_3 = \{X_1, X_3\}$ , la distribución conjunta factoriza como  $p(x_1, x_2, x_3) = \psi_1(x_1, x_2)\psi_2(x_2, x_3)\psi_3(x_1, x_3)$ , pero en este caso, la factorización presentada no muestra ninguna relación de independencia condicionada.

### 1.3.3. Propiedades de Markov sobre Grafos

En este Apartado se tratan las propiedades de Markov que se definen sobre los distintos grafos con los que se puede trabajar. A partir de estas propiedades, aparecen implicaciones sobre la factorización de la distribución conjunta, de forma que como se ha visto, a través de la factorización, se obtienen diversas relaciones de independencia condicionada que aparecen entre las variables de estudio.

Para ampliar la información acerca de las Propiedades de Markov que aquí se presenta, puede verse Cowell, et al. (1999).

#### Propiedades de Markov sobre grafos no dirigidos

La primera propiedad que se presenta sólo hace referencia a la factorización de la distribución conjunta y viene dada por

**(F)** *Propiedad de Markov de la factorización:* La distribución conjunta factoriza y por tanto la distribución de probabilidad conjunta es una distribución jerárquica.

Considerándose ahora un grafo no dirigido  $\mathcal{G} = (V, E)$ , asociadas a  $\mathcal{G}$  se tienen las siguientes propiedades:

**(P)** *Propiedad de Markov por pares:* Para cualquier par de vértices no adyacentes  $V_i$  y  $V_j$ , se tiene que  $V_i \perp\!\!\!\perp_{\mathcal{G}} V_j \mid (V \setminus \{V_i, V_j\})$ .

**(L)** *Propiedad local de Markov:* Para cualquier vértice  $V_i \in V$  se tiene que  $V_i \perp\!\!\!\perp_{\mathcal{G}} (V \setminus ci(V_i)) \mid frn(V_i)$ .

**(G)** *Propiedad global de Markov:* Para cualquier terna de conjuntos disjuntos de  $V$  dada por  $(A, B, S)$ , donde  $S$  separa  $A$  de  $B$  en  $\mathcal{G}$ , se tiene que  $A \perp\!\!\!\perp_{\mathcal{G}} B \mid S$ .

En general se tiene la siguiente implicación entre las propiedades presentadas:

$$(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)$$

En los dos teoremas que se muestran a continuación, aparecen nuevas implicaciones entre las propiedades de Markov en la distribución de probabilidad conjunta y su función de densidad o de masa asociada.

**Teorema 1.4**

Si una distribución de probabilidad de  $\mathbf{X}$  cumple la propiedad (v) de la independencia condicionada, denominada propiedad de la intersección, para todo par de subconjuntos disjuntos

$$(G) \Leftrightarrow (L) \Leftrightarrow (P)$$

**Demostración 1.4**

Puede verse en Pearl, et al. (1987). ■

**Teorema 1.5**

Una distribución de probabilidad definida sobre un espacio muestral discreto, con función de densidad o de masa conjunta estrictamente positiva, satisface la propiedad (P) si y solo si la distribución es jerárquica, es decir

$$(P) \Leftrightarrow (F)$$

**Demostración 1.5**

Véase Lauritzen (1996). ■

En general, sin la suposición de trabajar con una función de densidad o de masa conjunta positiva,  $(G) \not\Rightarrow (F)$ .

A continuación, se enuncian dos proposiciones que muestran las distintas propiedades que cumple la distribución de probabilidad conjunta para cualquier terna de conjuntos disjuntos  $(A, B, S)$  de  $V$ , que formen una descomposición del grafo  $\mathcal{G} = (V, E)$ .

**Proposición 1.1**

Sea  $(A, B, S)$  una descomposición de  $\mathcal{G}$ . La distribución de probabilidad factoriza con respecto a  $\mathcal{G}$ , si y solo si, las distribuciones de probabilidad de  $A \cup S$  y de  $B \cup S$  factorizan con respecto a  $\mathcal{G}_{A \cup S}$  y  $\mathcal{G}_{B \cup S}$  respectivamente y la distribución conjunta satisface

$$p(x) = \frac{p_{A \cup S}(x_{A \cup S})p_{B \cup S}(x_{B \cup S})}{p_S(x_S)}$$

**Demostración 1.1**

Puede verse en Cowell, et al. (1999). ■

**Proposición 1.2**

Sea  $(A, B, S)$  una descomposición de  $\mathcal{G}$  y sea un espacio muestral discreto.

La distribución de probabilidad cumple la propiedad global de Markov con respecto a  $\mathcal{G}$  si y solo si las distribuciones de probabilidad de  $A \cup S$  y de  $B \cup S$  cumplen dicha propiedad con respecto a  $\mathcal{G}_{A \cup S}$  y  $\mathcal{G}_{B \cup S}$  respectivamente, y

$$p(x) = \frac{p(x_{A \cup S})p(x_{B \cup S})}{p(x_S)}$$

### **Demostración 1.2**

Véase Lauritzen (1996). ■

Tras éstos resultados, se obtiene una nueva factorización de la distribución de probabilidad conjunta, a través de los ciclados del grafo, ya que si  $\mathcal{G}$  se puede descomponer, la aplicación recursiva presentada en la Proposición 1.1, muestra que

$$p(x) = \frac{\prod_{C \in \mathcal{C}} p(x_C)}{\prod_{S \in \mathcal{S}} p(x_S)}$$

siendo  $\mathcal{C}$  el conjunto de ciclados del grafo  $\mathcal{G}$ , y  $\mathcal{S}$  el conjunto *separadores*, es decir, el conjunto de elementos que pertenecen a la intersección de dos ciclados consecutivos, tal que  $S_i = C_i \cap C_{i-1}$ .

### **Propiedades de Markov sobre grafos acíclicos dirigidos**

Para introducir la propiedad de Markov de la factorización se ha de presentar la definición de factorización recursiva de un DAG. Además, se ha de obtener el grafo moral del grafo no dirigido asociado al grafo acíclico dirigido  $\mathcal{D} = (V, E)$ . Por tanto, un paso inicial básico será obtener un grafo moral no dirigido,  $\mathcal{D}^m$ , asociado a la estructura original del DAG.

A continuación, se presenta la definición de factorización recursiva de un DAG, ya que al trabajar con DAGs es posible obtener, de forma recursiva, una distribución de probabilidad conjunta asociada a las variables del problema, descomponiendo dicha probabilidad conjunta en el producto de las distribuciones de probabilidad condicionadas de cada una de las variables dada la ocurrencia de sus padres en el DAG.

### **Definición 1.18** (Factorización recursiva según un DAG)

Se dice que una distribución de probabilidad conjunta admite una factorización

recursiva según un DAG  $\mathcal{D}$ , si la distribución de probabilidad se puede expresar como

$$p(x) = \prod_{i=1}^n p(x_i | pa(X_i))$$

siendo  $p(x_i | pa(X_i))$  la distribución de probabilidad condicionada de  $X_i$  dados sus padres en  $\mathcal{D}$ ,  $pa(X_i)$ .

Recuérdese que los padres de un nodo  $X_i$  en  $\mathcal{D}$  son los nodos que envían aristas a dicho nodo  $X_i$ , y su notación viene dada por  $pa(X_i)$ .

La definición recogida es fundamental para introducir los resultados que se muestran a continuación, dados por las propiedades de Markov sobre DAGs. Dichas propiedades pueden ser ampliadas en Cowell, et al. (1999).

### Lema 1.1

*Cuando la distribución de probabilidad de un conjunto de variables representadas mediante un DAG,  $\mathcal{D}$ , admite una factorización recursiva según  $\mathcal{D}$ , entonces dicha distribución de probabilidad se factoriza tomando el grafo moral  $\mathcal{D}^m$ , asociado a  $\mathcal{D}$ , y además cumple la propiedad global de Markov relativa a dicho grafo moral.*

### Demostración 1.1

*La factorización viene dada por la construcción del grafo moral  $\mathcal{D}^m$  asociado al DAG,  $\mathcal{D}$ , ya que los conjuntos  $V_i \cup pa(V_i)$  son conjuntos completos en  $\mathcal{D}^m$ , entonces se tiene una función  $\psi_{V_i \cup pa(V_i)}$  definida para dicho conjunto. Por tanto, se cumple la propiedad global de Markov por la relación  $(F) \Rightarrow (G) \Rightarrow (L) \Rightarrow (P)$ . ■*

Teniendo en cuenta las implicaciones entre las propiedades de Markov, la propiedad local de Markov sobre el grafo moral  $\mathcal{D}^m$ , muestra que

$$V_i \perp\!\!\!\perp_{\mathcal{D}^m} (V \setminus V_i) \mid man(V_i)$$

siendo  $man(V_i)$  el *manto de Markov sobre el grafo moral de  $\mathcal{D}$* , esto es, el conjunto de vecinos de  $V_i$  en el grafo moral asociado a  $\mathcal{D}$ . Si en lugar de considerar el grafo moral  $\mathcal{D}^m$ , se trabaja directamente con el DAG  $\mathcal{D}$ , el manto de Markov sobre  $\mathcal{D}$  vendrá dada como  $man(V_i) = pa(V_i) \cup hi(V_i) \cup es(V_i)$ , donde  $es(V_i)$  representa a los *esposos de  $V_i$* , dados por aquellos nodos  $V_j$  tales que  $hi(V_j) \cap hi(V_i) \neq \emptyset$ .

Se tiene, por tanto,

### Proposición 1.3

Cuando la distribución de probabilidad de un conjunto de variables representadas mediante un DAG,  $\mathcal{D}$ , admite una factorización recursiva según  $\mathcal{D}$ , y  $C$  es un conjunto ancestral, de forma que  $C$  contiene todos los ascendientes de los nodos que conforman dicho conjunto, entonces la distribución de probabilidad marginal de  $C$  admite una factorización recursiva de acuerdo al subgrafo asociado a  $C$ , dado por el DAG  $\mathcal{D}_C$ .

Como resultado de la proposición, se tiene el siguiente corolario, que introduce la *propiedad global de Markov en DAGs*.

### Corolario 1.1

Sea una distribución de probabilidad que admite una factorización recursiva respecto a  $\mathcal{D}$ . Entonces,

$$A \perp\!\!\!\perp_{\mathcal{D}} B | S$$

siempre que  $A$  y  $B$  estén separados por  $S$  en el grafo moral del conjunto ancestral más pequeño que contiene la unión entre  $A$ ,  $B$  y  $S$ , es decir, siempre que  $A$  y  $B$  estén separados por  $S$  en el grafo moral de  $\mathcal{D}_{an(A \cup B \cup S)}$ , tal que  $\mathcal{D}_{an(A \cup B \cup S)}^m$ .

Todos los resultados expuestos anteriormente facilitan el encontrar una relación de independencia condicionada del tipo  $A \perp\!\!\!\perp_{\mathcal{D}} B | S$  en un DAG,  $\mathcal{D}$ , a partir del grafo estricto del conjunto ancestral más pequeño que contiene a los elementos  $A$ ,  $B$  y  $S$ . Conocer éstos resultados permite encontrar una distribución de probabilidad que admite una factorización recursiva.

### Ejemplo 1.1

Sea desea estudiar si  $A \perp\!\!\!\perp_{\mathcal{D}} B | S$  con  $S = \{X, Y\}$  en el DAG  $\mathcal{D}$  que se presenta en la Figura 1.16.

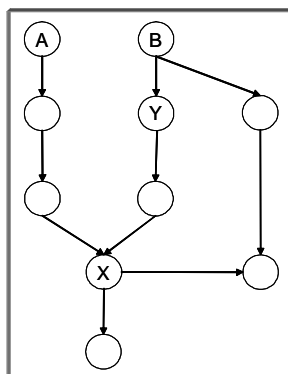


Figura 1.16. DAG  $\mathcal{D} = (V, E)$

Para conocer si existe una relación de independencia condicionada del tipo  $A \perp\!\!\!\perp_{\mathcal{D}} B | S$ , se obtiene el grafo moral del conjunto ancestral más pequeño que contiene la unión  $A \cup B \cup S$ , dado por  $\mathcal{D}_{an(A \cup B \cup S)}^m$  (véase la Figura 1.17). En dicho grafo, se observa que  $S = \{X, Y\}$  separa ambos conjuntos  $A$  y  $B$ , por lo tanto,  $A \perp\!\!\!\perp_{\mathcal{D}} B | S$ .

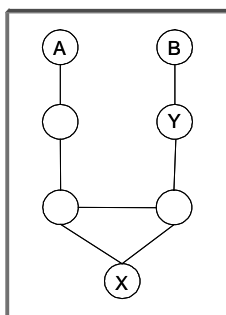


Figura 1.17. Grafo moral del conjunto ancestral más pequeño que contiene a  $A \cup B \cup S$  de  $\mathcal{D}$ , dado por  $\mathcal{D}_{an(A \cup B \cup S)}^m$

### Propiedades de Markov sobre grafos cadena

Se va a considerar ahora un grafo cadena  $\mathcal{K}$ , esto es un grafo mixto sin ciclos dirigidos parcialmente. Una distribución de probabilidad conjunta satisface las siguientes *propiedades de Markov sobre grafos cadena* en un grafo  $\mathcal{K}$ :

- (PC) *Propiedad de Markov por pares para grafos cadena:* Para cualquier par de vértices no adyacentes  $(V_i, V_j)$  con  $V_j \in nd(V_i)$ , se tiene que  $V_i \perp\!\!\!\perp_{\mathcal{K}} V_j | (nd(V_i) \setminus \{V_i, V_j\})$ .
- (LC) *Propiedad local de Markov para grafos cadena:* Para cualquier vértice  $V_i \in V$ , se tiene que  $V_i \perp\!\!\!\perp_{\mathcal{K}} (nd(V_i) \setminus frn(V_i)) | frn(V_i)$ .
- (GC) *Propiedad global de Markov para grafos cadena:* Para cualquier terna  $(A, B, S)$  de conjuntos disjuntos de  $V$ , donde  $S$  separa  $A$  de  $B$  en el grafo moral formado por el conjunto ancestral más pequeño que contiene  $A \cup B \cup S$ , dado por  $\mathcal{K}_{an(A \cup B \cup S)}^m$ , se tiene que  $A \perp\!\!\!\perp_{\mathcal{K}} B | S$ .

Las propiedades de Markov relativas a un grafo cadena  $\mathcal{K}$ , unifican las propiedades vistas para grafos no dirigidos y para grafos acíclicos dirigidos.

Cuando se interpretan las relaciones de independencia condicionada en  $\mathcal{K}$ , se utiliza una aproximación basada en la extensión de la propiedad de ordenación de Markov dada para DAGs, así, al tener un grafo  $\mathcal{K}$  mixto sin ciclos dirigidos parcialmente, el conjunto de nodos se puede particionar, tal que  $V = V(1) \cup \dots \cup V(T)$  de forma que cada conjunto  $V(t)$  o *componente cadena*, solo tenga aristas no dirigidas entre sus vértices; cuando aparezcan nodos cuyas únicas uniones sean aristas dirigidas y por tanto no estén en ninguna componente cadena, se formará una nueva componente cadena compuesta por el menor número de vértices posible, inicialmente unidos con una arista dirigida. Ésta partición se denota como *cadena dependiente*.

Sea  $C(t) = V(1) \cup \dots \cup V(t)$ . La distribución de probabilidad conjunta, se dice que satisface la *propiedad de bloques recursivos de Markov (BC)*, si para cualquier par de vértices no adyacentes  $(V_i, V_j)$ , se tiene que

$$V_i \perp\!\!\!\perp_{\mathcal{K}} V_j | (C(t^*) \setminus \{V_i, V_j\})$$

siendo  $t^*$  el menor  $t$  tal que  $\{V_i, V_j\} \subseteq C(t)$ .

### Teorema 1.6

*Si una distribución de probabilidad cumple la propiedad (v) de independencia condicionada para los subconjuntos de  $V$ , entonces*

$$(GC) \Leftrightarrow (LC) \Leftrightarrow (PC) \Leftrightarrow (BC)$$



**Demostración 1.6**

Véase Frydenberg (1990). ■

**Ejemplo 1.2**

Considérese el grafo cadena  $\mathcal{K}$  de la Figura 1.18, se desea estudiar si  $C \perp\!\!\!\perp_{\mathcal{K}} H | \{B, E\}$ .

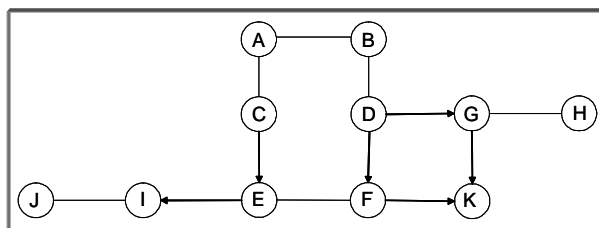


Figura 1.18. Grafo cadena  $\mathcal{K} = (V, E)$

Trabajando con la aproximación anteriormente expuesta, se observa que cada componente cadena o conjunto  $V(t)$  viene dado por  $\{A, B, C, D\}$ ,  $\{E, F\}$ ,  $\{I, J\}$ ,  $\{G, H\}$  y  $\{K\}$ .

El grafo moral formado por el conjunto ancestral más pequeño que contiene a los vértices  $\{C, H, B, E\}$ , que se muestra en la Figura 1.19, añade una arista entre  $C$  y  $D$  porque ambos tienen hijos en una componente cadena  $\{E, F\}$ . Como se puede observar en dicho grafo moral  $\mathcal{K}_{an(C \cup H \cup \{B, E\})}^m$ , no se puede concluir que  $C \perp\!\!\!\perp_{\mathcal{K}} H | \{B, E\}$ .

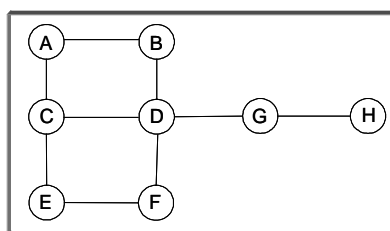


Figura 1.19. Grafo moral del conjunto ancestral más pequeño que contiene a  $C \cup H \cup \{B, E\}$  de  $\mathcal{K}$ , dado por  $\mathcal{K}_{an(C \cup H \cup \{B, E\})}^m$

Si en este ejemplo se quiere comprobar si  $C \perp\!\!\!\perp_{\mathcal{K}} H | B$ , se considera el grafo moral  $\mathcal{K}_{an(C \cup H \cup B)}^m$ , que contiene el conjunto ancestral más pequeño formado por

los vértices  $\{C, H, B\}$ , que se presenta en la Figura 1.20, y se concluye  $C$  y  $H$  son independientes condicionalmente dado  $B$ , es decir  $C \perp\!\!\!\perp_{\mathcal{K}} H|B$ .

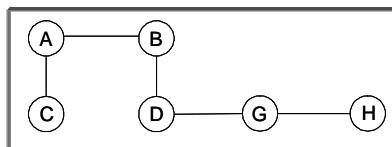


Figura 1.20. Grafo moral del conjunto ancestral más pequeño que contiene a  $C \cup H \cup B$  de  $\mathcal{K}$ , dado por  $\mathcal{K}_{an(C \cup H \cup B)}^m$

Una forma de construir una distribución de probabilidad conjunta que satisfaga las propiedades de Markov de los grafos cadena es mediante la factorización. Así, si  $V(1), \dots, V(T)$  es una cadena dependiente de  $\mathcal{K}$ , entonces cualquier distribución de probabilidad conjunta se factoriza de manera que

$$p(x) = \prod_{t=1}^T p(x_{V(t)} | x_{C(t-1)})$$

siendo  $C(t) = V(1) \cup \dots \cup V(t)$ .

Si  $B(t) = pa(V(t)) = frn(V(t))$  la factorización de la distribución de probabilidad toma la siguiente expresión

$$p(x) = \prod_{t=1}^T p(x_{V(t)} | x_{B(t)}).$$

Pero en los grafos cadena  $\mathcal{K}$ , la factorización no conduce a que se satisfagan todas las propiedades de Markov de grafos cadena. Para describir el resto de propiedades, se define  $\mathcal{K}^*(t)$  como un grafo no dirigido con conjunto de nodos  $V(t) \cup B(t)$ , de forma que  $V_i$  y  $V_j$  serán adyacentes en el nuevo grafo si  $(V_i, V_j) \in E$  o  $(V_j, V_i) \in E$  o si  $(V_i, V_j) \subseteq B(t)$ , y se añaden las aristas no dirigidas que faltan entre todos los elementos de  $B(t)$ , de manera que el conjunto  $B(t)$  sea completo en el nuevo grafo  $\mathcal{K}^*(t)$ .

Se han desarrollado más resultados asociados a las propiedades de Markov sobre grafos cadena que muestran la factorización de la distribución de probabilidad, cuando el conjunto de variables que forman el problema son discretas.

Véase Cowell, et al. (1999). Aquí se han recogido los resultados más interesantes para la introducción a los modelos gráficos probabilísticos formados con grafos cadena.

### 1.3.4. Separación gráfica

Para las propiedades de Markov sobre grafos, se van a recoger algunos criterios que surgen como formulación alternativa a las propiedades de Markov y que facilitan la obtención de relaciones de independencia condicionada entre las variables del problema a partir de la disposición de las mismas en el grafo. Esto permite describir relaciones de independencia condicionada en el grafo, al igual que la posibilidad de tratar localmente los cálculos que se han de realizar con la red de interés, como se verá en posteriores capítulos.

#### Separación en grafos no dirigidos

Cuando se representa un conjunto de variables mediante un grafo no dirigido  $\mathcal{G} = (V, E)$ , las aristas no dirigidas del grafo muestran las relaciones de dependencia entre las variables, así si dos variables  $X$  e  $Y$  son dependientes entonces ha de existir un camino entre los nodos  $X$  e  $Y$  que representan a dichas variables.

Si además existe dependencia indirecta entre dos variables  $X$  e  $Y$ , de forma que son dependientes dada una tercera  $Z$ , gráficamente  $Z$  no ha de cortar el camino existente entre  $X$  e  $Y$ . Esta idea conduce de forma directa al criterio de separación que se define en grafos no dirigidos, así si existe un nodo  $Z$  (o un conjunto de nodos) que corta todos los posibles caminos entre dos variables  $X$  e  $Y$  (o dos conjuntos de variables), se dice que  $X$  e  $Y$  están *separadas* dado  $Z$ .

Por tanto, las relaciones de independencia condicionada que se tienen entre un conjunto de variables, se pueden representar gráficamente teniendo en cuenta la siguiente definición.

**Definición 1.19** (*Separación*)

Sean  $X$ ,  $Y$  y  $Z$  tres conjuntos disjuntos de nodos de un grafo no dirigido  $\mathcal{G} = (V, E)$ . Se dice que  $Z$  separa  $X$  e  $Y$  en  $\mathcal{G}$ , y se escribe  $X \perp\!\!\!\perp_{\mathcal{G}} Y | Z$ , si y solo si cada camino entre  $X$  e  $Y$  contiene algún nodo de  $Z$ .

Esta definición auna las propiedades de Markov para grafos no dirigidos, mostrando las relaciones de independencia condicionada que hay entre un conjunto de variables representadas mediante un grafo no dirigido  $\mathcal{G}$ .

En el grafo de la Figura 1.21 se observa como  $X \perp\!\!\!\perp_{\mathcal{G}} W | \{Y, Z\}$ .

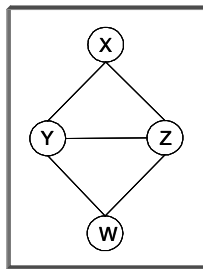


Figura 1.21. Grafo con relaciones de independencia condicionada

### Separación en grafos dirigidos acíclicos

Antes de introducir el criterio de separación gráfica en grafos acíclicos dirigidos es fundamental analizar los tipos de relaciones o conexiones que aparecen entre los nodos de un DAG  $\mathcal{D}$ .

1. *Conexión en serie*: Es cuando un nodo es padre de otro que a su vez es padre de un tercero. En el grafo dirigido (1) de la Figura 1.22 se puede ver dicha conexión.
2. *Conexión divergente*: Es cuando un nodo es padre de un conjunto de nodos no conectados entre sí. Véase el grafo dirigido (2) de la Figura 1.22.
3. *Conexión convergente*: Es cuando un conjunto de nodos no conectados entre sí son padres de un nodo concreto. En el grafo dirigido (3) de la Figura 1.22 se presenta una conexión convergente entre un conjunto de nodos.

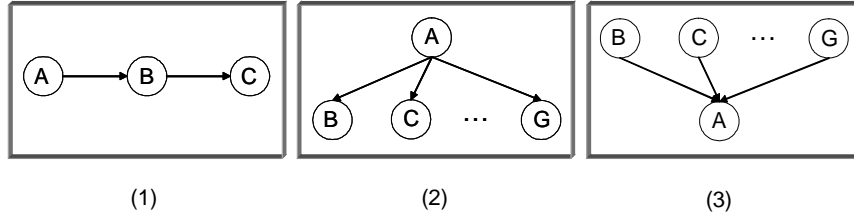


Figura 1.22. Los grafos dirigidos (1), (2) y (3) muestran conexiones en serie, divergente y convergente, respectivamente

Pearl (1986a) introduce la definición de separación gráfica en DAGs, llamada *d-separación*, como alternativa a las propiedades de Markov en grafos acíclicos dirigidos. Este concepto se trata formalmente en Verma, et al. (1990).

**Definición 1.20** (*d-separación*)

Sean  $X$ ,  $Y$  y  $Z$  tres conjuntos disjuntos de nodos de un grafo acíclico dirigido (DAG),  $\mathcal{D} = (V, E)$ . Se dice que  $Z$  *d-separa*  $X$  e  $Y$ , y se denota como  $X \perp\!\!\!\perp_{\mathcal{D}} Y | Z$ , si y solo si, para cualquier camino no dirigido entre un nodo de  $X$  y un nodo de  $Y$  existe un nodo intermedio  $V$  tal que

- (i) Existe una conexión convergente, siendo  $V$  el nodo al que convergen las aristas, y ni  $V$  ni los descendientes de  $V$  están en  $Z$ .
- (ii) La conexión es en serie y  $V$  es un nodo intermedio o la conexión es divergente siendo  $V$  el padre, y  $V$  está en  $Z$ .

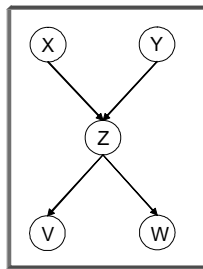


Figura 1.23. Grafo dirigido  $\mathcal{D}$  con relaciones de independencia condicionada

En la Figura 1.23 se muestra un grafo acíclico dirigido DAG  $\mathcal{D} = (V, E)$ , donde si no se consideran agrupaciones entre nodos se pueden ver algunas relaciones de independencia condicionada entre las variables. Así, se tiene que:

1.  $X \perp\!\!\!\perp_{\mathcal{D}} Y | \emptyset$  ya que  $X$  e  $Y$  aparecen en una conexión convergente y el único camino entre  $X$  e  $Y$  es  $X-Z-Y$  y ni  $Z$  ni sus descendientes están en  $\emptyset$ .
2.  $V \perp\!\!\!\perp_{\mathcal{D}} W | Z$ , aparecen en una conexión divergente donde el único camino entre  $V$  y  $W$  es  $V-Z-W$ , siendo  $Z$  el padre de  $V$  y  $W$  en la conexión del grafo.
3.  $(X \cup Y) \perp\!\!\!\perp_{\mathcal{D}} (V \cup W) | Z$ , teniendo en cuenta la propiedad de la independencia condicionada de la descomposición se tiene que  $X \perp\!\!\!\perp_{\mathcal{D}} V | Z$  y  $X \perp\!\!\!\perp_{\mathcal{D}} W | Z$  y  $Y \perp\!\!\!\perp_{\mathcal{D}} V | Z$  y  $Y \perp\!\!\!\perp_{\mathcal{D}} W | Z$ . En los cuatro casos presentados la conexión es en serie y en todos los casos,  $Z$  es el único nodo que aparece en el camino entre  $X$  y  $V$ ,  $X$  y  $W$ ,  $Y$  y  $V$ , y  $Y$  y  $W$ , siendo nodo intermedio para cualquiera de los caminos mostrados.

A continuación se introduce otra definición de separación en grafos dirigidos, presentada por Lauritzen, et al. (1990), que es la definición más utilizada de  $d$ -separación por ser más sencillo su manejo.

**Definición 1.21** ( *$d$ -separación*)

Sean  $X$ ,  $Y$  y  $Z$  tres conjuntos disjuntos en un DAG,  $\mathcal{D} = (V, E)$ . Se dice que  $Z$   $d$ -separa  $X$  e  $Y$ , y se denota como  $X \perp\!\!\!\perp_{\mathcal{D}} Y | Z$ , si y solo si  $Z$  separa  $X$  e  $Y$  en el grafo moral del menor subconjunto ancestral que contenga a los nodos de  $X$ ,  $Y$  y  $Z$ , dado por  $\mathcal{D}_{an(X \cup Y \cup Z)}^m$ .

Se puede observar como al obtener el grafo moral  $\mathcal{D}_{an(X \cup Y \cup Z)}^m$  del DAG  $\mathcal{D}$  original, todas las conexiones convergentes pasan a tener una arista entre los padres, de forma que habrá un camino no dirigido entre los padres que no será cortado por el hijo de la conexión. En el resto de conexiones, bien en serie, bien divergente,  $Z$  ha de separar  $X$  e  $Y$ , independientemente de que se considere el grafo moral del DAG  $\mathcal{D}$  inicial. Por tanto, esta segunda definición refleja las condiciones de la definición de  $d$ -separación introducida por Pearl (1986a).

### Separación en grafos cadena

Para los grafos cadena, el criterio de separación gráfica que refleja las propiedades de Markov en  $\mathcal{K}$  ha sido introducido por Studený, et al. (1998), definiendo el concepto de  $c$ -separación. Dicho concepto es una extensión de la  $d$ -separación para

DAGs y es equivalente a la propiedad de separación global de Markov presentada para grafos cadena en el tercer Punto del Apartado 1.3.3.

Al igual que en los DAGs, para enunciar el criterio de  $c$ -separación es fundamental introducir algunos conceptos especiales de grafos, que generan clasificaciones análogas a las conexiones entre los nodos.

Así, sea un grafo cadena  $\mathcal{K}$ , se define un *deslizamiento* del nodo  $X$  al nodo  $Y$  como un camino  $V_1, \dots, V_k$  con  $k \geq 2$  tal que  $X$  es la *cima del deslizamiento* siendo  $X = V_1 \longrightarrow V_2$ , además, para todo  $i = 2, \dots, k - 1$  la conexión entre los nodos es mediante una arista no dirigida, tal que  $V_i - V_j$ , y finalmente  $Y = V_k$ .

Sea un grafo cadena  $\mathcal{K}$ , se define un *corte de ruta*  $\sigma$ , dado por  $V_1, \dots, V_k$  con  $k \geq 1$ , como cualquier subcamino maximal de aristas no dirigidas tal que  $V_i - \dots - V_j$  con  $1 \leq i \leq j \leq k$ . Cualquier ruta o camino se descompone en partes. Así, los nodos  $V_i$  y  $V_j$  se denominan *terminales* del corte de ruta  $\sigma$  y se llama *cabeza del terminal* al nodo  $V_i$  (respecto a  $V_j$ ) si  $i > 1$  y  $V_{i-1} \longrightarrow V_i$  en  $\mathcal{K}$  (respectivamente  $j < k$  y  $V_j \longleftarrow V_{j+1}$  en  $\mathcal{K}$ ), y en caso contrario, es decir para  $V_i$  tal que o  $i = 1$  o  $V_{i-1} \longleftarrow V_i$  y para  $V_j$  tal que o  $j = k$  o  $V_j \longrightarrow V_{j+1}$ , se denomina *cola del terminal*.

Se dice que un corte de ruta  $\sigma$  es:

- *de cabeza a cabeza* si tiene dos cabezas del terminal.
- *de cabeza a cola* si tiene una cabeza del terminal y una cola del terminal.
- *de cola a cola* si tiene dos colas del terminal.

Nótese que si el corte de ruta tiene un solo nodo, tal que  $i = j$ , el nodo se considera dos veces como terminal, porque dos aristas del camino deben entrar en el nodo.

Para terminar con las definiciones previas antes de exponer el concepto de  $c$ -separación, se ha de introducir la idea de *bloqueo de un corte de ruta* como un conjunto de nodos  $Z$ , de forma que el corte está bloqueado por  $Z$  si se da una cualquiera de las siguientes condiciones:

1. El corte de ruta  $\sigma$  es de cabeza a cabeza respecto a la intersección entre el conjunto de nodos descendientes del corte y  $Z$  es vacía, es decir,  $de(\sigma) \cap Z = \emptyset$ .

2. El corte de ruta  $\sigma$  no es de cabeza a cabeza con respecto a un nodo de la ruta o camino, si está en  $Z$  y para al menos una de sus colas del terminal  $X$ , todos los deslizamientos en  $\mathcal{K}$  hacia  $X$  tienen un nodo en  $Z$ .

A continuación, se presenta el criterio y la definición de  $c$ -separación.

**Definición 1.22** (*Criterio de  $c$ -separación*)

Sea un grafo cadena  $\mathcal{K} = (V, E)$ . Se dice que un camino o ruta del grafo está  $c$ -separado por  $Z$  si al menos uno de sus cortes de ruta está bloqueado por  $Z$ .

**Definición 1.23** ( *$c$ -separación*)

Sean  $X$ ,  $Y$  y  $Z$  tres subconjuntos distintos de  $\mathcal{K} = (V, E)$ , se tiene que  $Z$   $c$ -separa  $X$  e  $Y$ , y se denota como  $X \perp\!\!\!\perp_{\mathcal{K}} Y | Z$ , si cualquier camino o ruta del grafo de  $X$  a  $Y$  cumple el criterio de  $c$ -separación dado  $Z$ .

A continuación, se introduce un ejemplo para determinar si se tiene independencia condicionada en un grafo cadena.

**Ejemplo 1.3**

Considérese el grafo cadena de la Figura 1.24. Se desea conocer si  $A$  es condicionalmente independiente de  $F$  dado  $\{C, E, G\}$ .

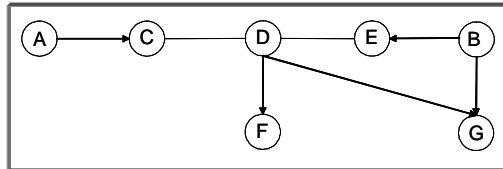


Figura 1.24. Grafo cadena  $\mathcal{K} = (V, E)$

Con el objetivo de determinar si  $A \perp\!\!\!\perp_{\mathcal{K}} F | \{C, E, G\}$ , se buscan las rutas de  $A$  a  $F$  y se estudia si se cumple el criterio de  $c$ -separación.

El camino  $A \rightarrow C \rightarrow D \rightarrow F$  está  $c$ -separado por  $\{C, E, G\}$ , porque el corte de cabeza a cola  $C \rightarrow D$  está bloqueado por  $\{C, E, G\}$ , ya que el nodo  $C$  está en  $\{C, E, G\}$  y desde la cola del terminal todos los deslizamientos hacia  $A$  tienen un nodo en  $\{C, E, G\}$ .

También se tiene un camino de  $A$  a  $F$  dado por  $A \rightarrow C \rightarrow D \rightarrow E \leftarrow B \rightarrow G \leftarrow D \rightarrow F$ , en este caso, el camino o ruta no está  $c$ -separado por  $\{C, E, G\}$ ,



porque si se consideran los cortes de cabeza a cabeza  $C-D-E$ , la intersección entre el conjunto descendiente del corte y  $\{C, E, G\}$  es  $G$ , y por tanto, no es vacía. Además, el resto de cortes no contienen los nodos  $\{C, E, G\}$ .

De esta forma, se puede concluir que  $A$  y  $F$  no están  $c$ -separados dado  $\{C, E, G\}$  y por tanto, no existe una relación de independencia condicionada entre los mismos.

Finalmente, se debe destacar que se han desarrollado diversos algoritmos que sirven para obtener las distintas relaciones de separación gráfica, bien en grafos no dirigidos como dirigidos. Alguno de estos algoritmos pueden ser encontrados en Castillo, et al. (1997a).

## 1.4. Modelos gráficos probabilísticos

Los *modelos gráficos probabilísticos* surgen como resultado de la unión entre la Teoría de Grafos y la Teoría de la Probabilidad, ya que cuando se construye un modelo matemático probabilístico, es fundamental tener en cuenta dos componentes importantes relativas a la información de la que se dispone: la información cualitativa y la información cuantitativa del problema.

La *información cualitativa* del problema introduce información asociada a las relaciones de dependencia entre las variables del modelo. Apoyándose en la Teoría de Grafos, esta información se puede resumir mediante un grafo, en el que los nodos representan a las variables del problema y las aristas del grafo las relaciones de dependencia y causalidad entre las mismas, de forma que la falta de aristas induce a relaciones de independencia.

Además de la información cualitativa, al construir un modelo matemático probabilístico se dispone de información relativa a la distribución de probabilidad de las variables del problema, también denominada *información cuantitativa* del problema. La Teoría de la Probabilidad es fundamental para obtener relaciones entre las distribuciones de probabilidad de las variables del problema; dichas distribuciones, pueden ser estimadas a partir de un conjunto de datos o mediante la información que los expertos tienen acerca del problema en estudio.

Por tanto, con la información cualitativa y cuantitativa del problema, se define el modelo gráfico probabilístico asociado al mismo, como el nexo entre la Teoría de Grafos y la Teoría de la Probabilidad, dado por un par  $(G, P)$  donde  $G$  es el grafo que representa la información cualitativa del problema, siendo los nodos las variables del modelo y las aristas las relaciones de dependencia entre dichas variables; y  $P$  es el conjunto de distribuciones, que pueden ser distribuciones condicionadas mediante las cuales se obtiene la distribución de probabilidad conjunta del problema.

Los tipos de variables que se utilizan son discretas o absolutamente continuas, así es que nos referiremos a distribuciones de probabilidad para recoger estos dos casos y referirnos a función de masa en el caso discreto y función de densidad en el caso continuo.

Muchos de los problemas clásicos de análisis multivariante estudiados en campos como la Estadística, la Ingeniería o la Teoría de la Información, en los que intervienen un gran número de variables con diversas relaciones de dependencia entre las mismas, se pueden representar mediante modelos gráficos probabilísticos, favoreciéndose una visión más intuitiva y sencilla de los mismos. Además, como se ha visto en la Sección 1.3, los grafos permiten describir y estudiar las relaciones de dependencia o independencia condicionada existentes entre las variables que componen el problema y teniendo en cuenta la Teoría de la Probabilidad, la obtención de la distribución conjunta del problema se simplifica debido a la factorización de la misma.

Así, se han utilizado modelos gráficos probabilísticos en áreas de Matemática Aplicada y Estadística (Whittaker, 1990) y en análisis de datos en general (Gilks, et al., 1993), en sistemas dinámicos y series temporales (Kjæruff, 1992; Dagum, et al., 1995), en técnicas de diagnóstico (Jensen, 2001) y en diversas aplicaciones de Ingeniería como sistemas expertos (Cowell, et al., 1999) y planificación y control (Dean, et al., 1991; Chan, et al., 1992). Los modelos gráficos probabilísticos, también juegan un importante papel en el diseño y análisis de sistemas de aprendizaje y de actualización de la información.

Investigadores de diversas áreas como Ciencias de la Computación, Ingenierías, Ciencias Sociales o Estadística entre otras, trabajan en el desarrollo y aplicación de los modelos gráficos probabilísticos.

En los últimos años, se han presentado operaciones gráficas para manipular la estructura del grafo manteniéndose la información relevante entre las variables que lo conforman. De esta manera, se puede obtener el árbol de unión de un grafo no dirigido  $\mathcal{G}$ , o de un DAG,  $\mathcal{D}$ , o de un grafo cadena,  $\mathcal{K}$ . Además, se han desarrollado múltiples algoritmos que facilitan el tratamiento del mismo y la actualización de la información cuantitativa del problema, incluyendo toda la nueva evidencia o información de que se disponga asociada a una variable o a un conjunto de variables del problema.

Dependiendo del tipo de grafo que se utilice al definir el problema, es decir, dependiendo del tipo de arista que se emplea en la representación gráfica del mismo, se han de distinguir tres tipos de modelos gráficos probabilísticos:

- Modelos gráficos probabilísticos no dirigidos. Redes de Markov.
- Modelos gráficos probabilísticos dirigidos. Redes Bayesianas.
- Modelos gráficos probabilísticos dirigidos y no dirigidos, o mixtos. Redes Cadena.

#### **1.4.1. Modelos gráficos probabilísticos no dirigidos. Redes de Markov**

Se trabaja con modelos gráficos probabilísticos no dirigidos, en los que el grafo que representa la información cualitativa del problema es un grafo no dirigido  $\mathcal{G}$ , cuando las relaciones de dependencia entre las variables de problema son relaciones de asociación o correlación, sin determinarse ninguna variable como causa o como efecto, de manera que la información de la que se dispone indica que un conjunto de variables presentan distintos niveles de asociación o correlación.

Una vez que se construye el grafo no dirigido  $\mathcal{G}$ , que representa el conjunto de variables y sus relaciones de asociación, se busca la distribución de probabilidad conjunta asociada a las variables del problema, como una factorización de funciones. Para ello, se introducen algunas ideas que pueden ser ampliadas en Pearl (1988) y Lauritzen, et al. (1988).

Con el objetivo de obtener la distribución de probabilidad conjunta como una distribución jerárquica, se triangula el grafo no dirigido original  $\mathcal{G}$ . Del grafo

triangulado se obtiene un conjunto de ciclados ordenados  $\{C_1, \dots, C_k\}$  que han de cumplir la propiedad de la intersección dinámica. Sean los separadores tales que  $S_i = C_i \cap (C_1 \cup \dots \cup C_{i-1})$  con  $i = 2, \dots, k$ , por la citada propiedad,  $S_i \subset C_i$  y en consecuencia se definen los conjuntos *residuales* como los elementos que están en  $C_i$  pero no están en  $S_i$ , tales que  $R_i = C_i \setminus S_i$  con  $i = 2, \dots, k$ . Como el conjunto residual  $R_i$  contiene todos los elementos de  $C_i$  que no están en  $C_1 \cup \dots \cup C_{i-1}$ , la distribución de probabilidad conjunta se puede factorizar mediante

$$p(x_1, \dots, x_n) = \prod_{i=1}^k p(r_i | r_1, \dots, r_{i-1}) = \prod_{i=1}^k p(r_i | s_i).$$

De esta forma, se obtiene la distribución de probabilidad conjunta como una distribución jerárquica mediante la factorización de distribuciones de probabilidad condicionada. Estas ideas se resumen en el siguiente teorema.

**Teorema 1.7**

*Si la distribución de probabilidad conjunta se puede descomponer según el grafo no dirigido  $\mathcal{G}$ , entonces dicha distribución se puede obtener como el producto de las distribuciones de probabilidad condicionada de los residuos obtenidos a partir de los ciclados del grafo  $\mathcal{G}$  y de sus separadores.*

**Demostración 1.7**

*Véase Lauritzen, et al. (1988). ■*

Una clase importante de modelos gráficos no dirigidos son las *Redes de Markov*, también denominadas Markov Random Fields, y se definen a continuación.

**Definición 1.24** (*Red de Markov*)

*Una Red de Markov es un par  $(\mathcal{G}, \Psi)$ , donde  $\mathcal{G}$  es un grafo no dirigido y  $\Psi = \{\psi_1(c_1), \dots, \psi_k(c_k)\}$  es un conjunto de  $k$  factores potenciales definidos en los ciclados  $C_1, \dots, C_k$  de  $\mathcal{G}$ . Así, el conjunto  $\Psi$  define una distribución de probabilidad conjunta como*

$$p(x) = \prod_{i=1}^k \psi_i(c_i).$$

*Si el grafo no dirigido  $\mathcal{G}$  es triangulado, la distribución de probabilidad conjunta también puede ser factorizada utilizando las distribuciones de probabilidad condicionada obtenidas a partir de los residuos y los separadores de los ciclados  $\{p(r_1|s_1), \dots, p(r_k|s_k)\}$ . En este caso, la Red de Markov viene dada por el par*

$(\mathcal{G}, P)$ , con  $P = \{p(r_1|s_1), \dots, p(r_k|s_k)\}$ , y la distribución de probabilidad conjunta se obtiene mediante el producto de todos los elementos de  $P$ , es decir

$$p(x) = \prod_{i=1}^k p(r_i|s_i).$$

Como se puede observar, la factorización de la distribución de probabilidad conjunta está basada en la triangulación del grafo  $\mathcal{G}$ , obteniéndose los ciclados que cumplen la propiedad de la intersección dinámica. Así, con los separadores y los residuos, los factores potenciales se pueden definir como distribuciones de probabilidad condicionada cuyo producto permite construir la distribución de probabilidad conjunta. A pesar de esto, no existe una relación directa entre estos factores potenciales y las distribuciones de probabilidad de las variables que conforman el problema.

Los Redes de Markov se utilizan básicamente en campos como la Física, la Robótica (Anguelov, et al., 2005), para el análisis de imágenes (Besag, 1974) y actualmente en el análisis de textos.

#### 1.4.2. Modelos gráficos probabilísticos dirigidos. Redes Bayesianas

Cuando las relaciones de dependencia entre las distintas variables del problema son de tipo causal, es decir, cuando se sabe que el efecto de una variable  $X_j$  es producido por otra variable  $X_i$ , se utiliza un grafo dirigido para representar la información cualitativa del problema, donde las aristas dirigidas indican la existencia de una relación causa-efecto entre las variables en estudio.

Teniendo en cuenta que los criterios de separación gráfica se enuncian para grafos acíclicos dirigidos (DAGs) y que la existencia de ciclos en el grafo dirigido dificulta la obtención de una distribución de probabilidad conjunta jerárquica consistente, se trabaja con DAGs para la representación de la información cualitativa del modelo gráfico probabilístico dirigido.

Además, como se observa en el teorema que se introduce a continuación, cuando la distribución de probabilidad conjunta se obtiene mediante una factorización recursiva, las relaciones de independencia condicionada en el grafo, obtenidas mediante el criterio de  $d$ -separación, son verificadas por las relaciones

de independencia condicionada de la distribución de probabilidad en estudio. Por tanto, dado un DAG,  $\mathcal{D}$ , se puede obtener la distribución de probabilidad conjunta mediante la factorización recursiva según  $\mathcal{D}$ , como el producto de las distribuciones condicionadas  $p(x_i|pa(X_i))$ , como se vió en la definición de factorización recursiva de un DAG. Este resultado, obtenido por Pearl (1988), se enuncia en el siguiente teorema.

**Teorema 1.8**

*Sea  $\mathcal{D}$  un DAG y sea  $p(x)$  una distribución de probabilidad conjunta de  $\mathbf{X}$ .*

*Entonces, las siguientes condiciones son equivalentes:*

- (i)  *$p(x)$  admite una factorización recursiva según  $\mathcal{D}$ .*
- (ii)  *$\mathcal{D}$  verifica todas las relaciones de dependencia de  $p(x)$ , de forma que la independencia condicionada gráfica coincide con las relaciones de independencia condicionada de  $p(x)$ .*

**Demostración 1.8**

*Puede verse en Pearl (1988). ■*

Una clase importante de modelos gráficos probabilísticos dirigidos es la constituida por las *Redes Bayesianas*.

Las Redes Bayesianas tienen una factorización recursiva según el DAG,  $\mathcal{D}$ , ya que por el teorema anteriormente expuesto, cualquier relación de independencia que se observe en el grafo  $\mathcal{D}$  utilizando el criterio de  $d$ -separación, también estará en el modelo probabilístico correspondiente, obteniéndose la distribución de probabilidad conjunta como el producto de las distribuciones de probabilidad condicionadas de cada uno de los nodos de  $\mathcal{D}$  dada la ocurrencia de sus padres en dicho grafo  $\mathcal{D}$ .

**Definición 1.25** (*Red Bayesiana*)

*Una Red Bayesiana es un par  $(\mathcal{D}, P)$ , donde  $\mathcal{D}$  es un DAG tal que los nodos representan las variables del problema  $\mathbf{X} = \{X_1, \dots, X_n\}$  y los arcos representan las dependencias probabilísticas, y  $P = \{p(x_1|pa(X_1)), \dots, p(x_n|pa(X_n))\}$  es un conjunto de  $n$  distribuciones de probabilidad condicionada, una para cada variable, siendo  $pa(X_i)$  el conjunto de padres del nodo  $X_i$  en el grafo  $\mathcal{D}$ .*

*Además, la distribución de probabilidad conjunta del problema se obtiene me-*

diante el producto de los elementos de  $P$ , tal que

$$p(x) = \prod_{i=1}^n p(x_i | pa(X_i)). \quad (1.1)$$

Si se comparan las Redes de Markov y las Redes Bayesianas, se observa que el proceso de construcción del modelo es más sencillo e intuitivo en el caso dirigido, ya que la factorización asociada a la distribución de probabilidad conjunta del problema requiere menos pasos iniciales. Así, cuando se trabaja con una Red Bayesiana, el DAG  $\mathcal{D}$  que representa el conjunto de variables y sus relaciones de dependencia, muestra gráficamente las relaciones de dependencia e independencia condicionada que se presentan en el propio modelo probabilístico.

Además, en las Redes Bayesianas la descomposición de la distribución de probabilidad conjunta se obtiene de forma directa tras definir las distribuciones condicionadas de los nodos, dados sus padres en  $\mathcal{D}$ .

En los últimos años ha aumentado la popularidad de las Redes Bayesianas y su campo de aplicación es cada vez más amplio. Fundamentalmente se trabaja con Redes Bayesianas en la construcción de sistemas expertos, utilizados en Inteligencia Artificial y en Estadística. Además, se aplican en diversos campos como por ejemplo, el diagnóstico médico, la bioinformática y la hidroeinformática, en sistemas dinámicos o en planificación y control.

### 1.4.3. Modelos gráficos probabilísticos mixtos. Redes Cadena

Puede darse el caso de que el problema de interés presente relaciones de dependencia de tipo causal y relaciones de asociación o correlación, entre las variables que lo conforman.

Cuando la información cualitativa del modelo indica la existencia de relaciones de dependencia causales y de asociación, el grafo adjunto al modelo gráfico probabilístico ha de ser un grafo mixto, con aristas dirigidas para representar las relaciones causales y aristas no dirigidas para mostrar las relaciones de asociación. En este caso, el grafo asociado al modelo probabilístico, es un grafo cadena  $\mathcal{K}$ .

Como se observa en la Definición 1.7, un grafo cadena  $\mathcal{K}$ , es un grafo mixto

sin ciclos dirigidos, en el que, por tanto, pueden existir ciclos formados sólo por aristas no dirigidas.

Para obtener con este tipo de grafos una distribución de probabilidad conjunta asociada a las variables del problema, y definir mediante el grafo las relaciones de independencia condicionada entre las variables en estudio, se trabaja con la definición de  $c$ -separación que combina conceptos de los modelos gráficos probabilísticos no dirigidos y dirigidos, que se han presentado previamente.

La obtención de la distribución de probabilidad conjunta mediante la factorización de un conjunto de distribuciones, depende del tipo de variables del problema (variables discretas, continuas: Gaussianas o ambas) y de como se agrupan en función de las aristas no dirigidas del grafo  $\mathcal{K}$ . Por tanto, aunque, como en las Redes de Markov, se obtienen factores potenciales para calcular la distribución de probabilidad conjunta mediante una factorización, su tratamiento depende de más características asociadas al problema en estudio.

Para ampliar la información presentada acerca de los modelos gráficos probabilísticos mixtos, puede verse Lauritzen, et al. (1984) que generalizan este tipo de grafos cadena partiendo de los grafos no dirigidos y dirigidos.

Una clase fundamental de modelos gráficos probabilísticos mixtos viene dada por las *Redes Cadena*.

Buntine (1994), basándose en la definición de componentes cadena presentada por Frydenberg (1990), define las Redes Cadena como una cadena de grafos no dirigidos, conectados mediante grafos dirigidos. La interpretación de dicha Red Cadena, se muestra como la de una Red Bayesiana definida sobre unas componentes cadena en lugar de sobre las variables originales del problema. Además, presenta la factorización de la probabilidad conjunta como un nexo entre el criterio de  $c$ -separación y las factorizaciones introducidas en Redes de Markov y Redes Bayesianas.

Para profundizar en la definición del modelo y en conceptos de independencia en el grafo cadena  $\mathcal{K}$ , puede verse Lauritzen, et al. (1989a) y (1989b) y Frydenberg (1990).

A continuación se muestra un ejemplo clásico de Redes Cadena, presentado por Cowell, et al. (1999) en el que todas las variables del problema son



categorías.

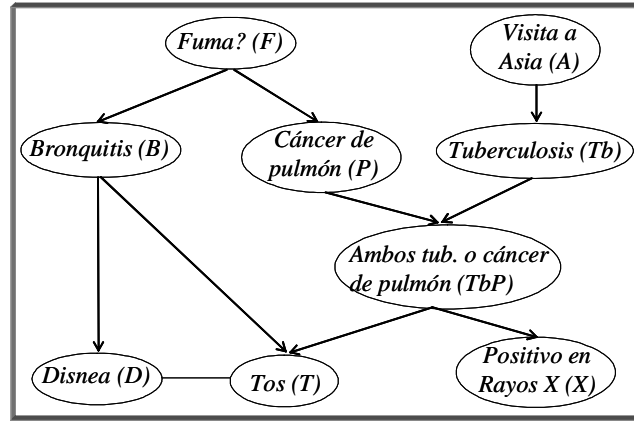


Figura 1.25. Red Cadena Ch-Asia

En este ejemplo solo hay una componente cadena formada por las variables  $D$  y  $T$ . Trabajando con las ideas anteriormente expuestas, la distribución de probabilidad conjunta se puede obtener como una factorización de distribuciones de probabilidad condicionadas, apoyándose en las Redes Bayesianas, donde las variables  $D$  y  $T$  aparecen como una misma componente. Dicha factorización es tal que,

$$\begin{aligned}
 P(F, A, B, P, Tb, TbP, D, T, X) &= \\
 &= P(F)P(A)P(B|F)P(P|F)P(Tb|A)P(TbP|P, Tb)P(D, T|B, TbP)P(X|TbP)
 \end{aligned}$$

Cowell, et al. (1999) desarrollan diversos algoritmos, para la obtención de la probabilidad conjunta, así como para el tratamiento de estos modelos, en función de la definición de las variables que intervienen en el problema.

Las posibles aplicaciones de las redes mixtas son muy amplias, aunque la dificultad de manejo y desarrollo de las mismas hacen laboriosa su definición inicial. Las Redes Cadena se utilizan principalmente en la descripción de sistemas expertos asociados al campo de la Inteligencia Artificial (Stephenson, et al., 2002).

## 2

# Redes Bayesianas

### 2.1. Introducción

En este Capítulo se estudian las Redes Bayesianas, y más concretamente las Redes Bayesianas Gaussianas, sus propiedades y los mecanismos de inferencia más característicos en ellas.

En la Sección 2.2, se introducen las Redes Bayesianas y los distintos tipos de redes que se obtienen en función de las variables del modelo. También, se presentan algunas aplicaciones de las Redes Bayesianas.

En la siguiente Sección, se presenta el mecanismo de inferencia de las Redes Bayesianas. Este proceso consiste en la actualización de la probabilidad de la red cuando se tiene información acerca del estado de alguna de las variables del problema, y se denomina propagación de la evidencia.

La red objeto de estudio, para la que se introduce en el Capítulo 3 el análisis de sensibilidad propuesto, es una Red Bayesiana Gaussiana. Por tanto, es necesario conocer bien su estructura antes de enunciar algunos conceptos fundamentales asociados a la sensibilidad de la red. Con este finalidad, en la Sección 2.4 se profundiza en las Redes Bayesianas Gaussianas y en sus características y se presenta una metodología sencilla para la realización de la propagación de la evidencia en este tipo de redes con estructura Gaussiana.

## 2.2. Redes Bayesianas

Las Redes Bayesianas son una clase de modelos gráficos probabilísticos dirigidos intrínsecamente, asociados a relaciones de dependencia de tipo causal.

Jensen (2001) introduce un ejemplo para distinguir las relaciones de tipo causal de las relaciones de asociación. Así, considerando las variables categóricas, fumar ( $F$ ) y cáncer de pulmón ( $C$ ), existe una relación causal entre las mismas cuando se hace la suposición de que fumar presenta un efecto directo sobre la incidencia de cáncer de pulmón. Considérese esta misma situación, pero ahora sabiendo que las variables  $F$  y  $C$  presentan una relación de asociación debido, por ejemplo, a la existencia de un gen - no explícito en el modelo - que se presenta en determinados individuos y que los predispone a fumar y también a padecer un cáncer de pulmón. En este caso,  $F$  y  $C$  presentan una relación de asociación o correlación, sin la existencia de causalidad.

Aunque las relaciones causales no son siempre obvias y el concepto de causalidad puede presentar dificultades en su interpretación, Jensen (2001) contribuye a determinar dicha relación, con un sencillo ejemplo. Sean  $X_i$  y  $X_j$  dos variables correladas de las que no se puede determinar si una es causa de la otra. Entonces, imagínese que un agente externo fija un estado de  $X_i$  y esto no cambia la incertidumbre acerca de  $X_j$ , entonces, se puede concluir que  $X_i$  no es una causa de  $X_j$ . Si con este sencillo test no se indica ninguna relación causal entre  $X_i$  y  $X_j$ , entonces se ha de buscar un evento que tenga un impacto sobre  $X_i$  y  $X_j$ . Así, si  $X_k$  es dicho evento, se ha de revisar si  $X_i$  y  $X_j$  se vuelven independientes dado  $X_k$ .

El tipo de relación de dependencia que se presenta entre las variables en estudio, es fundamental en el desarrollo del modelo gráfico probabilístico apropiado para la resolución del problema. De esta forma, en modelos gráficos probabilísticos dirigidos, la relación de dependencia causal conforma la propia definición del modelo.

Para un estudio más completo del concepto de causalidad, se puede ver Pearl (2000) o Spirtes, et al. (2000).

Como se ha introducido en el Capítulo 1, los modelos gráficos probabilísticos en general y las Redes Bayesianas en particular, se componen de una parte cualitativa y otra parte cuantitativa.

Para la especificación de la información cualitativa de la Red Bayesiana, se utiliza un grafo acíclico dirigido (DAG)  $\mathcal{D} = (V, E)$ , donde cada uno de los nodos de  $\mathcal{D}$  representa las variables del problema  $\mathbf{X} = \{X_1, \dots, X_n\}$ , siendo por tanto  $V = \{X_1, \dots, X_n\}$ , y las aristas dirigidas que están en  $E$  muestran las relaciones de tipo causal, siendo él o los nodos padre, la causa y él o los nodos hijos, el efecto.

En el primer caso del ejemplo del tabaco y el cáncer de pulmón, en el que se tiene una relación de tipo causal entre  $T$  y  $C$ , el DAG que representa la información cualitativa del problema tiene una arista dirigida que va de  $T$  a  $C$ .

La existencia de un DAG,  $\mathcal{D}$ , en la especificación del problema, facilita la definición de los tres tipos de conexiones básicas entre los nodos (conexión en serie, divergente y convergente) que son fundamentales al enunciar el concepto inicial de  $d$ -separación.

Recuérdese que la  $d$ -separación muestra criterios de separación gráfica en  $\mathcal{D}$  asociados a la idea de independencia condicionada. En el Apartado 1.3.4 se presentaron los criterios de separación gráfica, en función del tipo de grafo en estudio. Para los DAGs se enunciaron dos definiciones del concepto de  $d$ -separación, aunque es la definición introducida por Lauritzen, et al. (1990) la más utilizada a la hora de concluir si dos conjuntos disjuntos  $X$  e  $Y$  de  $\mathcal{D}$  son  $d$ -separados según el grafo  $\mathcal{D}$  dado un tercer conjunto  $Z$ , es decir  $X \perp\!\!\!\perp_{\mathcal{D}} Y|Z$ , ya que

$$X \perp\!\!\!\perp_{\mathcal{D}} Y|Z \Leftrightarrow Z \text{ separa } X \text{ e } Y \text{ en el grafo moral del menor subconjunto ancestral que contiene a } X, Y \text{ y } Z$$

Además, en las Redes Bayesianas, las relaciones de independencia condicionada fijadas por  $\mathcal{D}$ , aplicando el concepto de  $d$ -separación, coinciden con las relaciones de independencia condicionada que presenta el modelo probabilístico que se especifica mediante la información cuantitativa de la Red Bayesiana (véase el Teorema 1.8).

Como ya se indicó, la información cuantitativa de la Red Bayesiana viene dada por un conjunto de distribuciones de probabilidad condicionada  $P = \{p(x_1|pa(X_1)), \dots, p(x_n|pa(X_n))\}$ , de forma que para cada variable  $X_i \in \mathbf{X}$  se tendrá la distribución de probabilidad condicionada de  $X_i$  dada la ocurrencia de sus padres  $pa(X_i)$  en el grafo  $\mathcal{D}$ , denotada por  $p(x_i|pa(X_i))$ .

Por tanto, como herramienta fundamental para el manejo de la Red Bayesiana

se puede obtener la distribución de probabilidad conjunta asociada a las variables del problema como una distribución jerárquica, de forma que dicha probabilidad se calcula multiplicando todos los elementos que describen  $P$ , es decir

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | pa(X_i))$$

Los modelos gráficos probabilísticos, y en particular las Redes Bayesianas, se utilizan como mecanismo de inferencia, actualizando la probabilidad conjunta del modelo, cuando se dispone de información o *evidencia* acerca de alguna de las variables que componen el problema. Por tanto, en una Red Bayesiana se pueden incorporar variables observables y no observables, incluyendo en esta denominación las variables y los parámetros de los modelos de inferencia habituales, desde una perspectiva bayesiana, con tal de que se puedan establecer distribuciones iniciales sobre las variables que no tienen padres en el DAG, y distribuciones condicionadas de cada variable por sus padres, para el resto de variables.

Se dice que se tiene evidencia acerca de una de las variables que describen una Red Bayesiana, cuando se conoce el estado preciso de dicha variable aleatoria. Así, por ejemplo, considérese una Red Bayesiana diseñada para resolver problemas de diagnóstico médico, con una estructura de "síntomas-enfermedad". La red está formada por un conjunto de variables, de manera que hay  $n$  variables que representan la ausencia o presencia de determinados síntomas y  $m$  variables que representan la ausencia o presencia de varias enfermedades.

Cuando llega un paciente, que tiene una de las  $m$  enfermedades, describe los síntomas que sufre y por tanto se tiene evidencia de los síntomas que padece el paciente y de los que no padece. Con esa evidencia, la red actúa como un mecanismo de inferencia, actualizando la probabilidad conjunta, y por consiguiente, todas las probabilidades de la Red Bayesiana.

Este proceso, definido como *propagación de la evidencia*, está basado en el Teorema de Bayes actuando como transmisor de la información. De esta forma, tras conocerse los síntomas del paciente en el ejemplo anterior, se obtiene la probabilidad final o a posteriori asociada a sufrir una enfermedad, dada la evidencia acerca de los síntomas. Siendo la probabilidad inicial o a priori, la probabilidad de presencia de la enfermedad, y la verosimilitud, la probabilidad de que un paciente con una enfermedad tenga los síntomas descritos.

A continuación se introduce un ejemplo de Red Bayesiana muy utilizado, denominado Asia. Este ejemplo es una variación de la Red Cadena Ch-Asia introducida en el Capítulo 1. En realidad, el ejemplo presentado mediante una Red Cadena es una variación del ejemplo Asia que se muestra a continuación y fue introducido por Lauritzen, et al. (1988).

### Ejemplo 2.1

*Se sabe que una disnea puede ser producida por la tuberculosis, el cáncer de pulmón o la bronquitis, cualquier combinación entre las tres enfermedades o ninguna de las tres. Una visita reciente a Asia incrementa la posibilidad de tener tuberculosis, mientras que fumar es un factor de riesgo tanto para el cáncer de pulmón como para la bronquitis. El resultado de una prueba simple de rayos X no discrimina entre cáncer de pulmón y tuberculosis, al igual que entre presencia y ausencia de disnea.*

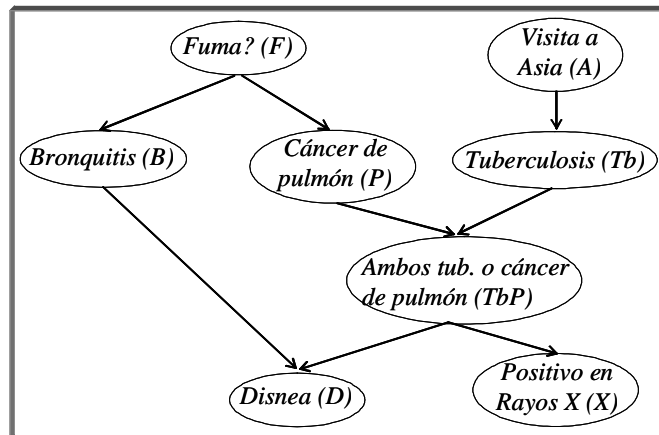


Figura 2.1. DAG asociado a la Red Bayesiana que modeliza el problema Asia

El DAG de la Figura 2.1, representa las variables del problema y las relaciones de dependencia que existen entre las mismas. La parte cuantitativa del problema viene dada por la tabla que se muestra a continuación.

|                                 |   |
|---------------------------------|---|
| $P(A = 1) = 0.01$               | $\Rightarrow P(A = 0) = 0.99$               |
| $P(F = 1) = 0.5$                | $\Rightarrow P(F = 0) = 0.5$                |
| $P(B = 1 F = 1) = 0.6$          | $\Rightarrow P(B = 0 F = 1) = 0.4$          |
| $P(B = 1 F = 0) = 0.3$          | $\Rightarrow P(B = 0 F = 0) = 0.7$          |
| $P(P = 1 F = 1) = 0.1$          | $\Rightarrow P(P = 0 F = 1) = 0.9$          |
| $P(P = 1 F = 0) = 0.01$         | $\Rightarrow P(P = 0 F = 0) = 0.99$         |
| $P(Tb = 1 A = 1) = 0.05$        | $\Rightarrow P(Tb = 0 A = 1) = 0.95$        |
| $P(Tb = 1 A = 0) = 0.01$        | $\Rightarrow P(Tb = 0 A = 0) = 0.99$        |
| $P(D = 1 B = 1, TbP = 1) = 0.9$ | $\Rightarrow P(D = 0 B = 1, TbP = 1) = 0.1$ |
| $P(D = 1 B = 0, TbP = 1) = 0.7$ | $\Rightarrow P(D = 0 B = 0, TbP = 1) = 0.3$ |
| $P(D = 1 B = 1, TbP = 0) = 0.8$ | $\Rightarrow P(D = 0 B = 1, TbP = 0) = 0.2$ |
| $P(D = 1 B = 0, TbP = 0) = 0.1$ | $\Rightarrow P(D = 0 B = 0, TbP = 0) = 0.9$ |
| $P(TbP = 1 P = 1, Tb = 1) = 1$  | $\Rightarrow P(TbP = 0 P = 1, Tb = 1) = 0$  |
| $P(TbP = 1 P = 0, Tb = 1) = 1$  | $\Rightarrow P(TbP = 0 P = 0, Tb = 1) = 0$  |
| $P(TbP = 1 P = 1, Tb = 0) = 1$  | $\Rightarrow P(TbP = 0 P = 1, Tb = 0) = 0$  |
| $P(TbP = 1 P = 0, Tb = 0) = 0$  | $\Rightarrow P(TbP = 0 P = 0, Tb = 0) = 1$  |
| $P(X = 1 TbP = 1) = 0.98$       | $\Rightarrow P(X = 0 TbP = 1) = 0.02$       |
| $P(X = 1 TbP = 0) = 0.05$       | $\Rightarrow P(X = 0 TbP = 0) = 0.95$       |

Tabla 2.1. Probabilidades condicionadas de las variables del Ejemplo 2.1

Se sabe que la distribución conjunta se puede calcular como el producto de las condicionadas, siendo

$$\begin{aligned}
 &P(F, A, B, P, Tb, TbP, D, T, X) = \\
 &= P(F)P(A)P(B|F)P(P|F)P(Tb|A)P(TbP|P, Tb)P(D|B, TbP)P(X|TbP)
 \end{aligned}$$

Además, si se tiene evidencia acerca de alguna de las variables del problema y por ejemplo se sabe que un paciente con los síntomas presentados es fumador, siendo  $F = 1$ , entonces, se puede actualizar cualquier probabilidad de interés de la red, dada la evidencia. Así, si se desea conocer la probabilidad de que el paciente fumador tenga disnea, se sabe que

$$P(D = 1|F = 1) = \frac{P(D = 1, F = 1)}{P(F = 1)} = 0.5528$$

de forma que  $P(D = 1, F = 1)$  y  $P(F = 1)$  se obtienen marginalizando la distribución de probabilidad conjunta.

El proceso de inferencia realizado mediante la propagación de la evidencia, puede tener como objetivo describir el conjunto de variables no evidenciales o describir el comportamiento de una única variable de interés  $X_i$ . En este último caso, que aparece con frecuencia en las Redes Bayesianas y en los modelos gráficos probabilísticos, la variable de interés se suele representar en el DAG mediante un nodo compuesto por dos círculos concéntricos (véase el Ejemplo 2.9) y el proceso de propagación de la evidencia se ve reducido a obtener respuestas sólo para dicha variable de interés.

Se han desarrollado una amplia colección de algoritmos para realizar el proceso de propagación de la evidencia en Redes Bayesianas. Entre otros autores se puede destacar a Kim, et al. (1983) que presentan una versión para actualizar la probabilidad en DAGs basada en el envío de mensajes por la Red Bayesiana. Shafer, et al. (1990) que desarrollan un método basado en la propagación en los árboles de unión. Lauritzen, et al. (1988) y Jensen, et al. (1990a) que proponen el método Hugin, que posteriormente será implementado creando el programa informático HUGIN (Andersen, et al., 1989), basado en la multiplicación de potenciales, de forma que los potenciales de los ciclados cambian dinámicamente. Gilks, et al. (1994) que desarrollan el sistema BUGS, basado en el muestreador Gibbs en Redes Bayesianas.

En la Sección 2.3, se define y se profundiza en este concepto de propagación de la evidencia.

### 2.2.1. Tipos de Redes Bayesianas

Existen diversos tipos de Redes Bayesianas dependiendo de las variables aleatorias que intervienen en el problema.

Así, si las variables del problema son todas discretas, el modelo asociado es una *Red Bayesiana Discreta* o *Red Bayesiana Multinomial*. Si las variables del problema son normales, la red es una *Red Bayesiana Gaussiana*, y aunque se trabaja en la definición de una Red Bayesiana para otro tipo de variables aleatorias continuas, todavía no está determinado el mecanismo de inferencia



cuando la red está formada por otro tipo de variables continuas no Gaussianas. Además, también existen las *Redes Bayesianas Mixtas* formadas por variables discretas y Gaussianas. A continuación se presentan estos tipos de redes.

### Redes Bayesianas Discretas

Las Redes Bayesianas Discretas se caracterizan porque todas las variables del modelo son discretas, de forma que cada variable sólo puede tomar un conjunto finito de valores.

Cuando además las variables del problema son binarias, respondiendo a los procesos de Bernoulli, la red se denota como Red Bayesiana Multinomial.

A continuación, se presenta un ejemplo de Red Bayesiana Discreta muy utilizado en la literatura. La red queda definida mediante el DAG de la Figura 2.2 y las distribuciones de probabilidad condicionada asociadas a los valores de las variables, dados sus padres en el DAG.

#### Ejemplo 2.2

*Una mañana al levantarse, el Sr. Sánchez comprueba que su césped está mojado. No sabe si habrá llovido durante la noche o si dejó encendido el aspersor de su jardín. Observando el jardín de su vecino el Sr. Pérez puede obtener alguna conclusión.*

El problema que se plantea puede modelizarse mediante una Red Bayesiana, siendo el DAG que representa la información cualitativa de la red, tal que

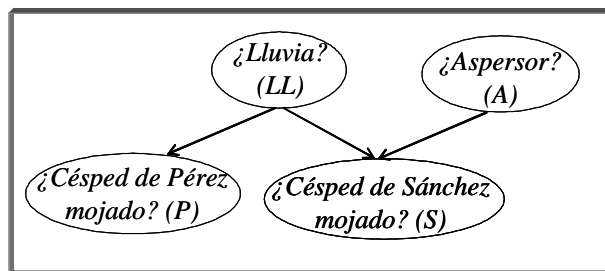


Figura 2.2. DAG asociado a la Red Bayesiana que modeliza el problema del césped mojado del Sr. Sanchez

Todas las variables del problema son dicotómicas, siendo 1 el valor asociado a la respuesta 'sí' y 0 a la respuesta 'no'.

Para determinar la Red Bayesiana Multinomial, se presentan en la Tabla 2.2 las probabilidades condicionadas  $P(LL)$ ,  $P(A)$ ,  $P(P|LL)$  y  $P(S|LL, A)$

|                                |  |
|--------------------------------|--|
| $P(LL = 1) = 0.2$              | $\Rightarrow P(LL = 0) = 0.8$              |
| $P(A = 1) = 0.1$               | $\Rightarrow P(A = 0) = 0.9$               |
| $P(P = 1 LL = 1) = 1$          | $\Rightarrow P(P = 0 LL = 1) = 0$          |
| $P(P = 1 LL = 0) = 0.2$        | $\Rightarrow P(P = 0 LL = 0) = 0.8$        |
| $P(S = 1 LL = 1, A = 1) = 1$   | $\Rightarrow P(S = 0 LL = 1, A = 1) = 0$   |
| $P(S = 1 LL = 1, A = 0) = 1$   | $\Rightarrow P(S = 0 LL = 1, A = 0) = 0$   |
| $P(S = 1 LL = 0, A = 1) = 0.9$ | $\Rightarrow P(S = 0 LL = 0, A = 1) = 0.1$ |
| $P(S = 1 LL = 0, A = 0) = 0$   | $\Rightarrow P(S = 0 LL = 0, A = 0) = 1$   |

Tabla 2.2. Probabilidades condicionadas de las variables del Ejemplo 2.2

Teniendo en cuenta la definición de Red Bayesiana, se puede obtener la probabilidad conjunta como el producto de todas las probabilidades condicionadas, de forma que

$$P(LL, A, P, S) = P(LL)P(A)P(P|LL)P(S|LL, A)$$

Con la distribución de probabilidad conjunta es posible calcular la distribución de las variables de interés, cuando se tiene evidencia acerca de alguna variable del problema. Así, en este caso, se sabe que el césped del Sr. Sanchez está mojado, por tanto, se puede calcular  $P(A|S = 1)$  y  $P(LL|S = 1)$ , siendo

$$\begin{aligned} P(A = 1|S = 1) &= 0.3382 \\ P(LL = 1|S = 1) &= 0.7353 \end{aligned}$$

Como se puede observar con las probabilidades obtenidas, una vez que se ha introducido evidencia en la red, siendo  $S = 1$ , tanto la probabilidad de que sea debido al aspersor,  $A = 1$ , como la de que sea debido a la lluvia,  $LL = 1$ , aumentan su valor respecto a las probabilidades iniciales.

Cuando se observa el césped del vecino, se puede calcular de nuevo la distribución de  $A$  y de  $LL$ , así, si por ejemplo el césped del vecino también está mojado, se tiene que

$$\begin{aligned}P(A = 1|S = 1, P = 1) &= 0.1604 \\P(LL = 1|S = 1, P = 1) &= 0.9328\end{aligned}$$

Ahora con esta nueva evidencia, disminuye la probabilidad asociada a  $A = 1$  y aumenta la probabilidad de  $LL = 1$ . Este efecto por el que las probabilidades condicionadas se ven notablemente alteradas al incorporar nueva evidencia, se conoce como *explaining away* y reproduce claramente el fenómeno de la transmisión de la evidencia entre los diferentes nodos de una red, no sólo entre los que existe un enlace.

## Redes Bayesianas Gaussianas

Cuando las variables aleatorias del problema son normales, la Red Bayesiana que lo modeliza es una Red Bayesiana Gaussiana.

En las Redes Bayesianas Gaussianas, la distribución conjunta de las variables del problema  $\mathbf{X} = \{X_1, \dots, X_n\}$  es también normal multivariante  $N(\boldsymbol{\mu}, \Sigma)$ , de forma que la función de densidad conjunta es

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

donde  $\boldsymbol{\mu}$  es el vector de medias de dimensión  $n$ ,  $\Sigma$  es la matriz de covarianzas, definida positiva de dimensión  $n \times n$ ,  $|\Sigma|$  denota el determinante de la matriz de covarianzas y  $(\mathbf{x} - \boldsymbol{\mu})^T$  el vector traspuesto de  $(\mathbf{x} - \boldsymbol{\mu})$ .

Cabe señalar que la condición de normalidad aplicada a las distribuciones condicionadas de cada hijo por sus padres no implica una distribución conjunta normal multivariante, salvo en el caso en que se exigen varianzas condicionadas constantes y regresiones lineales (Arnold, et al., 1999).

Por la definición de Red Bayesiana se ha de verificar que la probabilidad

conjunta es jerárquica, factorizándose mediante

$$f(\mathbf{x}) = \prod_{i=1}^n f(x_i | pa(X_i)).$$

Por tanto, partiendo de la densidad conjunta del problema  $N(\boldsymbol{\mu}, \Sigma)$ , la densidad condicionada  $f(x_i | pa(X_i)) \forall X_i$ , es también normal y para cada variable del problema  $X_i$  dados sus padres, viene dada por

$$f(x_i | pa(X_i)) \sim N \left( \mu_i + \sum_{j=1}^{i-1} \beta_{ij} (x_j - \mu_j), v_i \right)$$

donde  $\beta_{ij}$  es el coeficiente de regresión de  $X_j$  en la regresión de  $X_i$  sobre sus padres, y  $v_i$  es la varianza condicionada de  $X_i$  dados sus padres, siendo  $v_i = \Sigma_i - \Sigma_{ipa(X_i)} \Sigma_{pa(X_i)}^{-1} \Sigma_{ipa(X_i)}^T$ .

Nótese que el coeficiente de regresión es cero ( $\beta_{ij} = 0$ ) si y solo si no hay una arista dirigida de  $X_j$  a  $X_i$ .

Para ilustrar el concepto de Red Bayesiana Gaussiana, a continuación se introduce un ejemplo tomado de Castillo, et al. (2003).

### Ejemplo 2.3

*Se desea estudiar el caudal de un río; para ello se toman varios puntos en el río en los que se mide dicho caudal. Se asume que la distribución conjunta del caudal del agua del río es normal.*

Se puede modelizar el problema mediante una Red Bayesiana Gaussiana, donde las variables del problema son los distintos puntos del río, así  $\mathbf{X} = \{A, B, C, D\}$ .

El DAG que se muestra a continuación especifica una parte de la Red Bayesiana Gaussiana

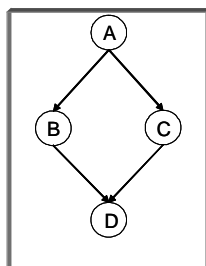


Figura 2.3. DAG asociado a la Red Bayesiana Gaussiana que modeliza el problema del caudal del río

Para continuar con la especificación de la Red Bayesiana Gaussiana, se han de definir los parámetros de la distribución conjunta  $N(\boldsymbol{\mu}, \Sigma)$ . En este ejemplo,

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 4 \\ 9 \\ 14 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 4 & 4 & 8 & 12 \\ 4 & 5 & 8 & 13 \\ 8 & 8 & 20 & 28 \\ 12 & 13 & 28 & 42 \end{pmatrix}$$

Verificándose que

$$f(a, b, c, d) = f(a)f(b|a)f(c|a)f(d|b, c)$$

Además, dichas densidades condicionadas han de responder a

$$\begin{aligned} f(a) &\sim N(\mu_A, v_A) \\ f(b|a) &\sim N(\mu_B + \beta_{BA}(a - \mu_A), v_B) \\ f(c|a) &\sim N(\mu_C + \beta_{CA}(a - \mu_A), v_C) \\ f(d|b, c) &\sim N(\mu_D + \beta_{DB}(b - \mu_b) + \beta_{DC}(c - \mu_C), v_D) \end{aligned}$$

Tabla 2.3. Distribución condicionada de las variables del Ejemplo 2.3

Se han desarrollado diversas técnicas para obtener el parámetro  $\Sigma$ , a partir de los parámetros  $\beta_{ij}$  y  $v_i$  que determinan la densidades condicionadas. La matriz de covarianzas  $\Sigma$  ha de ser una matriz definida positiva y ha de mantener en su estructura, las características de independenciam condicionada que muestra el DAG. Dichas técnicas y propiedades se exponen con detalle en la Sección 2.4.

## Redes Bayesianas Mixtas

Las Redes Bayesianas Mixtas, también denotadas como *Redes Bayesianas Discretas-Gaussianas*, se caracterizan por incluir variables discretas y continuas en el modelo gráfico probabilístico dirigido.

Para poder especificar el modelo, las variables discretas toman un número finito de estados y las variables continuas han de ser Gaussianas. Además, las variables discretas deben preceder a las continuas en el grafo.

En las Redes Bayesianas Mixtas, se particiona el conjunto de nodos  $V = \{X_1, \dots, X_n\}$  en función de si representan a variables discretas ( $\Delta$ ) o a variables continuas ( $\Gamma$ ), siendo  $V = \Delta \cup \Gamma$ . Así, se denota el conjunto de todas las variables aleatorias como

$$\mathbf{X} = (x)_{\alpha \in \mathbf{V}} = (i, \varsigma) = ((i_\delta)_{\delta \in \Delta}, (\varsigma_\gamma)_{\gamma \in \Gamma})$$

La distribución conjunta asociada a las variables que forman una Red Bayesiana Mixta, es la distribución condicionada Gaussiana, cuya densidad viene dada por

$$f(x) = f(i, \varsigma) = \exp \{g(i) + \mathbf{h}(i)^T \varsigma - \varsigma^T \mathbf{K}(i) \varsigma / 2\}$$

donde  $i$  representa las variables discretas y  $\varsigma$  las continuas,  $g(i)$  es un escalar,  $\mathbf{h}(i)$  un vector,  $\mathbf{K}(i)$  una matriz definida positiva y  $\mathbf{h}(i)^T$  denota el vector  $\mathbf{h}(i)$  traspuesto.

Existen diversas propiedades para que la densidad conjunta se presente de forma jerárquica y se pueda factorizar, considerando las densidades condicionadas de las variables dada la ocurrencia de sus padres en el DAG. Para ampliar los conceptos introducidos acerca de Redes Bayesianas Mixtas, véase Cowell, et al. (1999).

A continuación se muestra un ejemplo, introducido por Cowell, et al. (1999), relativo al control de la emisión de metales pesados en los residuos de una incineradora.

**Ejemplo 2.4**

*La emisión de residuos de una incineradora depende de los diferentes componentes de los residuos resultantes.*

*Otro factor importante es el residuo quemado que se mide en función de la concentración de  $\text{CO}_2$  en la emisión.*

*La eficiencia del filtro depende del estado técnico del electrofiltro y de la composición del residuo.*

*La emisión de metales pesados depende de la concentración de metal en los residuos resultantes y de la emisión de partículas de polvo. La emisión de partículas de polvo se monitoriza midiendo la penetrabilidad de la luz.*

Este problema se puede modelizar mediante una Red Bayesiana Mixta, ya que se observan variables de tipo discreto y de tipo continuo en la red. Tras la especificación del DAG se presentan las variables y sus distribuciones de probabilidad condicionada. Los valores de los parámetros de dichas distribuciones vienen especificadas por los expertos tras realizar un estudio minucioso de las variables.

El DAG que describe las relaciones de dependencia entre las variables en estudio, se muestra en la Figura 2.4.

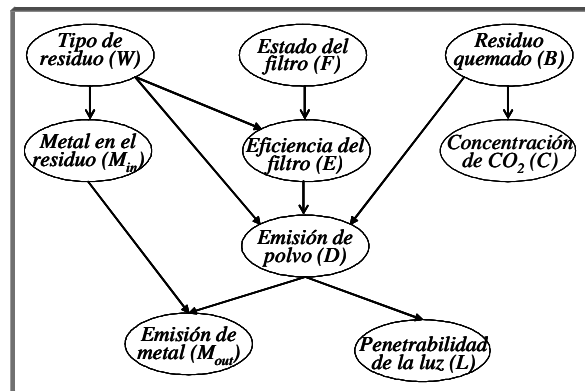


Figura 2.4. DAG asociado a la Red Bayesiana que modeliza el problema de la emisión de residuos

Las variables discretas del problema son  $W$ ,  $F$  y  $B$ , cuyas probabilidades condicionadas de interés se muestran en la Tabla 2.4.

|                                   |  |
|-----------------------------------|--|
| $P(W = industrial) = \frac{2}{7}$ | $\Rightarrow P(W = doméstico) = \frac{5}{7}$ |
| $P(F = eficiente) = 0.95$         | $\Rightarrow P(F = deficiente) = 0.05$       |
| $P(B = estable) = 0.85$           | $\Rightarrow P(B = inestable) = 0.15$        |

Tabla 2.4. Probabilidades de las variables discretas del Ejemplo 2.4

El resto de variables son normales y se representan en una escala logarítmica, especificando sus distribuciones condicionadas en la Tabla 2.5.

|   |
|---|
| $M_{in} industrial \sim N(0.5, 0.01)$             |
| $M_{in} doméstico \sim N(-0.5, 0.005)$            |
| $E eficiente, doméstico \sim N(-3.2, 0.00002)$    |
| $E deficiente, doméstico \sim N(-0.5, 0.0001)$    |
| $E eficiente, industrial \sim N(-3.9, 0.00002)$   |
| $E deficiente, industrial \sim N(-0.4, 0.0001)$   |
| $C estable \sim N(-2, 0.1)$                       |
| $C inestable \sim N(-1, 0.3)$                     |
| $D estable, industrial, e \sim N(6.5 + e, 0.03)$  |
| $D estable, doméstico, e \sim N(6 + e, 0.04)$     |
| $D inestable, industrial, e \sim N(7.5 + e, 0.1)$ |
| $D inestable, doméstico, e \sim N(7 + e, 0.1)$    |
| $M_{out} d, M_{in} \sim N(d + M_{in}, 0.002)$     |
| $L d \sim N(3 - d/2, 0.25)$                       |

Tabla 2.5. Distribuciones condicionadas de las variables del Ejemplo 2.4

La Red Bayesiana Mixta queda determinada mediante el DAG de la Figura 2.4 y las Tablas 2.4 y 2.5.

Se han realizado interesantes aplicaciones de Redes Bayesianas Mixtas. Por ejemplo, en modelos poligénicos para determinar el pedigrí genético (Sham, 1998), donde los genes mayores son variables discretas y los efectos poligénicos son las variables continuas.



Sin embargo, existen restricciones asociadas al modelo y a su diseño, que dificultan la propagación exacta de la evidencia (Cowell, et al., 1999).

### 2.2.2. Aplicaciones de las Redes Bayesianas

Como se ha ido viendo a la largo de la Sección, las aplicaciones de las Redes Bayesianas son muchas y muy diversas permitiendo modelizar problemas con estructuras complejas. Además, el mecanismo de propagación de la evidencia sirve para el aprendizaje de los parámetros y para realizar inferencia, a medida que se tiene información acerca de las variables de la red.

Algunos autores como Buntine (1994), Heckerman (1995), Lauritzen, et al. (2003), Dobra, et al. (2004) o Pérez, et al. (2006) entre otros, han aplicado las Redes Bayesianas en la resolución de problemas de distintas áreas.

En este Apartado se muestran algunos campos que más han contribuido al desarrollo de los modelos gráficos en general y dos aplicaciones de las Redes Bayesianas. La primera, de las más significativas, es el diagnóstico médico y la segunda, más actual, está en el campo de la genética.

Una de las áreas que más han influido en el desarrollo de los modelos gráficos, y en particular de la Redes Bayesianas, es la de la Inteligencia Artificial con los *sistemas expertos*.

Un sistema experto se utiliza para codificar el conocimiento y las habilidades de un grupo de expertos en una herramienta que pueda ser utilizada por individuos no expertos.

Los sistemas expertos están formados por dos partes, la base del conocimiento y el motor de inferencia. La base del conocimiento contiene el conocimiento específico de un problema, codificado. Y el motor de inferencia consiste en uno o varios algoritmos que procesan el conocimiento codificado de la base del conocimiento, junto con futura información específica introducida en la aplicación, para realizar inferencias y obtener resultados relativos a las variables del problema.

Inicialmente, los sistemas expertos se formaban con deducciones lógicas (reglas), aunque este proceso era insuficiente y limitado, ya que no siempre es posible

determinar con certeza la ocurrencia de un evento concreto y es este aspecto del manejo de la incertidumbre el que restaba una enorme eficiencia a los procedimientos basados en reglas.

Entonces, se introduce la idea de cuantificar la incertidumbre asociada a los resultados de las reglas lógicas y surgen los sistemas expertos probabilísticos, que describen las relaciones entre las variables mediante su distribución de probabilidad conjunta.

El razonamiento bayesiano es perfecto como motor de inferencia, ya que conociendo la probabilidad conjunta del modelo, se puede actualizar la información acerca de cualquier variable cuando se tiene evidencia sobre otra de las variables del problema.

Son muchas las aplicaciones que los sistemas expertos han tenido en el área del diagnóstico médico. Cada vez más problemas complejos descritos mediante un conjunto de variables relacionadas entre sí, se modelizan mediante Redes Bayesianas y la variedad de los mismos en el área del diagnóstico médico es cada vez mayor.

A continuación se incluye un ejemplo de Red Bayesiana, introducido inicialmente por Franklin, et al. (1991) denominado Child.

### **Ejemplo 2.5**

*El Hospital Great Ormond Street de enfermedades de niños en Londres, actúa como centro de referencia de niños prematuros con enfermedad coronaria congénita.*

*Se sospecha de la existencia de una enfermedad cardiovascular cuando se presenta cianosis (coloración azul de la piel del bebé), debido a algún fallo del corazón ocurrido inmediatamente tras el nacimiento del bebé. Es vital llevar al bebé a algún centro especialista de referencia, aunque antes de transportar al bebé debe comenzarse algún tratamiento apropiado que se consulta telefónicamente.*

*El diagnóstico no es sencillo y normalmente lo realizan médicos residentes no especialistas a cualquier hora del día. La decisión se basa en una descripción clínica dada por el pediatra de referencia y pruebas como tensión arterial, electrocardiograma y rayos X.*

Por la premura con la que se ha de tomar la decisión acerca del posible

diagnóstico, se construye una Red Bayesiana, con información y valoraciones aportadas por pediatras expertos, que ayude a los médicos a obtener un juicio sobre la enfermedad. La Red Bayesiana construida debe ser muy eficiente, sin embargo, se producirán errores si se permiten datos ausentes, presentaciones atípicas o errores en las interpretaciones de los signos clínicos.

Se consideran seis enfermedades distintas (PFC, TGA, tetralogía de Fallot, PAIVS, TAPVD y enfermedad pulmonar) que categorizan la variable Enfermedad.

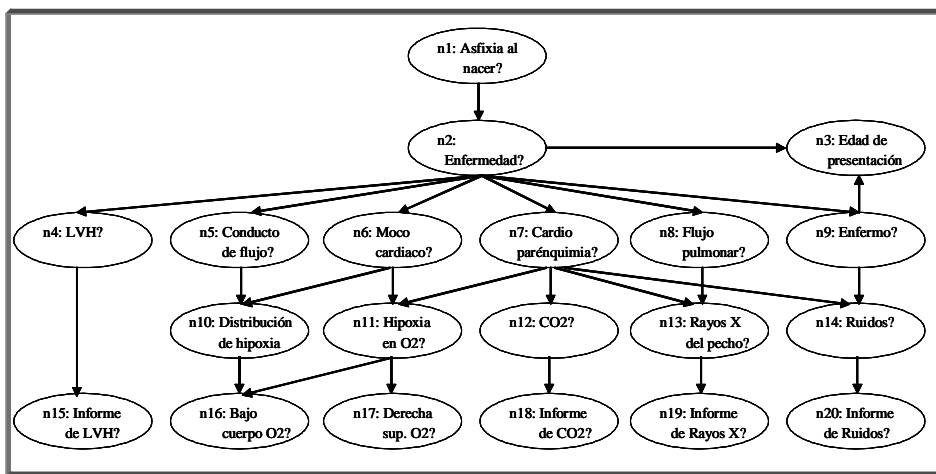


Figura 2.5. DAG asociado a la Red Bayesiana que modeliza el problema Child

En la Figura 2.5 se muestra el DAG que describe la Red Bayesiana del problema.

Las probabilidades condicionadas fueron asignadas por los expertos mediante la observación de las bases de datos existentes.

La Red Bayesiana del ejemplo Child es muy efectiva y se utiliza para describir la distribución de la variable Enfermedad del problema, en función de los síntomas mostrados por un niño concreto. De esta forma, se diagnostica la posible enfermedad y se realizan las pruebas adecuadas.

Otra aplicación importante de las Redes Bayesianas se encuentra en el campo de la genética. Esta aplicación surge debido a la elevada dimensión de los pro-

blemas genéticos, que hace intratable el problema computacionalmente. Con Redes Bayesianas, y modelos gráficos probabilísticos en general, se pueden realizar cálculos eficientes y simplificados, imponiendo hipótesis de independencia condicionada entre las variables del problema. Estas hipótesis, pueden reducir el problema a pequeñas componentes más manejables. Además, debido a la estructura compleja de los problemas genéticos, es posible representar el problema gráficamente mediante los grafos introducidos en la Sección 1.2.

Lauritzen, et al. (2003) describen una Red Bayesiana para tratar vínculos genéticos asociados con el pedigrí, entendiendo por pedigrí el conjunto de relaciones familiares entre un grupo de individuos. En los pedigrís, es frecuente trabajar con su representación gráfica (véase la Figura 2.6).

### Ejemplo 2.6

Se desea estudiar el pedigrí mostrado en la Figura 2.6. En dicho pedigrí, los fundadores iniciales del mismo son los individuos 1,2,3, y 4. Los individuos 5,8 y 9 son fundadores recientes que se han incorporado al pedigrí emparejándose con miembros del mismo. Finalmente, los individuos 11, 12, 13 y 14 son los últimos del pedigrí y no están emparejados.

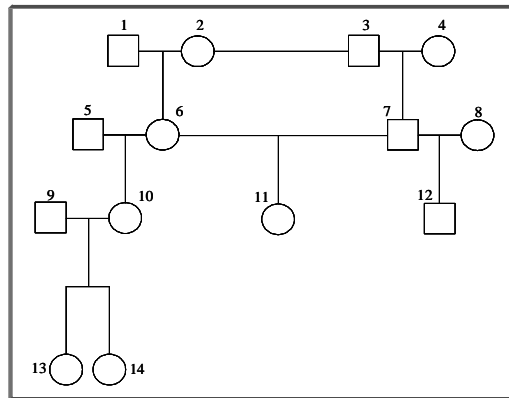


Figura 2.6. Representación gráfica estándar del Ejemplo 2.6, donde las hembras se muestran mediante un círculo y los machos mediante un cuadrado

Para expresar el pedigrí de la Figura 2.6 como una Red Bayesiana, los nodos del grafo deben representar variables aleatorias para las que se puedan definir una distribución de probabilidad conjunta que satisfaga la factorización (1.1).

Existen varias formas de diseñar la Red Bayesiana y sus propiedades asociadas. En este ejemplo, se describe la *red de segregación*, que es la más directa y que recoge todas las relaciones inherentes a las relaciones familiares del pedigrí.

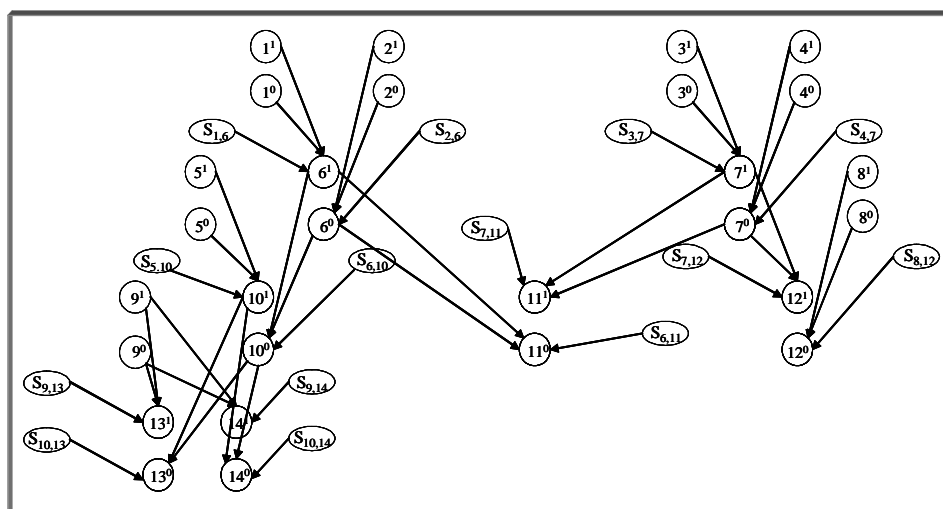


Figura 2.7. DAG asociado a la Red Bayesiana descrita para el problema del pedigrí del Ejemplo 2.6

La Red Bayesiana de la Figura 2.7, se construye de forma que para cada individuo  $i$ , se tengan dos nodos ( $i^0$  y  $i^1$ ) que representen la maternidad heredada en los genes de  $i$ , siendo 0 la etiqueta asociada a la maternidad heredada y 1 a la paternidad heredada. Las variables aleatorias subyacentes pueden tomar cualquiera de los tipos de alelos del gen  $a$  del sistema.

A cada miembro del pedigrí no fundador, le llegan dos aristas dirigidas de los dos genes del padre al gen paternidad y dos aristas dirigidas de los genes de la madre al gen maternidad.

Los nodos etiquetados como  $i^1$  se asocian a la variable aleatoria  $L_{i^1}$  asignando el tipo de alelo del gen heredado por el individuo  $i$  de su padre. Igualmente, se tiene para el gen heredado por  $i$  de su madre, siendo  $L_{i^0}$  la variable aleatoria.

El resto de nodos adicionales son indicadores de la meiosis (proceso de reducción cromática en el que se reduce a la mitad el número de cromosomas, gametos o células reproductoras) o la segregación; de forma que se añaden a los padres de cada nodo gen. Estos nodos son binarios, donde 1 denota que se

ha heredado una copia del gen de paternidad y 0 indica la herencia del gen de maternidad. De esta forma, cada tipo de alelo de los no fundadores es una función determinística de sus padres. Por tanto, para la herencia del gen de paternidad se tiene que

$$L_{i^1} = f(l_{p_i^1}, l_{p_i^0}, s_{p_i,i}) = \begin{cases} l_{p_i^1} & \text{si } s_{p_i,i} = 1 \\ l_{p_i^0} & \text{si } s_{p_i,i} = 0 \end{cases}$$

Y de forma similar para el gen de maternidad heredada

$$L_{i^0} = f(l_{m_i^1}, l_{m_i^0}, s_{m_i,i}) = \begin{cases} l_{m_i^1} & \text{si } s_{m_i,i} = 1 \\ l_{m_i^0} & \text{si } s_{m_i,i} = 0 \end{cases}$$

donde  $m_i$  y  $p_i$  son etiquetas ligadas a la madre y el padre del individuo  $i$ , y  $s_{p_i,i}$  y  $s_{m_i,i}$  son variables binarias que asignan indicadores de segregaciones de  $i$  del padre y de la madre, respectivamente.

Las leyes de la herencia se pueden codificar dejando que los indicadores de segregación sean independientes, con probabilidades de transmisión dadas por

$$P(S_{p_i,i} = 1) = p_1 \quad \text{y} \quad P(S_{m_i,i} = 1) = p_0$$

siendo para el caso más simple de la herencia Mendeliana  $p_1 = p_0 = 1/2$ .

En la Red Bayesiana de la Figura 2.7 también se supone la unión aleatoria de los gametos (células que, en la reproducción sexual, se unen a otra célula para dar origen a un nuevo ser), además de mostrarse la independencia entre genes fundadores del resto y de los indicadores de segregación.

Para obtener más detalles acerca de la obtención de la Red Bayesiana que describe el problema del pedigrí presentado, puede verse Lauritzen, et al. (2003). También Dobra, et al. (2004) contiene aplicaciones de las Redes Bayesianas, para la descripción y resolución de problemas de detección de patrones en las expresiones genéticas obtenidas mediante *microarrays*.

### 2.3. Propagación de la evidencia en Redes Bayesianas

En esta Sección se presenta el proceso de *propagación de la evidencia*, que es una de las características fundamentales asociadas a las Redes Bayesianas.

Este proceso, también denominado *propagación de incertidumbre* o *inferencia probabilística*, consiste en actualizar la información probabilística de las variables de la red cuando se tiene información o evidencia acerca del estado o del valor de alguna de las variables que componen el problema.

Como se ha comentado en el Capítulo anterior, estos métodos de propagación de la evidencia se basan principalmente en el Teorema de Bayes, que permite calcular las probabilidades a posteriori de ocurrencia de un determinado suceso, dada la probabilidad a priori de dicho suceso y la verosimilitud de los datos. La información actualizada de las variables es la probabilidad a posteriori de las mismas.

De hecho, las Redes Bayesianas reciben dicho nombre por la utilización del Teorema de Bayes en la actualización de las probabilidades de la red. Sin embargo, para Redes Bayesianas con un gran número de variables se ha de recurrir a métodos de propagación eficientes, como los que se describirán posteriormente, que generalizan la idea del Teorema de Bayes.

Los métodos de propagación de la evidencia que se introducen en esta Sección se han desarrollado para el caso discreto y son básicos para entender la idea fundamental de propagación de la evidencia basada en el envío de mensajes y la multiplicación de potenciales para actualizar las probabilidades de la red.

Además, se presentan dos algoritmos de propagación de la evidencia. El primero se basa en el envío de mensajes cuando la estructura del DAG es simple, un poliárbol. El segundo algoritmo muestra la misma idea pero cuando la estructura del DAG es más compleja y existe un árbol de unión asociado a dicho DAG. Este último algoritmo es básico para el conocimiento de las Redes Bayesianas y es el más utilizado; además, se ha implementado en programas específicos para el tratamiento de Redes Bayesianas, como el programa HUGIN<sup>1</sup> (Andersen, et al., 1989).

En esta Sección se presenta la definición de evidencia y los algoritmos de propagación comentados. Dado que en la siguiente Sección se estudian a fondo las Redes Bayesianas Gaussianas, es en dicho punto donde se introduce un algoritmo de propagación de la evidencia para el modelo Gaussiano.

---

<sup>1</sup>Página web del programa <http://www.hugin.com>

### 2.3.1. Evidencia y propagación

En las Redes Bayesianas se tiene *evidencia* acerca de una variable aleatoria, cuando para una situación particular, se conoce el estado de dicha variable. Así, se define el concepto de evidencia como

**Definición 2.1** (*Evidencia*)

Sea una Red Bayesiana donde  $\mathbf{X} = \{X_1, \dots, X_n\}$  es el conjunto de variables del problema. Se dice que existe evidencia acerca de un subconjunto de variables  $\mathbf{E} \subset \mathbf{X}$ , cuando se conocen los valores exactos que toman dichas variables, siendo  $\mathbf{E} = \{X_{e_1} = e_1, \dots, X_{e_k} = e_k\}$  con  $k < n$ .

El conjunto de evidencias también se denota como  $\mathbf{E} = \mathbf{e}$ .

La evidencia es un elemento dinámico que varía en función de la situación de la red en el momento. Así, en Redes Bayesianas diseñadas por ejemplo para el diagnóstico médico, con una estructura de "síntomas-enfermedad", dependiendo de los síntomas del paciente, se tendrá una evidencia u otra, y por tanto una probabilidad distinta asociada a las posibles enfermedades causantes de dichos síntomas.

Inicialmente, cuando se describe una Red Bayesiana no se tiene evidencia acerca de ninguna de las variables del problema, por tanto en este caso el proceso de propagación de la evidencia consiste en calcular las probabilidades marginales para cada variable  $X_i \in \mathbf{X}$ , obteniéndose así una primera información sobre la distribución marginal de cada variable del problema.

Cuando para un caso particular, se tiene evidencia acerca de la ocurrencia de alguna de las variables del problema, el proceso de propagación de la evidencia consiste en calcular las probabilidades condicionadas de cada variable  $X_i \notin E$  dada la evidencia  $E = e$ , obteniéndose  $p(x_i|e)$  en el caso discreto o  $f(x_i|e)$  en el continuo, para todo  $X_i \notin E$ .

Obviamente, cuando no se dispone de evidencia, siendo  $E = \phi$ , las probabilidades condicionadas son las probabilidades marginales obtenidas al comienzo.

Considerando como se indicó anteriormente una Red Bayesiana Discreta, en el proceso de propagación de la evidencia se puede obtener la probabilidad condi-



cionada de interés  $p(x_i|e)$  mediante su propia definición, de forma que

$$p(x_i|e) = \frac{p(x_i, e)}{p(e)} \propto p(x_i, e) \quad (2.1)$$

siendo  $\frac{1}{p(e)}$  la constante de proporcionalidad.

Como existe una estructura de dependencia entre las variables del problema, es fundamental trabajar teniendo en cuenta las distintas relaciones de independencia que se tienen entre dichas variables. Así, la distribución de probabilidad conjunta puede obtenerse mediante la expresión de la definición de Red Bayesiana, en (1.1), donde la probabilidad conjunta viene dada por el producto de todas las probabilidades condicionadas dada la ocurrencia de sus padres en el DAG, es decir

$$p(x) = \prod_{i=1}^n p(x_i|pa(X_i))$$

Para realizar de forma eficiente el proceso de propagación de la evidencia, y por tanto los distintos cálculos asociados a la probabilidad conjunta y a la probabilidad condicionada, es fundamental trabajar con la estructura de dependencia que presenta la Red Bayesiana en el DAG. De no ser así, el proceso de propagación de la evidencia sería ineficiente ya que el tiempo de procesamiento necesario para realizar los cálculos crece exponencialmente a medida que aumenta el número de variables del problema.

### 2.3.2. Algoritmos de propagación de la evidencia

Existen diversos algoritmos de propagación de la evidencia en Redes Bayesianas y se pueden clasificar en algoritmos exactos y aproximados. Para los algoritmos de propagación exactos, se obtienen las probabilidades de los nodos de forma precisa y sin error. Los algoritmos de propagación aproximados<sup>2</sup> utilizan técnicas de simulación para obtener valores aproximados de las probabilidades y se aplican cuando los exactos son muy costosos o inaplicables.

Diversos autores han desarrollado algoritmos de propagación de la evidencia. Así, en Redes Bayesianas Discretas destacan Olmsted (1983) y Shachter (1988)

---

<sup>2</sup>Véase Saul, et al. (1996) y Jaakkola, et al. (1996)

que desarrollan un algoritmo basado en hacer las aristas del DAG reversibles, hasta que la respuesta a la pregunta probabilística pueda ser leída directamente del grafo, de forma que la probabilidad de interés asociada a cada arista reversible se obtiene tras la aplicación del Teorema de Bayes. Kim, et al. (1983) y Pearl (1986b) desarrollan el esquema de "paso-de-mensaje" que actualiza la distribución de probabilidad de cada nodo de la Red Bayesiana cuando se tiene evidencia acerca de una o más variables. Lauritzen, et al. (1988), Jensen, et al. (1990a), Jensen, et al. (1990b) y Dawid (1992) desarrollan un algoritmo basado en transformar el DAG inicial, que representa la Red Bayesiana, en un árbol donde cada nodo del árbol está formado por un subconjunto de variables de  $\mathbf{X}$ ; además, este algoritmo explota diversas propiedades matemáticas del árbol para realizar el proceso de propagación de la evidencia. Posteriormente, D'Ambrosio (1991) desarrolla un algoritmo de propagación de la evidencia que simplifica alguno de los cálculos del algoritmo especificado anteriormente, aunque sigue siendo el algoritmo descrito por Lauritzen, et al. (1988), Jensen, et al. (1990a), Jensen, et al. (1990b) y Dawid (1992) el más utilizado en Redes Bayesianas Discretas.

Para las Redes Bayesianas Gaussianas o Redes Bayesianas Mixtas, los algoritmos desarrollados utilizan resultados asociados a la independencia condicionada, para simplificar el proceso de inferencia. En dichas redes Normand, et al. (1992) y Lauritzen (1992), respectivamente, desarrollan dos algoritmos de propagación exactos; el primero basándose en la propagación en poliárboles y el segundo en la propagación en árboles de unión. Lauritzen, et al. (2001) desarrollan un algoritmo alternativo a Lauritzen (1992) en el que se estabiliza el esquema de los cálculos locales y Cowell (2005) presenta un esquema de propagación en Redes Bayesianas Gaussianas sobre el grafo descomponible asociado, describiendo un algoritmo para realizar los cálculos locales en dichos modelos que combina la aproximación de Lauritzen, et al. (2001) con algunos elementos descritos por Shachter, et al. (1989) para modelos gráficos Gaussianos.

A pesar de trabajar con las relaciones de independencia condicionada para simplificar el proceso de propagación de la evidencia, la propagación exacta en Redes Bayesianas es un problema NP-duro (Cooper, 1990). Incluso los algoritmos de propagación aproximados son también NP-duros, véase Dagum, et al. (1993). Esto es debido a la estructura gráfica de la Red Bayesiana que a pesar de ser un DAG, en la propagación de la evidencia se modifica dicha estructura gráfica pudiendo existir ciclos no dirigidos, que hacen intratable el proceso de propagación de la evidencia. No obstante en la mayoría de los casos particulares

se llega a una solución mediante un proceso eficiente.

En esta Memoria se trabaja con un algoritmo de propagación exacto para Redes Bayesianas Gaussianas que se detalla posteriormente en el Apartado 2.4.4 de la Sección 2.4.

En los siguientes puntos se muestra el desarrollo del proceso de propagación de la evidencia desde que Kim, et al. (1983) y Pearl (1986b) presentan su algoritmo basado en "paso-de-mensaje" pensado para la *propagación en poliárboles*, hasta el algoritmo más frecuentemente utilizado en la propagación en Redes Bayesianas Discretas, desarrollado por diversos autores como Lauritzen, et al. (1988), Jensen, et al. (1990a) y Dawid (1992), denominado *propagación en árboles de unión o universos del conocimiento* (Jensen, et al., 1990b).

El objetivo de este Apartado es conocer e implementar el proceso de propagación de la evidencia en Redes Bayesianas, pasando de una estructura gráfica sencilla de DAG, los poliárboles, a una estructura más compleja y general de DAG. A pesar de esto, se han introducido diversos métodos de propagación eficientes considerando otras estructuras gráficas características aunque el método de propagación en árboles de unión es el más generalizado.

Este proceso se muestra para un conjunto de variables discretas ya que fueron las Redes Bayesianas Discretas las primeras en describirse y presentan claramente la idea que subyace en el proceso de propagación de la evidencia basándose en el Teorema de Bayes.

A pesar de realizarse los procesos de propagación cuando se tiene evidencia acerca de alguna de las variables de la red, la propagación de la evidencia también se puede efectuar en una fase inicial del problema, cuando no se tiene evidencia de ninguna variable. En este caso se obtienen las probabilidades marginales de cada una de las variables que forman la red,  $p(x_i)$ , a partir de la definición de Red Bayesiana, con  $P$  dado por  $p(x_i|pa(X_i))$  para todo  $X_i$ .

### **Propagación en Poliárboles**

En este Punto se muestra la idea inicial de propagación de la evidencia presentada por Kim, et al. (1983) y Pearl (1986b) basada en el envío de un mensaje a

través de la red, siendo dicho mensaje el resultado de algunos cálculos locales. Además, se introduce el algoritmo asociado de propagación, que destaca por tener una complejidad lineal en el número de nodos y aristas que componen la Red Bayesiana. El grafo que representa la Red Bayesiana en estudio es un poliárbol.

Como se ha visto en el Capítulo 1, un poliárbol es una estructura simple que se caracteriza por tener un único camino entre cada par de nodos, pudiendo existir nodos con más de un padre. Por tanto, cada nodo  $X_i$  divide al poliárbol en dos poliárboles inconexos: uno que contiene a sus padres y a los nodos que son accesibles desde  $X_i$  a través de sus padres y otro que contiene a sus hijos y a los nodos que son accesibles desde  $X_i$  a través de sus hijos.

Sea  $\mathbf{E} = \mathbf{e}$  la evidencia de la Red Bayesiana. Teniendo en cuenta la estructura del poliárbol, para calcular la probabilidad de interés  $p(x_i|\mathbf{e})$  para todo  $x_i$  de  $X_i \notin \mathbf{E}$ , se descompone la evidencia  $\mathbf{E}$  en dos subconjuntos disjuntos, de forma que cada subconjunto está contenido en uno de los poliárboles obtenidos al separar el poliárbol original por el nodo  $X_i$ , quedando

$\mathbf{E}_i^+$ : subconjunto de  $\mathbf{E}$  accesible desde  $X_i$  a través de sus padres

$\mathbf{E}_i^-$ : subconjunto de  $\mathbf{E}$  accesible desde  $X_i$  a través de sus hijos

Donde  $\mathbf{E} = \mathbf{E}_i^+ \cup \mathbf{E}_i^-$ .

Aplicando la definición de probabilidad condicionada (2.1) se tiene que

$$p(x_i|\mathbf{e}) = p(x_i|\mathbf{e}_i^+, \mathbf{e}_i^-) = \frac{p(x_i, \mathbf{e}_i^+, \mathbf{e}_i^-)}{p(\mathbf{e}_i^+, \mathbf{e}_i^-)} = \frac{p(\mathbf{e}_i^+, \mathbf{e}_i^-|x_i)p(x_i)}{p(\mathbf{e}_i^+, \mathbf{e}_i^-)}$$

Como en el poliárbol inicial el nodo  $X_i$  separa  $\mathbf{E}_i^+$  de  $\mathbf{E}_i^-$ , entonces  $\mathbf{E}_i^+ \perp\!\!\!\perp \mathbf{E}_i^- | X_i$ , siendo la probabilidad de interés

$$p(x_i|\mathbf{e}) = \frac{p(\mathbf{e}_i^+|x_i)p(\mathbf{e}_i^-|x_i)p(x_i)}{p(\mathbf{e}_i^+, \mathbf{e}_i^-)} = \frac{p(\mathbf{e}_i^+, x_i)p(\mathbf{e}_i^-|x_i)}{p(\mathbf{e}_i^+, \mathbf{e}_i^-)}$$

Considerando  $k = \frac{1}{p(\mathbf{e}_i^+, \mathbf{e}_i^-)}$  la constante de normalización y siendo  $\lambda_i(x_i) = p(\mathbf{e}_i^-|x_i)$  la probabilidad de tener evidencia procedente de los hijos de  $X_i$  y  $\rho_i(x_i) = p(\mathbf{e}_i^+, x_i)$  la probabilidad conjunta asociada a la evidencia procedente de los padres de  $X_i$ , se tiene que

$$p(x_i|\mathbf{e}) = k\lambda_i(x_i)\rho_i(x_i) = k\beta_i(x_i) \propto \beta_i(x_i)$$

donde  $\beta_i(x_i) = \lambda_i(x_i)\rho_i(x_i)$

Las funciones  $\lambda_i(x_i)$  y  $\rho_i(x_i)$  son los mensajes de envío entre pares de nodos, por tanto, para su cálculo se considera la siguiente situación mostrada en la Figura 2.8 donde el nodo  $X_i$  tiene  $p$  padres y  $h$  hijos, siendo  $\mathbf{U} = \{U_1, \dots, U_p\}$  el conjunto de padres de  $X_i$  y  $\mathbf{L} = \{L_1, \dots, L_h\}$  el conjunto de hijos de  $X_i$ .

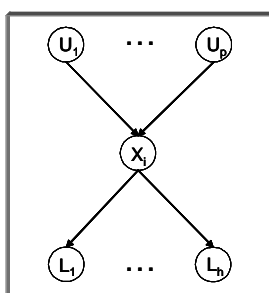


Figura 2.8. Padres e hijos del nodo  $X_i$

De esta forma, el conjunto  $\mathbf{E}_i^+$  de evidencias accesibles desde  $X_i$  a través de sus padres y el conjunto  $\mathbf{E}_i^-$  de evidencias accesibles desde  $X_i$  a través de sus hijos, pueden venir dados como

$$\begin{aligned}\mathbf{E}_i^+ &= \{E_{U_1 X_i}^+, \dots, E_{U_p X_i}^+\} \\ \mathbf{E}_i^- &= \{E_{X_i L_1}^-, \dots, E_{X_i L_h}^-\}\end{aligned}$$

donde  $E_{U_j X_i}^+$  es el subconjunto de  $\mathbf{E}_i^+$  contenido en el grafo asociado al nodo  $U_j$  cuando se elimina la arista  $U_j \longrightarrow X_i$  y  $E_{X_i L_j}^-$  es el subconjunto de  $\mathbf{E}_i^-$  contenido en el grafo asociado al nodo  $L_j$  cuando se elimina la arista  $X_i \longrightarrow L_j$ .

Para obtener  $\rho_i(x_i)$  se considera una realización de los padres del nodo  $X_i$  tal que  $\mathbf{u} = \{u_1, \dots, u_p\}$  y se trabaja con la relación de independencia condicionada que se tiene entre  $\{U_j, E_{U_j X_i}^+\}$  y  $\{U_k, E_{U_k X_i}^+\} \quad \forall j \neq k$ .

$$\begin{aligned}
\rho_i(x_i) &= p(\mathbf{e}_i^+, x_i) = \sum_{\mathbf{u}} p(\mathbf{u} \cup \mathbf{e}_i^+, x_i) = \sum_{\mathbf{u}} p(x_i | \mathbf{u} \cup \mathbf{e}_i^+) p(\mathbf{u} \cup \mathbf{e}_i^+) = \\
&= \sum_{\mathbf{u}} p(x_i | \mathbf{u} \cup \mathbf{e}_i^+) p(\mathbf{u} \cup e_{U_1 X_i}^+ \cup \dots \cup e_{U_p X_i}^+) = \\
&= \sum_{\mathbf{u}} p(x_i | \mathbf{u} \cup \mathbf{e}_i^+) \prod_{j=1}^p p(u_j \cup e_{U_j X_i}^+) = \\
&= \sum_{\mathbf{u}} p(x_i | \mathbf{u} \cup \mathbf{e}_i^+) \prod_{j=1}^p \rho_{U_j X_i}(u_j) \tag{2.2}
\end{aligned}$$

siendo  $\rho_{U_j X_i}(u_j) = p(u_j \cup e_{U_j X_i}^+)$  el mensaje  $\rho$  que  $X_i$  recibe de su padre  $U_j$ . Por tanto, la función  $\rho_i(x_i)$  se puede obtener en cuanto el nodo  $X_i$  haya recibido los mensajes  $\rho$  de todos sus padres.

Si  $U_j$  fuese una variable evidencial, con  $u_j = e_j$ , entonces el mensaje que envía  $U_j$  a su hijo  $X_i$  es

$$\rho_{U_j X_i}(u_j) = \begin{cases} 1 & \text{si } u_j = e_j \\ 0 & \text{si } u_j \neq e_j \end{cases}$$

Para calcular la función  $\lambda_i(x_i)$  se tiene en cuenta que los elementos de  $\mathbf{E}_i^-$  están  $d$ -separados dado  $X_i$ . Por tanto se obtiene  $\lambda_i(x_i)$  como

$$\lambda_i(x_i) = p(\mathbf{e}_i^- | x_i) = p(e_{X_i L_1}^-, \dots, e_{X_i L_h}^- | x_i) = \prod_{j=1}^h p(e_{X_i L_j}^- | x_i) = \prod_{j=1}^h \lambda_{L_j X_i}(x_i) \tag{2.3}$$

donde  $\lambda_{X_i L_j}(x_i) = p(e_{X_i L_j}^- | x_i)$  es el mensaje  $\lambda$  que  $X_i$  recibe de su hijo  $L_j$ . Por lo que  $\lambda_i(x_i)$  se puede calcular en cuanto  $X_i$  haya recibido los mensajes  $\lambda$  de todos sus hijos.

La probabilidad de interés viene dada por la siguiente expresión

$$p(x_i | \mathbf{e}) \propto \beta_i(x_i) = \left( \prod_{j=1}^h \lambda_{L_j X_i}(x_i) \right) \left( \sum_{\mathbf{u}} p(x_i | \mathbf{u} \cup \mathbf{e}_i^+) \prod_{j=1}^p \rho_{U_j X_i}(u_j) \right)$$

Una vez que  $X_i$  recibe todos los mensajes de sus padres y de sus hijos, es fundamental conocer el mensaje que el nodo  $X_i$  envía a un hijo cualquiera  $L_j$ , es decir  $\rho_{X_i L_j}(x_i)$ . Para ello se considera el nodo  $X_i$  y su hijo  $L_j$ . Teniendo en cuenta que la evidencia accesible desde  $L_j$  a través de su padre  $X_i$  viene dada por la evidencia accesible desde  $X_i$  a través de sus padres ( $\mathbf{E}_i^+$ ) y por la evidencia accesible desde  $X_i$  a través del resto de sus hijos  $L_k$  con  $k \neq j$ , siendo

$$E_{X_i L_j}^+ = \mathbf{E}_i^+ \cup_{k \neq j} E_{X_i L_k}^-$$

se puede calcular el mensaje que  $X_i$  envía a su hijo  $L_j$  como  $\rho_{X_i L_j}(x_i)$ , tal que

$$\begin{aligned} \rho_{X_i L_j}(x_i) &= p(x_i \cup e_{X_i L_j}^+) = p\left(x_i \cup \left(\mathbf{e}_i^+ \cup_{k \neq j} e_{X_i L_k}^- \right)\right) = \\ &= p\left(\mathbf{e}_i^+ | x_i \cup_{k \neq j} e_{X_i L_k}^- \right) p\left(x_i \cup_{k \neq j} e_{X_i L_k}^- \right) = \\ &= p(\mathbf{e}_i^+ | x_i) p\left(\bigcup_{k \neq j} e_{X_i L_k}^- | x_i\right) p(x_i) \\ &\propto p(x_i | \mathbf{e}_i^+) \prod_{k \neq j} p(e_{X_i L_k}^- | x_i) \propto \rho_i(x_i) \prod_{k \neq j} \lambda_{L_k X_i}(x_i) \quad (2.4) \end{aligned}$$

En este caso, el mensaje  $\rho_{X_i L_j}(x_i)$  puede ser enviado del nodo  $X_i$  a su hijo  $L_j$  tan pronto se haya calculado su función  $\rho_i(x_i)$  y recibido los mensajes  $\lambda$  del resto de sus hijos, es decir, en cuanto  $X_i$  haya recibido todos los mensajes de sus padres y del resto de sus hijos.

Del mismo modo, si  $X_i$  es un nodo con evidencia, entonces el mensaje de  $X_i$  a  $L_j$  es

$$\rho_{X_i L_j}(x_i) = \begin{cases} 1 & \text{si } x_i = e_i \\ 0 & \text{si } x_i \neq e_i \end{cases}$$

Finalmente para calcular el mensaje  $\lambda$  que  $L_j$  manda a su padre  $X_i$  ( $\lambda_{L_j X_i}(x_i)$ ) se considera el conjunto de todos los padres de  $L_j$  distintos de  $X_i$  como  $\mathbf{V} =$

$\{V_1, \dots, V_q\}$ , de forma que el nodo  $L_j$  tiene  $q + 1$  padres. Entonces la evidencia accesible desde  $X_i$  a través de su hijo  $L_j$  es tal que

$$e_{X_i L_j}^- = e_{L_j}^- \cup e_{V L_j}^+$$

siendo  $e_{V L_j}^+$  la evidencia que  $L_j$  obtiene a través de todos sus padres, excepto del nodo  $X_i$ .

De esta forma,

$$\begin{aligned} \lambda_{L_j X_i}(x_i) &= p(e_{X_i L_j}^- | x_i) = \sum_{l_j, \mathbf{v}} p(l_j, \mathbf{v}, e_{X_i L_j}^- | x_i) = \sum_{l_j, \mathbf{v}} p(l_j, \mathbf{v}, e_{L_j}^-, e_{V L_j}^+ | x_i) = \\ &= \sum_{l_j, \mathbf{v}} p(e_{L_j}^- | l_j, \mathbf{v}, e_{V L_j}^+, x_i) p(l_j | \mathbf{v}, e_{V L_j}^+, x_i) p(\mathbf{v}, e_{V L_j}^+ | x_i) = \\ &= \sum_{l_j} p(e_{L_j}^- | l_j) \sum_{\mathbf{v}} p(l_j | \mathbf{v}, x_i) p(\mathbf{v}, e_{V L_j}^+) = \\ &= \sum_{l_j} \lambda_{L_j}(l_j) \sum_{\mathbf{v}} p(l_j | pa(L_j)) \prod_{k=1}^q \rho_{V_k L_j} p(v_k) \end{aligned} \quad (2.5)$$

Por tanto, el nodo  $X_i$  puede enviar el mensaje  $\lambda$  a su padre  $U_j$  ( $\lambda_{X_i U_j}(u_j)$ ) en cuanto haya calculado la función  $\lambda_i(x_i)$  y haya recibido los mensajes  $\rho$  del resto de los padres, es decir, en cuanto  $X_i$  haya recibido todos los mensajes de sus hijos y del resto de sus padres.

A continuación se introduce un algoritmo, basado en lo anterior, para calcular las probabilidades  $p(x_i | \mathbf{e})$  para todos sus nodos no evidenciales  $X_i \notin \mathbf{E}$ . En dicho algoritmo, para no tener problema con los mensajes que se envían, se generaliza la notación dada a cualquier variable  $X_i$  del conjunto de variables  $\mathbf{X}$ , de forma que  $X_i$  puede ser variable evidencial  $X_i \in \mathbf{E}$  y variable no evidencial, tal que  $X_i \notin \mathbf{E}$ . Para profundizar en la idea y en el algoritmo de propagación en poliárboles véase Kim, et al. (1983), Pearl (1986b) o Castillo, et al. (1997a).



**Algoritmo de propagación en poliárboles**

ENTRADA: La Red Bayesiana  $(\mathcal{D}, P)$  donde  $\mathcal{D}$  es un poliárbol,  $\mathbf{X} = \{X_1, \dots, X_n\}$  es el conjunto de variables de la red y  $\mathbf{E} = \mathbf{e}$  es un conjunto de evidencias.

SALIDA: Las distribuciones de probabilidad condicionadas  $p(x_i|\mathbf{e})$  para todo  $X_i \notin \mathbf{E}$ .

Paso 1. Inicializar los mensajes de envío para todas las variables evidenciales  $X_i \in \mathbf{E}$  tal que

$$\begin{aligned}\rho_i(x_i) &= \begin{cases} 1 & \text{si } x_i = e_i \\ 0 & \text{si } x_i \neq e_i \end{cases} \\ \lambda_i(x_i) &= \begin{cases} 1 & \text{si } x_i = e_i \\ 0 & \text{si } x_i \neq e_i \end{cases}\end{aligned}$$

Paso 2. Para las variables no evidenciales  $X_i \notin \mathbf{E}$  que no tengan padres en el DAG, inicializar el mensaje  $\rho_i(x_i)$  como

$$\rho_i(x_i) = p(x_i)$$

Paso 3. Para las variables no evidenciales  $X_i \notin \mathbf{E}$  que no tengan hijos en el DAG, inicializar el mensaje  $\lambda_i(x_i)$  como

$$\lambda_i(x_i) = 1$$

Paso 4. Para cada variable no evidencial  $X_i \notin E$  calcular

(a) Si  $X_i$  a recibido los mensajes  $\rho$  de todos sus padres entonces  $\rho_i(x_i)$  se calcula como en (2.2), tal que

$$\rho_i(x_i) = \sum_{\mathbf{u}} p(x_i|\mathbf{u} \cup \mathbf{e}_i^+) \prod_{j=1}^p \rho_{U_j X_i}(u_j)$$

(b) Si  $X_i$  ha recibido los mensajes  $\lambda$  de todos sus hijos, entonces  $\lambda_i(x_i)$  se calcula mediante la siguiente expresión, presentada en (2.3)

$$\lambda_i(x_i) = \prod_{j=1}^h \lambda_{L_j X_i}(x_i)$$

(c) Si ya se ha calculado  $\rho_i(x_i)$ , entonces, para cada hijo  $L_j$  de  $X_i$ , tal que  $X_i$  ya ha recibido los mensajes  $\lambda$  del resto de sus hijos, calcular y enviar el mensaje  $\rho_{X_i L_j}(x_i)$  apoyándose en (2.4).

Si  $X_i$  ha recibido los mensajes  $\lambda$  de todos sus hijos, entonces ya puede enviar todos los mensajes  $\rho$ .

(d) Si ya se ha calculado  $\lambda_i(x_i)$ , entonces, para cada padre  $U_j$  de  $X_i$ , tal que  $X_i$  ya ha recibido los mensajes  $\rho$  del resto de sus padres, calcular y enviar el mensaje  $\lambda_{X_i U_j}(u_i)$  apoyándose en (2.5).

De igual forma, si  $X_i$  ha recibido los mensajes  $\rho$  de todos sus padres, entonces ya puede enviar todos los mensajes  $\lambda$ .

Paso 5. Repetir el Paso 4 tantas veces sea necesario hasta obtener las funciones  $\rho$  y  $\lambda$  de todas las variables no evidenciales  $X_i \notin \mathbf{E}$ , de forma que no exista ningún nuevo mensaje en una iteración completa.

Paso 6. Calcular  $\beta_i(x_i)$  para cada variable no evidencial  $X_i \notin \mathbf{E}$ , siendo  $\beta_i(x_i) \propto p(x_i|\mathbf{e})$ , teniendo en cuenta que

$$\beta_i(x_i) = \lambda_i(x_i)\rho_i(x_i)$$

Paso 7. Calcular  $p(x_i|\mathbf{e})$  para cada variable no evidencial  $X_i \notin \mathbf{E}$ , normalizando la función  $\beta_i(x_i)$ , de forma que

$$p(x_i|\mathbf{e}) = \frac{\beta_i(x_i)}{\sum_{x_i} \beta_i(x_i)}$$

Cuando solo se tenga interés en una variable  $X_i$  del problema, siendo dicha variable la variable objetivo o de interés, no es necesario realizar todo el proceso de propagación hasta encontrar las funciones  $\rho$  y  $\lambda$  ya que el objetivo es conocer  $\rho_i(x_i)$  y  $\lambda_i(x_i)$ .

### Propagación en árboles de unión

En este Punto se introduce la idea de transformar la estructura gráfica del DAG en un árbol de unión de ciclados, para obtener de forma eficiente la distribución

de probabilidad de las variables del problema, cuando se tiene evidencia acerca de la ocurrencia de alguna de las mismas.

Al igual que en el método de propagación en poliárboles, el algoritmo que se presenta se basa en el envío de mensajes, pero ahora la estructura gráfica es la de un árbol de unión de ciclados construido a partir del DAG, que representa la Red Bayesiana en estudio.

Como se ha expuesto en el Capítulo 1, la estructura del DAG que representa la red del problema puede manipularse hasta obtener un árbol de unión de ciclados. Para ello, se moraliza el grafo no dirigido asociado al DAG. Posteriormente, se triangula el grafo resultante, obteniendo así el árbol de unión de ciclados<sup>3</sup>. En la Figura 2.9 se muestra un DAG y su transformación en un árbol de unión de ciclados.

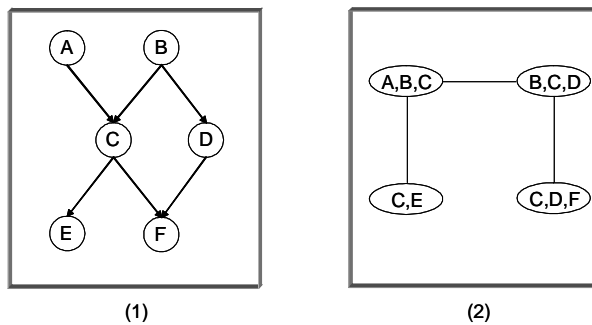


Figura 2.9. DAG (1) y árbol de unión asociado (2)

Considérese un árbol de unión de ciclados en el que cada nodo representa un ciclado  $C = \{C_1, \dots, C_k\}$ . El conjunto de potenciales asociados a los ciclados del grafo viene dado por  $\{\psi_1(c_1), \dots, \psi_k(c_k)\}$ , siendo

$$\psi_i(c_i) = \prod_{x_i \in C_i} p(x_i | pa(X_i))$$

Sean  $S_{ij}$ , o  $S_{ji}$ , el conjunto separador de los ciclados  $C_i$  y  $C_j$ , dado por  $S_{ij} = C_i \cap C_j$ . En la Figura 2.10 se representa el árbol de unión (2) de la Figura 2.9 con los separadores de los ciclados.

<sup>3</sup>Para ampliar la información acerca del proceso de manipulación del DAG hasta obtener un árbol de unión, véase el Apartado 1.2.5. Estructuras gráficas de interés.

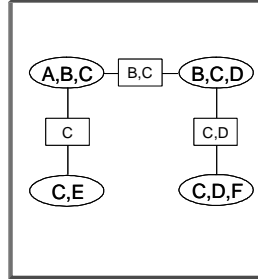


Figura 2.10. Árbol de unión con el conjunto de separadores

Teniendo en cuenta la estructura del árbol de unión, el proceso de propagación de la evidencia, consiste en realizar los cálculos locales necesarios en cada ciclado enviando y propagando mensajes entre los ciclados vecinos del árbol de unión, de manera que se efectúe la propagación de la evidencia de forma análoga a la propagación en poliárboles.

Sea  $\mathbf{E} = \mathbf{e}$  el conjunto de evidencias. Se sabe que la probabilidad de interés  $p(x_i|\mathbf{e})$  de un nodo cualquiera  $X_i \notin \mathbf{E}$ , se puede obtener aplicando el Teorema de Bayes, considerando las funciones potenciales de los ciclados. Aunque el primer paso consiste en que las funciones potenciales *absorban la evidencia*, de forma que si una variable  $X_i$  es evidencial, entonces se asocia dicha variable a uno de los ciclados  $C_i$  a los que pertenece y se modifica la función potencial de dicho ciclado, de manera que

$$\psi_i^*(c_i) = \begin{cases} 0 & \text{si algún valor de } c_i \text{ no es consistente con } \mathbf{e} \\ \psi_i(c_i) & \text{en otro caso} \end{cases}$$

El resto de potenciales no se modifican.

En adelante, se denotan todos los potenciales de igual forma,  $\psi_i(c_i)$ , aunque algunos se hayan modificado en el proceso de absorción de la evidencia.

El paso siguiente, consiste en calcular los mensajes que se envían entre ciclados vecinos en el árbol de unión, trabajando con el producto de las funciones potenciales de los vecinos.

Teniendo en cuenta la información de los mensajes recibidos por cada ciclado, se calcula la distribución de probabilidad del separador y del ciclado, como la distribución marginal de la probabilidad conjunta, quedando dicha probabilidad

en función de los mensajes que se envían entre ciclados. Posteriormente, se marginaliza la distribución de probabilidad del ciclado  $p(c_i)$  que contenga a la variable  $X_i$ , sobre el resto de variables del ciclado, para obtener la probabilidad de interés  $p(x_i|\mathbf{e})$ ; generalmente se considera en este proceso el ciclado de menor tamaño que contiene a la variable  $X_i$ .

Sea  $C_i$  un ciclado cualquiera del árbol de unión y sean  $\{B_1, \dots, B_q\}$  sus ciclados vecinos. Si se elimina la arista que une los ciclados  $C_i$  y  $B_j$  se obtienen dos subgrafos disjuntos, uno en el que están  $C_i$  y sus ciclados vecinos, excepto  $B_j$  y otro en el que está  $B_j$ . Considérese en esta situación los siguientes conjuntos:

$C_{ij}$  : conjunto de ciclados asociados a  $C_i$ . Al eliminarse la arista entre  $C_i$  y  $B_j$  se tiene que  $B_j \notin C_{ij}$ .

$X_{ij}$ : conjunto de todos los nodos contenidos en  $C_{ij}$ .

De forma que los conjuntos  $C_{ij}$  y  $C_{ji}$  son complementarios, al igual que  $X_{ij}$  y  $X_{ji}$ , siendo el conjunto de todas las variables del problema tal que  $\mathbf{X} = C_{ij} \cup C_{ji} = X_{ij} \cup X_{ji}$ .

Con el grafo separado, obtenido tras eliminar la arista  $C_i - B_j$ , se procede al cálculo de las distribuciones de probabilidad conjuntas tanto de cada ciclado, como de cada separador. Para ello, se comienza con la distribución de probabilidad de los separadores.

Para obtener la distribución de probabilidad de un conjunto separador  $S_{ij}$ , se ha de tener en cuenta que

$$\mathbf{X} \setminus S_{ij} = (X_{ij} \cup X_{ji}) \setminus S_{ij} = (X_{ij} \setminus S_{ij}) \cup (X_{ji} \setminus S_{ij}) = R_{ij} \cup R_{ji}$$

donde  $R_{ij} = X_{ij} \setminus S_{ij}$  son los residuos dados por el conjunto de variables contenidas en el subárbol asociado a  $C_i$  pero que no están en  $B_j$  cuando se elimina la arista  $C_i - B_j$ .

Si un nodo está contenido en dos ciclados distintos, entonces tiene que estar también contenido en todos los ciclados que haya en el camino que une ambos ciclados, así, los nodos comunes entre  $X_{ij}$  y  $X_{ji}$  tienen que estar contenidos en el separador  $S_{ij}$ , por tanto los residuos  $R_{ij}$  y  $R_{ji}$  son subconjuntos disjuntos.

Trabajando con esta idea se obtiene la distribución de probabilidad conjunta de  $S_{ij}$  como

$$\begin{aligned}
 p(s_{ij}) &= \sum_{x \setminus s_{ij}} \prod_{l=1}^k \psi_l(c_l) = \sum_{r_{ij} \cup r_{ji}} \prod_{l=1}^k \psi_l(c_l) = \\
 &= \left( \sum_{r_{ij}} \prod_{c_l \in C_{ij}} \psi_l(c_l) \right) \left( \sum_{r_{ji}} \prod_{c_l \in C_{ji}} \psi_l(c_l) \right) = \\
 &= M_{ij}(s_{ij}) M_{ji}(s_{ij})
 \end{aligned}$$

donde  $M_{ij}(s_{ij}) = \sum_{r_{ij}} \prod_{c_l \in C_{ij}} \psi_l(c_l)$  es el mensaje que envía el ciclado  $C_i$  al ciclado vecino  $B_j$  y  $M_{ji}(s_{ij}) = \sum_{r_{ji}} \prod_{c_l \in C_{ji}} \psi_l(c_l)$  es el mensaje que envía el ciclado  $B_j$  a  $C_i$ .

Por tanto, la distribución de probabilidad conjunta del separador  $S_{ij}$  viene dada por el producto de los dos mensajes que se envían entre los ciclados  $C_i$  y  $B_j$ . Nótese que la información necesaria para calcular cada uno de los mensajes, está en uno de los dos subgrafos obtenidos al eliminar la arista  $C_i-B_j$ , como son subgrafos separados del árbol de unión de ciclados, estos mensajes propagan la información de una parte del grafo en la otra parte. Además los mensajes pueden calcularse de forma independiente, lo que facilita la implementación del algoritmo.

Para obtener la distribución de probabilidad del ciclado  $C_i$ , se procede de igual manera que con los separadores, de forma que

$$\mathbf{X} \setminus C_i = \left( \bigcup_{l=1}^q X_{li} \right) \setminus C_i = \bigcup_{l=1}^q (X_{li} \setminus C_i) = \bigcup_{l=1}^q R_{li}$$

siendo  $R_{li} = X_{li} \setminus C_i$ , ya que según la propiedad de los árboles de unión que dice que cada variable de  $X_{li}$  que está contenida en  $C_i$  también está contenida en  $S_{li}$ , de manera que  $R_{li} = X_{li} \setminus C_i = X_{li} \setminus S_{li}$ .

De esta forma, se puede calcular la distribución de probabilidad del ciclado

$C_i$  como

$$\begin{aligned}
p(c_i) &= \sum_{x \setminus c_i} \prod_{j=1}^k \psi_j(c_j) = \psi_i(c_i) \sum_{x \setminus c_i} \prod_{j \neq i} \psi_j(c_j) = \\
&= \psi_i(c_i) \sum_{r_{1i} \cup \dots \cup r_{qi} \neq i} \prod \psi_j(c_j) = \\
&= \psi_i(c_i) \left( \sum_{r_{1i}} \prod_{c_l \in C_{1i}} \psi_l(c_l) \right) \dots \left( \sum_{r_{qi}} \prod_{c_l \in C_{qi}} \psi_l(c_l) \right) = \\
&= \psi_i(c_i) \prod_{j=1}^q M_{ji}(s_{ji})
\end{aligned}$$

donde  $M_{ji}(s_{ji})$  es el mensaje que envía el ciclado  $B_j$  al ciclado vecino  $C_i$ .

Por tanto, la distribución de probabilidad del ciclado  $p(c_i)$  se puede calcular en cuanto  $C_i$  haya recibido todos los mensajes de los ciclados vecinos. Y la distribución de probabilidad conjunta de un ciclado cualquiera se obtiene cuando se han calculado y enviado todos los mensajes.

Si  $p(c_i)$  no acumula una probabilidad total de 1, entonces se divide cada una de las probabilidades de todos los posibles valores entre la suma que acumula  $p(c_i)$ , de forma que la probabilidad del ciclado valga 1.

Se puede simplificar el cálculo de los mensajes  $M_{ij}(s_{ij})$ , teniendo en cuenta que

$$X_{ij} \setminus S_{ij} = (C_i \setminus S_{ij}) \cup \left( \bigcup_{l \neq j} X_{li} \setminus S_{li} \right).$$

Así, se obtiene que el mensaje que envía el ciclado  $C_i$  al ciclado vecino  $B_j$  como

$$\begin{aligned}
M_{ij}(s_{ij}) &= \sum_{x_{ij} \setminus s_{ij}} \prod_{c_s \in C_{ij}} \psi_s(c_s) = \sum_{c_i \setminus s_{ij}} \sum_{x_{li} \setminus s_{li}, l \neq j} \prod_{c_s \in C_{ij}} \psi_s(c_s) = \\
&= \sum_{c_i \setminus s_{ij}} \psi_i(c_i) \prod_{l \neq j} \sum_{x_{li} \setminus s_{li}} \prod_{c_s \in C_{li}} \psi_s(c_s) = \\
&= \sum_{c_i \setminus s_{ij}} \psi_i(c_i) \prod_{l \neq j} M_{li}(s_{li}). \tag{2.6}
\end{aligned}$$

Por tanto, el mensaje  $M_{ij}(s_{ij})$  que envía el ciclado  $C_i$  a su vecino  $B_j$  se puede calcular en el momento que  $C_i$  haya recibido todos los mensajes  $M_{li}(s_{li})$  de sus ciclados vecinos.

Como se ha comentado anteriormente, cuando se tiene la distribución de probabilidad conjunta de todos los ciclados, se marginaliza la distribución de probabilidad del ciclado que contenga a  $X_i$  para obtener la probabilidad de interés  $p(x_i|\mathbf{e})$ . Cuando el nodo  $X_i$  esté contenido en más de un ciclado, se puede tomar cualquier ciclado para su obtención, aunque para reducir cálculos es más adecuado tomar el ciclado de menor tamaño, como se hace en el algoritmo que se presenta a continuación. Dicho algoritmo sirve para obtener la probabilidad de interés  $p(x_i|\mathbf{e})$ , de todas las variables no evidenciales, cuando se introducen los datos de una Red Bayesiana y se puede obtener el árbol de unión asociado al DAG que describe la red.

### Algoritmo de propagación en árboles de unión

ENTRADA: La Red Bayesiana  $(\mathcal{D}, P)$ , siendo  $\mathbf{X} = \{X_1, \dots, X_n\}$  el conjunto de variables de la red y  $\mathbf{E} = \mathbf{e}$  el conjunto de evidencias. El árbol de unión asociado al DAG  $\mathcal{D}$ , donde cada nodo es un ciclado  $C_i$ , siendo  $C = \{C_1, \dots, C_k\}$ .

SALIDA: Las distribuciones de probabilidad condicionadas  $p(x_i|\mathbf{e})$  para todo  $X_i \notin \mathbf{E}$ .

Paso 1. Para cada ciclado  $C_i$ , calcular

$$\psi_i(c_i) = \prod_{x_i \in C_i} p(x_i | pa(X_i))$$

Paso 2. Introducir la evidencia en las funciones potenciales mediante el proceso de absorción de la evidencia, de forma que cada variable evidencial se asigna a uno de los ciclados  $C_i$  al que pertenece, definiéndose la función potencial de  $C_i$  como

$$\psi_i^*(c_i) = \begin{cases} 0 & \text{si algún valor de } c_i \text{ no es consistente con } \mathbf{e} \\ \psi_i(c_i) & \text{en otro caso} \end{cases}$$



Para el resto de ciclados no es necesario realizar ningún cambio sobre su función potencial  $\psi_i(c_i)$ .

Paso 3. Para todos los ciclados, con  $i = 1, \dots, k$ , se calcula el mensaje  $M_{ij}(s_{ij})$  que el ciclado  $C_i$  envía a su vecino  $B_j$  como en (2.6), tal que

$$M_{ij}(s_{ij}) = \sum_{c_i \setminus s_{ij}} \psi_i(c_i) \prod_{l \neq j} M_{li}(s_{li})$$

En este paso se pueden dar tres situaciones distintas:

- (a) Que el ciclado  $C_i$  haya recibido los mensajes de todos sus vecinos. En este caso  $C_i$  puede calcular y enviar los mensajes a todos sus vecinos.
- (b) Que el ciclado  $C_i$  haya recibido los mensajes de todos los vecinos excepto de  $B_j$ . En este caso  $C_i$  solo puede calcular y enviar su mensaje al ciclado  $B_j$ .
- (c) Que el ciclado  $C_i$  no haya recibido los mensajes de dos o más vecinos. En este caso todavía no se puede calcular ningún mensaje.

Paso 4. Repetir el Paso 3 hasta que no se obtenga ningún mensaje nuevo.

Paso 5. Para cada ciclado  $C_i$ , calcular la distribución de probabilidad del ciclado, mediante

$$p(c_i) = \psi_i(c_i) \prod_k M_{ki}(s_{ki})$$

Paso 6. Para cada variable  $X_i$  de la red, calcular la probabilidad condicionada de interés, utilizando la siguiente expresión

$$p(x_i | \mathbf{e}) = \sum_{c_l \setminus x_i} p(c_l)$$

siendo  $C_l$  el ciclado de menor tamaño que contiene a  $X_i$ .

Existen muchas variaciones sobre este algoritmo que reducen los cálculos o muestran soluciones para DAGs en los que es difícil encontrar el árbol de unión asociado, aunque la popularidad del mismo hace que siga siendo éste un algoritmo básico de propagación de la evidencia en Redes Bayesianas Discretas.

El ejemplo que se muestra a continuación, introducido por Castillo, et al. (1997a), ilustra el algoritmo presentado. En el Ejemplo 2.7, se realiza el algoritmo en una fase inicial del problema, donde se busca conocer las probabilidades

asociadas a los ciclados del grafo que representa la Red Bayesiana antes de introducir la evidencia en la red. Tras obtenerse dichas probabilidades, es posible calcular la probabilidad marginal de cada una de las variables de la red, marginalizando la probabilidad del ciclado. En el Ejemplo 2.8 se introduce evidencia acerca de dos variables de la red y se realiza el algoritmo para propagar dicha evidencia, con el fin de determinar las probabilidades de interés de cada variable  $X_i \notin \mathbf{E}$  dada la evidencia,  $p(x_i|\mathbf{e})$ .

**Ejemplo 2.7**

Considérese la siguiente Red Bayesiana Discreta definida por el DAG  $\mathcal{D}$  de la Figura 2.11 y el conjunto de probabilidades condicionadas  $p(x_i|pa(X_i))$  de la Tabla 2.6.

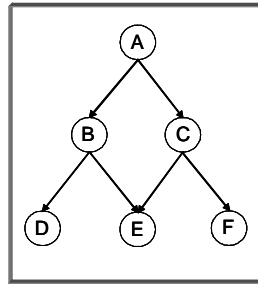


Figura 2.11. DAG de la Red Bayesiana descrita para el Ejemplo 2.7

|                               |   |
|-------------------------------|---|
| $P(A = 1) = 0.7$              | $\Rightarrow P(A = 0) = 0.3$              |
| $P(B = 1 A = 1) = 0.9$        | $\Rightarrow P(B = 0 A = 1) = 0.1$        |
| $P(B = 1 A = 0) = 0.6$        | $\Rightarrow P(B = 0 A = 0) = 0.4$        |
| $P(C = 1 A = 1) = 0.5$        | $\Rightarrow P(C = 0 A = 1) = 0.5$        |
| $P(C = 1 A = 0) = 0.8$        | $\Rightarrow P(C = 0 A = 0) = 0.2$        |
| $P(E = 1 B = 1, C = 1) = 0.8$ | $\Rightarrow P(E = 0 B = 1, C = 1) = 0.2$ |
| $P(E = 1 B = 1, C = 0) = 0.3$ | $\Rightarrow P(E = 0 B = 1, C = 0) = 0.7$ |
| $P(E = 1 B = 0, C = 1) = 0.5$ | $\Rightarrow P(E = 0 B = 0, C = 1) = 0.5$ |
| $P(E = 1 B = 0, C = 0) = 0.6$ | $\Rightarrow P(E = 0 B = 0, C = 0) = 0.4$ |
| $P(D = 1 B = 1) = 0.8$        | $\Rightarrow P(D = 0 B = 1) = 0.2$        |
| $P(D = 1 B = 0) = 0.7$        | $\Rightarrow P(D = 0 B = 0) = 0.3$        |
| $P(F = 1 C = 1) = 0.6$        | $\Rightarrow P(F = 0 C = 1) = 0.4$        |
| $P(F = 1 C = 0) = 0.9$        | $\Rightarrow P(F = 0 C = 0) = 0.1$        |

Tabla 2.6. Probabilidades condicionadas que definen la Red Bayesiana

Para obtener la distribución de probabilidad conjunta de cada ciclado, se aplica el algoritmo presentado. Para ello, es necesario obtener el árbol de unión asociado al DAG. En la Figura 2.12 se muestra un árbol de unión asociado al grafo que describe la Red Bayesiana.

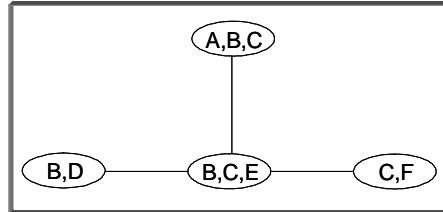


Figura 2.12. Árbol de unión asociado al DAG del Ejemplo 2.7

Siendo los ciclados del grafo  $C_1 = \{A, B, C\}$ ,  $C_2 = \{B, C, E\}$ ,  $C_3 = \{B, D\}$  y  $C_4 = \{C, F\}$ .

En el primer paso del algoritmo se calculan las funciones potenciales de los ciclados. Las Tablas 2.7 y 2.8 muestran el cálculo y los valores numéricos de dichas funciones.

$$\begin{aligned}
 \psi_1(a, b, c) &= p(a)p(b|a)p(c|a) \\
 \psi_2(b, c, e) &= p(e|b, c) \\
 \psi_3(b, d) &= p(d|b) \\
 \psi_4(c, f) &= p(f|c)
 \end{aligned}$$

Tabla 2.7. Funciones potenciales de la Red Bayesiana

Nótese que la distribución de probabilidad conjunta de la red, se puede calcular como el producto de las funciones potenciales mostradas.

|                           |                         |
|---------------------------|-------------------------|
| $\psi_1(1, 1, 1) = 0.315$ | $\psi_2(1, 1, 1) = 0.8$ |
| $\psi_1(1, 1, 0) = 0.315$ | $\psi_2(1, 1, 0) = 0.2$ |
| $\psi_1(1, 0, 1) = 0.035$ | $\psi_2(1, 0, 1) = 0.3$ |
| $\psi_1(1, 0, 0) = 0.035$ | $\psi_2(1, 0, 0) = 0.7$ |
| $\psi_1(0, 1, 1) = 0.144$ | $\psi_2(0, 1, 1) = 0.5$ |
| $\psi_1(0, 1, 0) = 0.036$ | $\psi_2(0, 1, 0) = 0.5$ |
| $\psi_1(0, 0, 1) = 0.096$ | $\psi_2(0, 0, 1) = 0.6$ |
| $\psi_1(0, 0, 0) = 0.024$ | $\psi_2(0, 0, 0) = 0.4$ |
| $\psi_3(1, 1) = 0.8$      | $\psi_4(1, 1) = 0.6$    |
| $\psi_3(1, 0) = 0.2$      | $\psi_4(1, 0) = 0.4$    |
| $\psi_3(0, 1) = 0.7$      | $\psi_4(0, 1) = 0.9$    |
| $\psi_3(0, 0) = 0.3$      | $\psi_4(0, 0) = 0.1$    |

Tabla 2.8. Valores asociados a las funciones potenciales de la Red Bayesiana

En el Paso 3 del algoritmo, se calculan los mensajes que se van a enviar desde cada ciclado.

- Como se puede observar, el ciclado  $C_1$  solo tiene un vecino ( $C_2$ ), por tanto se puede calcular el mensaje que  $C_1$  manda a  $C_2$  y enviarlo. Aplicando (2.6) se obtiene que

$$M_{12}(s_{12}) = M_{12}(b, c) = \sum_{c_1 \setminus s_{12}} \psi_1(c_1) = \sum_a \psi_1(a, b, c).$$

En función de los distintos valores de  $B$  y  $C$  se tiene que

$$\begin{aligned} M_{12}(1, 1) &= \psi_1(0, 1, 1) + \psi_1(1, 1, 1) = 0.459 \\ M_{12}(1, 0) &= \psi_1(0, 1, 0) + \psi_1(1, 1, 0) = 0.351 \\ M_{12}(0, 1) &= \psi_1(0, 0, 1) + \psi_1(1, 0, 1) = 0.131 \\ M_{12}(0, 0) &= \psi_1(0, 0, 0) + \psi_1(1, 0, 0) = 0.059 \end{aligned}$$

- El ciclado  $C_2$  tiene tres vecinos y solo ha recibido el mensaje de  $C_1$ , por tanto todavía no puede enviar ningún mensaje.

- El ciclado  $C_3$  tiene un único vecino,  $C_2$ , así que se puede calcular y enviar el mensaje aplicando (2.6)

$$M_{32}(s_{23}) = M_{32}(b) = \sum_{c_3 \setminus s_{23}} \psi_3(c_3) = \sum_d \psi_3(b, d).$$

Siendo,

$$\begin{aligned} M_{32}(1) &= 1 \\ M_{32}(0) &= 1 \end{aligned}$$

- El ciclado  $C_4$  también puede calcular y enviar su mensaje, ya que solo tiene por vecino al ciclado  $C_2$ . El mensaje que  $C_4$  envía a  $C_2$  es tal que

$$M_{42}(s_{24}) = M_{42}(c) = \sum_{c_4 \setminus s_{24}} \psi_4(c_4) = \sum_f \psi_4(c, f).$$

Por lo tanto,

$$\begin{aligned} M_{42}(1) &= 1 \\ M_{42}(0) &= 1 \end{aligned}$$

Como se han obtenido varios mensajes en esta iteración, se repite el Paso 3.

Ahora el ciclado  $C_2$  ha recibido todos los mensajes de sus vecinos, por lo que ya se pueden calcular los mensajes que  $C_2$  mandará a  $C_1$ ,  $C_3$  y  $C_4$ .

- El mensaje que  $C_2$  manda al ciclado vecino  $C_1$  es

$$M_{21}(s_{12}) = M_{21}(b, c) = \sum_{c_2 \setminus s_{12}} \psi_2(c_2) \prod_{l \neq 1} M_{l2}(s_{l2}) = \sum_e \psi_2(b, c, e) M_{32}(b) M_{42}(c).$$

Por tanto,

$$\begin{aligned} M_{21}(1, 1) &= \psi_2(1, 1, 0) M_{32}(1) M_{42}(1) + \psi_2(1, 1, 1) M_{32}(1) M_{42}(1) = 1 \\ M_{21}(1, 0) &= \psi_2(1, 0, 0) M_{32}(1) M_{42}(0) + \psi_2(1, 0, 1) M_{32}(1) M_{42}(0) = 1 \\ M_{21}(0, 1) &= \psi_2(0, 1, 0) M_{32}(0) M_{42}(1) + \psi_2(0, 1, 1) M_{32}(0) M_{42}(1) = 1 \\ M_{21}(0, 0) &= \psi_2(0, 0, 0) M_{32}(0) M_{42}(0) + \psi_2(0, 0, 1) M_{32}(0) M_{42}(0) = 1 \end{aligned}$$

- El mensaje que  $C_2$  manda al ciclado  $C_3$  se obtiene mediante

$$M_{23}(s_{23}) = M_{23}(b) = \sum_{c_2 \setminus s_{23}} \psi_2(c_2) \prod_{l \neq 3} M_{l2}(s_{l2}) = \sum_{c,e} \psi_2(b, c, e) M_{12}(b, c) M_{42}(c).$$

De esta forma cada mensaje viene dado por

$$M_{23}(1) = \sum_{c,e} \psi_2(1, c, e) M_{12}(1, c) M_{42}(c) = 0.81$$

$$M_{23}(0) = \sum_{c,e} \psi_2(0, c, e) M_{12}(0, c) M_{42}(c) = 0.19$$

- Finalmente, el mensaje que  $C_2$  manda a su vecino  $C_4$  es

$$M_{24}(s_{24}) = M_{24}(c) = \sum_{c_2 \setminus s_{24}} \psi_2(c_2) \prod_{l \neq 4} M_{l2}(s_{l2}) = \sum_{b,e} \psi_2(b, c, e) M_{12}(b, c) M_{32}(b).$$

Por consiguiente,

$$M_{24}(1) = \sum_{b,e} \psi_2(b, 1, e) M_{12}(b, 1) M_{32}(b) = 0.59$$

$$M_{24}(0) = \sum_{b,e} \psi_2(b, 0, e) M_{12}(b, 0) M_{32}(b) = 0.41$$

La Figura 2.13 muestra el orden en el que se han calculado y enviado los mensajes entre ciclados, tras realizarse el Paso 3 completo.

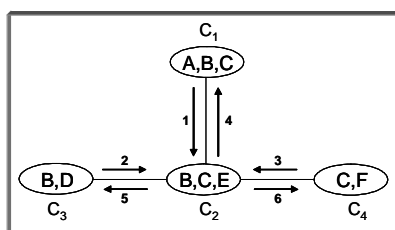


Figura 2.13. Orden en el que se calculan y envían los mensajes

Con todos los mensajes enviados entre los ciclados, el algoritmo llega al Paso 5 en el que se calculan las distribuciones de probabilidad de los ciclados, objetivo inicial del problema. Dichas distribuciones de probabilidad se presentan en las Tablas 2.9 y 2.10.

$$\begin{aligned}
p(c_1) &= p(a, b, c) = \psi_1(a, b, c)M_{21}(b, c) \\
p(c_2) &= p(b, c, e) = \psi_2(b, c, e)M_{12}(b, c)M_{32}(b)M_{42}(c) \\
p(c_3) &= p(b, d) = \psi_3(b, d)M_{23}(b) \\
p(c_4) &= p(c, f) = \psi_4(c, f)M_{24}(c)
\end{aligned}$$

Tabla 2.9. Cálculo de las probabilidades de los ciclados

|                                  |                                   |
|----------------------------------|-----------------------------------|
| $p(A = 1, B = 1, C = 1) = 0.315$ | $p(B = 1, C = 1, E = 1) = 0.3672$ |
| $p(A = 1, B = 1, C = 0) = 0.315$ | $p(B = 1, C = 1, E = 0) = 0.0918$ |
| $p(A = 1, B = 0, C = 1) = 0.035$ | $p(B = 1, C = 0, E = 1) = 0.1053$ |
| $p(A = 1, B = 0, C = 0) = 0.035$ | $p(B = 1, C = 0, E = 0) = 0.2457$ |
| $p(A = 0, B = 1, C = 1) = 0.144$ | $p(B = 0, C = 1, E = 1) = 0.0655$ |
| $p(A = 0, B = 1, C = 0) = 0.036$ | $p(B = 0, C = 1, E = 0) = 0.0655$ |
| $p(A = 0, B = 0, C = 1) = 0.096$ | $p(B = 0, C = 0, E = 1) = 0.0354$ |
| $p(A = 0, B = 0, C = 0) = 0.024$ | $p(B = 0, C = 0, E = 0) = 0.0236$ |
| $p(B = 1, D = 1) = 0.648$        | $p(C = 1, F = 1) = 0.354$         |
| $p(B = 1, D = 0) = 0.162$        | $p(C = 1, F = 0) = 0.236$         |
| $p(B = 0, D = 1) = 0.133$        | $p(C = 0, F = 1) = 0.369$         |
| $p(B = 0, D = 0) = 0.057$        | $p(C = 0, F = 0) = 0.041$         |

Tabla 2.10. Valores de las probabilidades de los ciclados

**Ejemplo 2.8**

Considerando la Red Bayesiana del Ejemplo 2.7, obtener los mensajes que se envían a través de la red cuando se tiene evidencia acerca de las variables  $C$  y  $D$ , siendo  $\mathbf{E} = \{C = 1, D = 1\}$ .

La evidencia  $C = 1$  se puede introducir en el ciclado  $C_1$  y  $D = 1$  en el ciclado  $C_3$ , por tanto, en el Paso 2 del algoritmo, se modifican las funciones potenciales de  $C_1$  y  $C_3$  quedando

---



---

|                             |
|-----------------------------|
| $\psi_1^*(1, 1, 1) = 0.315$ |
| $\psi_1^*(1, 1, 0) = 0$     |
| $\psi_1^*(1, 0, 1) = 0.035$ |
| $\psi_1^*(1, 0, 0) = 0$     |
| $\psi_1^*(0, 1, 1) = 0.144$ |
| $\psi_1^*(0, 1, 0) = 0$     |
| $\psi_1^*(0, 0, 1) = 0.096$ |
| $\psi_1^*(0, 0, 0) = 0$     |
| <hr/>                       |
| $\psi_3^*(1, 1) = 0.8$      |
| $\psi_3^*(1, 0) = 0$        |
| $\psi_3^*(0, 1) = 0.7$      |
| $\psi_3^*(0, 0) = 0$        |

---



---

Tabla 2.11. Nuevas funciones potenciales de los ciclados  $C_1$  y  $C_3$

Las funciones potenciales de los ciclados  $C_2$  y  $C_4$  se mantienen como en la Tabla 2.8. A partir de los potenciales  $\psi_1^*(c_1)$ ,  $\psi_2(c_2)$ ,  $\psi_3^*(c_3)$  y  $\psi_4(c_4)$ , se procede al cálculo de los mensajes.

Igual que en el ejemplo anterior, el orden de envío de los mensajes, es el que se muestra en la Figura 2.13 ya que  $C_1$ ,  $C_3$  y  $C_4$  solo tienen un vecino (el ciclado  $C_2$ ) y por tanto se puede calcular el mensaje que se envía desde dichos ciclados a  $C_2$ . Una vez que  $C_2$  haya recibido todos los mensajes de sus vecinos, se puede calcular el mensaje que  $C_2$  les envía.

- El mensaje que  $C_1$  envía a su único vecino  $C_2$  se obtiene, igual que en el ejemplo anterior, aplicando (2.6), mediante

$$M_{12}(s_{12}) = M_{12}(b, c) = \sum_{c_1 \setminus s_{12}} \psi_1^*(c_1) = \sum_a \psi_1^*(a, b, c).$$

En función de los distintos valores de  $B$  y  $C$  se tiene que

$$\begin{aligned} M_{12}(1, 1) &= \psi_1^*(0, 1, 1) + \psi_1^*(1, 1, 1) = 0.459 \\ M_{12}(1, 0) &= \psi_1^*(0, 1, 0) + \psi_1^*(1, 1, 0) = 0 \\ M_{12}(0, 1) &= \psi_1^*(0, 0, 1) + \psi_1^*(1, 0, 1) = 0.131 \\ M_{12}(0, 0) &= \psi_1^*(0, 0, 0) + \psi_1^*(1, 0, 0) = 0 \end{aligned}$$



- El mensaje que  $C_3$  envía a  $C_2$  se calcula mediante

$$M_{32}(s_{23}) = M_{32}(b) = \sum_{c_3 \setminus s_{23}} \psi_3^*(c_3) = \sum_d \psi_3^*(b, d).$$

Por tanto,

$$\begin{aligned} M_{32}(1) &= \psi_3^*(1, 1) + \psi_3^*(1, 0) = 0.8 \\ M_{32}(0) &= \psi_3^*(0, 1) + \psi_3^*(0, 0) = 0.7 \end{aligned}$$

- El mensaje que  $C_4$  envía a  $C_2$  es

$$M_{42}(s_{24}) = M_{42}(c) = \sum_{c_4 \setminus s_{24}} \psi_4(c_4) = \sum_f \psi_4(c, f).$$

Por lo que el mensaje es

$$\begin{aligned} M_{42}(1) &= 1 \\ M_{42}(0) &= 1 \end{aligned}$$

Se repite el Paso 3 del algoritmo para obtener los mensajes que  $C_2$  envía a sus vecinos, ya que el ciclado  $C_2$  ha recibido todos los mensajes de sus vecinos.

- Así, el mensaje que  $C_2$  envía al ciclado  $C_1$  viene dado por

$$M_{21}(s_{12}) = M_{21}(b, c) = \sum_{c_2 \setminus s_{12}} \psi_2(c_2) \prod_{l \neq 1} M_{l2}(s_{l2}) = \sum_e \psi_2(b, c, e) M_{32}(b) M_{42}(c).$$

Por lo tanto,

$$\begin{aligned} M_{21}(1, 1) &= \psi_2(1, 1, 0) M_{32}(1) M_{42}(1) + \psi_2(1, 1, 1) M_{32}(1) M_{42}(1) = 0.8 \\ M_{21}(1, 0) &= \psi_2(1, 0, 0) M_{32}(1) M_{42}(0) + \psi_2(1, 0, 1) M_{32}(1) M_{42}(0) = 0.8 \\ M_{21}(0, 1) &= \psi_2(0, 1, 0) M_{32}(0) M_{42}(1) + \psi_2(0, 1, 1) M_{32}(0) M_{42}(1) = 0.7 \\ M_{21}(0, 0) &= \psi_2(0, 0, 0) M_{32}(0) M_{42}(0) + \psi_2(0, 0, 1) M_{32}(0) M_{42}(0) = 0.7 \end{aligned}$$

- El mensaje que  $C_2$  manda al ciclado  $C_3$  es

$$M_{23}(s_{23}) = M_{23}(b) = \sum_{c_2 \setminus s_{23}} \psi_2(c_2) \prod_{l \neq 3} M_{l2}(s_{l2}) = \sum_{c, e} \psi_2(b, c, e) M_{12}(b, c) M_{42}(c).$$

De esta forma cada mensaje viene dado por

$$\begin{aligned} M_{23}(1) &= \sum_{c, e} \psi_2(1, c, e) M_{12}(1, c) M_{42}(c) = 0.459 \\ M_{23}(0) &= \sum_{c, e} \psi_2(0, c, e) M_{12}(0, c) M_{42}(c) = 0.131 \end{aligned}$$

- Finalmente, el mensaje que  $C_2$  manda a su vecino  $C_4$  se calcula a partir de la siguiente expresión

$$M_{24}(s_{24}) = M_{24}(c) = \sum_{c_2 \setminus s_{24}} \psi_2(c_2) \prod_{l \neq 4} M_{l2}(s_{l2}) = \sum_{b,e} \psi_2(b, c, e) M_{12}(b, c) M_{32}(b).$$

Siendo,

$$M_{24}(1) = \sum_{b,e} \psi_2(b, 1, e) M_{12}(b, 1) M_{32}(b) = 0.4589$$

$$M_{24}(0) = \sum_{b,e} \psi_2(b, 0, e) M_{12}(b, 0) M_{32}(b) = 0$$

Con todos los mensajes enviados, se puede calcular la probabilidad de cualquier ciclado y posteriormente cualquier probabilidad de interés. En este caso, la probabilidad de cada ciclado suma 0.4589, por tanto dividiendo entre dicha cantidad cada una de las probabilidades asociadas a los distintos valores de las variables del ciclado se obtiene una probabilidad conjunta que suma 1.

Posteriormente, se puede calcular cada probabilidad de interés  $p(x_i|\mathbf{e})$ , marginalizando la distribución conjunta de cada ciclado. En la Figura 2.14 se muestran las probabilidades condicionadas, dada la evidencia  $\mathbf{E} = \{C = 1, D = 1\}$ , de cada una de las variables de la red.

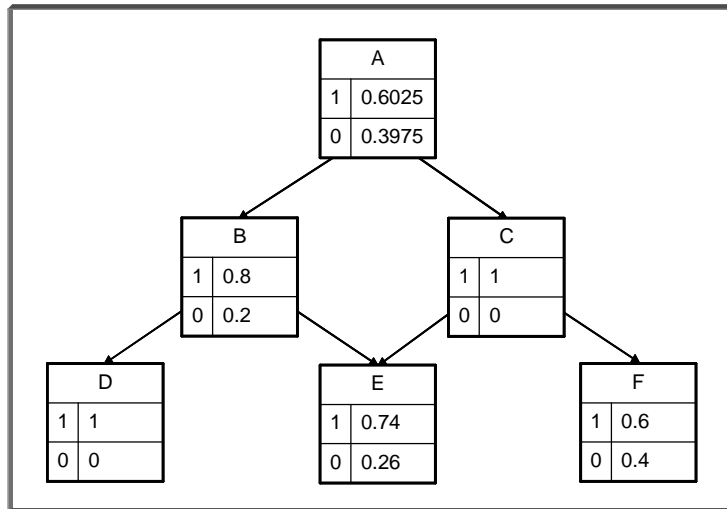


Figura 2.14. Probabilidades de interés de las variables de la red, dada la evidencia  $\mathbf{E} = \{C = 1, D = 1\}$

## 2.4. Redes Bayesianas Gaussianas

Inicialmente, las Redes Bayesianas se definieron para un conjunto finito de variables aleatorias discretas de las que se conocía su distribución de probabilidad condicionada, dada la ocurrencia de sus padres en el DAG. Aplicando conceptos básicos del Cálculo de Probabilidades, la obtención de las probabilidades finales de interés,  $p(x_i|\mathbf{e})$ , tras presentarse evidencia, es un paso directo, como se ha visto en las técnicas de propagación de la evidencia presentadas en la Sección anterior.

Posteriormente, se busca ampliar el concepto de Red Bayesiana a variables aleatorias continuas, aunque la distribución más frecuente con propiedades de interés para la descripción de una Red Bayesiana Continua, cuya distribución es cerrada frente al cálculo de la distribución condicionada y marginal, es la distribución Normal.

Dempster (1972) introduce la teoría de los modelos gráficos Gaussianos, también denominados *modelos de selección de la covarianza*, en los que se relaciona un conjunto de variables  $\mathbf{X}$  con distribución normal multivariante con un grafo  $\mathcal{G}$  que representa las relaciones entre las variables del problema. Existe un desarrollo teórico para grafos descomponibles (Definición 1.13) que generaliza algunos de los conceptos presentados en esta Memoria, como se puede ver en Giudici, et al. (1999). Además, estos modelos gráficos Gaussianos también han sido estudiados por autores como Wermuth (1980), Whittaker (1990), Andersen, et al. (1995) o Lauritzen (1996), entre otros.

Con el interés puesto en los modelos cuya representación gráfica viene dada por un DAG, se definen las Redes Bayesianas Mixtas como una combinación entre las Redes Bayesianas Discretas y las Gaussianas.

Actualmente, se trabaja en la descripción de modelos gráficos con distribuciones continuas distintas de la Normal, como es el caso de las variaciones de la normal asimétrica o Skew-Normal (Azzalini, 1985) que presenta Capitanio, et al. (2003), aunque aparecen dificultades asociadas al concepto de independencia condicionada y a la propagación de la evidencia.

En esta Memoria, el interés se centra en las Redes Bayesianas Gaussianas. Se define una Red Bayesiana Gaussiana como una Red Bayesiana, dada por el par  $(\mathcal{D}, P)$  donde  $\mathcal{D}$  es un DAG, que representa las variables del problema y su estructura de dependencia, y  $P = \{f(x_1|pa(X_1)), \dots, f(x_n|pa(X_n))\}$  es el conjunto de distribuciones condicionadas de cada una de las variables, dada la ocurrencia de sus padres en el DAG. Además, la hipótesis del modelo en las Redes Bayesianas Gaussianas es que su distribución de probabilidad conjunta viene dada por una distribución normal multivariante.

**Definición 2.2** (*Red Bayesiana Gaussiana*)

Se dice que una Red Bayesiana es una Red Bayesiana Gaussiana, cuando la distribución conjunta asociada a las variables de la red  $\mathbf{X} = \{X_1, \dots, X_n\}$  es una distribución normal multivariante, de forma que la función de densidad de  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  viene dada por

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (2.7)$$

donde  $\boldsymbol{\mu}$  es el vector de medias de dimensión  $n$ ,  $\Sigma$  la matriz de covarianzas, definida positiva de dimensión  $n \times n$ , por  $|\Sigma|$  se denota el determinante de la matriz de covarianzas y por  $(\mathbf{x} - \boldsymbol{\mu})^T$  el vector traspuesto de  $(\mathbf{x} - \boldsymbol{\mu})$ . A la matriz  $\Sigma^{-1}$  se la denomina matriz de precisión o concentración.

Además, en las Redes Bayesianas la distribución de probabilidad conjunta de la red es el producto de los elementos de  $P$ , que son, en el caso de las Redes Bayesianas Gaussianas, las funciones de densidad condicionada dada la ocurrencia de los nodos padres en el DAG,  $f(x_i|pa(X_i))$ . Por tanto,

$$f(\mathbf{x}) = \prod_{i=1}^n f(x_i|pa(X_i)). \quad (2.8)$$

La distribución normal multivariante tiene la ventaja de que las densidades condicionadas y marginales se distribuyen también normalmente. Otras propiedades como la independencia de las variables normales incorreladas o como la determinación de independencias condicionadas por los ceros en la inversa de la matriz de covarianzas, hacen más fácil el cálculo y el tratamiento local de las variables de la red. En el Apartado 2.4.3 se describen algunas de estas propiedades de la distribución normal multivariante, que facilitan el tratamiento de una Red Bayesiana Gaussiana.

Se describe la distribución condicionada de cada una de las variables de la red, dada la ocurrencia de sus padres en el DAG, como una distribución normal univariante tal que

$$f(x_i|pa(X_i)) \sim N\left(\mu_i + \sum_{j=1}^{i-1} \beta_{ij}(x_j - \mu_j), v_i\right) \quad (2.9)$$

donde  $\beta_{ij}$  con  $j < i$  es el coeficiente de regresión de  $X_j$  en la regresión de  $X_i$  sobre sus padres, y  $v_i$  es la varianza condicionada de  $X_i$  dados sus padres en el DAG.

Nótese que el coeficiente de regresión es cero ( $\beta_{ij} = 0$ ), si y solo si no hay una arista dirigida del nodo  $X_j$  al  $X_i$ , por tanto para Redes Bayesianas Gaussianas descritas con valores de  $\beta_{ij} \neq 0$  se sabe que existe en el DAG una arista entre dichas variables.

La varianza condicionada  $v_i$ , se puede calcular en función de los valores de la matriz de covarianzas  $\Sigma$ , mediante la siguiente expresión

$$v_i = \Sigma_i - \Sigma_{ipa(X_i)} \Sigma_{pa(X_i)}^{-1} \Sigma_{ipa(X_i)}^T$$

siendo  $\Sigma_i$  la varianza de  $X_i$ ,  $\Sigma_{ipa(X_i)}$  la covarianza entre  $X_i$  y sus padres y  $\Sigma_{pa(X_i)}^{-1}$  la inversa de la matriz de covarianzas de los padres del nodo  $X_i$  en el DAG.

Los parámetros de la expresión (2.9) reflejan una dependencia en la media condicionada de los valores de los padres  $x_j \in pa(X_i)$ , y sin embargo, la varianza condicionada es independiente de dichos valores. Además, se observa una ordenación entre los nodos del grafo; esta ordenación o numeración perfecta de los nodos del grafo permite la triangulación del mismo, así como la obtención del grafo descomponible, como se puede ver en los Teoremas 1.1 y 1.2. Estas características facilitan el tratamiento de la red localmente.

En esta Sección se muestra la construcción de una Red Bayesiana Gaussiana en función de los parámetros que se describen. También, se introducen las propiedades de independencia condicionada que presenta la matriz de covarianzas  $\Sigma$  a través de su inversa y se muestra un algoritmo de propagación de la evidencia en Redes Bayesianas Gaussianas, que se utilizará en el Capítulo 3 para el desarrollo del análisis de sensibilidad.

### 2.4.1. Construcción de una Red Bayesiana Gaussiana

Cuando se desea modelizar un problema mediante una Red Bayesiana, un grupo de expertos en el problema en estudio ha de consensuar las variables del problema, el conjunto de relaciones de dependencia que se tiene entre las mismas, y las distribuciones de probabilidad condicionadas de cada una de las variables, dados sus padres en el DAG. Dichas distribuciones se pueden obtener de un análisis previo de los datos, o los mismos expertos pueden concluir subjetivamente las distintas distribuciones condicionadas necesarias para definir la red.

Cuando las variables aleatorias del problema son discretas y son pocas sus categorías, es posible ir asignando probabilidades a los distintos valores de las variables, dados sus padres en el DAG, pero en el caso Gaussiano esta asignación es algo más compleja.

Como se ha enunciado anteriormente, una Red Bayesiana Gaussiana puede presentarse según su propia definición, es decir, mediante los parámetros asociados a la distribución normal multivariante,  $\boldsymbol{\mu}$  vector de medias y  $\Sigma$  matriz de covarianzas. Y también se puede construir una Red Bayesiana Gaussiana mediante la definición de Red Bayesiana, esto es, teniendo en cuenta las distribuciones condicionadas de cada variable dada la ocurrencia de sus padres en el DAG  $f(x_i|pa(X_i))$ , para todo  $X_i$ . Por lo tanto, se tienen dos formas alternativas de obtener la especificación cuantitativa de una Red Bayesiana Gaussiana.

A continuación se muestran los parámetros necesarios para construir una Red Bayesiana Gaussiana, considerando las dos alternativas citadas.

#### Construcción de una Red Bayesiana Gaussiana con las distribuciones de probabilidad condicionadas

Por la definición de Red Bayesiana, se sabe que dicho modelo se especifica en función del DAG que presenta las variables del problema con sus relaciones de dependencia y un conjunto de distribuciones de probabilidad  $P$  dadas mediante  $f(x_i|pa(X_i))$  para todo  $X_i$ .

En el caso de las Redes Gaussianas, la distribución condicionada de cada variable, dada la ocurrencia de sus padres, es normal univariante como se puede

ver en la expresión (2.9), siendo

$$\begin{aligned}E[X_i|pa(X_i)] &= \mu_i + \sum_{j=1}^{i-1} \beta_{ij}(x_j - \mu_j) \\V[X_i|pa(X_i)] &= v_i\end{aligned}$$

Por lo tanto, para obtener la Red Bayesiana Gaussiana descrita a partir de la distribución condicionada de cada una de las variables, es necesario especificar los siguientes parámetros

- $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$

donde  $\mu_i$  es la media marginal de la variable  $X_i$ , para todo  $i = 1, \dots, n$ .

- $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$

donde  $v_i$  es la varianza condicionada de  $X_i$  por sus padres en el DAG, con  $i = 1, \dots, n$ .

- $\mathbf{B} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \beta_{21} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \dots & 0 \end{pmatrix}$

siendo  $\beta_{ij}$ , con  $j < i$ , el coeficiente de regresión de  $X_j$  en el modelo de regresión de  $X_i$  sobre  $X_j$ , con  $i = 1, \dots, n$ .

Con esta información, obtenida bien por observación e información de diversos estudios, bien por un análisis de datos previo, es posible definir una Red Bayesiana Gaussiana calculando la distribución conjunta normal multivariante como el producto de las densidades condicionadas.

A continuación, se analiza la construcción de una Red Bayesiana Gaussiana diseñada para estudiar el funcionamiento de una máquina en la que intervienen cinco componentes.

**Ejemplo 2.9**

Se desea estudiar, mediante una Red Bayesiana Gaussiana, el funcionamiento de una máquina formada por cinco componentes, conectadas como muestra el DAG de la Figura 2.15, donde  $X_5$  es la variable de interés que presenta el resultado final del proceso realizado por la máquina.

Se sabe que el funcionamiento de cada componente se distribuye normalmente. Para especificar la probabilidad de la red, los expertos determinan los parámetros  $\mu_i, \beta_{ij}$  y  $v_i$  para todo  $X_i$  dados sus padres  $X_j \quad \forall j < i$ , en el DAG.

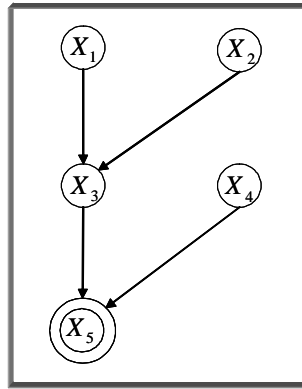


Figura 2.15. DAG asociado a la Red Bayesiana Gaussiana descrita para el Ejemplo 2.9

El funcionamiento de la máquina queda recogido por la variable de interés  $X_5$ . Por consiguiente, la variable de interés  $X_5$  se especifica en el DAG mediante un nodo formado por dos círculos concéntricos. En este caso, los algoritmos de propagación de la evidencia se pueden reducir, obteniendo como único resultado  $f(x_i|e)$ .

Se sabe que

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_1)f(x_2)f(x_3|x_1, x_2)f(x_4)f(x_5|x_3, x_4)$$

donde



$$\begin{aligned}X_1 &\sim N(\mu_1, v_1) \\X_2 &\sim N(\mu_2, v_2) \\X_3|X_1, X_2 &\sim N(\mu_3 + \beta_{31}(x_1 - \mu_1) + \beta_{32}(x_2 - \mu_2), v_3) \\X_4 &\sim N(\mu_4, v_4) \\X_5|X_3, X_4 &\sim N(\mu_5 + \beta_{53}(x_3 - \mu_3) + \beta_{54}(x_4 - \mu_4), v_5)\end{aligned}$$

Por consiguiente, la red queda definida mediante los parámetros  $\mu_i$ ,  $v_i$  y  $\beta_{ij}$  con  $j < i$  para todas las variables de la red, siendo

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \mathbf{v} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \\ 3 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \end{pmatrix}$$

La obtención de los parámetros que describen la distribución conjunta normal multivariante de las variables de la red, se realiza de forma directa en el caso del vector de medias y para la matriz de covarianzas se aplica la transformación enunciada en el Apartado 2.4.2 desarrollada por Shachter, et al. (1989).

### **Construcción de una Red Bayesiana Gaussiana con la distribución de probabilidad conjunta**

Se trata de especificar una distribución de probabilidad conjunta normal multivariante  $N(\boldsymbol{\mu}, \Sigma)$ .

En este caso, los expertos han de especificar los valores que componen el vector de medias  $\boldsymbol{\mu}$  y la matriz de covarianzas  $\Sigma$ , bajo la condición de ser dicha matriz una matriz definida positiva. Por consiguiente, para definir la Red Bayesiana Gaussiana que describe un problema, se deben fijar los elementos de

- $\boldsymbol{\mu}$  : vector de medias, dado por  $\mu_i$  para todas las variables  $X_i$  de la red.
- $\Sigma$  : matriz de covarianzas simétrica definida positiva, donde  $\sigma_{ii}$  es la varianza de  $X_i$  y  $\sigma_{ij}$  es la covarianza entre  $X_i$  y  $X_j$ .

Obsérvese que la varianza condicionada de  $X_i$  por sus padres,  $v_i$ , se puede presentar en función de los elementos de la matriz  $\Sigma$  mediante

$$v_i = \Sigma_i - \Sigma_{ipa(X_i)} \Sigma_{pa(X_i)}^{-1} \Sigma_{ipa(X_i)}^T$$

donde  $\Sigma_i$  es la varianza de  $X_i$ ,  $\Sigma_{ipa(X_i)}$  es la covarianza entre  $X_i$  y sus padres y  $\Sigma_{pa(X_i)}^{-1}$  es la inversa de la matriz de covarianzas de los padres del nodo  $X_i$  en el DAG.

En el siguiente Apartado se implementa como obtener la matriz de covarianzas  $\Sigma$  a partir de los parámetros que describen la distribución condicionada.

A continuación, se muestran el vector de medias y la matriz de covarianzas asociados a la distribución conjunta  $N(\boldsymbol{\mu}, \Sigma)$  de los datos del problema presentado en el Ejemplo 2.9.

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 0 & 6 & 0 & 6 \\ 0 & 2 & 2 & 0 & 2 \\ 6 & 2 & 15 & 0 & 15 \\ 0 & 0 & 0 & 2 & 4 \\ 6 & 2 & 15 & 4 & 26 \end{pmatrix}$$

### 2.4.2. Estructura de dependencia

Shachter, et al. (1989) estudian la tratabilidad de los modelos gráficos Gaussianos en una estructura gráfica concreta, el diagrama de influencia, describiendo una transformación de los parámetros condicionados  $\{v_1, \dots, v_n\}$  y  $\{\beta_{ij}$  con  $j < i\}$  para obtener la matriz de covarianzas  $\Sigma$  definida positiva. Con esta finalidad, desarrollan un algoritmo que muestra la relación entre el modelo gráfico Gaussiano y la representación de la matriz de covarianzas para una distribución normal.

En los modelos gráficos Gaussianos, la matriz de covarianzas  $\Sigma$  es de gran interés, ya que recoge la estructura de dependencia de las variables del problema mediante la matriz inversa de  $\Sigma$ , como se verá en el siguiente Apartado. Algunos desarrollos teóricos de estos modelos no exigen que la matriz de covarianzas  $\Sigma$

sea definida positiva, por ello en esta Memoria el interés se centra en el desarrollo de Shachter, et al. (1989) que especifican una matriz definida positiva, que será necesaria para posteriores desarrollos en el modelo gráfico dado por una Red Bayesiana Gaussiana.

Apoyándose en resultados de Yule (1907) y en la descomposición de Cholesky, Shachter, et al. (1989) concluyen que la matriz de covarianzas  $\Sigma$  se puede descomponer como el producto de diversas matrices, siendo dicha matriz  $\Sigma$  definida positiva e invertible o regular. Con esta finalidad, se definen las siguientes matrices:

- $\mathbf{D}$  : matriz diagonal formada por las varianzas condicionadas  $v_i$ , siendo  $\mathbf{D} = \text{diag}(v_i)$
- $\mathbf{B}$  : matriz triangular inferiormente formada por los coeficientes de regresión  $\beta_{ij}$  para todo  $j < i$ .

Como  $\mathbf{B}$  es triangular inferiormente,  $(\mathbf{I} - \mathbf{B})$  es invertible y por tanto, se puede definir la matriz  $\mathbf{U} = (\mathbf{I} - \mathbf{B}^T)^{-1}$ . De esta forma, la matriz de covarianzas  $\Sigma$  se obtiene como

$$\Sigma = \mathbf{U}^T \mathbf{D} \mathbf{U} = [(\mathbf{I} - \mathbf{B}^T)^{-1}]^T \mathbf{D} (\mathbf{I} - \mathbf{B}^T)^{-1} \quad (2.10)$$

Como la matriz  $\mathbf{U}$  es regular, la transformación de  $\Sigma$  dada en (2.10) es congruente, y por tanto, si todas las varianzas condicionadas son distintas de 0, la matriz  $\Sigma$  es definida positiva. En caso de existir algún  $v_i = 0$ , entonces dicha matriz es semidefinida positiva.

Sea la *matriz de precisión*  $\mathbf{K}$ , la matriz de covarianzas inversa dada por  $\mathbf{K} = \Sigma^{-1}$ . Se sabe que si  $\Sigma$  es regular, entonces dicha matriz de precisión se puede obtener como

$$\mathbf{K} = \Sigma^{-1} = \mathbf{U}^{-T} \mathbf{D}^{-1} \mathbf{U}^{-1} = (\mathbf{I} - \mathbf{B}^T)^T \mathbf{D}^{-1} (\mathbf{I} - \mathbf{B}^T)$$

La matriz de precisión  $\mathbf{K}$  es de gran interés en el estudio de los modelos gráficos Gaussianos, por reflejar relaciones de independencia entre las variables del problema. En el caso de las Redes Bayesianas Gaussianas dicha matriz permite

obtener relaciones de independencia condicionada entre las variables, como se enuncia en el siguiente Apartado.

El algoritmo desarrollado por Shachter, et al. (1989), permite calcular de forma recursiva la matriz de precisión a partir de los parámetros  $\{v_1, \dots, v_n\}$  y  $\{\beta_{ij}$  con  $j < i\}$ , mediante la siguiente expresión

$$\mathbf{K}(i+1) = \begin{pmatrix} \mathbf{K}(i) + \frac{\beta_{i+1}\beta_{i+1}^T}{v_{i+1}} & -\frac{\beta_{i+1}}{v_{i+1}} \\ -\frac{\beta_{i+1}^T}{v_{i+1}} & \frac{1}{v_{i+1}} \end{pmatrix}$$

siendo  $\mathbf{K}(1) = \frac{1}{v_1}$  (2.11)

donde  $\mathbf{K}(i)$  es la matriz superior izquierda  $i \times i$  de la matriz de precisión  $\mathbf{K}$  y  $\beta_i$  es un vector columna que representa los coeficientes de regresión  $\beta_{ij}$  con  $j < i$ .

La matriz de covarianzas  $\Sigma$  se obtiene directamente tras calcular la matriz inversa de  $\mathbf{K}$ .

La matriz de covarianzas  $\Sigma$  del Ejemplo 2.9, obtenida a partir de los parámetros  $\{v_1, \dots, v_n\}$  y  $\{\beta_{ij}$  con  $j < i\}$  se obtiene igualmente aplicando (2.10) o (2.11).

### 2.4.3. Independencia condicionada

Han sido muchos los autores que han trabajado con la distribución normal multivariante. En este Apartado, se presentan dos resultados conocidos de dicha distribución que relacionan la independencia condicionada con la matriz de covarianzas  $\Sigma$ . Las demostraciones de las proposiciones propuestas pueden verse en Lauritzen (1996) o Anderson (2003).

Como la matriz de covarianzas  $\Sigma$  está formada por varianzas en la diagonal, y por covarianzas en el resto de elementos, se puede determinar la independencia entre dos variables en función de los ceros de la matriz  $\Sigma$ . La siguiente proposición, refleja esta idea.

**Proposición 2.1**

Sea  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  tal que  $\mathbf{X}$  se particiona en dos componentes  $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2\}$  siendo  $\boldsymbol{\mu}$  y  $\Sigma$

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Entonces  $\mathbf{X}_1$  y  $\mathbf{X}_2$  son independientes si y solo si  $\Sigma_{12}$  es la matriz 0.

Si la matriz de covarianzas  $\Sigma$  es regular, es decir existe su inversa, entonces esta proposición se puede enunciar también si y solo si  $\mathbf{K}_{12} = 0$ .

**Demostración 2.1**

Puede verse Anderson (2003). ■

Con este resultado, a continuación se presenta una proposición que fija la independencia condicionada de las variables con la distribución normal multivariante, en función de los ceros que aparecen en la matriz de precisión  $\mathbf{K} = \Sigma^{-1}$  de dicha distribución. Esta proposición es consecuencia directa de la estructura de la distribución normal condicionada.

**Proposición 2.2**

Sea  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , donde  $\Sigma$  es una matriz regular (o invertible). Entonces, las variables  $X_i$  y  $X_j$  son condicionalmente independientes dado el resto de las variables de  $\mathbf{X}$ , si y solo si el valor de la posición  $(i, j)$  de la matriz  $\mathbf{K}$  es cero, es decir

$$X_i \perp\!\!\!\perp X_j \mid (\mathbf{X} \setminus \{X_i, X_j\}) \iff k_{ij} = 0$$

donde  $\mathbf{K} = \{k_{ij}\}_{X_i, X_j \in \mathbf{X}} = \Sigma^{-1}$  es la matriz de precisión de la distribución.

**Demostración 2.2**

Puede verse Anderson (2003). ■

Por tanto, a partir de la matriz de covarianzas se pueden conocer las relaciones de independencia condicionada existentes entre las variables del problema, identificando los ceros en la matriz de precisión  $\mathbf{K}$ . Como se vio en el Capítulo 1, estas relaciones de independencia condicionada se corresponden con las observadas gráficamente aplicando las propiedades de Markov sobre grafos y por tanto, mediante los conceptos de separación gráfica.

A continuación, se presentan las relaciones de independencia condicionada entre las variables del Ejemplo 2.9, en función de los resultados anteriormente expuestos.

### Ejemplo 2.10

Se van a describir las relaciones de independencia condicionada existentes entre las variables  $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5\}$  del Ejemplo 2.9, siendo la matriz de covarianzas que describe la distribución conjunta normal multivariante

$$\Sigma = \begin{pmatrix} 3 & 0 & 6 & 0 & 6 \\ 0 & 2 & 2 & 0 & 2 \\ 6 & 2 & 15 & 0 & 15 \\ 0 & 0 & 0 & 2 & 4 \\ 6 & 2 & 15 & 4 & 26 \end{pmatrix}$$

Con la matriz de covarianzas se observa independencia entre las variables  $X_1$  y  $X_2$ ,  $X_1$  y  $X_4$ ,  $X_2$  y  $X_4$  y entre  $X_3$  y  $X_4$ . Este mismo resultado se obtiene estudiando las conexiones entre los nodos del DAG de la Figura 2.15. Así, por ejemplo, en el DAG se tiene que  $X_1 \perp\!\!\!\perp_{\mathcal{D}} X_2 | \emptyset$ , ya que  $X_1$  y  $X_2$  aparecen en una conexión convergente como padres de  $X_3$ .

Por el contrario, si se estudian las relaciones de independencia condicionada que presenta el DAG, se observa que  $X_1$  y  $X_2$  son dependientes condicionalmente, ya que dada la variable  $X_3$ , la conexión convergente entre  $X_1$  y  $X_2$  hace que dichas variables sean dependientes condicionalmente. Estas mismas relaciones de dependencia e independencia condicionada, también se pueden concluir mediante la matriz de precisión  $\mathbf{K}$

$$\mathbf{K} = \Sigma^{-1} = \begin{pmatrix} \frac{13}{3} & 2 & -2 & 0 & 0 \\ 2 & \frac{3}{2} & -1 & 0 & 0 \\ -2 & -1 & \frac{4}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & 0 & \frac{2}{3} & \frac{11}{6} & -\frac{2}{3} \\ 0 & 0 & -\frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Por lo tanto, las relaciones de independencia condicionada entre las variables

del problema son

$$\begin{aligned} X_1 &\perp\!\!\!\perp X_4 \mid \{X_2, X_3, X_5\} \\ X_1 &\perp\!\!\!\perp X_5 \mid \{X_2, X_3, X_4\} \\ X_2 &\perp\!\!\!\perp X_4 \mid \{X_1, X_3, X_5\} \\ X_2 &\perp\!\!\!\perp X_5 \mid \{X_1, X_3, X_4\} \end{aligned}$$

Se habría concluido este mismo resultado si se estudian la relaciones de independencia condicionada del DAG de la Figura 2.15, mediante el criterio de  $d$ -separación enunciado en el Apartado 1.3.4. Así, efectivamente al conocerse por ejemplo el valor de  $X_3$ , las variables  $X_1$  y  $X_5$  que aparecen en una conexión en serie en el DAG, tal que  $X_1 \longrightarrow X_3 \longrightarrow X_5$ , pasan a ser independientes condicionalmente, siendo  $X_1 \perp\!\!\!\perp X_5 \mid \{X_2, X_3, X_4\}$ . Lo mismo sucede para el resto de relaciones de independencia condicionada obtenidas mediante  $\mathbf{K}$ .

#### 2.4.4. Propagación de la evidencia en Redes Bayesianas Gaussianas

En la Sección 2.3 se han introducido los algoritmos de propagación más significativos desarrollados para Redes Bayesianas Discretas. Para Redes Bayesianas Gaussianas, Normand, et al. (1992) desarrollan un algoritmo de propagación apoyándose en las ideas del algoritmo de propagación en poliárboles mostrado en el Apartado 2.3.2. También Lauritzen (1992), basándose en el algoritmo de propagación en árboles de unión, desarrolla un algoritmo de propagación calculando medias y varianzas localmente, aunque dicho algoritmo es inestable numéricamente. Años más tarde, Lauritzen, et al. (2001) desarrollan un algoritmo alternativo al anteriormente citado, bajo el mismo esquema de computación local del árbol de unión para las Redes Bayesianas Gaussianas. Sin embargo, la ejecución de éste algoritmo es compleja porque requiere evaluaciones de las matrices de precisión y combinaciones recursivas de las funciones potenciales. Recientemente, Cowell (2005) presenta una alternativa a estos algoritmos, basada en realizar los cálculos locales, no sobre el árbol de unión sino sobre el grafo descomponible, eliminándose las manipulaciones de la matrices y las operaciones complejas con las funciones potenciales.

Algunos de éstos algoritmos utilizan las distribuciones de probabilidad condi-

cionadas, como punto de partida para realizar la propagación de la evidencia, y otros, comienzan con los parámetros  $\boldsymbol{\mu}$  vector de medias y  $\Sigma$  matriz de covarianzas, que definen la distribución conjunta normal multivariante que describe la red.

En este Apartado, se introduce un algoritmo conceptualmente distinto a los presentados en el Apartado 2.3.2, que utiliza como punto de partida los parámetros de la distribución normal multivariante conjunta. Este algoritmo se apoya en resultados básicos propios de los modelos Gaussianos asociados a la idea de distribución condicionada normal multivariante. Así, si se realiza de forma recursiva el cálculo de la distribución condicionada dada la evidencia, considerando en cada paso la existencia de una única variable evidencial, se tiene un algoritmo que actualiza las probabilidades no evidenciales de la red, en tiempo lineal, dada la evidencia.

El siguiente teorema presenta un resultado conocido asociado a la distribución condicionada de una distribución normal multivariante.

### Teorema 2.1

Sea  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  tal que  $\mathbf{X}$  se particiona en dos componentes  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ , siendo los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  tales que

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}} \\ \boldsymbol{\mu}_{\mathbf{E}} \end{pmatrix} \quad \text{y} \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} & \Sigma_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

donde  $\boldsymbol{\mu}_{\mathbf{Y}}$  y  $\Sigma_{\mathbf{Y}\mathbf{Y}}$  son el vector de medias y la matriz de covarianzas de  $\mathbf{Y}$ ,  $\boldsymbol{\mu}_{\mathbf{E}}$  y  $\Sigma_{\mathbf{E}\mathbf{E}}$  son el vector de medias y la matriz de covarianzas de  $\mathbf{E}$  y  $\Sigma_{\mathbf{Y}\mathbf{E}}$  es la matriz de covarianzas de  $\mathbf{Y}$  y  $\mathbf{E}$ .

Entonces, si  $\Sigma_{\mathbf{E}\mathbf{E}}$  es regular o invertible, la distribución de probabilidad condicionada de  $\mathbf{Y}$  dado  $\mathbf{E}$  es normal multivariante de parámetros  $\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  vector de medias y  $\Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  matriz de covarianzas, donde

$$\begin{aligned} \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}} &= \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{Y}\mathbf{E}}\Sigma_{\mathbf{E}\mathbf{E}}^{-1}(\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}}) \\ \Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}} &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{E}}\Sigma_{\mathbf{E}\mathbf{E}}^{-1}\Sigma_{\mathbf{E}\mathbf{Y}} \end{aligned}$$

### Demostración 2.1

Véase Anderson (2003). ■



Por tanto, considerando  $\mathbf{Y}$  el conjunto de variables no evidenciales del problema y  $\mathbf{E}$  el conjunto de variables evidenciales, se puede propagar la evidencia en Redes Bayesianas Gaussianas, aplicando el Teorema 2.1. De esta forma, se pueden obtener los parámetros que describen las probabilidades de interés, dadas por  $f(x_i|e)$  para todo  $X_i \in \mathbf{Y}$ .

El algoritmo que se introduce en este Apartado considera que el conjunto  $\mathbf{E}$  está formado por un único elemento, siendo  $E = \{X_e = e\}$ . La realización iterativa del cálculo de la distribución condicionada normal multivariante permite actualizar, en tiempo lineal, las probabilidades de las variables no evidenciales, dada la evidencia. Así, el algoritmo de propagación define los parámetros condicionados de la distribución normal multivariante de  $\mathbf{Y}|E = e$  como

$$\begin{aligned}\boldsymbol{\mu}^{\mathbf{Y}|E=e} &= \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{Y}E} \frac{1}{\sigma_{ee}} (e - \mu_E) \\ \Sigma^{\mathbf{Y}|E=e} &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}E} \frac{1}{\sigma_{ee}} \Sigma_{E\mathbf{Y}}\end{aligned}\quad (2.12)$$

siendo  $E = e$  el valor que toma dicha variable  $X_e = e$ ,  $\sigma_{ee}$  la varianza de la variable evidencial y  $\mathbf{Y} = \mathbf{X} \setminus E$  las variables no evidenciales.

Tras realizarse este proceso de propagación de la evidencia, se obtienen los parámetros condicionados dados por el vector de medias y la matriz de covarianzas de todas las variables de  $\mathbf{Y}|E = e$ , de forma que, además de poder calcular  $f(x_i|e)$  para todo  $X_i \in \mathbf{Y}$ , se conocen todas las relaciones de dependencia entre las variables de  $\mathbf{Y}$ , información que otros algoritmos no presentan tras su realización.

En esta Memoria se trabaja con este algoritmo porque al obtener el parámetro  $\Sigma^{\mathbf{Y}|E=e}$ , se puede calcular la matriz de precisión  $\mathbf{K}^{\mathbf{Y}|E=e}$ , que determina la nueva estructura de dependencias del resto de variables, dada la evidencia. Esto permite describir nuevas relaciones entre las variables que intervienen en el problema, en las condiciones introducidas por la evidencia actual.

Considerando dicho proceso de propagación de la evidencia como un proceso recursivo en el que se actualiza en cada paso una variable evidencial, el número de operaciones necesarias para actualizar la distribución de probabilidad de las variables de  $\mathbf{Y}$ , es lineal en el número de variables de  $\mathbf{X}$ . Por tanto, este proceso es

sencillo y eficiente para propagar la evidencia en Redes Bayesianas Gaussianas y ha sido utilizado por diversos autores como Castillo (1997b) en la implementación de un método de propagación simbólica.

### Algoritmo de propagación en Redes Bayesianas Gaussianas

ENTRADA: Los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  que definen la Red Bayesiana Gaussiana mediante la distribución conjunta normal multivariante  $N(\boldsymbol{\mu}, \Sigma)$ , siendo  $\mathbf{X} = \{X_1, \dots, X_n\}$  el conjunto de variables de la red, que se particiona tal que  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$  donde  $\mathbf{Y}$  es el conjunto de variables no evidenciales y  $\mathbf{E}$  el conjunto de variables evidenciales.

SALIDA: Los parámetros de la distribución conjunta de las variables no evidenciales de la red, dada la evidencia, es decir,  $\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  y  $\Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  tales que  $\mathbf{Y}|\mathbf{E} = \mathbf{e} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}, \Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}})$ .

Paso 1. Para cada variable evidencial  $X_e \in \mathbf{E}$  calcular los parámetros de la distribución condicionada aplicando (2.12), de forma que el vector de medias de las variables no evidenciales condicionadas a la evidencia  $X_e = e$  es

$$\boldsymbol{\mu}^{\mathbf{Y}|E=e} = \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{Y}E} \frac{1}{\sigma_{ee}} (e - \mu_E)$$

Y la matriz de covarianzas condicionada a la evidencia  $X_e = e$  se calcula como

$$\Sigma^{\mathbf{Y}|E=e} = \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}E} \frac{1}{\sigma_{ee}} \Sigma_{E\mathbf{Y}}$$

Paso 2. Repetir el Paso 1 hasta que no quede ninguna variable evidencial en la red

Paso 3. Para cada variable  $X_i \in \mathbf{Y}$  se tiene que

$$X_i|E \sim N(\mu_i^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}, \sigma_{ii}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}})$$

A continuación, se muestra una aplicación del algoritmo para la Red Bayesiana Gaussiana introducida en el Ejemplo 2.9.

### Ejemplo 2.11

Considérese la Red Bayesiana Gaussiana descrita en el Ejemplo 2.9. Se sabe que la variable  $X_2 = 4$ , por tanto  $E = \{X_2 = 4\}$  y  $\mathbf{Y} = \{X_1, X_3, X_4, X_5\}$ . Se desea calcular la distribución condicionada del resto de variables no evidenciales de la red, dada la evidencia.

En este caso, sólo se tiene evidencia acerca de una de las variables del problema, por tanto, sólo es necesario realizar el Paso 1 del algoritmo una vez. Calculando los parámetros de la distribución condicionada dada la evidencia con (2.12), se obtiene que  $\mathbf{Y}|E \sim N(\boldsymbol{\mu}^{\mathbf{Y}|E=e}, \Sigma^{\mathbf{Y}|E=e})$ , donde

$$\boldsymbol{\mu}^{\mathbf{Y}|E=e} = \begin{pmatrix} 2 \\ 4 \\ 4 \\ 6 \end{pmatrix} \quad \Sigma^{\mathbf{Y}|E=e} = \begin{pmatrix} 3 & 6 & 0 & 6 \\ 6 & 13 & 0 & 13 \\ 0 & 0 & 2 & 4 \\ 6 & 13 & 4 & 24 \end{pmatrix}$$

Teniendo en cuenta que el objeto de estudio de esta red es la variable  $X_5$ , se tiene que, la variable de interés dada la evidencia es normal multivariante, tal que

$$X_5|X_2 = 4 \sim N(6, 24)$$

Como se puede observar en la matriz de covarianzas  $\Sigma^{\mathbf{Y}|E=e}$ , se mantienen las relaciones de independencia entre las variables  $X_1$  y  $X_4$  y entre  $X_3$  y  $X_4$ .

A continuación se presenta la matriz de precisión  $\mathbf{K}^{\mathbf{Y}|E=e}$ , que muestra las relaciones de independencia condicionada que se tienen entre las variables de  $\mathbf{Y}$  dada la evidencia  $E$ .

$$\mathbf{K}^{\mathbf{Y}|E=e} = \begin{pmatrix} \frac{13}{3} & -2 & 0 & 0 \\ -2 & \frac{4}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & \frac{2}{3} & \frac{11}{6} & -\frac{2}{3} \\ 0 & -\frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}$$

Se tiene que dada la evidencia acerca de  $X_2$  las variables  $X_1$  y  $X_4$  siguen siendo condicionalmente independientes dados  $X_3$  y  $X_5$ , al igual que  $X_1$  y  $X_5$  dados  $X_3$  y  $X_4$ .

Los resultados concluidos para las variables de la red, de independencia e independencia condicionada, pueden obtenerse igualmente aplicando el criterio de  $d$ -separación al DAG que representa la red cuando se tiene evidencia acerca del valor que toma la variable  $X_2$ .



# Sensibilidad en Redes Bayesianas Gaussianas

## 3.1. Introducción

Como se ha enunciado en el Capítulo 2, el proceso de construcción de una Red Bayesiana requiere de la ayuda de los expertos en el campo de aplicación de la red. Es necesario especificar las dependencias entre las variables del problema para diseñar el DAG, e indicar la parte cuantitativa de la red introduciendo los valores convenientes de los parámetros, ya sean de las distribuciones condicionadas o de la distribución conjunta, que determinan la red.

A pesar de que en la especificación de la parte cuantitativa de la red es suficiente con describir los parámetros individualmente para obtener la distribución conjunta, esta asignación puede presentar consecuencias poco intuitivas sobre las variables cuyos parámetros no se definen directamente.

Por tanto, este proceso de diseño y definición de la Red Bayesiana suscita la posibilidad de asignar erróneamente los parámetros y obtener, por consiguiente, resultados inadecuados, tras realizar el proceso de propagación de la evidencia. Dichos resultados vienen dados por la distribución de probabilidad de interés, siendo una distribución final de una variable de interés o una distribución final del conjunto de variables de interés. Esta distribución final se calcula cuando se conocen los valores que toman un conjunto de variables en un caso específico, siendo éstas las variables evidenciales.

Como consecuencia, se muestra conveniente la realización de un análisis de sensibilidad, que indique la sensibilidad de los resultados a perturbaciones en los parámetros, con el objeto de precisar la distribución inicial de aquellas variables con más impacto sobre el resultado de interés en la red.

En este Capítulo se enumeran algunos análisis de sensibilidad propuestos para Redes Bayesianas en general y posteriormente, se desarrolla el análisis de sensibilidad que se propone para Redes Bayesianas Gaussianas.

En Redes Bayesianas Gaussianas con una variable de interés  $X_i$ , se realiza un análisis de sensibilidad valorando el efecto de los parámetros inciertos en la distribución obtenida tras propagarse la evidencia,  $f(x_i|e)$ . En la Sección 3.3, el análisis propuesto consiste en calcular una medida de sensibilidad, con la divergencia de Kullback-Leibler, que compara dos funciones de densidad, la densidad final normal multivariante del modelo original,  $f(x_i|e)$  y esa misma densidad obtenida para un modelo en el que se cuantifica la incertidumbre acerca de los parámetros que describen la red inicialmente.

En este contexto, se estudia el comportamiento de la medida de sensibilidad para situaciones extremas, dadas cuando los parámetros son muy distintos o muy similares a los introducidos y se concluyen resultados intuitivamente ajustados. También, se estudia el comportamiento de la medida de sensibilidad cuando la relación entre la variable de interés y la variable evidencial es extrema.

Posteriormente, como una generalización del análisis de sensibilidad desarrollado, en la Sección 3.5 se presenta una metodología para estudiar la sensibilidad de un conjunto de variables de interés y un conjunto de variables evidenciales.

Con el análisis de sensibilidad descrito, surge la idea de estudiar la robustez de una Red Bayesiana Gaussiana. Para ello, se propone una generalización de la metodología propuesta al estudiar la sensibilidad de las Redes Bayesianas Gaussianas, que sirve para determinar la robustez de la red frente a ciertos tipos de perturbaciones sobre los parámetros.

A lo largo de todo este Capítulo, los resultados se presentan sobre un ejemplo específico que sirve para ilustrar los distintos conceptos introducidos.

Finalmente, se enuncian algunas conclusiones y comentarios de los análisis

propuestos, así como futuras líneas de investigación relacionadas con el estudio de la sensibilidad en Redes Bayesianas Gaussianas.

### **3.2. Análisis de sensibilidad para Redes Bayesianas: aproximación histórica**

En los últimos años, se han desarrollado diversas técnicas para estudiar la sensibilidad de una Red Bayesiana, debido a que el proceso de construcción del modelo requiere que un grupo de expertos en el problema de interés enumeren los parámetros que definen la red. Esta tarea es compleja porque generalmente el problema está formado por un gran número de parámetros y se tiene un conocimiento parcial de los datos. Además, como cita Coupé, et al. (2000), "la experiencia muestra que los expertos son reacios a asignar los parámetros requeridos porque creen que no son capaces de asignarlos con un alto grado de precisión". Por lo tanto, esta asignación de valores a los parámetros puede ser inexacta. Un ejemplo de las dificultades que aparecen al concretar una Red Bayesiana, puede verse en Onisko, et al. (1999).

Como consecuencia de una asignación inadecuada de los valores de los parámetros, los resultados obtenidos de la red, tras actualizar la información que se tiene acerca de las variables evidenciales del problema, pueden ser también inexactos, dependiendo de la sensibilidad del modelo.

En esta Sección se recogen los análisis de sensibilidad más destacados desarrollados para Redes Bayesianas.

La mayoría de los análisis de sensibilidad propuestos, se han desarrollado para Redes Bayesianas Discretas. En este caso, los parámetros que tienen que asignar los expertos, son los valores de las probabilidades condicionadas a la ocurrencia de los padres en el DAG, es decir  $p(x_i|pa(X_i))$  para todas las variables  $X_i$  del problema.

A continuación, se citan los análisis de sensibilidad más destacados que se han introducido para Redes Bayesianas Discretas.

Laskey (1995) desarrolla una metodología para estudiar la sensibilidad de las Redes Bayesianas Discretas. En su desarrollo, trabaja con la sensibilidad sobre



una única variable de interés  $X_i$  en el modelo y basa el análisis de sensibilidad propuesto en calcular, para cada parámetro, una medida denominada *valor de sensibilidad*, de forma que se calcula el valor de sensibilidad de la salida de la red, tras conocerse la evidencia, considerando cada vez un parámetro distinto.

Los valores de sensibilidad de la red se calculan mediante las derivadas parciales de la salida de la red respecto a las entradas del modelo. Por tanto, se estudia la sensibilidad mediante las derivadas de la probabilidad final de la variable de interés  $p(x_i|e)$ , con respecto a los parámetros que definen la red, que son las probabilidades condicionadas por sus padres en el DAG, dadas por  $p(x_j|pa(X_j))$ , para todo  $X_j$ .

Los valores de sensibilidad, por calcularse con las derivadas parciales, miden el impacto de cambios pequeños en los parámetros de la red sobre la salida de la misma, recogida ésta por la probabilidad final de interés.

Coupé, et al. (2000) desarrollan un algoritmo para construir una Red Bayesiana Discreta y a la vez estudiar la sensibilidad de los resultados a los parámetros que se van introduciendo en dicha red. Para ello, se centran en una Red Bayesiana Discreta con una variable de interés  $X_i$ , construyen inicialmente una red con valores asociados a los parámetros que serán mejorados. En cada paso del algoritmo varían un único parámetro, con lo cual se modifica la distribución condicionada inicial.

En este desarrollo, Coupé, et al. (2000) argumentan que se pueden obtener los parámetros que han de fijar los expertos, realizando, en un proceso iterativo, un análisis de sensibilidad de la red que se está construyendo. El análisis finaliza cuando se obtiene una Red Bayesiana Discreta que es satisfactoria para los expertos.

Por tanto, la idea básica del análisis de sensibilidad propuesto consiste en variar sistemáticamente la asignación inicial de uno de los parámetros de la red, sobre un intervalo de valores plausible, y estudiar el efecto sobre los resultados de la red.

Desafortunadamente, el análisis de sensibilidad desarrollado está basado en la modificación de los algoritmos de propagación y esto requiere mucho tiempo de realización debido a la gran cantidad de cálculos necesarios.

Posteriormente, Coupé, et al. (2002) mejoran la eficiencia del análisis de

sensibilidad propuesto, introduciendo en el algoritmo algunas propiedades que convierten el análisis de sensibilidad en un análisis más sencillo computacionalmente. De nuevo, el análisis de sensibilidad desarrollado considera una variable de interés  $X_i$  en el modelo y modifica en cada paso un único parámetro.

En este caso, se reduce la complejidad computacional del algoritmo ya que se identifican los parámetros del modelo, dados por las probabilidades condicionadas por sus padres, cuyas modificaciones no influyen en la probabilidad final de interés, dada por  $p(X_i|e)$ . También, se reduce el cálculo de la probabilidad de interés, al cociente de dos funciones lineales en el parámetro de estudio. Ambas propiedades permiten reducir considerablemente el tiempo de ejecución del análisis de sensibilidad desarrollado.

En el análisis de sensibilidad propuesto por Chan, et al. (2004), se varía un conjunto de parámetros a la vez, teniendo que tratar con las probabilidades condicionadas de un conjunto de variables, dada la ocurrencia de sus padres en el DAG, en lugar de con la probabilidad condicionada de una única variable. Además, muestran como encontrar el parámetro que más perturba los resultados de la red, dados por la probabilidad final de interés condicionada a la evidencia en la red. De nuevo se trabaja con una variable de interés  $X_i$ , aunque ahora se estudian simultáneamente un conjunto de parámetros inexactos.

Chan, et al. (2005) introducen un nuevo análisis de sensibilidad para Redes Bayesianas Discretas, basado en calcular una medida que evalúa la distancia entre dos distribuciones de probabilidad. La medida propuesta consiste en comparar los parámetros que describen la red calculando la distancia entre el máximo y el mínimo de la razón entre dichos parámetros. Además, dicha medida puede utilizarse para limitar los cambios resultantes de cada perturbación local de los parámetros de la red.

Otros autores como Castillo, et al. (1997c), Kjærulff, et al. (2000) o Bednarski, et al. (2004) también han estudiado la sensibilidad de las Redes Bayesianas Discretas, diseñando en cada caso un análisis con características propias.

Así, Castillo, et al. (1997c) desarrollan un método eficiente computacionalmente, basado en explotar la estructura de la probabilidad de interés, dada por la probabilidad final de la variable de interés, cuando se ha introducido evidencia. Para ello, apoyado en un método de propagación simbólica (propagando valores desconocidos) obtiene unas cotas para las probabilidades marginales de

las variables de la red.

Kjærulff, et al. (2000) presentan un análisis de sensibilidad que permite introducir más de un parámetro inexacto a la vez, para conocer como afectan dichos parámetros a todos los posibles resultados obtenidos de la red, tras propagar la evidencia. Así, generalizan otros análisis considerando más de una variable de interés y más de un parámetro inexacto a la vez.

Y Bednarski, et al. (2004) centran su análisis de sensibilidad en identificar el conjunto de sensibilidades que afecta más a la variable de interés, consiguiendo así reducir cálculos.

Cuando se trabaja con Redes Bayesianas Gaussianas, los parámetros de entrada que definen la red cambian. En el modelo Gaussiano, dichos parámetros no son las probabilidades condicionadas, sino los elementos del vector de medias  $\boldsymbol{\mu}$  y la matriz de covarianzas  $\Sigma$  que definen la distribución conjunta de las variables del modelo, o la media y la varianza de cada una de las variables condicionadas por los padres en el DAG.

La literatura sobre sensibilidad en Redes Bayesianas Gaussianas ha sido más bien escasa hasta la fecha.

Se dispone de la metodología propuesta por Castillo, et al. (1997b), basada en la propagación simbólica. Este concepto de propagación simbólica, enunciado al citar el estudio en Redes Bayesianas Discretas presentado por Castillo, et al. (1997c), consiste en realizar la propagación de la evidencia partiendo de una distribución con parámetros desconocidos o inciertos representados mediante un símbolo general, de forma que puede tomar cualquier valor. En este contexto se muestra la propagación simbólica sobre los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$ , descritos con algunos elementos conocidos definidos por un valor numérico y otros inciertos definidos simbólicamente. Los resultados básicos para efectuar el posterior análisis permiten representar los parámetros de la distribución condicionada  $\mathbf{Y}|\mathbf{E} = \mathbf{e}$  como funciones racionales de los valores numéricos y simbólicos introducidos.

Así, Castillo, et al. (1997b) muestran como obtener el vector de medias  $\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  y la matriz de covarianzas  $\Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}$  final, tras realizarse la propagación de la evidencia simbólicamente, utilizando el algoritmo de propagación en Redes Bayesianas Gaussianas descrito en el Apartado 2.4.4, que es sencillo de manejar

y simplifica los cálculos. Tras efectuar la propagación simbólica de la evidencia se tiene como salida de la red la distribución de  $\mathbf{Y}|\mathbf{E} = \mathbf{e} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}, \Sigma^{\mathbf{Y}|\mathbf{E}=\mathbf{e}})$  y si se trabaja con una variable de interés en el problema, se tiene que  $X_i|E = e \sim N(\mu_i^{\mathbf{Y}|\mathbf{E}=\mathbf{e}}, \sigma_{ii}^{\mathbf{Y}|\mathbf{E}=\mathbf{e}})$ , apareciendo los parámetros en función de los valores inciertos introducidos simbólicamente en  $\boldsymbol{\mu}$  y  $\Sigma$ .

Con este resultado, se posibilita el desarrollo de un análisis de sensibilidad basado en la estructura de los parámetros finales que describen a las variables de interés, no evidenciales.

Este análisis de sensibilidad se concreta en posteriores trabajos de los autores. Así, Castillo, et al. (2001) trabajan con la estructura algebraica de las medias y las varianzas condicionadas, dadas por funciones cuadráticas de los parámetros, para estudiar la sensibilidad de la red, calculando los valores de sensibilidad, definidos por Laskey (1995) para estudiar la sensibilidad en Redes Bayesianas Discretas.

Con esta idea, en Castillo, et al. (2003) se fija una variable de interés  $X_i$  en la Red Bayesiana Gaussiana y se estudia la salida de la red, recogida por la distribución final de  $X_i$ , condicionada a la evidencia introducida en el modelo, para cada parámetro incierto que se ha introducido al efectuar la propagación simbólica. Dichos parámetros, para los cuales se realiza el cálculo de los valores de sensibilidad, son los elementos desconocidos representados simbólicamente que aparecen en el vector de medias  $\boldsymbol{\mu}$  y en la matriz de covarianzas  $\Sigma$  y que describen la distribución conjunta de la red al inicio.

Por tanto, apoyados en los resultados obtenidos en trabajos previos, desarrollan un análisis de sensibilidad basado en el cálculo de los valores de sensibilidad de la red, dados por las derivadas parciales de la distribución final de la variable de interés, para cada parámetro incierto que describe la red.

Los valores de sensibilidad, obtenidos tras calcularse todas las derivadas parciales, son expresiones cerradas de la distribución de probabilidad final de interés con respecto a los parámetros. Sin embargo, al igual que en el desarrollo presentado por Laskey (1995), son medidas de sensibilidad locales.

Por consiguiente, el análisis de sensibilidad propuesto presenta una desventaja al centrarse sólo en pequeños cambios de los parámetros y no permitir estudiar la salida de la red para otro tipo de cambios.

En la siguiente Sección se desarrolla un análisis de sensibilidad para Redes Bayesianas Gaussianas basado en el cálculo de una medida de divergencia. En este caso, se mide el impacto de distintas perturbaciones en los parámetros  $\mu$  y  $\Sigma$ , que definen la distribución conjunta inicial de la Red Bayesiana Gaussiana, desde una perspectiva global cuantificando su efecto en la distribución de probabilidad final.

### 3.3. Análisis de sensibilidad de una vía para Redes Bayesianas Gaussianas

Como se ha introducido anteriormente, un *análisis de sensibilidad* se puede definir como una técnica que sirve para estudiar los efectos de perturbaciones en los parámetros que definen un modelo matemático sobre las salidas de dicho modelo (Saltelli, 2004).

En las Redes Bayesianas, el análisis de sensibilidad estudia el efecto de valores inciertos o inexactos en los parámetros que definen la red sobre las salidas de la misma, dadas por la distribución final de la variable o variables de interés del problema, conocida la evidencia. Para ello, los valores inciertos de los parámetros se modifican o perturban y es el efecto de dicha perturbación sobre la salida de la red lo que cuantifica el análisis de sensibilidad para una Red Bayesiana. Esta es la aproximación que se propone en esta Memoria.

Se pueden introducir distintos tipos de análisis de sensibilidad en función de los parámetros inciertos que se modifican cada vez, al estudiar la sensibilidad del modelo.

El análisis de sensibilidad más simple, para una Red Bayesiana, se denomina *análisis de sensibilidad de una vía* y consiste en modificar uno sólo de los parámetros que describen la red, manteniendo fijos el resto de parámetros, siendo dichos parámetros los elementos que intervienen en  $\mu$  y  $\Sigma$ . De esta forma, el análisis de sensibilidad revela el efecto de un parámetro incierto, que ha sido variado, sobre la probabilidad final de interés.

Son análisis de sensibilidad de una vía los descritos por autores como Laskey (1995), Coupé, et al. (2000) o Castillo, et al. (2003).

En un *análisis de sensibilidad de dos vías* de una Red Bayesiana, dos son los parámetros que se consideran inciertos al definir el modelo y se modifican a la vez. En este caso, el análisis de sensibilidad estudia el efecto separado de la variación de cada uno de estos parámetros, además del efecto de su variación conjunta, sobre la salida de la red.

Aunque a medida que aumenta el número de parámetros inciertos, es más difícil la interpretación del análisis de sensibilidad, se define el *análisis de sensibilidad de n vías* cuando se modifican a la vez un conjunto de parámetros, estudiando la sensibilidad de la salida de la red a las variaciones individuales y conjuntas de los parámetros inciertos del problema. Siendo en todos los casos dicha salida, la distribución final de interés obtenida tras la propagación de la evidencia.

Autores citados en la Sección anterior como Kjærulff, et al. (2000) o Chan, et al. (2004), desarrollan análisis de sensibilidad de n vías para Redes Bayesianas.

En esta Sección se desarrolla un nuevo análisis de sensibilidad de una vía para Redes Bayesianas Gaussianas. Dicho análisis constituye uno de los elementos inéditos introducidos en esta Memoria. Algunos resultados fundamentales del mismo han sido aceptados para su publicación en Gómez-Villegas, et al. (2007).

Se considera una variable de interés  $X_i$ , de forma que el análisis de sensibilidad estudia el efecto producido al perturbar un parámetro incierto de la red, siendo dicho parámetro un elemento de  $\boldsymbol{\mu}$  y  $\Sigma$ , sobre la salida de la red dada por la función de probabilidad final de la variable de interés tras propagarse la evidencia, es decir  $f(x_i|e)$ .

Para obtener la salida de la red, se aplica el algoritmo de propagación descrito en el Apartado 2.4.4 para propagar la evidencia en la Red Bayesiana Gaussiana de estudio, en las condiciones que se detallan.

El análisis de sensibilidad que se propone en esta Memoria, consiste en calcular sistemáticamente, considerando incierto un único parámetro de la red en cada paso y el resto fijos, una medida de sensibilidad que se define en el Apartado 3.3.1 y que compara la distribución de probabilidad final de la variable de interés, dada la evidencia, obtenida para dos modelos distintos: el modelo original y el modelo perturbado.

Se define el *modelo original* como la Red Bayesiana Gaussiana descrita por los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  con los valores asignados inicialmente por los expertos.

Se define el *modelo perturbado* como la Red Bayesiana Gaussiana descrita por los parámetros perturbados  $\boldsymbol{\mu}^\delta$  o  $\Sigma^\delta$  que se tiene tras modificar alguno de sus elementos; esta modificación se obtiene añadiéndole la perturbación  $\delta \in \mathbb{R}$  al parámetro incierto. Por tanto, para cada elemento del vector de medias y la matriz de covarianzas, se considera el modelo perturbado como la Red Bayesiana Gaussiana que se obtiene al modificar dicho elemento cuando se le suma la perturbación  $\delta$ . Esta perturbación se obtiene de los expertos, y cuantifica la inexactitud asociada al parámetro incierto. En caso de no poder cuantificarse dicha perturbación, se estudian y comparan las distintas medidas de sensibilidad obtenidas para diferentes valores de  $\delta \in \mathbb{R}$ , representando las medidas de sensibilidad en función de  $\delta$ , lo que permite también valorar cualitativamente el efecto de la perturbación.

Con la idea del análisis de sensibilidad introducida, en el Apartado 3.3.2 se profundiza en el método, calculando la medida de sensibilidad para cada uno de los parámetros que describen la distribución normal multivariante que define la red. Cabe señalar que algunos parámetros inciertos no influyen en la variable de interés, por estar trabajando con el modelo conjunto dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$  que describe la Red Bayesiana Gaussiana y por el método de propagación utilizado.

En el Apartado 3.3.3 se estudian las medidas de sensibilidad resultantes, cuando la variable de interés  $X_i$  y la variable evidencial  $X_e$  son dependientes o independientes. En estos casos, se simplifica notablemente la expresión asociada a las distintas medidas de sensibilidad calculadas.

Finalmente en el Apartado 3.3.4 se introduce un Algoritmo, implementado en Susi (2006), que se ejecuta en tiempo lineal, para realizar el análisis de sensibilidad propuesto. La utilización del mismo facilita el estudio de la sensibilidad para cualquier Red Bayesiana Gaussiana.

Uno de los objetivos de este análisis de sensibilidad es mostrar la perturbación del parámetro o parámetros inciertos que afecta más a la salida del modelo, debiéndose, por consiguiente, poner más cuidado al definir estos elementos en la red. Si los resultados del análisis de sensibilidad muestran cierta insensibilidad de la distribución final de interés, se pueden obtener conclusiones acerca de la

robustez del modelo, como se verá en la Sección 3.6.

### 3.3.1. Medida de sensibilidad

El análisis de sensibilidad que se propone en esta Memoria está basado en el cálculo de una medida que se ha utilizado en otros contextos para medir la discrepancia entre dos distribuciones de probabilidad y que se propone aquí para comparar la salida de la red, considerando el modelo original y el modelo perturbado de la Red Bayesiana Gaussiana en estudio.

La medida es la divergencia de Kulback-Leibler. Esta medida de divergencia, introducida por Kullback, et al. (1951), es la medida de discrepancia más comúnmente utilizada para comparar dos distribuciones estructuralmente y se define como

$$KL(f(w), f'(w)) = \int_{-\infty}^{\infty} f(w) \ln \frac{f(w)}{f'(w)} dw$$

donde las funciones  $f(w)$  y  $f'(w)$  son dos funciones de densidad definidas para el mismo conjunto de variables.

En el estudio de la sensibilidad de una Red Bayesiana Gaussiana con una variable de interés  $X_i$ , la salida de la red tras propagarse la evidencia viene dada por la distribución final de interés, es decir, por la función de densidad de la variable de interés condicionada a la evidencia introducida en la red, dada por  $f(x_i|e)$ .

Cuando se consideran los modelos original y perturbado de la Red Bayesiana Gaussiana en estudio, se calcula la discrepancia entre las salidas de la red, dadas por  $f(x_i|e)$  para el modelo original y  $f(x_i|e, \delta)$  para el modelo perturbado. Lo que se propone es calcular como medida de sensibilidad, la divergencia de Kulback-Leibler entre las densidades finales de interés obtenidas para el modelo original y el modelo perturbado.

**Definición 3.1** (*Medida de sensibilidad*)

Sea  $(\mathcal{D}, P)$  una Red Bayesiana Gaussiana con distribución conjunta  $N(\boldsymbol{\mu}, \Sigma)$ . Sea  $f(x_i|e)$  la densidad final de interés obtenida tras propagar la evidencia y sea



$f(x_i|e, \delta)$  esa misma densidad, obtenida tras añadir la perturbación  $\delta$  a uno de los elementos que componen los parámetros de la distribución inicial conjunta.

Se define la medida de sensibilidad como

$$S^{p_j}(f(x_i|e), f(x_i|e, \delta)) = \int_{-\infty}^{\infty} f(x_i|e) \ln \frac{f(x_i|e)}{f(x_i|e, \delta)} dx_i \quad (3.1)$$

donde el subíndice  $p_j$  indica el elemento que ha sido modificado añadiéndole la perturbación  $\delta$ , siendo dicho elemento en el modelo perturbado  $p_j^\delta = p_j + \delta$ .

En el siguiente Apartado se calcula la expresión que toma la medida de sensibilidad considerando inexactos, de forma sistemática, cada uno de los parámetros o elementos de  $\boldsymbol{\mu}$  y  $\Sigma$ .

### 3.3.2. Análisis de sensibilidad de una vía

En este Apartado se detalla el análisis de sensibilidad de una vía propuesto para estudiar la sensibilidad de una Red Bayesiana Gaussiana con una variable de interés  $X_i$ .

El mecanismo de análisis consiste en un proceso iterativo que compara en cada paso, considerando la especificación cuantitativa de una Red Bayesiana Gaussiana, el modelo original dado por la distribución normal multivariante de las variables  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$ , con el modelo perturbado, también normal multivariante, obtenido tras añadir una perturbación  $\delta \in \mathbb{R}$  a uno de los elementos de los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$ .

Por tanto, lo primero que se ha de saber al realizar el análisis de sensibilidad son los parámetros inexactos de la red. De esta forma, en cada paso del análisis de sensibilidad se considera un parámetro inexacto y por tanto un modelo perturbado concreto, obtenido tras sumarle la perturbación  $\delta$  a dicho parámetro del modelo original.

Para comparar en cada paso ambos modelos se calcula la medida de sensibilidad, que compara la salida de la red dada por la densidad final de interés  $f(x_i|e)$ , obtenida tras propagar la evidencia sobre el modelo original, con la densidad final de interés  $f(x_i|e, \delta)$ , obtenida tras propagar la evidencia sobre el modelo perturbado considerado.

Por lo tanto, para obtener estas salidas de la red, es necesario propagar la evidencia en ambos modelos, el original y el perturbado. Para ello, se trabaja con el algoritmo de propagación para Redes Bayesianas Gaussianas, introducido en el Capítulo 2 en el Apartado 2.4.4. Es importante recordar, que es un proceso iterativo basado en el cálculo de la distribución condicionada normal multivariante, considerándose en cada paso una única variable evidencial  $X_e = e$ .

Con la propagación de la evidencia se calculan los parámetros de la distribución condicionada normal multivariante del resto de las variables no evidenciales,  $\mathbf{Y}|\mathbf{E}$ . Así, los nuevos parámetros de las variables no evidenciales  $\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=e}$  y  $\Sigma^{\mathbf{Y}|\mathbf{E}=e}$ , que se obtienen aplicando las expresiones descritas en (2.12), son

$$\begin{aligned}\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}=e} &= \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{Y}\mathbf{E}} \frac{1}{\sigma_{ee}} (e - \mu_{\mathbf{E}}) \\ \Sigma^{\mathbf{Y}|\mathbf{E}=e} &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{E}} \frac{1}{\sigma_{ee}} \Sigma_{\mathbf{E}\mathbf{Y}}\end{aligned}$$

Aunque se dispone de toda la información que describe la distribución final de las variables no evidenciales  $\mathbf{Y}|\mathbf{E}$ , la Red Bayesiana Gaussiana en estudio tiene una variable de interés  $X_i$ . En este caso, la salida de interés de la red viene dada por la distribución final de  $X_i$  tras propagarse la evidencia, es decir

$$X_i|E = e \sim N(\mu_i^{\mathbf{Y}|\mathbf{E}=e}, \sigma_{ii}^{\mathbf{Y}|\mathbf{E}=e})$$

donde

$$\begin{aligned}\mu_i^{\mathbf{Y}|\mathbf{E}=e} &= \mu_i + \frac{\sigma_{ie}}{\sigma_{ee}} (e - \mu_e) \\ \sigma_{ii}^{\mathbf{Y}|\mathbf{E}=e} &= \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}}\end{aligned}$$

siendo  $\mu_i$  y  $\sigma_{ii}$  la media y varianza iniciales, respectivamente, que describen la variable  $X_i$  cuando se detalla la Red Bayesiana Gaussiana,  $\sigma_{ie}$  la covarianza entre la variable de interés  $X_i$  y la variable evidencial  $X_e$ , y  $\mu_e$  y  $\sigma_{ee}$  la media y varianza iniciales, respectivamente, que describen la variable evidencial  $X_e$ .

Como se puede observar la distribución final de la variable de interés, sólo depende de la evidencia  $e$ , de los parámetros iniciales de la variable evidencial  $X_e$  y de la propia variable de interés  $X_i$ . Por tanto, cuando los parámetros inciertos sean de variables no evidenciales distintas de la variable de interés, tales que

$X_j \in \mathbf{Y}$  para todo  $j \neq i$ , dicha incertidumbre no afecta a la distribución final de la variable de interés.

Así, al calcular la medida de sensibilidad para cada posible parámetro incierto, se ha de diferenciar entre los parámetros de la variable de interés  $X_i$ , los parámetros de la variable evidencial  $X_e$  y los parámetros del resto de variables no evidenciales  $X_j \in \mathbf{Y}$  para todo  $j \neq i$ . Por tanto, cuando se modeliza un problema con una variable de interés mediante una Red Bayesiana Gaussiana, todas aquellas variables que no puedan ser evidenciales no influyen en la salida de la red.

Al realizarse un análisis de sensibilidad de una vía, se obtiene una medida de sensibilidad para cada modelo perturbado considerado. Posteriormente, se pueden comparar las medidas obtenidas, bien con el valor de las mismas, si se han cuantificado las perturbaciones asociadas a cada parámetro inexacto, bien gráficamente, presentando las medidas en función de la perturbación  $\delta \in \mathbb{R}$ .

En el Apartado 3.3.4 se presenta un Algoritmo para efectuar el análisis de sensibilidad propuesto. Dicho Algoritmo se basa en el cálculo de la medida de sensibilidad para cada uno de los parámetros inciertos que influyen en la salida, dada por la distribución de la variable  $X_i$  conocida la evidencia. El Algoritmo presentado, compara el modelo original con el modelo perturbado, introduciendo las perturbaciones asociadas a cada modelo perturbado en un solo paso, mediante un vector y una matriz que muestran las perturbaciones asociadas a cada elemento incierto de los parámetros que describen el modelo original.

A continuación, se presenta el cálculo de la medida de sensibilidad, diferenciando los resultados en función del parámetro incierto que se está perturbando, es decir, dependiendo de si la perturbación  $\delta$  se está sumando a un elemento del vector de medias  $\boldsymbol{\mu}$  o a un elemento de la matriz de covarianzas  $\Sigma$ .

Dentro de los casos descritos cuando se perturba un elemento del vector de medias, cabe señalar la expresión de la medida de sensibilidad cuando la media de la variable evidencial  $\mu_e$  es incierta. Esta expresión, coincide con la medida de sensibilidad que se obtiene cuando se considera inexacto el valor de la evidencia  $e$ . Por tanto, aunque el análisis de sensibilidad que se presenta estudia los parámetros inciertos que más afectan a los resultados de la red, también se puede considerar simétricamente el caso asociado a un valor de la evidencia  $e$  variable, calculando la medida de sensibilidad con la expresión (3.3) asociada a

la incertidumbre en  $\mu_e$ .

Además, para los casos de incertidumbre en los elementos de la matriz de covarianzas, el parámetro  $\delta$  no puede tomar cualquier valor de la recta real, ya que es necesario que la matriz de covarianzas perturbada sea definida positiva.

### Sensibilidad para el vector de medias

Cuando el parámetro inexacto es un elemento del vector de medias  $\boldsymbol{\mu}$ , el modelo perturbado que describe la parte cuantitativa de la Red Bayesiana Gaussiana, viene dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^\delta, \Sigma)$ , donde  $\boldsymbol{\mu}^\delta$  puede considerarse de tres formas distintas, en función de las siguientes situaciones:

1. La media de la variable de interés  $X_i$  es inexacta, por tanto, en el modelo perturbado dicho parámetro pasa a ser

$$\mu_i^\delta = \mu_i + \delta : \forall \delta \in \mathbb{R}$$

El modelo perturbado es tal que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^\delta, \Sigma)$ , donde el vector de medias perturbado se especifica mediante  $\boldsymbol{\mu}^\delta = (\mu_1, \dots, \mu_i + \delta, \dots, \mu_n)^\mathbf{T}$ .

2. La media de la variable evidencial  $X_e$  es incierta, de forma que en el modelo perturbado el parámetro es tal que

$$\mu_e^\delta = \mu_e + \delta : \forall \delta \in \mathbb{R}$$

El modelo perturbado, dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^\delta, \Sigma)$ , tiene por vector de medias  $\boldsymbol{\mu}^\delta = (\mu_1, \dots, \mu_i, \dots, \mu_e + \delta, \dots, \mu_n)^\mathbf{T}$ .

3. La media de cualquier variable no evidencial  $X_j$ , distinta de la variable de interés, es incierta, siendo el parámetro en el modelo perturbado

$$\mu_j^\delta = \mu_j + \delta : \forall \delta \in \mathbb{R}$$

El modelo perturbado es tal que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^\delta, \Sigma)$ , donde el vector de medias perturbado es  $\boldsymbol{\mu}^\delta = (\mu_1, \dots, \mu_j + \delta, \dots, \mu_i, \dots, \mu_e, \dots, \mu_n)^\mathbf{T}$ .

En la Proposición 3.1 se calcula la medida de sensibilidad considerando que la relación lineal entre  $X_i$  y  $X_e$  es tal que el coeficiente de correlación lineal no toma valores extremos, siendo  $\rho_{ie}^2 \in (0, 1)$ , para los tres casos anteriormente expuestos.

**Proposición 3.1**

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $X_i$  variable de interés y  $X_e$  variable evidencial y sea  $\rho_{ie}^2 \in (0, 1)$ . Si se considera la perturbación  $\delta \in \mathbb{R}$  añadida a cualquier elemento del vector de medias  $\boldsymbol{\mu}$ , la medida de sensibilidad (3.1) es tal que

1. Cuando se suma la perturbación a la media de  $X_i$ , siendo  $\mu_i^\delta = \mu_i + \delta$ , la distribución final de la variable de interés tras la propagación de la evidencia es

$$X_i|E = e, \delta \sim N(\mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e})$$

con

$$\mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} + \delta$$

La medida de sensibilidad es tal que

$$S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \quad (3.2)$$

2. Cuando se suma la perturbación a la media de la variable evidencial  $X_e$ , con  $\mu_e^\delta = \mu_e + \delta$ , la distribución final de la variable de interés, tras la propagación de la evidencia, es

$$X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e}\right)$$

donde

$$\mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} - \frac{\sigma_{ie}}{\sigma_{ee}} \delta$$

La medida de sensibilidad viene dada por la siguiente expresión

$$S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) = \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \left(\frac{\sigma_{ie}}{\sigma_{ee}}\right)^2 \quad (3.3)$$

3. Si la perturbación  $\delta$  se suma a la media de cualquier otra variable no evidencial distinta de la variable de interés, entonces, dicha perturbación no influye sobre la variable de interés  $X_i$ , siendo  $f(x_i|e, \delta) = f(x_i|e)$ . Por tanto, la medida de sensibilidad es cero.

**Demostración 3.1**

Los parámetros finales de modelo perturbado se obtienen directamente aplicando (2.12). Se ha de tener en cuenta en cada caso un modelo inicial perturbado.

Para calcular la medida de sensibilidad se trabaja con la divergencia de Kullback-Leibler. Con funciones de densidad normales se tiene que

$$KL(f(w), f'(w)) = \frac{1}{2} \left[ \ln \left( \frac{\sigma^{2'}}{\sigma^2} \right) + \frac{\sigma^2}{\sigma^{2'}} + \frac{(\mu' - \mu)^2}{\sigma^{2'}} - 1 \right] \quad (3.4)$$

donde  $f(w)$  es una distribución  $N(\mu, \sigma^2)$  y  $f'(w)$  es una  $N(\mu', \sigma^{2'})$ .

Entonces, la medida de sensibilidad se puede calcular como

$$S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e, \delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e, \delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_{ii}^{Y|E=e, \delta}} - 1 \right]$$

Para cada caso presentado en la demostración, se tiene que

$$\begin{aligned} 1. S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) &= \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e, \delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e, \delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_{ii}^{Y|E=e, \delta}} - 1 \right] \\ &= \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \end{aligned}$$

$$\text{siendo } \mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} + \delta \quad \text{y} \quad \sigma_{ii}^{Y|E=e, \delta} = \sigma_{ii}^{Y|E=e}$$

$$\begin{aligned} 2. S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) &= \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e, \delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e, \delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_{ii}^{Y|E=e, \delta}} - 1 \right] \\ &= \frac{1}{2} \left[ \frac{\left( -\frac{\sigma_{ie}}{\sigma_{ee}} \delta \right)^2}{\sigma_{ii}^{Y|E=e, \delta}} \right] \\ &= \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \left( \frac{\sigma_{ie}}{\sigma_{ee}} \right)^2 \end{aligned}$$

$$\text{donde } \mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} - \frac{\sigma_{ie}}{\sigma_{ee}} \delta \quad \text{y} \quad \sigma_{ii}^{Y|E=e, \delta} = \sigma_{ii}^{Y|E=e}$$

3. Tras propagarse la evidencia se comprueba que  $f(x_i|e, \delta) = f(x_i|e)$ . En este caso, la medida de sensibilidad es cero. ■

Las expresiones obtenidas para la medida de sensibilidad, (3.2) y (3.3), son cuadráticas en función del valor de la perturbación  $\delta$  y dichas expresiones se pueden comparar mediante el cociente  $\left(\frac{\sigma_{ie}}{\sigma_{ee}}\right)^2$ .

Como se ha señalado anteriormente, cuando existe incertidumbre acerca de la evidencia asociada a la variable  $X_e$ , es decir, cuando la evidencia  $e$  es inexacta, siendo  $e^\delta = e + \delta$ , se obtiene la medida de sensibilidad  $S^{\mu_e}(f(x_i|e), f(x_i|e, \delta))$  para  $\mu_e = e$ . Por tanto, se estudia este caso asociado a variaciones en la evidencia cuando se trabaja con la incertidumbre acerca de la media de la variable evidencial. Así, en caso de considerarse incierta la evidencia acerca de la variable evidencial o si interesa conocer el efecto de la observación de valores extremos, se puede realizar igualmente el análisis de sensibilidad calculando la medida de sensibilidad, mediante la expresión (3.3). Dicho resultado se recoge en la Proposición 3.2.

### Proposición 3.2

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $X_i$  variable de interés y  $X_e$  variable evidencial. Considérese la evidencia  $e$  que se tiene acerca de  $X_e$ . Perturbando dicha evidencia, de forma que la nueva evidencia sea  $e^\delta = e + \delta$ , se obtiene que el modelo perturbado es

$$X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e}\right)$$

donde

$$\mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} + \frac{\sigma_{ie}}{\sigma_{ee}} \delta$$

La medida de sensibilidad obtenida en este caso coincide con la media de sensibilidad de la expresión (3.3), siendo

$$S^e(f(x_i|e), f(x_i|e, \delta)) = \frac{\delta^2}{2\sigma_{ii}^{Y|E=e}} \left(\frac{\sigma_{ie}}{\sigma_{ee}}\right)^2 \quad (3.5)$$

### Demostración 3.2

Análoga a la demostración de la Proposición 3.1, Apartado 2. ■

A continuación, se presentan las medidas de sensibilidad obtenidas cuando la incertidumbre se tiene en los parámetros que definen la matriz de covarianzas.

### Sensibilidad para la matriz de covarianzas

Si se considera incierto o inexacto un elemento de la matriz de covarianzas  $\Sigma$ , el modelo perturbado que describe la red en estudio, viene dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^\delta)$ , donde  $\Sigma^\delta$  puede expresarse mediante

1. La varianza de la variable de interés  $X_i$  es incierta, siendo dicha varianza en el modelo perturbado tal que

$$\sigma_{ii}^\delta = \sigma_{ii} + \delta$$

con

$$\delta > -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}}$$

(se ha de imponer dicha restricción a la perturbación  $\delta$  para que la varianza final de la variable de interés sea positiva).

En este caso,  $\Sigma^\delta$  coincide con la matriz de covarianzas del modelo original en todos sus elementos salvo en la varianza de  $X_i$ .

2. La varianza de la variable evidencial  $X_e$  es inexacta, por tanto, en el modelo perturbado dicha varianza se expresa por

$$\sigma_{ee}^\delta = \sigma_{ee} + \delta$$

siendo

$$\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$$

donde  $\rho_{je}$  es el coeficiente de correlación lineal entre la variable evidencial  $X_e$  y cualquier variable  $X_j$  no evidencial.

El modelo perturbado  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^\delta)$  se define como el modelo original salvo en el parámetro asociado a la varianza de la variable evidencial, que en este caso es  $\sigma_{ee}^\delta$ .



3. La varianza de cualquier variable no evidencial  $X_j \in \mathbf{Y}$  con  $j \neq i$  viene dada por

$$\sigma_{jj}^\delta = \sigma_{jj} + \delta$$

con

$$\delta > -\sigma_{jj} + \frac{\sigma_{je}^2}{\sigma_{ee}}$$

De forma que  $\Sigma^\delta$  coincide con  $\Sigma$  en todos sus elementos, salvo en la varianza de  $X_j$ .

4. La covarianza entre la variable de interés  $X_i$  y la variable evidencial  $X_e$  queda modificada, de forma que en el modelo perturbado los elementos de la matriz de covarianzas,  $\sigma_{ie}^\delta$  y  $\sigma_{ei}^\delta$ , son

$$\sigma_{ie}^\delta = \sigma_{ie} + \delta = \sigma_{ei}^\delta$$

donde

$$-\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}} < \delta < -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$$

El modelo perturbado  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^\delta)$  queda como el modelo original salvo en el parámetro asociado a la covarianza entre  $X_i$  y  $X_e$ , es decir se modifican los elementos de la matriz de covarianzas de las posiciones  $(i, e)$  y  $(e, i)$ .

5. Cualquier otra covarianza queda modificada. Es decir, la covarianza entre la variable evidencial  $X_i$  y cualquier otra variable no evidencial  $X_j \in \mathbf{Y}$  con  $j \neq i$  es incierta, o la covarianza entre cualquier  $X_j \in \mathbf{Y}$  con  $j \neq i$  y la variable evidencial  $X_e$  cambia. En el modelo perturbado se tiene que

$$\sigma_{ij}^\delta = \sigma_{ij} + \delta = \sigma_{ij}^\delta$$

o se considera

$$\sigma_{je}^\delta = \sigma_{je} + \delta = \sigma_{ej}^\delta$$

siendo

$$-\sigma_{je} - \sqrt{\sigma_{jj}\sigma_{ee}} < \delta < -\sigma_{je} + \sqrt{\sigma_{jj}\sigma_{ee}}$$

Ahora  $\Sigma^\delta$  es igual a  $\Sigma$  salvo para los elementos de las posiciones  $(i, j)$  y  $(j, i)$  o  $(j, e)$  y  $(e, j)$ .

En todos los casos se deben imponer restricciones al valor de la perturbación  $\delta$ , con el fin de mantener positivas las varianzas del modelo perturbado.

En la Proposición 3.3 se incluyen las medidas de sensibilidad calculadas para los distintos casos introducidos anteriormente. De nuevo, se trabaja considerando un coeficiente de correlación lineal entre  $X_i$  y  $X_e$  tal que  $\rho_{ie}^2 \in (0, 1)$ .

**Proposición 3.3**

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $X_i$  variable de interés y  $X_e$  variable evidencial. Cuando se modifica la matriz de covarianzas  $\Sigma$ , sumándole a alguno de sus elementos la perturbación  $\delta$ , y se considera  $\rho_{ie}^2 \in (0, 1)$ , la medida de sensibilidad (3.1) obtenida para cada uno de los casos posibles es tal que

1. Si se suma la perturbación a la varianza de la variable de interés, siendo  $\sigma_{ii}^\delta = \sigma_{ii} + \delta$  para  $\delta > -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}}$ , tras realizarse la propagación de la evidencia, la distribución final de la variable de interés es

$$X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e, \delta}\right)$$

donde

$$\sigma_{ii}^{Y|E=e, \delta} = \sigma_{ii}^{Y|E=e} + \delta$$

La expresión dada para medida de sensibilidad en este caso es

$$S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 + \frac{\delta}{\sigma_{ii}^{Y|E=e}} \right) - \frac{\delta}{\sigma_{ii}^{Y|E=e, \delta}} \right] \quad (3.6)$$

2. Cuando la perturbación  $\delta$  se suma a la varianza de  $X_e$ , siendo  $\sigma_{ee}^\delta = \sigma_{ee} + \delta$  con  $\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$ , donde  $\rho_{je}$  es el coeficiente de correlación lineal entre  $X_j$  y  $X_e$ , la distribución final de interés es

$$X_i|E = e, \delta \sim N\left(\mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e, \delta}\right)$$

con

$$\begin{aligned} \mu_i^{Y|E=e, \delta} &= \mu_i + \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta}(e - \mu_e) \\ \sigma_{ii}^{Y|E=e, \delta} &= \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta} \end{aligned}$$

La medida de sensibilidad viene dada por

$$S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e, \delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( \frac{-\delta}{\sigma_{ee} + \delta} \right) \left( 1 + (e - \mu_e)^2 \left( \frac{-\delta}{(\sigma_{ee} + \delta)\sigma_{ee}} \right) \right)}{\sigma_{ii}^{Y|E=e, \delta}} \right] \quad (3.7)$$

3. Si la perturbación  $\delta$  se suma a la varianza de cualquier variable no evidencial  $X_j \in \mathbf{Y}$  con  $j \neq i$ , siendo  $\sigma_{jj}^\delta = \sigma_{jj} + \delta$ , dicha perturbación no afecta a la variable de interés  $X_i$ , quedando  $f(x_i|e, \delta) = f(x_i|e)$ . Por tanto, la medida de sensibilidad es cero.
4. Cuando la perturbación está en la covarianza entre  $X_i$  y  $X_e$ , es decir  $\sigma_{ie}^\delta = \sigma_{ie} + \delta = \sigma_{ei}^\delta$ , siendo  $-\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}} < \delta < -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$ , la distribución final de interés es

$$X_i|E = e, \delta \sim N \left( \mu_i^{Y|E=e, \delta}, \sigma_{ii}^{Y|E=e, \delta} \right)$$

donde

$$\begin{aligned} \mu_i^{Y|E=e, \delta} &= \mu_i^{Y|E=e} + \frac{\delta}{\sigma_{ee}}(e - \mu_e) \\ \sigma_{ii}^{Y|E=e, \delta} &= \sigma_{ii} - \frac{(\sigma_{ie} + \delta)^2}{\sigma_{ee}} \end{aligned}$$

La medida de sensibilidad viene dada por la siguiente expresión

$$S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 - \frac{\delta^2 + 2\sigma_{ie}\delta}{\sigma_{ee}\sigma_{ii}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \frac{\delta}{\sigma_{ee}}(e - \mu_e) \right)^2}{\sigma_{ii}^{Y|E=e, \delta}} - 1 \right] \quad (3.8)$$

5. Cuando la perturbación se supone en cualquier otra covarianza de la matriz de covarianzas, es decir, en las covarianzas entre  $X_i$  y cualquier variable no evidencial  $X_j$  o entre la variable evidencial  $X_e$  y  $X_j \in \mathbf{Y}$  para todo  $j \neq i$ , entonces, dicha perturbación no influye en la variable de interés, quedando  $f(x_i|e, \delta) = f(x_i|e)$ . Por tanto, la medida de sensibilidad es cero.

**Demostración 3.3**

Los parámetros finales del modelo perturbado se obtienen tras aplicar (2.12), siendo  $\mu_i^{Y|E=e,\delta}$  el elemento  $i$ -ésimo del vector de medias y  $\sigma_{ii}^{Y|E=e,\delta}$  el elemento  $i$ -ésimo de la matriz de covarianzas. En cada caso se considera una única perturbación  $\delta$ .

Trabajando con la expresión (3.4) se calcula la medida de sensibilidad para los casos presentados. Así,

$$\begin{aligned}
 1. S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) &= \\
 &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e,\delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_{ii}^{Y|E=e,\delta}} - 1 \right] \\
 &= \frac{1}{2} \left[ \ln \left( 1 + \frac{\delta}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e}}{\sigma_{ii}^{Y|E=e} + \delta} - 1 \right] \\
 &= \frac{1}{2} \left[ \ln \left( 1 + \frac{\delta}{\sigma_{ii}^{Y|E=e}} \right) - \frac{\delta}{\sigma_{ii}^{Y|E=e,\delta}} \right]
 \end{aligned}$$

considerando  $\mu_i^{Y|E=e,\delta} = \mu_i^{Y|E=e}$  y  $\sigma_{ii}^{Y|E=e,\delta} = \sigma_{ii}^{Y|E=e} + \delta$ .

Con la condición de que

$$\begin{aligned}
 \sigma_{ii}^{Y|E=e,\delta} &> 0 \Rightarrow \\
 \sigma_{ii}^{Y|E=e} + \delta &> 0 \Rightarrow \\
 \delta &> -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}}
 \end{aligned}$$

$$2. S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) =$$

$$\begin{aligned}
 &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e,\delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_{ii}^{Y|E=e,\delta}} - 1 \right] \\
 &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ii}^{Y|E=e} + \left( \mu_i^{Y|E=e,\delta} - \mu_i^{Y|E=e} \right)^2 - \sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e,\delta}} \right]
 \end{aligned}$$

considerando los parámetros finales del modelo perturbado

$$\mu_i^{Y|E=e,\delta} = \mu_i + \frac{\sigma_{ie}}{\sigma_{ee} + \delta}(e - \mu_e) \quad \text{y} \quad \sigma_{ii}^{Y|E=e,\delta} = \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta};$$

y los parámetros finales del modelo original

$$\mu_i^{Y|E=e} = \mu_i + \frac{\sigma_{ie}}{\sigma_{ee}}(e - \mu_e) \quad \text{y} \quad \sigma_{ii}^{Y|E=e} = \sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}}$$

Por tanto, la medida de sensibilidad queda

$$\begin{aligned} &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{-\frac{\sigma_{ie}^2}{\sigma_{ee}} + \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta} + \left( \frac{\sigma_{ie}}{\sigma_{ee} + \delta}(e - \mu_e) - \frac{\sigma_{ie}}{\sigma_{ee}}(e - \mu_e) \right)^2}{\sigma_{ii}^{Y|E=e,\delta}} \right] \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( \frac{-\delta}{\sigma_{ee} + \delta} \right) + \left( \sigma_{ie}(e - \mu_e) \left( \frac{-\delta}{(\sigma_{ee} + \delta)\sigma_{ee}} \right) \right)^2}{\sigma_{ii}^{Y|E=e,\delta}} \right] \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E=e,\delta}}{\sigma_{ii}^{Y|E=e}} \right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( \frac{-\delta}{\sigma_{ee} + \delta} \right) \left( 1 + (e - \mu_e)^2 \left( \frac{-\delta}{(\sigma_{ee} + \delta)\sigma_{ee}} \right) \right)}{\sigma_{ii}^{Y|E=e,\delta}} \right] \end{aligned}$$

En este caso hay que imponer que para cualquier variable  $X_j$

$$\begin{aligned} \sigma_{jj}^{Y|E=e,\delta} &> 0 \Rightarrow \\ \sigma_{jj} - \frac{\sigma_{je}^2}{\sigma_{ee} + \delta} &> 0 \Rightarrow \\ \sigma_{ee} + \delta &> \frac{\sigma_{je}^2}{\sigma_{jj}} \end{aligned}$$

Trabajando con el coeficiente de correlación lineal  $\rho_{je}^2 = \frac{\sigma_{je}^2}{\sigma_{jj}\sigma_{ee}}$

$$\begin{aligned} \sigma_{ee} + \delta &> \rho_{je}^2 \sigma_{ee} \Rightarrow \\ \delta &> \sigma_{ee}(\rho_{je}^2 - 1) \end{aligned}$$

Como todos los elementos de la diagonal se ven afectados por la perturbación introducida, se ha de exigir que

$$\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$$

3. La perturbación no afecta a la salida de interés del modelo perturbado, siendo  $f(x_i|e, \delta) = f(x_i|e)$ . Entonces, la medida de sensibilidad es cero.
4.  $S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) =$

$$\begin{aligned}
 &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_i^{Y|E=e, \delta}}{\sigma_i^{Y|E=e}} \right) + \frac{\sigma_i^{Y|E=e} + \left( \mu_i^{Y|E=e, \delta} - \mu_i^{Y|E=e} \right)^2}{\sigma_i^{Y|E=e, \delta}} - 1 \right] \\
 &= \frac{1}{2} \left[ \ln \left( 1 - \frac{\delta^2 + 2\sigma_{ii}\delta}{\sigma_{ee}\sigma_i^{Y|E=e}} \right) + \frac{\sigma_i^{Y|E=e} + \left( \frac{\delta}{\sigma_{ee}}(e - \mu_e) \right)^2}{\sigma_i^{Y|E=e, \delta}} - 1 \right]
 \end{aligned}$$

siendo  $\mu_i^{Y|E=e, \delta} = \mu_i^{Y|E=e} + \frac{\delta}{\sigma_{ee}}(e - \mu_e)$  y  $\sigma_{ii}^{Y|E=e, \delta} = \sigma_{ii}^{Y|E=e} - \frac{\delta^2 + 2\sigma_{ii}\delta}{\sigma_{ee}}$

En este caso para que la varianza final de interés sea mayor que cero se tiene

$$\begin{aligned}
 \sigma_{ii}^{Y|E=e, \delta} &> 0 \Rightarrow \\
 \sigma_{ii} - \frac{(\sigma_{ie} + \delta)^2}{\sigma_{ee}} &> 0 \Rightarrow \\
 (\sigma_{ie} + \delta)^2 &< \sigma_{ii}\sigma_{ee} \Rightarrow \\
 -\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}} &< \delta < -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}
 \end{aligned}$$

5. Tras propagarse la evidencia se comprueba que  $f(x_i|e, \delta) = f(x_i|e)$ , para cualquiera de los posibles modelos perturbados. Entonces, la medida de sensibilidad es cero. ■

Cuando la perturbación se presenta en la matriz de covarianzas, las medidas de sensibilidad obtenidas muestran expresiones similares respecto a la perturbación  $\delta$ .

Como se ha visto en el Apartado 2.4.3 del Capítulo 2, la matriz de covarianzas refleja las relaciones de independencia condicionada del problema mediante su inversa, la matriz de precisión. Cuando se modifica la matriz de covarianzas, también se modifica su matriz inversa, pudiendo aparecer nuevas relaciones de independencia o dependencia condicionada. Por tanto, al modificar las varianzas

y covarianzas, se puede modificar la estructura de dependencias que presenta la Red Bayesiana Gaussiana. Para estudiar las modificaciones presentes se puede calcular, para cada modelo perturbado, la matriz inversa de  $\Sigma^\delta$ . En caso de presentarse cambios importantes en la estructura de la red, los expertos han de estudiar en profundidad la variable que provoca esta modificación, a fin de describirla con la mayor precisión posible.

A continuación, se introduce un ejemplo que pone de manifiesto las ideas anteriores.

### Ejemplo 3.1

Considérese la Red Bayesiana Gaussiana introducida en el Ejemplo 2.9, cuya definición viene dada por el DAG de la Figura 3.1 y la distribución conjunta normal multivariante dada a continuación y se tiene evidencia acerca de  $X_2$ , siendo  $E = \{X_2 = 4\}$ .

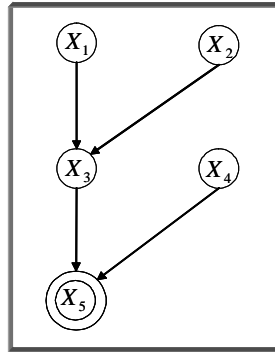


Figura 3.1. DAG asociado a la Red Bayesiana Gaussiana descrita para el Ejemplo 3.1

Donde  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  donde

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 0 & 6 & 0 & 6 \\ 0 & 2 & 2 & 0 & 2 \\ 6 & 2 & 15 & 0 & 15 \\ 0 & 0 & 0 & 2 & 4 \\ 6 & 2 & 15 & 4 & 26 \end{pmatrix}$$

Se tiene que la media y la varianza de la variable de interés  $X_5$  puede ser  $\mu_5^{\delta_5} = 7 = \mu_5 + \delta_5$  (donde  $\delta_5 = 2$ ) y  $\sigma_{55}^{\delta_{55}} = 24 = \sigma_{55} + \delta_{55}$  (con  $\delta_{55} = -2$ ).

También que los parámetros que definen la variable evidencial  $X_2$  pueden ser  $\mu_2^{\delta_2} = 6 = \mu_2 + \delta_2$  (con  $\delta_2 = 3$ ) y  $\sigma_{22}^{\delta_{22}} = 5 = \sigma_{22} + \delta_{22}$  (donde  $\delta_{22} = 3$ ).

La covarianza entre  $X_5$  y  $X_2$  puede venir dada por  $\sigma_{52}^{\delta_{52}} = 3 = \sigma_{52} + \delta_{52} = \sigma_{25}^{\delta_{25}}$  (donde  $\delta_{52} = 1 = \delta_{25}$ ).

Finalmente, la variable  $X_3$  puede definirse con  $\mu_3^{\delta_3} = 2 = \mu_3 + \delta_3$  (siendo  $\delta_3 = -1$ ) y  $\sigma_{33}^{\delta_{33}} = 16 = \sigma_{33} + \delta_{33}$  (con  $\delta_{33} = 1$ ).

Se desea estudiar la repercusión de las perturbaciones propuestas sobre la salida de la red, tras propagarse la evidencia  $E = \{X_2 = 4\}$ .

Como se vio en el Ejemplo 2.11, tras realizarse la propagación de la evidencia para el modelo original se obtiene que la variable de interés se distribuye según

$$X_5|X_2 = 4 \sim N(6, 24)$$

Para estudiar el efecto de la incertidumbre cuantificada por los expertos, se realiza el análisis de sensibilidad presentado, calculando para cada parámetro incierto la medida de sensibilidad.

Antes de obtener las distintas medidas de sensibilidad es necesario comprobar que los valores de las perturbaciones asociadas a la matriz de covarianzas,  $\delta_{22}$ ,  $\delta_{33}$ ,  $\delta_{55}$  y  $\delta_{52}$  hacen que dicha matriz perturbada  $\Sigma^\delta$ , sea definida positiva. En este caso, las perturbaciones supuestas en el ejemplo hacen que la matriz inicial perturbada sea definida positiva.

Existe incertidumbre acerca de los parámetros iniciales de las variables  $X_2$ ,  $X_3$  y  $X_5$ , siendo  $X_2$  la variable evidencial y  $X_5$  la de interés.

Por los resultados obtenidos al plantear el análisis, se sabe que la incertidumbre asociada a cualquier parámetro de una variable no evidencial, distinta de la de interés, en este caso  $X_3$ , no afecta a la salida de la red, dada por la distribución de  $X_5|X_2$ . Por tanto, no se calcula la medida de sensibilidad para dichos parámetros asociados a  $X_3$ , por ser cero.

Calculando la medida de sensibilidad para las perturbaciones asociadas a los parámetros de las variables evidencial  $X_2$  y de interés  $X_5$  se obtiene que:



- Si la media de la variable de interés es inexacta, tras la propagación de la evidencia, la variable de interés es  $X_5|X_2, \delta_5 = 4 \sim N(8, 24)$ . La medida de sensibilidad calculada mediante la expresión (3.2) es

$$S^{\mu_5}(f(x_5|e), f(x_5|e, \delta_5)) = 0.083$$

- Cuando se perturba la media de la variable evidencial, la distribución final de la variable de interés es normal, siendo en el modelo perturbado  $X_5|X_2 = 4, \delta_2 \sim N(3, 24)$ . Calculando la medida de sensibilidad con (3.3) se obtiene que

$$S^{\mu_2}(f(x_5|e), f(x_5|e, \delta_2)) = 0.1875$$

En este ejemplo, la medida de sensibilidad calculada para cualquier perturbación cuando la media de la variable evidencial es incierta, coincide con la medida de sensibilidad obtenida para la media de la variable de interés, debido a que los valores iniciales de la covarianza entre  $X_5$  y  $X_2$ ,  $\sigma_{52}$ , y de la varianza evidencial  $\sigma_{22}$  coinciden. Por lo tanto, por los valores de los parámetros asignados inicialmente, la medida de sensibilidad  $S^{\mu_5}(f(x_5|e), f(x_5|e, \delta))$  coincide con  $S^{\mu_2}(f(x_5|e), f(x_5|e, \delta))$  cuando se considera la misma perturbación  $\delta \in \mathbb{R}$ .

- Perturbando la varianza de la variable de interés, se tiene que tras la propagación de la evidencia,  $X_5|X_2 = 4, \delta_{55} \sim N(6, 22)$ . En este caso, la media final de interés se mantiene igual que en el modelo original y cambia la varianza final de interés, reduciendo su valor. Esta variación en la distribución de interés parece pequeña, por tanto, la medida de sensibilidad ha de ser también pequeña.

Con la expresión (3.5) se calcula la medida de sensibilidad obteniéndose

$$S^{\sigma_{55}}(f(x_5|e), f(x_5|e, \delta_{55})) = 0.0019$$

- Para el caso de perturbar la varianza evidencial, tras realizarse la propagación para el modelo perturbado, se tiene que  $X_5|X_2 = 4, \delta_{22} \sim N(5.8, 25.2)$ . Ambos parámetros finales del modelo perturbado cambian respecto al modelo original, reduciéndose muy poco el valor de la media y aumentando el valor de la varianza.

La medida de sensibilidad obtenida tras aplicar la expresión (3.6) es tal que

$$S^{\sigma_{22}}(f(x_5|e), f(x_5|e, \delta_{22})) = 0.0077$$

- Finalmente, si se perturba la covarianza entre la variable de interés  $X_5$  y la variable evidencial  $X_2$ , la distribución final de interés para el modelo perturbado es  $X_5|X_2 = 4, \delta_{52} \sim N(6.5, 21.5)$ . Dicha incertidumbre afecta a la media de interés, aumentando su valor, y a la varianza, que se ve reducida.

La medida de sensibilidad resultante al aplicar la expresión (3.7) viene dada por

$$S^{\sigma_{52}}(f(x_5|e), f(x_5|e, \delta_{52})) = 0.0089$$

Nótese que cuando se modifica algún elemento del vector de medias, dicha perturbación solo afecta a la media final de la variable de interés. Por el contrario, cuando se perturba algún elemento de la matriz de covarianzas, la perturbación afecta en determinados casos a la media y la varianza final de la variable de interés.

Los expertos en el problema han de determinar si los cambios producidos sobre la variable de interés y las medidas de sensibilidad obtenidas, son relevantes en función de las unidades de medida. No obstante los valores resultantes son pequeños y por tanto, se puede concluir que las incertidumbres acerca de los parámetros de la red no afectan en gran medida a la salida de la red, de forma que la red no es sensible a los cambios propuestos. Aún así, se puede destacar que el mayor valor de la medida de sensibilidad se obtiene cuando la media de la variable evidencial se perturba, siendo dicho valor 0.1875.

Los valores obtenidos de las medidas de sensibilidad son mayores cuando se modifica un elemento del vector de medias que cuando se modifica un elemento de la matriz de covarianzas.

Si los expertos consideran que los valores de las medidas de sensibilidad superan lo esperado, entonces deben revisar las variables cuyos parámetros presentan un mayor valor en la medida de sensibilidad.

Considérese ahora la situación en la que los expertos no pueden cuantificar la incertidumbre asociada a las variables  $X_2$ ,  $X_3$  y  $X_5$ . En este caso, se puede calcular la medida de sensibilidad, para cada uno de los parámetros inciertos en función de la perturbación  $\delta$ , obteniéndose

$$S^{\mu_5}(f(x_5|e), f(x_5|e, \delta)) = \frac{\delta^2}{48}$$

$$S^{\mu_2}(f(x_5|e), f(x_5|e, \delta)) = \frac{\delta^2}{48}$$

$$S^{\sigma_{55}}(f(x_5|e), f(x_5|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 + \frac{\delta}{24} \right) - \frac{\delta}{24+\delta} \right]$$

$$S^{\sigma_{22}}(f(x_5|e), f(x_5|e, \delta)) = \frac{1}{2} \left[ \ln \left( \frac{26 - \frac{4}{2+\delta}}{24} \right) + \frac{2 \left( \frac{-\delta}{2+\delta} \right) \left( 1 + \left( \frac{-\delta}{(2+\delta)^2} \right) \right)}{26 - \frac{4}{2+\delta}} \right]$$

$$S^{\sigma_{52}}(f(x_5|e), f(x_5|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 - \frac{\delta^2 + 4\delta}{48} \right) + \frac{24 + \left( \frac{\delta}{2} \right)^2}{26 - \frac{(2+\delta)^2}{2}} - 1 \right]$$

Si se muestran las medidas de sensibilidad en un gráfico en función del valor de  $\delta$ , se pueden comparar las medidas de sensibilidad y estudiar qué parámetro incierto puede afectar en mayor medida a los resultados de la red. En la Figura 3.2 se presentan dichas medidas para cualquier valor de  $\delta$ , aunque cuando la perturbación se añade a la matriz de covarianzas dicha perturbación no podrá tomar cualquier valor posible, puesto que la matriz de covarianzas inicial perturbada debe ser definida positiva.

Con la Figura 3.2 se observa como se pueden variar los parámetros sin producirse grandes cambios en la salida de la red, trabajando con valores de la medida de sensibilidad próximos a cero.

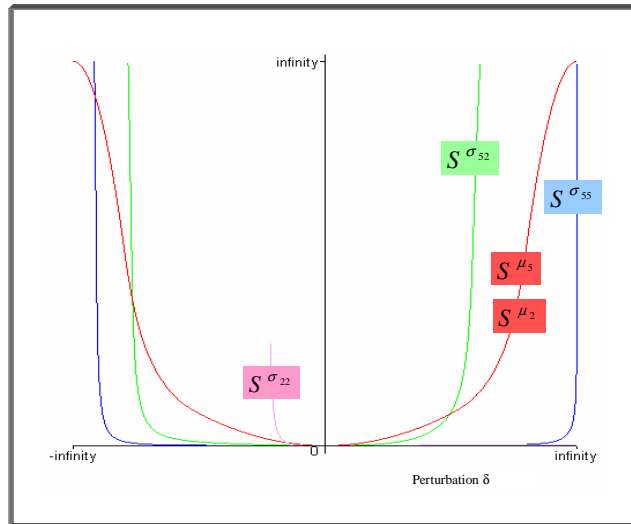


Figura 3.2. Medidas de sensibilidad obtenidas para cualquier valor de la perturbación  $\delta$  para el Ejemplo 3.1

Si se desea estudiar la medida de sensibilidad para pequeñas perturbaciones, también se pueden dibujar gráficamente las distintas medidas de sensibilidad para valores acotados de  $\delta$ . Así, en la Figura 3.3 se presentan dichas medidas cuando  $\delta \in [-4, 4]$ .

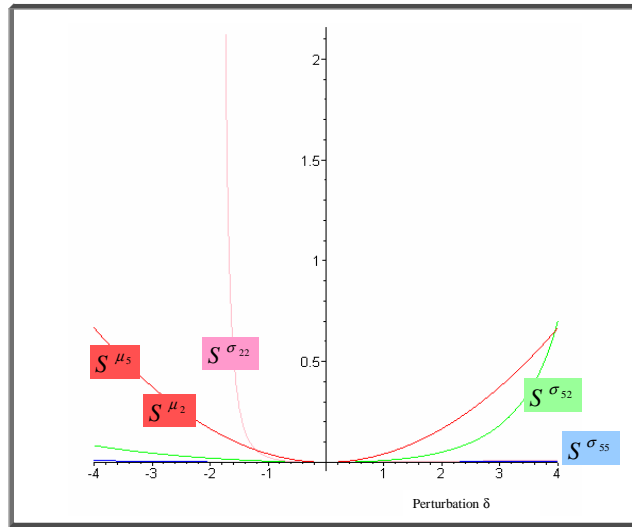


Figura 3.3. Medidas de sensibilidad obtenidas cuando la perturbación  $\delta \in [-4, 4]$  para el Ejemplo 3.1

Los resultados de este análisis son fundamentales a la hora de definir una Red Bayesiana Gaussiana, para comprobar si los parámetros introducidos y las posibles incertidumbres asociadas a los mismos cambian mucho los resultados que se obtienen acerca de la variable de interés. En caso de obtener valores de la medida de sensibilidad grandes, se deben estudiar de nuevo las variables que producen dichos valores para definirlos con mayor precisión y así garantizar que la red define correctamente el problema que se está tratando.

### 3.3.3. Casos extremos en la relación entre la variable de interés y la variable evidencial

Los resultados del análisis de sensibilidad introducidos se obtienen cuando el grado de relación lineal entre la variable de interés  $X_i$  y la variable evidencial  $X_e$ , no está en los extremos.

En este Apartado se analizan dos situaciones particulares de las medidas de sensibilidad, descritas en función de los valores extremos del coeficiente de correlación lineal  $\rho_{ie}^2$ . Ambas situaciones se particularizan en el grafo con una conexión concreta entre los nodos que representan a  $X_i$  y a  $X_e$ .

Así, si las variables  $X_i$  y  $X_e$  son independientes, el coeficiente de correlación lineal es  $\rho_{ie}^2 = 0$ . En este caso, sólo es posible una conexión convergente entre los nodos que representan dichas variables en el DAG. Por el contrario, cuando las variables  $X_i$  y  $X_e$  son linealmente dependientes con un coeficiente de correlación dado por  $\rho_{ie}^2 = 1$ , ambas variables en el DAG presentan una conexión en serie o divergente.

Trabajando con estas relaciones entre la variable de interés  $X_i$  y la variable evidencial  $X_e$ , se simplifican las expresiones de las medidas de sensibilidad expuestas en el Apartado anterior. En las Proposiciones 3.4 y 3.5 se presentan las medidas de sensibilidad obtenidas, considerando las situaciones extremas mostradas por el coeficiente de correlación lineal  $\rho_{ie}^2$ .

### Proposición 3.4

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $X_i$  variable de interés y  $X_e$  variable evidencial. Considérese una conexión convergente entre los nodos del DAG que representan dichas variables, siendo  $\rho_{ie}^2 = 0$ . Entonces, tras la propagación de la evidencia, la salida de la red no se ve influida por la evidencia introducida, siendo  $\mu_i^{Y|E=e} = \mu_i$  y  $\sigma_{ii}^{Y|E=e} = \sigma_{ii}$ .

Por tanto, solo influyen en la salida de la red las perturbaciones asociadas a los parámetros que describen a  $X_i$ ,  $\mu_i$  y  $\sigma_{ii}$ , y a la covarianza entre  $X_i$  y  $X_e$ ,  $\sigma_{ie}$ , modificándose en este último caso la relación de dependencia entre ambas variables. Para cada uno de los parámetros inciertos se tiene que:

1. Cuando se añade la perturbación  $\delta$  a la media de la variable de interés, entonces la distribución final de interés, tras la propagación de la evidencia, es  $X_i|E=e, \delta \sim N(\mu_i + \delta, \sigma_{ii})$ .

La medida de sensibilidad viene dada por

$$S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = \frac{\delta^2}{2\sigma_{ii}} \quad (3.9)$$

2. Si la perturbación  $\delta$  se suma a la varianza de la variable de interés, siendo

$\sigma_{ii}^\delta = \sigma_{ii} + \delta$  con  $\delta > -\sigma_{ii}$ , entonces la distribución final de interés es tal que  $X_i|E = e, \delta \sim N(\mu_i, \sigma_{ii} + \delta)$ .

La medida de sensibilidad viene dada por la siguiente expresión

$$S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 + \frac{\delta}{\sigma_{ii}} \right) - \frac{\delta}{\sigma_{ii} + \delta} \right] \quad (3.10)$$

3. Si se considera que puede existir alguna relación entre las variables  $X_i$  y  $X_e$ , se ha de modificar la conexión convergente del DAG y se pasa a tener una covarianza distinta de cero, dada por  $\sigma_{ie}^\delta = \sigma_{ie} + \delta = \delta = \sigma_{ei}^\delta$ , siendo  $\delta$  tal que  $-\sqrt{\sigma_{ii}\sigma_{ee}} < \delta < \sqrt{\sigma_{ii}\sigma_{ee}}$ . Entonces, la distribución final de interés es tal que  $X_i|E = e, \delta \sim N\left(\mu_i + \frac{\delta}{\sigma_{ee}}(e - \mu_e), \sigma_{ii} - \frac{\delta^2}{\sigma_{ee}}\right)$ .

La medida de sensibilidad es

$$S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( 1 - \frac{\delta^2}{\sigma_{ee}\sigma_{ii}} \right) + \frac{\frac{\delta^2}{\sigma_{ee}} \left( \frac{(e-\mu_e)^2}{\sigma_{ee}} + 1 \right)}{\sigma_{ii} - \frac{\delta^2}{\sigma_{ee}}} \right] \quad (3.11)$$

4. Para cualquier otra posible perturbación, la variable de interés no se ve modificada, siendo  $f(x_i|e, \delta) = f(x_i|e)$  y por tanto la medida de sensibilidad es cero.

### Demostración 3.4

Las expresiones de las medidas de sensibilidad obtenidas se calculan de forma directa trabajando con la relación de independencia entre  $X_i$  y  $X_e$ , dada por  $\sigma_{ie} = 0$ , en las Proposiciones 3.1 y 3.2. ■

Por tanto, cuando las variables  $X_i$  y  $X_e$  son independientes y se realiza el proceso de propagación de la evidencia, la información  $e$  acerca de  $X_e$  no afecta a la variable de interés, de forma que los parámetros finales de la variable de interés coinciden con dichos parámetros iniciales, siendo  $\mu_i^{Y|E=e} = \mu_i$  y  $\sigma_{ii}^{Y|E=e} = \sigma_{ii}$ .

Así, al estudiar la sensibilidad de la red, solo la incertidumbre acerca de los parámetros de la variable  $X_i$  puede afectar a la salida de la misma, recogida en la distribución final de la variable de interés. También se simplifica la medida de sensibilidad cuando se realizan modificaciones en la covarianza entre  $X_i$  y  $X_e$ , pasando de tener una relación de independencia entre ambas variables a una

relación de dependencia, dada por  $\sigma_{ie} = \delta$ . En este caso, el parámetro introducido señala que se ha de modificar el DAG, que representa las variables del problema, añadiendo una arista entre  $X_i$  y  $X_e$ . Cualquier otra perturbación posible de los parámetros de la red no influye sobre la distribución de la variable de interés, al igual que tampoco influye la incertidumbre acerca del valor de la evidencia  $e$ , tal que  $X_e = e$ , por tanto, la medida de sensibilidad también es cero.

Como representación de los resultados anteriores, se introduce un ejemplo en el que la variable de interés y la variable evidencial son independientes.

### **Ejemplo 3.2**

*Considérese la Red Bayesiana Gaussiana del Ejemplo 3.1. Sea ahora la variable de interés  $X_1$  y la variable evidencial  $X_2$ . Se pretende determinar que parámetros afectan a la salida de la red y como se realizaría el análisis de sensibilidad.*

Como se puede ver, ambas variables aparecen en una conexión convergente en el DAG, de forma que  $X_1$  y  $X_2$  son independientes.

En este caso, sólo influyen sobre la salida de la red los parámetros asociados a la variable de interés  $\mu_1$  y  $\sigma_{11}$  y la covarianza entre  $X_1$  y  $X_2$ , dada por  $\sigma_{12}$ . Para el resto de posibles parámetros inciertos, dicha incertidumbre no afecta a la salida de la red, siendo dicha salida, la distribución final de  $X_1$  tras la propagación de la evidencia.

Al realizarse el análisis de sensibilidad solo es necesario calcular las medidas de sensibilidad mostradas en la Proposición 3.4.

Por tanto, al especificarse una Red Bayesiana Gaussiana con una variable de interés  $X_i$  y una variable evidencial  $X_e$ , independiente de  $X_i$ , es muy importante ser preciso al asignar los parámetros a la variable de interés  $X_i$ .

La otra relación extrema considerada entre  $X_i$  y  $X_e$  se tiene, cuando el coeficiente de correlación es tal que  $\rho_{ie}^2 = 1$ . En este caso se tiene una conexión en serie o divergente entre los nodos del DAG que representan a  $X_i$  y  $X_e$ , de forma que existe una relación lineal entre ambas variables.

En la Proposición 3.5, se muestran las diferentes expresiones asociadas a la medida de sensibilidad que se obtienen para este caso extremo dado cuando el

coeficiente de correlación es  $\rho_{ie}^2 = 1$ .

### Proposición 3.5

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ ,  $X_i$  variable de interés y  $X_e$  variable evidencial, donde existe una relación lineal entre ambas variables, siendo  $\rho_{ie}^2 = 1$ . Entonces, tras la propagación de la evidencia, la varianza final de interés  $\sigma_{ii}^{Y|E=e}$  es cero.

La medida de sensibilidad para cualquier perturbación añadida a los parámetros de  $X_i$  o de  $X_e$  es igual a infinito.

Por tanto, cuando existe una relación lineal entre las variables de interés y evidencial, la medida de sensibilidad es extrema.

### Demostración 3.5

El resultado mostrado se obtiene de forma directa trabajando con las Proposiciones 3.1 y 3.2 siendo  $\rho_{ie}^2 = \frac{\sigma_{ie}^2}{\sigma_{ii}\sigma_{ee}} = 1$ . ■

Los resultados obtenidos en este caso en que  $\rho_{ie}^2 = 1$ , son los esperados, ya que cualquier perturbación asociada a los parámetros iniciales de  $X_i$  o  $X_e$  cambian mucho los resultados acerca de la variable de interés que depende de dichos parámetros, por tanto, la medida de sensibilidad resulta ser extrema.

### 3.3.4. Algoritmo para el cálculo de la medida de sensibilidad

A continuación, se introduce el Algoritmo diseñado para realizar el análisis de sensibilidad de una vía propuesto en esta Sección. El Algoritmo que se presenta, calcula en tiempo lineal las medidas de sensibilidad asociadas al conjunto de parámetros inciertos que describen la Red Bayesiana Gaussiana. La Red Bayesiana Gaussiana en estudio, ha de tener una variable de interés  $X_i$  y al menos evidencia acerca de una de las variables del problema,  $X_e$ .

En el Algoritmo diseñado, se introduce el modelo original, un vector  $\delta$  y una matriz simétrica  $\Delta$  con todos los valores de las perturbaciones asociadas a los parámetros inciertos del modelo original, siendo el vector  $\delta$  el *vector de perturbación de medias* y la matriz simétrica  $\Delta$  la *matriz de perturbación de covarianzas*, que describen las incertidumbres acerca de los parámetros del problema. Cuando



un elemento de  $\boldsymbol{\mu}$  o  $\Sigma$  no es incierto, entonces, en  $\delta$  o  $\Delta$  aparece un cero en la posición de dicho elemento, respectivamente. De esta forma, se genera un modelo perturbado para cada elemento de  $\delta$  y  $\Delta$  distinto de cero, que pueda afectar a la variable de interés  $X_i$ . Dicho modelo, se compara con el modelo original, para calcular la medida de sensibilidad en cada caso.

Nótese que los distintos cálculos que se deben realizar en el Algoritmo, se centran en los parámetros inciertos de la variable de interés y de la variable evidencial. Además, se obtienen todas las expresiones de las medidas de sensibilidad, en función de los elementos de  $\boldsymbol{\mu}$  y  $\Sigma$  que definen el modelo original. También, se puede obtener la medida de sensibilidad cuando la evidencia  $e$  es incierta, mediante la expresión (3.3) asociada a la incertidumbre en la media evidencial.

### Algoritmo

ENTRADA: Los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  que definen el modelo original dado por la Red Bayesiana Gaussiana con distribución conjunta normal multivariante  $N(\boldsymbol{\mu}, \Sigma)$ . La variable de interés  $X_i$ , la variable evidencial  $X_e$  y la evidencia  $e$  tal que  $X_e = e$ . Un vector de perturbación de medias  $\delta$ , siendo el elemento  $j$ -ésimo del vector la perturbación que se añade a la media de la variable  $X_j$ . Una matriz simétrica de perturbación de covarianzas  $\Delta$ , en el que el elemento  $(j, j)$  muestra la perturbación que se suma a la varianza de la variable  $X_j$  y el elemento  $(j, k)$  muestra la perturbación asociada a la covarianza entre  $X_j$  y  $X_k$ .

SALIDA: Las medidas de sensibilidad asociadas a los elementos inciertos que describen la salida de la red.

Paso 1. Calcular el coeficiente de correlación entre  $X_i$  y  $X_e$ , dado por la siguiente expresión

$$\rho_{ie}^2 = \frac{\sigma_{ie}^2}{\sigma_{ii}\sigma_{ee}}$$

Entonces, si  $\rho_{ie}^2 \in (0, 1)$  ir al Paso 2, si  $\rho_{ie}^2 = 0$  ir al Paso 3 y si  $\rho_{ie}^2 = 1$  ir al Paso 4.

Paso 2. Con  $\rho_{ie}^2 \in (0, 1)$  calcular:

- (a) Las medidas de sensibilidad de los parámetros inciertos del vector de medias, de forma que

- Si el elemento  $i$ -ésimo del vector de perturbación de medias  $\delta$  es distinto de cero, entonces, calcular  $S^{\mu_i}(f(x_i|e), f(x_i|e, \delta))$  con la expresión (3.2).
  - Si el elemento  $e$ -ésimo del vector de perturbación de medias  $\delta$  es distinto de cero, entonces, calcular  $S^{\mu_e}(f(x_i|e), f(x_i|e, \delta))$  mediante (3.3).
  - En otro caso, la medida de sensibilidad es cero.
- (b) Las medidas de sensibilidad de los parámetros inciertos de la matriz de covarianzas, teniendo en cuenta que
- Si el elemento  $i$ -ésimo de la diagonal de la matriz de perturbación de covarianzas  $\Delta$  es distinto de cero, entonces, calcular  $S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta))$  mediante (3.6).
  - Si el elemento  $e$ -ésimo de la diagonal de la matriz de perturbación de covarianzas  $\Delta$  es distinto de cero, entonces, calcular  $S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta))$  con (3.7).
  - Si los elementos  $(i, e)$  o  $(e, i)$  de la matriz de perturbación de covarianzas  $\Delta$  son distintos de cero, entonces, calcular  $S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta))$  mediante la expresión (3.8).
  - En otro caso la medida de sensibilidad es cero.

Paso 3. Con  $\rho_{ie}^2 = 0$  calcular:

- (a) La medida de sensibilidad asociada al vector de medias, cuando el elemento  $i$ -ésimo del vector  $\delta$  es distinto de cero, obteniéndose  $S^{\mu_i}(f(x_i|e), f(x_i|e, \delta))$  mediante la expresión (3.9).  
Para el resto de los casos, la medida de sensibilidad es cero.
- (b) Las medidas de sensibilidad asociadas a la matriz de covarianzas a través de
- Si el elemento  $i$ -ésimo de la diagonal de la matriz de perturbación de covarianzas  $\Delta$  es distinto de cero, entonces, calcular  $S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta))$  mediante (3.10).
  - Si los elementos  $(i, e)$  o  $(e, i)$  de la matriz de perturbación de covarianzas  $\Delta$  es distinto de cero, entonces, calcular  $S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta))$  con (3.11).
  - Para el resto de los casos, la medida de sensibilidad es cero.

Paso 4. Con  $\rho_{ie}^2 = 1$ , la medida de sensibilidad para cualquier elemento  $i$ -ésimo y/o  $e$ -ésimo de  $\delta$  o  $\Delta$ , es igual a infinito.

Para el resto de los casos, la medida de sensibilidad es cero.

Con este Algoritmo es posible calcular de forma eficiente las medidas de sensibilidad de una Red Bayesiana Gaussiana, que recogen la incertidumbre existente en la asignación de los parámetros que describen la misma. El Algoritmo está implementado en Susi (2006).

### 3.4. Sensibilidad para perturbaciones extremas

Una característica importante del análisis de sensibilidad presentado, es que permite estudiar la sensibilidad de una Red Bayesiana Gaussiana desde una perspectiva global, valorando la diferencia entre las distribuciones de probabilidad y no sólo entre algunas de sus características. Además, es posible asociar distintos grados de perturbación a los parámetros que describen el modelo original, al contrario de algunos análisis de sensibilidad desarrollados, como el descrito por Laskey (1995) para Redes Bayesianas Discretas, o el introducido por Castillo, et al. (2003) para Redes Bayesianas Gaussianas, que centran los análisis de sensibilidad en el estudio de perturbaciones locales sobre el parámetro descrito por los expertos.

Con este análisis de sensibilidad, también es posible valorar el efecto de posibles evidencias anómalas, por extremas, en las salidas de la red al propagar la evidencia.

En las Propositiones 3.6 y 3.7 de esta Sección, se estudia el comportamiento de las medidas de sensibilidad obtenidas, cuando la perturbación añadida al parámetro incierto es extrema.

El objetivo de este análisis es determinar el efecto de perturbaciones extremas acerca de los parámetros que describen la red inicialmente y comprobar si las medidas de sensibilidad, desarrolladas para cada caso, reflejan el comportamiento extremo de la perturbación. Para ello, se estudia la medida de sensibilidad para perturbaciones extremas calculando el límite de la medida de sensibilidad cuando

$\delta \longrightarrow u_\delta$  o  $\delta \longrightarrow l_\delta$ , donde  $u_\delta$  y  $l_\delta$  son las perturbaciones máxima y mínima posibles sobre el parámetro, y  $\delta \longrightarrow 0$ . En este caso, el coeficiente de correlación entre  $X_i$  y  $X_e$  es tal que  $\rho_{ie}^2 \in (0, 1)$ .

Los resultados presentados en esta Sección, han sido publicados en Gómez-Villegas, et al. (2006).

### Proposición 3.6

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $X_i$  variable de interés y  $X_e$  variable evidencial. Sea el coeficiente de correlación lineal entre ambas variables tal que  $\rho_{ie}^2 \in (0, 1)$ . Se desea estudiar la sensibilidad de la red. Entonces, cuando la perturbación  $\delta$  que se añade al vector de medias  $\boldsymbol{\mu}$  es extrema, la medida de sensibilidad también es extrema. De forma que,

1.

$$(a) \lim_{\delta \rightarrow \pm\infty} S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = \infty$$

$$(b) \lim_{\delta \rightarrow 0} S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = 0$$

2.

$$(a) \lim_{\delta \rightarrow \pm\infty} S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) = \infty$$

$$(b) \lim_{\delta \rightarrow 0} S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) = 0$$

### Demostración 3.6

Las expresiones resultantes de los límites se obtienen de forma directa. ■

Por tanto, cuando la perturbación es mínima, pudiendo  $\delta \rightarrow 0$ , la medida de sensibilidad es cero y cuando la perturbación es máxima, con  $\delta \rightarrow \pm\infty$ , la medida de sensibilidad es infinito. En ambas situaciones la medida de sensibilidad refleja el comportamiento extremo de la perturbación  $\delta$  que cuantifica la incertidumbre acerca de cualquier parámetro del vector de medias  $\boldsymbol{\mu}$  que influya en la salida de  $X_i$ .

Los resultados obtenidos para la media evidencial  $\mu_e$  acerca de la sensibilidad cuando la perturbación es extrema (2.(a) y 2.(b) de la Proposición 3.6), coinciden con el comportamiento de la medida de sensibilidad calculada cuando la evidencia  $e$ , tal que  $X_e = e$ , es extrema.

En la Proposición 3.7 se estudia el comportamiento de incertidumbres extremas, cuando la perturbación  $\delta$  aparece en la matriz de covarianzas.

**Proposición 3.7**

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $X_i$  variable de interés y  $X_e$  variable evidencial. Sea el coeficiente de correlación lineal entre ambas variables tal que  $\rho_{ie}^2 \in (0, 1)$ . Cuando la perturbación  $\delta$  que se añade a los diferentes elementos de la matriz de covarianzas  $\Sigma$  es extrema, la medida de sensibilidad es tal que

1.

$$(a) \lim_{\delta \rightarrow \infty} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \infty,$$

aunque  $S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = o(\delta)$

$$(b) \lim_{\delta \rightarrow M_{ii}} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \infty,$$

siendo  $M_{ii} = -\sigma_{ii} + \frac{\sigma_{ie}^2}{\sigma_{ee}} = -\sigma_{ii}(1 - \rho_{ie}^2)$  el menor valor posible para  $\delta$ .

$$(c) \lim_{\delta \rightarrow 0} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = 0$$

2.

$$(a) \lim_{\delta \rightarrow \infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ -\ln(1 - \rho_{ie}^2) - \rho_{ie}^2 \left( 1 - \frac{(e - \mu_e)^2}{\sigma_{ee}} \right) \right]$$

$$(b) \lim_{\delta \rightarrow M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) =$$

$$= \frac{1}{2} \left[ \ln \left( \frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1 - \rho_{ie}^2)} \right) + \frac{\rho_{ie}^2(1 - M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2} \left( 1 + \frac{(e - \mu_e)^2}{\sigma_{ee}} \left( \frac{1 - M_{ee}^*}{M_{ee}^*} \right) \right) \right]$$

donde el menor valor posible para la perturbación es

$$M_{ee} = -\sigma_{ee}(1 - M_{ee}^*) \text{ con } M_{ee}^* = \max_{X_j \in \mathbf{Y}} \rho_{je}^2$$

$$(c) \lim_{\delta \rightarrow 0} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = 0$$

3.

$$(a) \lim_{\delta \rightarrow M_{ie}^1} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \infty,$$

siendo  $M_{ie}^1 = -\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}}$  el límite inferior de  $\delta$ .

- (b)  $\lim_{\delta \rightarrow M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \infty$ ,  
 siendo  $M_{ie}^2 = -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$  el límite superior de  $\delta$ .
- (c)  $\lim_{\delta \rightarrow 0} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = 0$

**Demostración 3.7**

Para cada uno de los casos, se tiene que:

1.

- (a) Se calcula el límite de forma directa.
- (b) Cuando  $\sigma_{ii}^\delta = \sigma_{ii} + \delta$ , la varianza final de  $X_i$  es tal que  $\sigma_{ii}^{\mathbf{Y}|E=e, \delta} = \sigma_{ii}^{\mathbf{Y}|E=e} + \delta$ . Como  $\sigma_{ii}^{\mathbf{Y}|E=e, \delta} > 0$  la perturbación ha de ser tal que

$$\delta > -\sigma_{ii}^{\mathbf{Y}|E=e}$$

Sea  $M_{ii} = -\sigma_{ii}^{\mathbf{Y}|E=e}$  y sea  $x = \sigma_{ii}^{\mathbf{Y}|E=e} + \delta$  entonces

$$\begin{aligned} \lim_{\delta \rightarrow M_{ii}} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) &= \\ &= \lim_{x \rightarrow 0} \frac{1}{2} \left[ \ln x - \ln \sigma_{ii}^{\mathbf{Y}|E=e} - \frac{x - \sigma_{ii}^{\mathbf{Y}|E=e}}{x} \right] = \infty \end{aligned}$$

- (c) Se calcula el límite de forma directa.

2.

$$(a) \lim_{\delta \rightarrow \infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}}{\sigma_{ii}^{\mathbf{Y}|E=e}} \right) + \frac{-\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( 1 - \frac{(e - \mu_e)^2}{\sigma_{ee}} \right)}{\sigma_{ii}} \right]$$

Teniendo en cuenta que

$$\sigma_{ii}^{\mathbf{Y}|E=e} = \sigma_{ii}(1 - \rho_{ie}^2) \quad y \quad \rho_{ie}^2 = \frac{\sigma_{ie}^2}{\sigma_{ii}\sigma_{ee}}$$

Entonces,

$$\lim_{\delta \rightarrow \infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ -\ln(1 - \rho_{ie}^2) - \rho_{ie}^2 \left( 1 - \frac{(e - \mu_e)^2}{\sigma_{ee}} \right) \right]$$

(b) Para estudiar la sensibilidad respecto a la varianza evidencial, siendo  $\sigma_{ee}^\delta = \sigma_{ee} + \delta$ , la varianza final de interés para cualquier variable no evidencial es  $\sigma_{jj}^{Y|E=e,\delta} = \sigma_{jj} - \frac{\sigma_{je}^2}{\sigma_{ee} + \delta}$ .

Por tanto, se ha de imponer que  $\sigma_{jj}^{Y|E=e,\delta} > 0$  para todo  $X_j \in \mathbf{Y}$ .

Entonces, la perturbación  $\delta$  ha de satisfacer

$$\delta > -\sigma_{ee}(1 - \max_{X_j \in \mathbf{Y}} \rho_{je}^2)$$

Denotando por  $M_{ee}^* = \max_{X_j \in \mathbf{Y}} \rho_{je}^2$  y por  $M_{ee} = -\sigma_{ee}(1 - M_{ee}^*)$ , se tiene

$$\begin{aligned} \lim_{\delta \rightarrow M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) &= \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta}}{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee}}} \right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( \frac{-\delta}{\sigma_{ee} + \delta} \right) \left( 1 + (e - \mu_e)^2 \left( \frac{-\delta}{(\sigma_{ee} + \delta)\sigma_{ee}} \right) \right)}{\sigma_{ii} - \frac{\sigma_{ie}^2}{\sigma_{ee} + \delta}} \right] \\ &= \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}\sigma_{ee}M_{ee}^* - \sigma_{ie}^2}{M_{ee}^*(\sigma_{ii}\sigma_{ee} - \sigma_{ie}^2)} \right) + \frac{\frac{\sigma_{ie}^2}{\sigma_{ee}} \left( \frac{1 - M_{ee}^*}{M_{ee}^*} \right) \left( 1 + \frac{(e - \mu_e)^2}{\sigma_{ee}} \left( \frac{1 - M_{ee}^*}{M_{ee}^*} \right) \right)}{M_{ee}^* - \rho_{ie}^2} \right] \\ &= \frac{1}{2} \left[ \ln \left( \frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1 - \rho_{ie}^2)} \right) + \frac{\rho_{ie}^2(1 - M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2} \left( 1 + \frac{(e - \mu_e)^2}{\sigma_{ee}} \left( \frac{1 - M_{ee}^*}{M_{ee}^*} \right) \right) \right] \end{aligned}$$

(c) Se calcula el límite de forma directa.

3.

(a) Para estudiar la sensibilidad respecto a la covarianza entre  $X_i$  y  $X_e$ , siendo  $\sigma_{ie}^\delta = \sigma_{ie} + \delta$ , la varianza final de interés es tal que

$$\sigma_{ii}^{Y|E=e,\delta} = \sigma_{ii} - \frac{(\sigma_{ie} + \delta)^2}{\sigma_{ee}}$$

Para que dicha varianza sea positiva, la perturbación  $\delta$  ha de ser tal que

$$-\sigma_{ie} - \sqrt{\sigma_{ii}\sigma_{ee}} < \delta < -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$$

Sea  $M_{ie}^2 = -\sigma_{ie} + \sqrt{\sigma_{ii}\sigma_{ee}}$ , entonces, es posible calcular el límite  $\lim_{\delta \rightarrow M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta))$ . Aunque si se considera  $\delta \rightarrow M_{ie}^2$ , esto es equivalente a tener

$$(\delta^2 + 2\sigma_{ie}\delta) \rightarrow \sigma_{ee}\sigma_{ii}^{Y|E=e}$$

Por lo tanto, la medida de sensibilidad en este caso es

$$S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ee}\sigma_{ii}^{Y|E=e} - (\delta^2 + 2\sigma_{ie}\delta)}{\sigma_{ee}\sigma_{ii}^{Y|E=e}} \right) + \frac{\sigma_{ee}\sigma_{ii}^{Y|E=e} + \left( \frac{\delta}{\sigma_{ee}}(e - \mu_e) \right)^2}{\sigma_{ee}\sigma_{ii}^{Y|E=e} - (\delta^2 + 2\sigma_{ie}\delta)} - 1 \right]$$

Como  $\lim_{x \rightarrow 0} \left[ \ln x + \frac{k}{x} \right] = \infty$  para cualquier valor de  $k$ , entonces

$$\lim_{\delta \rightarrow M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \infty$$

(b) Se calcula de forma análoga al punto anterior 3.(a).

(c) Se calcula el límite de forma directa. ■

Los resultados obtenidos son intuitivos, ya que la medida de sensibilidad refleja el comportamiento extremo de la perturbación, salvo, en el caso de incertidumbre extrema acerca de la varianza evidencial, para el que existe un límite finito de la medida de sensibilidad.

Esto sucede porque cuando se tiene evidencia acerca de la variable evidencial  $X_e$ , su varianza tiene un efecto reducido sobre la variable de interés  $X_i$  ya que la salida del modelo perturbado  $f(x_i|e, \delta)$  no es muy distinta de la obtenida para el modelo original, dada por  $f(x_i|e)$ . Por tanto, aunque pueda existir una incertidumbre extrema acerca de la varianza evidencial, la medida de sensibilidad tiende a un valor finito.

Cuando se desee realizar un análisis de sensibilidad con la metodología propuesta en la Sección anterior y se considere incierta la varianza de la variable evidencial, entonces, es conveniente calcular el valor asociado al límite de la medida de sensibilidad cuando  $\delta \rightarrow \infty$  y cuando  $\delta \rightarrow M_{ee}$  para estudiar si la perturbación propuesta es extrema.

### Ejemplo 3.3

Considérese la Red Bayesiana Gaussiana definida en el Ejemplo 3.1. El grupo de expertos que ha asignado los valores de los parámetros, están en gran desacuerdo con los mismos y desean estudiar la sensibilidad de la red para valores extremos de las perturbaciones.



Para ello, consideran los parámetros de la variable de interés  $X_5$ , tales que,  $\mu_5^{\delta_5} = -20 = \mu_5 + \delta_5$ , siendo la perturbación  $\delta_5 = -25$ , y  $\sigma_{55}^{\delta_{55}} = 3$ , con  $\delta_{55} = -23$ .

Por otro lado, consideran la media y la varianza de la variable evidencial  $X_2$  tales que  $\mu_2^{\delta_2} = 30 = \mu_2 + \delta_2$ , con  $\delta_2 = 27$ , y  $\sigma_{22}^{\delta_{22}} = 0.27$ , donde  $\delta_{22} = -1.73$ .

Finalmente, fijan la covarianza entre  $X_5$  y  $X_2$  en  $\sigma_{52}^{\delta_{52}} = 3$ , con  $\delta_{52} = 1$ .

Se hace notar que aunque pudiese existir más incertidumbre en otros parámetros, dicha incertidumbre no afecta a la salida de la red, recogida por la distribución final de la variable de interés, que se obtiene tras la propagación de la evidencia.

Lo primero que se ha de comprobar es que los nuevos parámetros mantienen la matriz de covarianzas como definida positiva. Esto permite trabajar con las perturbaciones  $\delta_{55}$ ,  $\delta_{22}$  y  $\delta_{52}$ , para los tres posibles modelos perturbados compatibles con dichas perturbaciones.

Las medidas de sensibilidad para las perturbaciones propuestas, vienen recogidas por:

$$S^{\mu_5}(f(x_5|e), f(x_5|e, \delta_5)) = 13.02$$

$$S^{\sigma_{55}}(f(x_5|e), f(x_5|e, \delta_{55})) = 9.91$$

$$S^{\mu_2}(f(x_5|e), f(x_5|e, \delta_2)) = 15.19$$

$$S^{\sigma_{22}}(f(x_5|e), f(x_5|e, \delta_{22})) = 2.03$$

$$S^{\sigma_{52}}(f(x_5|e), f(x_5|e, \delta_{52})) = 0.009$$

En el caso de la medida de sensibilidad obtenida cuando hay incertidumbre en la varianza evidencial, es conveniente calcular el límite de la medida de sensibilidad cuando la perturbación  $\delta_{22}$  tiende hacia sus extremos. En este caso, la perturbación es negativa, por tanto, se ha de estudiar el límite de  $S^{\sigma_{22}}(f(x_5|e), f(x_5|e, \delta_{22}))$  cuando la perturbación tiende hacia el mínimo  $M_{ee}$ , donde  $M_{ee}$  viene dado por

$$M_{ee} = -\sigma_{ee}(1 - M_{ee}^*) = -1.7333$$

con

$$M_{ee}^* = \max_{X_j \in \mathbf{Y}} \rho_{je}^2 = 0.1333.$$

En este caso,

$$\begin{aligned} \lim_{\delta \rightarrow M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) &= \\ &= \frac{1}{2} \left[ \ln \left( \frac{M_{ee}^* - \rho_{ie}^2}{M_{ee}^*(1 - \rho_{ie}^2)} \right) + \frac{\rho_{ie}^2(1 - M_{ee}^*)}{M_{ee}^* - \rho_{ie}^2} \left( 1 + \frac{(e - \mu_e)^2}{\sigma_{ee}} \left( \frac{1 - M_{ee}^*}{M_{ee}^*} \right) \right) \right] \\ &= 2.1213 \end{aligned}$$

Por tanto, el valor obtenido para la medida de sensibilidad cuando la varianza evidencial es incierta,  $S^{\sigma_{22}}(f(x_5|e), f(x_5|e, \delta_{22})) = 2.03$ , es próximo a su límite.

Como se puede observar en el ejemplo, la medida de sensibilidad crece a medida que aumenta el valor de la perturbación. Véase la Figura 3.2 para comprobar el comportamiento de las medidas de sensibilidad obtenidas para cualquier perturbación  $\delta$  posible.

Con los resultados mostrados se puede estudiar la robustez de la Red Bayesiana Gaussiana del problema, exigiendo valores pequeños de la medida de sensibilidad. Así, fijando un umbral de 0.1, tal que  $S^{p_j} < 0.1$ , para las diferencias aceptables sobre los parámetros, con  $p_j \in \{\mu_5, \sigma_{55}, \mu_2, \sigma_{52}\}$ , las perturbaciones han de tomar valores en los siguientes intervalos:

- Para que  $S^{\mu_5} < 0.1$  y  $S^{\mu_2} < 0.1$ , las perturbaciones  $\delta_5$  y  $\delta_2$  han de definirse tal que  $\delta_5 \in (-1.55, 1.55)$  y  $\delta_2 \in (-1.55, 1.55)$
- Para conseguir que la medida de sensibilidad  $S^{\sigma_{55}}$  sea  $S^{\sigma_{55}} < 0.1$ , la perturbación  $\delta_{55}$  ha de ser  $\delta_{55} \in (-10.46, 24.66)$
- Para que  $S^{\sigma_{52}} < 0.1$ , la perturbación asociada ha de definirse mediante  $\delta_{52} \in (-4.31, 2.51)$ .
- Finalmente, para que la medida de sensibilidad  $S^{\sigma_{22}} < 0.1$ , la perturbación  $\delta_{22}$  ha de ser  $\delta_{22} > -1.32$ . Cuando  $\delta_{22}$  tiende a infinito, el límite de  $S^{\sigma_{22}}$  es 0.0208, por tanto, no es necesario imponer ninguna otra restricción sobre  $\delta_{22}$ .

A continuación, se estudia el comportamiento de la medida de sensibilidad para perturbaciones extremas, cuando la relación entre la variable de interés  $X_i$  y la variable evidencial  $X_e$  es también extrema. Para estudiar dicha relación entre las variables, se trabaja con el coeficiente de correlación lineal, considerándose una relación extrema cuando  $\rho_{ie}^2 = 0$  o  $\rho_{ie}^2 = 1$ .

### 3.4.1. Casos extremos en la relación entre la variable de interés y la variable evidencial

En las Proposiciones 3.8 y 3.9 se muestra el comportamiento de la medida de sensibilidad para perturbaciones extremas, cuando el grado de dependencia entre las variables  $X_i$  y  $X_e$  es también extremo, siendo el coeficiente de correlación lineal tal que  $\rho_{ie}^2 = 0$  y  $\rho_{ie}^2 = 1$ . En cada proposición se estudia para cada parámetro  $\mu$  o  $\Sigma$  el límite de la medida de sensibilidad cuando  $X_i$  y  $X_e$  son linealmente independientes y cuando presentan dependencia lineal.

Como variables que se representan en un DAG, estas situaciones se corresponden con los siguientes casos:

- $X_i$  y  $X_e$  independientes, por tanto sólo pueden estar situadas como ancestros, sin conexión, de conexiones convergentes.
- $X_i$  y  $X_e$  dependientes entonces pueden estar conectadas por conexiones seriales o divergentes con una relación lineal.

#### Proposición 3.8

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$ , con  $X_i$  variable de interés y  $X_e$  variable evidencial. Para estudiar la sensibilidad de la red, cuando las variables  $X_i$  y  $X_e$  son independientes,  $\rho_{ie}^2 = 0$ , siendo extrema la perturbación  $\delta$  asociada al parámetro incierto del vector de medias, la medida de sensibilidad viene dada por

1.

$$(a) \lim_{\delta \rightarrow \pm\infty} S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = \infty$$

$$(b) \lim_{\delta \rightarrow 0} S^{\mu_i}(f(x_i|e), f(x_i|e, \delta)) = 0$$

2.

$$(a) \lim_{\delta \rightarrow \pm\infty} S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) = \lim_{\delta \rightarrow 0} S^{\mu_e}(f(x_i|e), f(x_i|e, \delta)) = 0$$

Cuando hay una relación lineal entre  $X_i$  y  $X_e$ ,  $\rho_{ie}^2 = 1$ , la medida de sensibilidad obtenida para cualquier perturbación extrema, también es extrema.

### Demostración 3.8

Los límites se calculan de forma directa, considerando cuando hay independencia que  $\sigma_{ie} = \rho_{ie}^2 = 0$  y cuando hay dependencia lineal la Proposición 3.5, con  $\rho_{ie}^2 = 1$ . ■

La siguiente proposición muestra los mismos resultados que la proposición anterior, pero considerando ahora que se perturba la matriz de covarianzas  $\Sigma$ .

### Proposición 3.9

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$ , con  $X_i$  variable de interés y  $X_e$  variable evidencial. Se desea estudiar la sensibilidad de la red. Entonces, cuando las variables  $X_i$  y  $X_e$  son independientes, con  $\rho_{ie}^2 = 0$ , siendo extrema la perturbación  $\delta$  asociada al parámetro incierto de la matriz de covarianzas, la medida de sensibilidad es tal que

1.

$$(a) \lim_{\delta \rightarrow \infty} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \infty,$$

aunque  $S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = o(\delta)$

$$(b) \lim_{\delta \rightarrow M_{ii}} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = \infty,$$

siendo  $M_{ii} = -\sigma_{ii}$

$$(c) \lim_{\delta \rightarrow 0} S^{\sigma_{ii}}(f(x_i|e), f(x_i|e, \delta)) = 0$$

2.

$$(a) \lim_{\delta \rightarrow \infty} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = \lim_{\delta \rightarrow M_{ee}} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) =$$

$$\lim_{\delta \rightarrow 0} S^{\sigma_{ee}}(f(x_i|e), f(x_i|e, \delta)) = 0$$

donde  $M_{ee} = -\sigma_{ee}(1 - M_{ee}^*)$  con  $M_{ee}^* = \max_{X_j \in \mathbf{Y}} \rho_{je}^2$

3.

$$(a) \lim_{\delta \rightarrow M_{ie}^1} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \infty \text{ siendo } M_{ie}^1 = -\sqrt{\sigma_{ii}\sigma_{ee}}$$

$$(b) \lim_{\delta \rightarrow M_{ie}^2} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = \infty \text{ con } M_{ie}^2 = +\sqrt{\sigma_{ii}\sigma_{ee}}$$
$$(c) \lim_{\delta \rightarrow 0} S^{\sigma_{ie}}(f(x_i|e), f(x_i|e, \delta)) = 0$$

Cuando existe una relación lineal entre  $X_i$  y  $X_e$ , con  $\rho_{ie}^2 = 1$ , la medida de sensibilidad obtenida para cualquier perturbación extrema, es infinita, para todos los casos en los que la perturbación influye en la salida de la red.

### **Demostración 3.9**

Los límites se calculan de forma directa considerando la Proposición 3.4 y 3.5. Además, cuando hay independencia entre  $X_i$  y  $X_e$  se tiene que  $\sigma_{ie} = \rho_{ie}^2 = 0$  y con la dependencia lineal se trabaja con  $\rho_{ie}^2 = 1$ . ■

Como se puede ver en las proposiciones presentadas, cuando  $X_i$  y  $X_e$  son independientes, la salida de la red dada por la distribución final de  $X_i$ , tras la propagación de la evidencia, sólo se ve afectada por los parámetros inciertos de  $X_i$ , de forma que si la perturbación es extrema, la medida de sensibilidad también es extrema. Sin embargo, cuando existe una dependencia lineal entre  $X_i$  y  $X_e$ , con  $\rho_{ie}^2 = 1$ , cualquier perturbación añadida a los parámetros de  $X_i$  o  $X_e$  influye notablemente en la salida de la red, siendo infinito el valor de las medidas de sensibilidad calculadas.

## **3.5. Análisis de Sensibilidad de n vías para Redes Bayesianas Gaussianas**

En el análisis de sensibilidad de una vía introducido en la Sección 3.3 se trabaja con un único parámetro incierto en cada paso del análisis y se evalúa la salida de la red, trabajando con una única variable de interés  $X_i$ .

En esta Sección, se generaliza el análisis de sensibilidad de una vía, de forma que se trabaja con un conjunto de variables de interés. Además, se evalúa más de un parámetro incierto a la vez, por tanto, el análisis de sensibilidad que se presenta es un análisis de sensibilidad de n vías para una Red Bayesiana Gaussiana definida con una o varias variables de interés. Así, al trabajar con más de un parámetro incierto a la vez, en el análisis de sensibilidad se recogen tanto las variaciones producidas de forma individual por cada parámetro, como las variaciones debidas

al conjunto de parámetros.

En el análisis de sensibilidad que se desarrolla en esta Sección, se considera una partición del conjunto inicial de variables, dada por  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ , donde  $\mathbf{Y}$  está formado por  $q$  variables de interés y  $\mathbf{E}$  son  $n - q$  variables evidenciales. Ahora se tiene un conjunto de variables evidenciales, y aunque la propagación de la evidencia en Redes Bayesianas Gaussianas se realiza con el algoritmo presentado en el Apartado 2.4.4, introduciendo en cada paso una única variable evidencial, de forma que la ejecución del mismo se realiza en tiempo lineal, tras la introducción paso a paso de cada variable evidencial, se toma como salida de la red la distribución final del conjunto de variables de interés dada la evidencia,  $\mathbf{Y}|\mathbf{E}$ .

Los resultados que se presentan en esta Sección pueden extenderse a una Red Bayesiana Gaussiana en la que el conjunto de variables se particiona de forma que  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}, \mathbf{R}\}$ , donde  $\mathbf{Y}$  es el conjunto de variables de interés,  $\mathbf{E}$  son las variables evidenciales y  $\mathbf{R}$  son el resto de variables. Así, para el caso de existir incertidumbre acerca de los parámetros de  $\mathbf{R}$ , dichas imprecisiones no influyen en la distribución final de las variables de interés, siendo cero el valor de la medida de sensibilidad.

De nuevo, mediante el análisis de sensibilidad, se busca estudiar el efecto producido al perturbar al inicio un conjunto de parámetros inciertos, sobre la distribución final de la red, dada por la distribución normal multivariante  $\mathbf{Y}|\mathbf{E} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \boldsymbol{\Sigma}^{\mathbf{Y}|\mathbf{E}})$ . Los parámetros se estudian agrupados, considerando así el efecto de tener incertidumbre en las variables de interés o en las variables evidenciales. También se obtienen resultados por separado para los elementos de cada parámetro, dados por el vector de medias y la matriz de covarianzas.

El mecanismo de análisis para estudiar la sensibilidad de la red es similar al presentado en la Sección 3.3, de forma que se basa en comparar las salidas finales de la red, obtenidas tras propagar la evidencia, en el modelo original y en el modelo perturbado.

El modelo original es la Red Bayesiana Gaussiana definida por los parámetros asignados inicialmente, de manera que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

El modelo perturbado viene dado por el modelo original, salvo un conjunto de parámetros que se presentan perturbados porque se ha sumado, al valor inicial de los parámetros, una perturbación que cuantifica la incertidumbre acerca de los

mismos.

Ahora la perturbación no es un escalar  $\delta$ , sino un *vector de perturbación de medias*  $\delta$  y una *matriz de perturbación de covarianzas*  $\Delta$ . Teniendo en cuenta que se estudian las perturbaciones en función de si están asociadas a una variable de interés o a una variable evidencial, las perturbaciones que se introducen en esta Sección, dadas por el vector de perturbación de medias  $\delta$  y la matriz de perturbación de covarianzas  $\Delta$ , se particionan de forma que

$$\delta = \begin{pmatrix} \delta_{\mathbf{Y}} \\ \delta_{\mathbf{E}} \end{pmatrix} \quad \Delta = \begin{pmatrix} \Delta_{\mathbf{Y}\mathbf{Y}} & \Delta_{\mathbf{Y}\mathbf{E}} \\ \Delta_{\mathbf{E}\mathbf{Y}} & \Delta_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

El análisis de sensibilidad que se propone en este Apartado, consiste en calcular sistemáticamente una medida de sensibilidad para distribuciones multivariantes, generalización de la medida definida en el Apartado 3.3.1, considerando incierto en cada paso un único subconjunto de la partición de  $\delta$  o de  $\Delta$  presentada. Dicha medida de sensibilidad compara la salida final de la red, dada por la distribución de las variables de interés que se obtiene tras propagar la evidencia, para el modelo original y para el modelo perturbado.

Tras realizarse el análisis de sensibilidad, se obtienen cinco expresiones distintas asociadas a la medida de sensibilidad para distribuciones multivariantes, que sirven para calcular dicha medida dependiendo de los parámetros inciertos del modelo. En función de los resultados obtenidos para una Red Bayesiana Gaussiana concreta, se concluye él o los conjuntos de parámetros inciertos que han de ser revisados para enunciarlos con mayor precisión.

Al igual que en el análisis de sensibilidad de una vía presentado anteriormente, cuando los valores de las medidas de sensibilidad obtenidos son pequeños, se puede concluir que la red es robusta. Esta idea se concreta en la Sección 3.6.

A continuación, se introduce un Apartado en el que se define la medida de sensibilidad para distribuciones multivariantes, trabajando con un conjunto de parámetros inciertos, al igual que con un conjunto de variables de interés y un conjunto de variables evidenciales. En el siguiente Apartado, se detalla el análisis de sensibilidad que sirve para estudiar la Red Bayesiana Gaussiana. Para ello, se enuncian diversas proposiciones, en función de si son inciertos los parámetros de  $\mu$  o de  $\Sigma$ . Finaliza la Sección, ilustrando con un ejemplo los conceptos introducidos.

### 3.5.1. Medida de sensibilidad

En el análisis de sensibilidad de  $n$  vías que se desarrolla en esta Memoria, al igual que en el análisis de sensibilidad de una vía presentado anteriormente, se estudia la sensibilidad de la Red Bayesiana Gaussiana mediante una medida de sensibilidad. En este caso, aplicada a distribuciones multivariantes.

La medida de sensibilidad que se define a continuación, utiliza la medida de divergencia de Kullback-Leibler para comparar dos distribuciones normales multivariantes definidas sobre el mismo conjunto de variables. Dicha divergencia viene dada por la siguiente expresión

$$KL(f, f') = \frac{1}{2} \left[ \ln \frac{|\Sigma'|}{|\Sigma|} + \text{tr}(\Sigma \Sigma'^{-1}) + (\boldsymbol{\mu}' - \boldsymbol{\mu})^T \Sigma'^{-1} (\boldsymbol{\mu}' - \boldsymbol{\mu}) - \dim(\mathbf{X}) \right]$$

donde  $f$  es la densidad de  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$  y  $f'$  la de  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}', \Sigma')$ .

En el estudio de la sensibilidad de una Red Bayesiana Gaussiana que se desarrolla, se busca comparar las distribuciones finales de la red del modelo original con las del modelo perturbado. La distribución final de interés es una distribución normal multivariante condicionada a la evidencia, formada ahora por un conjunto de variables evidenciales  $\mathbf{E}$ .

Para el modelo original la salida final de interés es una distribución normal multivariante, dada por  $\mathbf{Y}|\mathbf{E} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , y para el modelo perturbado dicha salida es también normal multivariante, siendo los parámetros finales, los obtenidos tras realizarse la propagación de la evidencia para el modelo perturbado. Ambas salidas, se obtienen considerando el modelo inicial que describe la red como  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$ , o como dicho modelo con un conjunto de parámetros de  $\boldsymbol{\mu}$  o  $\Sigma$  modificados, en función de los parámetros inciertos del problema, para el modelo perturbado. De forma que, el modelo perturbado se obtiene tras sumar un vector o una matriz de perturbaciones, con las particiones originales asociadas al vector de perturbación de medias  $\boldsymbol{\delta}$  y a la matriz de perturbación de covarianzas  $\Delta$ . Dicha partición se suma a  $\boldsymbol{\mu}_{\mathbf{Y}}$ ,  $\boldsymbol{\mu}_{\mathbf{E}}$ ,  $\Sigma_{\mathbf{Y}\mathbf{Y}}$ ,  $\Sigma_{\mathbf{E}\mathbf{E}}$  o  $\Sigma_{\mathbf{Y}\mathbf{E}}$ , en cada caso, definiéndose así cinco posibles modelos perturbados. En el Apartado siguiente se profundiza en el proceso de ejecución del análisis de sensibilidad de  $n$  vías.

A continuación, se define la medida de sensibilidad para distribuciones multi-



variantes sobre una Red Bayesiana Gaussiana descrita con un conjunto de variables de interés. La expresión resultante obtenida para la medida de sensibilidad es complicada, aunque la aplicación de la misma no presenta ninguna dificultad.

**Definición 3.2** (*Medida de sensibilidad para distribuciones multivariantes*)

Sea  $(\mathcal{D}, P)$  una Red Bayesiana Gaussiana con distribución conjunta  $N(\boldsymbol{\mu}, \Sigma)$ . Sea la salida final de la red, tras propagar la evidencia en el modelo original,  $\mathbf{Y}|\mathbf{E} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , con  $f$  densidad asociada, y sea la salida final de la red, tras propagar la evidencia en el modelo perturbado,  $\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j}, \Sigma^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j})$ , con  $f^{\mathbf{P}_j}$  densidad asociada.

Se define la medida de sensibilidad para distribuciones multivariantes como

$$\begin{aligned} S^{\mathbf{P}_j}(f, f^{\mathbf{P}_j}) &= E_f \left[ \ln \frac{f}{f^{\mathbf{P}_j}} \right] \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\ &\quad + \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_j} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right] \end{aligned} \quad (3.12)$$

donde el subíndice  $\mathbf{p}_j$  indica el conjunto de parámetros inciertos para los que se calcula la medida, siendo  $\mathbf{p}_j^{\boldsymbol{\delta}_j} = \mathbf{p}_j + \boldsymbol{\delta}_j$ .

La medida de sensibilidad introducida sirve para estudiar la sensibilidad cuando la salida de interés viene dada por la distribución final de un conjunto de variables.

Con la medida de sensibilidad definida para distribuciones multivariantes, en el siguiente Apartado, se especifican los pasos que se han de seguir al realizar el análisis de sensibilidad de  $n$  vías propuesto.

### 3.5.2. Análisis de sensibilidad de n vías

En este Apartado se muestra con detalle el proceso a realizar para estudiar la sensibilidad de una Red Bayesiana Gaussiana formada por un conjunto de variables de interés.

El mecanismo de análisis es un proceso iterativo que compara en cada paso el modelo original con el modelo perturbado, mediante la medida de sensibilidad para distribuciones multivariantes. De forma que ambos modelos definen la red inicial y se comparan finalmente cuando se ha realizado la propagación de la evidencia.

Sea  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$  el conjunto de variables modelizadas con la Red Bayesiana Gaussiana en estudio. Se sabe que la distribución inicial conjunta del modelo original es normal multivariante, de manera que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$ . Al tener particionado el conjunto de variables, los parámetros  $\boldsymbol{\mu}$  y  $\Sigma$  también se pueden particionar, de manera que

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}} \\ \boldsymbol{\mu}_{\mathbf{E}} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} & \Sigma_{\mathbf{E}\mathbf{E}} \end{pmatrix}.$$

El modelo perturbado se obtiene tras añadir un conjunto de perturbaciones a cualquiera de los parámetros particionados mostrados, es decir, si por ejemplo se consideran inexactos los valores de las medias de interés, entonces, el modelo perturbado inicial viene dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^{\delta_{\mathbf{Y}}}, \Sigma)$ , donde

$$\boldsymbol{\mu}^{\delta_{\mathbf{Y}}} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}} + \boldsymbol{\delta}_{\mathbf{Y}} \\ \boldsymbol{\mu}_{\mathbf{E}} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} & \Sigma_{\mathbf{E}\mathbf{E}} \end{pmatrix}.$$

De esta forma se pueden definir cinco modelos perturbados distintos, en función de donde se añadan las perturbaciones, es decir, en función de si se perturba  $\boldsymbol{\mu}_{\mathbf{Y}}$ ,  $\boldsymbol{\mu}_{\mathbf{E}}$ ,  $\Sigma_{\mathbf{Y}\mathbf{Y}}$ ,  $\Sigma_{\mathbf{E}\mathbf{E}}$  o  $\Sigma_{\mathbf{Y}\mathbf{E}}$ .

Además, como se ha introducido al comienzo de la Sección, las perturbaciones se presentan mediante un vector de perturbación de medias  $\boldsymbol{\delta}$  y mediante la

matriz de perturbación de covarianzas  $\Sigma$ . De forma que el elemento  $j$ -ésimo del vector representa una cuantificación de la incertidumbre en la media de  $X_j$ , el elemento  $(j, j)$  cuantifica la incertidumbre acerca de la varianza de  $X_j$  y el elemento  $(j, k) = (k, j)$  representa la incertidumbre acerca de la covarianza entre  $X_j$  y  $X_k$ . Además, tanto el vector de perturbación de medias  $\delta$  como la matriz de perturbación de covarianzas  $\Delta$  se representan particionados, siendo el elemento  $\delta_j$  de la Definición 3.2, un elemento de la partición de  $\delta$  o de  $\Delta$ , de forma que

$$\delta = \begin{pmatrix} \delta_{\mathbf{Y}} \\ \delta_{\mathbf{E}} \end{pmatrix} \quad \Delta = \begin{pmatrix} \Delta_{\mathbf{Y}\mathbf{Y}} & \Delta_{\mathbf{Y}\mathbf{E}} \\ \Delta_{\mathbf{E}\mathbf{Y}} & \Delta_{\mathbf{E}\mathbf{E}} \end{pmatrix}.$$

Cuando no existe incertidumbre acerca de un parámetro concreto, se le asigna el valor cero al elemento correspondiente del vector de perturbación de medias  $\delta$  o de la matriz de perturbación de varianzas  $\Delta$ .

Recuérdese que el modelo original y el modelo perturbado describen la Red Bayesiana Gaussiana inicial, y que el interés de la red se centra en las salidas finales de los modelos, tras realizarse la propagación de la evidencia. Es decir, para el modelo original y para el perturbado se obtiene la distribución final de interés como la distribución de  $\mathbf{Y}|\mathbf{E}$  y  $\mathbf{Y}|\mathbf{E}, \delta_j$ , respectivamente, una vez que se ha propagado la evidencia, siendo  $\delta_j$  un elemento de la partición de  $\delta$  o de  $\Delta$ . Posteriormente, se comparan ambas distribuciones calculando para cada modelo perturbado las distintas medidas de sensibilidad.

Es importante señalar que el proceso de propagación de la evidencia se realiza completo, es decir se introduce una a una la evidencia en la red y una vez que se ha introducido toda la evidencia acerca de todas las variables evidenciales de la misma se calcula la distribución final de interés dada por una distribución conjunta condicionada a  $\mathbf{E}$ .

En este análisis de sensibilidad, se pueden obtener cinco medidas de sensibilidad distintas. Dependiendo de los valores de las mismas, se puede determinar el conjunto de variables que han de ser revisadas para asignar sus parámetros con mayor precisión.

Los resultados de las medidas de sensibilidad, en cada caso, se muestran en los siguientes puntos, reflejando separadamente el comportamiento de la medida

de sensibilidad cuando se perturba el vector de medias y cuando se perturba la matriz de covarianzas.

### Sensibilidad para el vector de medias

Cuando el conjunto de parámetros inexactos son elementos del vector de medias, se han de considerar dos situaciones distintas al especificar el modelo perturbado que describe la parte cuantitativa de la Red Bayesiana Gaussiana. Dichas situaciones dependen de los parámetros inexactos y aparecen cuando:

1. Se quiere estudiar la sensibilidad respecto a algunas medias de las variables de interés  $\mathbf{Y}$ . En este caso, el modelo perturbado es normal multivariante, de forma que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^{\delta_{\mathbf{Y}}}, \Sigma)$ , siendo

$$\boldsymbol{\mu}^{\delta_{\mathbf{Y}}} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}} + \boldsymbol{\delta}_{\mathbf{Y}} \\ \boldsymbol{\mu}_{\mathbf{E}} \end{pmatrix}$$

con  $\boldsymbol{\delta}_{\mathbf{Y}}$  perturbaciones de las medias de interés del vector de perturbación de medias  $\boldsymbol{\delta}$ .

2. Se quiere estudiar la sensibilidad respecto a algunas medias de las variables evidenciales  $\mathbf{E}$ . Por tanto, el modelo perturbado viene dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}^{\delta_{\mathbf{E}}}, \Sigma)$ , donde

$$\boldsymbol{\mu}^{\delta_{\mathbf{E}}} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{Y}} \\ \boldsymbol{\mu}_{\mathbf{E}} + \boldsymbol{\delta}_{\mathbf{E}} \end{pmatrix}$$

con  $\boldsymbol{\delta}_{\mathbf{E}}$  perturbaciones de las medias de interés del vector de perturbación de medias  $\boldsymbol{\delta}$ .

En la Proposición 3.10 se calcula la medida de sensibilidad para distribuciones multivariantes en los dos casos anteriormente expuestos.

### Proposición 3.10

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , tal que  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ , siendo  $\mathbf{Y}$  el conjunto de variables de interés y  $\mathbf{E}$  el conjunto de variables evidenciales. Si se considera el vector de perturbación de medias  $\boldsymbol{\delta}$  particionado tal que  $\boldsymbol{\delta} = (\boldsymbol{\delta}_{\mathbf{Y}}, \boldsymbol{\delta}_{\mathbf{E}})^T$ , entonces la medida de sensibilidad (3.12) es tal que

1. Cuando se suma la perturbación  $\delta_{\mathbf{Y}}$  a la media de las variables de interés  $\mathbf{Y}$ , la distribución final de interés del modelo perturbado, tras la propagación de la evidencia, es

$$\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{Y}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{Y}}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$$

con

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{Y}}} = \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} + \delta_{\mathbf{Y}}$$

La medida de sensibilidad viene dada por la siguiente expresión

$$S^{\mu_{\mathbf{Y}}}(f, f^{\mu_{\mathbf{Y}}}) = \frac{1}{2} \left[ \delta_{\mathbf{Y}}^T (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} \delta_{\mathbf{Y}} \right] \quad (3.13)$$

2. Cuando se aplica la perturbación  $\delta_{\mathbf{E}}$  a la media de las variables evidenciales  $\mathbf{E}$ , la distribución final de interés del modelo perturbado, tras la propagación de la evidencia, es

$$\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{E}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{E}}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$$

donde

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{E}}} = \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} - \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \delta_{\mathbf{E}}$$

La medida de sensibilidad es

$$S^{\mu_{\mathbf{E}}}(f, f^{\mu_{\mathbf{E}}}) = \frac{1}{2} \left[ \delta_{\mathbf{E}}^T (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1})^T (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1}) \delta_{\mathbf{E}} \right] \quad (3.14)$$

### Demostración 3.10

Los parámetros finales del modelo perturbado se obtienen directamente teniendo en cuenta el proceso de propagación de la evidencia.

En cada caso se trabaja con un modelo perturbado inicial distinto.

Se calcula la medida de sensibilidad para distribuciones multivariantes directamente aplicando la expresión asociada a dicha medida. Así,

1. Para calcular  $S^{\mu_{\mathbf{Y}}}(f, f^{\mu_{\mathbf{Y}}})$  se sabe que la varianza final obtenida para el modelo perturbado, coincide con dicha varianza para el modelo original, por tanto  $\text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} \right) = \dim(\mathbf{Y})$  y la medida de sensibilidad es

$$\begin{aligned}
 S^{\mu_{\mathbf{Y}}}(f, f^{\mu_{\mathbf{Y}}}) &= \\
 &= \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{Y}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{Y}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right] \\
 &= \frac{1}{2} \left[ \boldsymbol{\delta}_{\mathbf{Y}}^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \right)^{-1} \boldsymbol{\delta}_{\mathbf{Y}} \right]
 \end{aligned}$$

2. En el caso de  $S^{\mu_{\mathbf{E}}}(f, f^{\mu_{\mathbf{E}}})$  también coinciden las varianzas finales del modelo original y del perturbado, entonces  $\text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \right)^{-1} \right) = \dim(\mathbf{Y})$ . Por tanto,

$$\begin{aligned}
 S^{\mu_{\mathbf{E}}}(f, f^{\mu_{\mathbf{E}}}) &= \\
 &= \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right] \\
 &= \frac{1}{2} \left[ \boldsymbol{\delta}_{\mathbf{E}}^T \left( \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \right)^{-1} \left( \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \right) \boldsymbol{\delta}_{\mathbf{E}} \right]
 \end{aligned}$$

■

Al igual que en el caso del análisis de sensibilidad de una vía, cuando existe incertidumbre acerca de algunas variables evidenciales  $\mathbf{E}$ , si se añade un *vector de perturbación de evidencias*  $\boldsymbol{\delta}_{\mathbf{e}}$  al conjunto de evidencias, siendo  $\mathbf{e}^{\delta_{\mathbf{e}}} = \mathbf{e} + \boldsymbol{\delta}_{\mathbf{e}}$ , se obtiene la medida de sensibilidad calculada para  $S^{\mu_{\mathbf{E}}}(f, f^{\mu_{\mathbf{E}}})$ . Por tanto, este caso se puede estudiar trabajando con la medida de sensibilidad obtenida cuando se perturban las medias de las variables evidenciales. La siguiente proposición, recoge el resultado.

### Proposición 3.11

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , tal que  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ , siendo  $\mathbf{Y}$  el conjunto de variables de interés y  $\mathbf{E}$  el conjunto de variables evidenciales. Considérense inexactos algunos valores evidenciales. Se define el vector de perturbación de evidencias  $\boldsymbol{\delta}_{\mathbf{e}}$ , de forma que para el modelo perturbado se trabaja con las nuevas evidencias,  $\mathbf{e}^{\delta_{\mathbf{e}}} = \mathbf{e} + \boldsymbol{\delta}_{\mathbf{e}}$ . La salida de la red recogida en la distribución final de interés para el modelo perturbado es

$$\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}_{\mathbf{e}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{e}}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$$

donde

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \delta_{\mathbf{e}}} = \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} + \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \boldsymbol{\delta}_{\mathbf{e}}$$

La medida de sensibilidad obtenida coincide con la medida de sensibilidad de la expresión (3.14), siendo

$$S^e(f, f^e) = \frac{1}{2} \left[ \delta_e^T (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1})^T (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1}) \delta_e \right] \quad (3.15)$$

### Demostración 3.11

Los parámetros de la distribución del modelo perturbado se obtienen directamente, tras realizarse la propagación de la evidencia.

La medida de sensibilidad se calcula directamente aplicando (3.12). Así, teniendo en cuenta que coinciden las varianzas finales del modelo original y del perturbado y por tanto  $\text{tr}(\Sigma^{\mathbf{Y}|\mathbf{E}} (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1}) = \dim(\mathbf{Y})$ , entonces

$$\begin{aligned} S^e(f, f^e) &= \\ &= \frac{1}{2} \left[ \left( \mu^{\mathbf{Y}|\mathbf{E}, \delta_e} - \mu^{\mathbf{Y}|\mathbf{E}} \right)^T (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} \left( \mu^{\mathbf{Y}|\mathbf{E}, \delta_e} - \mu^{\mathbf{Y}|\mathbf{E}} \right) \right] \\ &= \frac{1}{2} \left[ \delta_e^T (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1})^T (\Sigma^{\mathbf{Y}|\mathbf{E}})^{-1} (\Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1}) \delta_e \right] \end{aligned}$$

■

Nótese que en la Proposición 3.11 se trabaja con las perturbaciones de los valores evidenciales  $\delta_e$  y en (3.14) se consideran las perturbaciones de las medias correspondientes a las variables sobre las que se introduce evidencia, denotadas como  $\delta_{\mathbf{E}}$ .

A continuación, se calcula la medida de sensibilidad cuando existe incertidumbre en la matriz de covarianzas.

### Sensibilidad para la matriz de covarianzas

Para el estudio de la sensibilidad en la matriz de covarianzas, se deben considerar tres situaciones distintas.

1. Se tiene incertidumbre acerca de algunas varianzas y covarianzas de las variables de interés  $\mathbf{Y}$ . Entonces, el modelo perturbado es normal multivariante, tal que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^{\Delta_{\mathbf{Y}\mathbf{Y}}})$ , donde

$$\Sigma^{\Delta_{\mathbf{Y}\mathbf{Y}}} = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} + \Delta_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} & \Sigma_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

siendo  $\Delta_{\mathbf{Y}\mathbf{Y}}$  la perturbación introducida sobre las varianzas y covarianzas de  $\mathbf{Y}$ . Los elementos de  $\Delta_{\mathbf{Y}\mathbf{Y}}$  han de verificar que la matriz  $\Sigma^{\Delta_{\mathbf{Y}\mathbf{Y}}}$  sea definida positiva y que la matriz de covarianzas final, calculada tras la propagación de la evidencia, sea también definida positiva.

2. Estudiar la sensibilidad respecto a las varianzas y covarianzas de las variables evidenciales  $\mathbf{E}$ . En este caso, el modelo perturbado es normal multivariante, dado por  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^{\Delta_{\mathbf{E}\mathbf{E}}})$ , donde

$$\Sigma^{\Delta_{\mathbf{E}\mathbf{E}}} = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} & \Sigma_{\mathbf{E}\mathbf{E}} + \Delta_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

siendo  $\Delta_{\mathbf{E}\mathbf{E}}$  las perturbaciones asociadas a las varianzas y covarianzas de  $\mathbf{E}$ . Los elementos de  $\Delta_{\mathbf{E}\mathbf{E}}$  han de verificar que la matriz  $\Sigma^{\Delta_{\mathbf{E}\mathbf{E}}}$  sea definida positiva y que la matriz de covarianzas final, calculada tras la propagación de la evidencia, también sea definida positiva.

3. Estudiar la sensibilidad respecto a las covarianzas entre las variables de interés y las variables evidenciales. El modelo perturbado se define tal que  $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma^{\Delta_{\mathbf{Y}\mathbf{E}}})$ , donde

$$\Sigma^{\Delta_{\mathbf{Y}\mathbf{E}}} = \begin{pmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{E}} \\ \Sigma_{\mathbf{E}\mathbf{Y}} + \Delta_{\mathbf{Y}\mathbf{E}}^{\mathbf{T}} & \Sigma_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

siendo  $\Delta_{\mathbf{Y}\mathbf{E}}$  las perturbaciones asociadas a las covarianzas entre las variables de  $\mathbf{Y}$  y las de  $\mathbf{E}$ . Los elementos de  $\Delta_{\mathbf{Y}\mathbf{E}}$  han de verificar que la matriz  $\Sigma^{\Delta_{\mathbf{Y}\mathbf{E}}}$  sea definida positiva y que, tras la propagación de la evidencia, la matriz de covarianzas final sea también definida positiva.

En la Proposición 3.12 se recogen las medidas de sensibilidad para los casos anteriormente citados.



**Proposición 3.12**

Sea una Red Bayesiana Gaussiana  $(\mathcal{D}, P)$  con  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ , tal que  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ , donde  $\mathbf{Y}$  es el conjunto de variables de interés y  $\mathbf{E}$  es el conjunto de variables evidenciales. Considérese la matriz de perturbación de covarianzas  $\Delta$  formada por las perturbaciones asociadas a los parámetros inciertos de  $\Sigma$ , tal que

$$\Delta = \begin{pmatrix} \Delta_{\mathbf{Y}\mathbf{Y}} & \Delta_{\mathbf{Y}\mathbf{E}} \\ \Delta_{\mathbf{E}\mathbf{Y}} & \Delta_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$

La medida de sensibilidad (3.12) viene dada por

1. Cuando se suma la perturbación  $\Delta_{\mathbf{Y}\mathbf{Y}}$  a las varianzas y covarianzas de las variables de interés  $\mathbf{Y}$ , la distribución final de interés del modelo perturbado, tras la propagación de la evidencia es

$$\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{Y}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{Y}}})$$

donde

$$\Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{Y}}} = \Sigma^{\mathbf{Y}|\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{Y}}$$

La medida de sensibilidad es

$$\begin{aligned} S^{\Sigma_{\mathbf{Y}\mathbf{Y}}}(f, f^{\Sigma_{\mathbf{Y}\mathbf{Y}}}) &= \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{Y}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{Y}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] \end{aligned} \quad (3.16)$$

2. Para estudiar la sensibilidad respecto a las varianzas y covarianzas de las variables evidenciales  $\mathbf{E}$ , se suma la perturbación  $\Delta_{\mathbf{E}\mathbf{E}}$ . La distribución final de interés del modelo perturbado, tras la propagación de la evidencia es

$$\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}})$$

con

$$\begin{aligned} \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} &= \boldsymbol{\mu}_{\mathbf{Y}} + \Sigma_{\mathbf{Y}\mathbf{E}} (\Sigma_{\mathbf{E}\mathbf{E}} + \Delta_{\mathbf{E}\mathbf{E}})^{-1} (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}}) \\ \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} &= \Sigma_{\mathbf{Y}\mathbf{Y}} - \Sigma_{\mathbf{Y}\mathbf{E}} (\Sigma_{\mathbf{E}\mathbf{E}} + \Delta_{\mathbf{E}\mathbf{E}})^{-1} \Sigma_{\mathbf{E}\mathbf{Y}} \end{aligned}$$

La medida de sensibilidad viene dada por

$$\begin{aligned}
 S^{\Sigma_{\mathbf{E}\mathbf{E}}}(f, f^{\Sigma_{\mathbf{E}\mathbf{E}}}) &= \\
 &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\
 &\quad + \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{E}\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right]
 \end{aligned} \tag{3.17}$$

3. Cuando se añaden las perturbaciones  $\Delta_{\mathbf{Y}\mathbf{E}}$  y  $\Delta_{\mathbf{E}\mathbf{Y}}$  a las covarianzas entre las variables de interés  $\mathbf{Y}$  y las variables evidenciales  $\mathbf{E}$ , la distribución final de interés del modelo perturbado, tras la propagación de la evidencia, es normal multivariante

$$\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}})$$

donde

$$\begin{aligned}
 \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} &= \boldsymbol{\mu}_{\mathbf{Y}} + (\Sigma_{\mathbf{Y}\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{E}}) \Sigma_{\mathbf{E}\mathbf{E}}^{-1} (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}}) \\
 \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} &= \Sigma_{\mathbf{Y}\mathbf{Y}} - (\Sigma_{\mathbf{Y}\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{E}}) \Sigma_{\mathbf{E}\mathbf{E}}^{-1} (\Sigma_{\mathbf{E}\mathbf{Y}} + \Delta_{\mathbf{E}\mathbf{Y}})
 \end{aligned}$$

La medida de sensibilidad viene dada por

$$\begin{aligned}
 S^{\Sigma_{\mathbf{Y}\mathbf{E}}}(f, f^{\Sigma_{\mathbf{Y}\mathbf{E}}}) &= \\
 &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}})|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}}) \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\
 &\quad + \frac{1}{2} \left[ (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}})^T (\Sigma_{\mathbf{E}\mathbf{E}}^{-1})^T \Delta_{\mathbf{Y}\mathbf{E}}^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}}) \right)^{-1} \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}}) \right]
 \end{aligned} \tag{3.18}$$

siendo  $M(\Delta_{\mathbf{Y}\mathbf{E}}) = \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Sigma_{\mathbf{Y}\mathbf{E}}^T + \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Delta_{\mathbf{E}\mathbf{Y}} + \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Delta_{\mathbf{E}\mathbf{Y}}$ .

Para todos los casos presentados la matriz de covarianzas inicial del modelo perturbado ha de ser definida positiva, al igual que la matriz de covarianzas final calculada para el modelo perturbado.

### Demostración 3.12

Los parámetros finales del modelo perturbado se calculan directamente.

En cada caso se trabaja con un modelo inicial perturbado distinto.

El cálculo de la medida de sensibilidad es directo, aplicando la expresión asociada a dicha medida (3.12). Así,

1. En este caso la media final de interés del modelo perturbado coincide con dicha media para el modelo original, lo que simplifica la expresión de la medida. Teniendo esto en cuenta, se tiene que

$$\begin{aligned} S^{\Sigma_{\mathbf{Y}\mathbf{Y}}}(f, f^{\Sigma_{\mathbf{Y}\mathbf{Y}}}) &= \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{Y}}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{Y}}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{Y}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{Y}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] \end{aligned}$$

2. En este caso, la expresión asociada a la medida de sensibilidad coincide con la descrita para la medida de sensibilidad en (3.12).

3. La medida de sensibilidad de interés viene dada por

$$\begin{aligned} S^{\Sigma_{\mathbf{Y}\mathbf{E}}}(f, f^{\Sigma_{\mathbf{Y}\mathbf{E}}}) &= \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\ &\quad + \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right] \\ &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}})|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}}) \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\ &\quad + \frac{1}{2} \left[ \left( \mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}} \right)^T \left( \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \right)^T \Delta_{\mathbf{Y}\mathbf{E}}^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}}) \right)^{-1} \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \left( \mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}} \right) \right] \end{aligned}$$

siendo

$$\Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{Y}\mathbf{E}}} = \Sigma_{\mathbf{Y}\mathbf{Y}} - (\Sigma_{\mathbf{Y}\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{E}}) \Sigma_{\mathbf{E}\mathbf{E}}^{-1} (\Sigma_{\mathbf{E}\mathbf{Y}} + \Delta_{\mathbf{E}\mathbf{Y}}) = \Sigma^{\mathbf{Y}|\mathbf{E}} - M(\Delta_{\mathbf{Y}\mathbf{E}})$$

$$\text{con } M(\Delta_{\mathbf{Y}\mathbf{E}}) = \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Sigma_{\mathbf{Y}\mathbf{E}}^T + \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Delta_{\mathbf{E}\mathbf{Y}} + \Delta_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} \Delta_{\mathbf{E}\mathbf{Y}}. \blacksquare$$

En los resultados presentados se habla de 'algunos parámetros inexactos', porque puede existir certeza en la asignación de varios de los parámetros de la red. Cuando existe consenso y certeza acerca del valor de un parámetro, entonces, en el vector de perturbación de medias  $\boldsymbol{\delta}$  o en la matriz de perturbación de covarianzas  $\Delta$  se asigna un cero a la posición asociada al parámetro cierto.

A continuación, se presenta un ejemplo del análisis de sensibilidad de  $n$  vías introducido en este Apartado.

**Ejemplo 3.4**

Considérese la Red Bayesiana Gaussiana recogida en el Ejemplo 3.1. Ahora el interés se centra en las variables  $X_3, X_4$  y  $X_5$ , siendo  $\mathbf{Y} = \{X_3, X_4, X_5\}$ . La evidencia viene dada por  $\mathbf{E} = \{X_1 = 2, X_2 = 4\}$ . En la Figura 3.4, se muestra el DAG con la nueva situación, y a continuación se especifican los parámetros que describen la red.

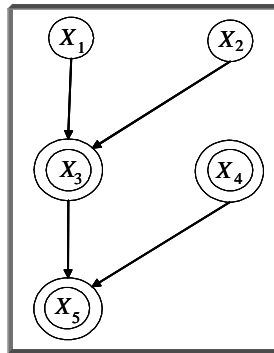


Figura 3.4. DAG asociado a la Red Bayesiana Gaussiana descrita para el Ejemplo 3.4

Siendo  $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$  donde

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 3 & 0 & 6 & 0 & 6 \\ 0 & 2 & 2 & 0 & 2 \\ 6 & 2 & 15 & 0 & 15 \\ 0 & 0 & 0 & 2 & 4 \\ 6 & 2 & 15 & 4 & 26 \end{pmatrix}$$

Si se estudia la sensibilidad caracterizada mediante

$$\begin{aligned} \boldsymbol{\delta}_{\mathbf{Y}} &= \begin{pmatrix} -2 \\ -1 \\ 0 \end{pmatrix} & \boldsymbol{\delta}_{\mathbf{E}} &= \begin{pmatrix} 0 \\ 2 \end{pmatrix} & \Delta_{\mathbf{Y}\mathbf{Y}} &= \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{pmatrix} \\ \Delta_{\mathbf{E}\mathbf{E}} &= \begin{pmatrix} 0 & 0 \\ 0 & 3 \end{pmatrix} & \Delta_{\mathbf{Y}\mathbf{E}} &= \begin{pmatrix} -1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

se obtienen los siguientes resultados.

Lo primero que se ha de señalar es que para obtener los valores del vector de perturbación de medias  $\delta$  y de la matriz de perturbación de covarianzas  $\Delta$ , es necesario que los expertos expresen sus incertidumbres acerca de los parámetros asignados para el modelo original.

Además es necesario que, para todas las perturbaciones asociadas a la matriz de covarianzas, la matriz de covarianzas inicial y final del modelo perturbado sean definidas positivas. En este ejemplo, para las perturbaciones presentadas se cumple dicha hipótesis.

Como se puede ver con la información del problema, se observa que no existe incertidumbre acerca de la media o la varianza de la variable evidencial  $X_1$  y sólo se declara incierta la covarianza entre  $X_1$  y  $X_3$ .

Lo primero que se va a calcular en esta red es la salida de interés final para el modelo original, tras realizarse la propagación de la evidencia. Para ello, se sabe que la distribución de interés final es normal multivariante, tal que  $\mathbf{Y}|\mathbf{E} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , donde

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 4 \\ 4 \\ 6 \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 4 \\ 1 & 4 & 12 \end{pmatrix}$$

con  $\mathbf{Y} = \{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$  conjunto de variables de interés y  $\mathbf{E} = \{\mathbf{X}_1 = 2, \mathbf{X}_2 = 4\}$  el conjunto de variables evidenciales de la Red Bayesiana Gaussiana que describe el problema de interés.

Una vez que se ha comprobado que las perturbaciones propuestas para la matriz de covarianzas, mantienen las matrices de covarianzas perturbadas definidas positivas, se puede calcular el valor de la medida de sensibilidad aplicando las Propositiones 3.10 y 3.12. Para obtener algunas expresiones de la medida de sensibilidad es necesario calcular la distribución final de interés para el modelo perturbado, como es el caso de  $S^{\Sigma_{\mathbf{E}\mathbf{E}}}(f, f^{\Sigma_{\mathbf{E}\mathbf{E}}})$  y  $S^{\Sigma_{\mathbf{Y}\mathbf{E}}}(f, f^{\Sigma_{\mathbf{Y}\mathbf{E}}})$ . Para estas medidas, dichos modelos son tales que

$$\begin{aligned} \mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{EE}} &\sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{EE}}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{EE}}}) \\ \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{EE}}} &= \begin{pmatrix} \frac{17}{5} \\ 4 \\ \frac{27}{5} \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{EE}}} = \begin{pmatrix} \frac{11}{5} & 0 & \frac{11}{5} \\ 0 & 2 & 4 \\ \frac{11}{5} & 4 & \frac{66}{5} \end{pmatrix} \end{aligned}$$

y

$$\begin{aligned} \mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{YE}} &\sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{YE}}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{YE}}}) \\ \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{YE}}} &= \begin{pmatrix} 4 \\ 4 \\ \frac{13}{2} \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta_{\mathbf{YE}}} = \begin{pmatrix} \frac{14}{3} & 0 & 2 \\ 0 & 2 & 4 \\ 2 & 4 & \frac{19}{2} \end{pmatrix} \end{aligned}$$

La medida de sensibilidad obtenida para cada conjunto de perturbaciones es

$$S^{\boldsymbol{\mu}^{\mathbf{Y}}}(f, f^{\boldsymbol{\mu}^{\mathbf{Y}}}) = 4.916$$

$$S^{\boldsymbol{\mu}^{\mathbf{E}}}(f, f^{\boldsymbol{\mu}^{\mathbf{E}}}) = 2$$

$$S^{\Sigma^{\mathbf{Y}\mathbf{Y}}}(f, f^{\Sigma^{\mathbf{Y}\mathbf{Y}}}) = 0.491$$

$$S^{\Sigma^{\mathbf{EE}}}(f, f^{\Sigma^{\mathbf{EE}}}) = 0.203$$

$$S^{\Sigma^{\mathbf{YE}}}(f, f^{\Sigma^{\mathbf{YE}}}) = 1.889$$

Con las medidas de sensibilidad obtenidas, se observa como la incertidumbre que más afecta a la salida de la red es la referente a la media de las variables de interés.

Para las incertidumbres acerca de la matriz de covarianzas se observa como la medida de sensibilidad mayor se corresponde con incertidumbre en las covarianzas entre  $\mathbf{Y}$  y  $\mathbf{E}$ . Además, los valores de las medidas de sensibilidad obtenidas para perturbaciones en la matriz de covarianzas son menores, porque para todos los casos, salvo la covarianza entre  $X_3$  y  $X_1$ , el aumento de la variabilidad de las variables no hace que cambie demasiado la distribución de interés. Cabe señalar el valor obtenido para la sensibilidad cuando la varianza evidencial es incierta, que como se puede observar es el valor más pequeño de las medidas de sensibilidad calculadas, por lo que, al igual que en el caso del análisis de sensibilidad de una vía, la medida de sensibilidad para incertidumbres en las varianzas-covarianzas de

las variables evidenciales es pequeña ya que al tener evidencia sobre las variables, la incertidumbre en su matriz de covarianzas no influye demasiado en las salidas finales de las variables de interés.

### 3.6. Robustez en Redes Bayesianas Gaussianas

La idea de concluir resultados acerca de la robustez de una Red Bayesiana Gaussiana surge asociada a los distintos análisis de sensibilidad presentados. De esta forma, se puede tratar la robustez de la red, respecto a las perturbaciones propuestas, en función de los valores de las medidas de sensibilidad. Así, cuando estos valores son pequeños, menores que un umbral previamente establecido, se puede concluir que la red es robusta, en otro caso se tiene que la red no es robusta.

En esta Sección se recoge una aproximación para estudiar la robustez cuantitativa, a partir de la metodología propuesta para realizar los análisis de sensibilidad desarrollados. Con este objetivo, se define una *medida de robustez*, también basada en la divergencia de Kullback-Leibler para distribuciones normales multivariantes, de forma que, para todos las posibles perturbaciones de los parámetros inexactos de la red se calcula una única medida. Dicha medida de robustez compara el modelo original con el modelo perturbado, tras la propagación de la evidencia, y en función de su valor se concluye si la Red Bayesiana Gaussiana es o no robusta, para las perturbaciones propuestas.

En este caso, se trabaja con una Red Bayesiana Gaussiana definida en su forma más generalizada, es decir, con un conjunto de variables de interés y un conjunto de variables evidenciales, siendo  $\mathbf{X} = \{\mathbf{Y}, \mathbf{E}\}$ . Tanto el modelo original como el perturbado se describen inicialmente, antes de introducir evidencia en la red, y se estudian sus salidas finales, dadas por las distribuciones de interés, tras la propagación de la evidencia. Además, el modelo perturbado se obtiene sumando en un único paso el vector de perturbación de medias  $\boldsymbol{\delta}$  y la matriz de perturbación de covarianzas  $\Delta$ , de manera que,  $\boldsymbol{\delta}$  y  $\Delta$  están compuestas por todas las perturbaciones indicadas por los expertos asociadas a los parámetros que describen la red.

A continuación, se define la medida de robustez para una Red Bayesiana

Gaussiana.

**Definición 3.3** (*Medida de robustez*)

Sea  $(\mathcal{D}, P)$  una Red Bayesiana Gaussiana con distribución conjunta  $N(\boldsymbol{\mu}, \Sigma)$ . Sea la salida final de la red, tras propagar la evidencia en el modelo original, tal que  $\mathbf{Y}|\mathbf{E} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , con función de densidad asociada  $f$ , y sea la salida final de la red, tras propagar la evidencia en el modelo perturbado, tal que  $\mathbf{Y}|\mathbf{E}, \mathbf{p} \sim N(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \mathbf{p}}, \Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}})$ , con función de densidad asociada  $f^{\mathbf{p}}$ .

Se define la medida de robustez para distribuciones multivariantes como

$$\begin{aligned}
 R^{\mathbf{p}}(f, f^{\mathbf{p}}) &= E_f \left[ \ln \frac{f}{f^{\mathbf{p}}} \right] \\
 &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} \right)^{-1} \right) - \dim(\mathbf{Y}) \right] + \\
 &\quad + \frac{1}{2} \left[ \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} \right)^{-1} \left( \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} - \boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} \right) \right]
 \end{aligned} \tag{3.19}$$

donde el subíndice  $\mathbf{p}$  indica las perturbaciones asociadas a los parámetros inciertos, de forma que, si se perturba el vector de medias  $\boldsymbol{\mu}$  y la matriz de covarianzas  $\Sigma$ , entonces  $\mathbf{p} = (\boldsymbol{\delta}, \Delta)$ , siendo en el modelo perturbado  $\boldsymbol{\mu}^{\boldsymbol{\delta}} = \boldsymbol{\mu} + \boldsymbol{\delta}$  y  $\Sigma^{\Delta} = \Sigma + \Delta$ .

Alternativamente se tiene

$$\begin{aligned}
 R^{\mathbf{p}}(f, f^{\mathbf{p}}) &= \\
 &= \frac{1}{2} \left[ \ln \frac{|\Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}}|}{|\Sigma^{\mathbf{Y}|\mathbf{E}}|} + \text{tr} \left( \Sigma^{\mathbf{Y}|\mathbf{E}} \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} \right)^{-1} \right) - \dim(\mathbf{Y}) + M(\boldsymbol{\delta}, \Delta)^T \left( \Sigma^{\mathbf{Y}|\mathbf{E}, \mathbf{p}} \right)^{-1} M(\boldsymbol{\delta}, \Delta) \right]
 \end{aligned} \tag{3.20}$$

donde

$$M(\boldsymbol{\delta}, \Delta) = \boldsymbol{\delta}_{\mathbf{Y}} + (\Sigma_{\mathbf{Y}\mathbf{E}} + \Delta_{\mathbf{Y}\mathbf{E}}) (\Sigma_{\mathbf{E}\mathbf{E}} + \Delta_{\mathbf{E}\mathbf{E}})^{-1} (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}} - \boldsymbol{\delta}_{\mathbf{E}}) - \Sigma_{\mathbf{Y}\mathbf{E}} \Sigma_{\mathbf{E}\mathbf{E}}^{-1} (\mathbf{e} - \boldsymbol{\mu}_{\mathbf{E}})$$

y el vector de perturbación de medias  $\boldsymbol{\delta}$  y la matriz de perturbación de covarianzas  $\Delta$  se particionan tal que

$$\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_{\mathbf{Y}} \\ \boldsymbol{\delta}_{\mathbf{E}} \end{pmatrix} \quad \Delta = \begin{pmatrix} \Delta_{\mathbf{Y}\mathbf{Y}} & \Delta_{\mathbf{Y}\mathbf{E}} \\ \Delta_{\mathbf{E}\mathbf{Y}} & \Delta_{\mathbf{E}\mathbf{E}} \end{pmatrix}$$



Como se puede observar, la definición asociada a la medida de robustez (3.19) coincide con la medida de sensibilidad (3.12), aunque la aplicación de la misma al estudiar la robustez, cambia respecto del análisis de sensibilidad, ya que ahora se obtiene una única medida que evalúa las diferencias entre las salidas finales para el modelo original y esas mismas salidas para el modelo perturbado.

Para mantener el concepto de Red Bayesiana Gaussiana, al estudiar una aproximación a la robustez de dicho modelo es necesario que la matriz de covarianzas inicial del modelo perturbado,  $\Sigma^\Delta = \Sigma + \Delta$ , y la matriz de covarianzas final,  $\Sigma^{\mathbf{Y}|\mathbf{E},\Delta}$ , sean definidas positivas.

Tras realizarse el proceso de propagación de la evidencia, para el modelo original y para el modelo perturbado, se obtienen las distribuciones finales de interés para ambos modelos y se calcula la medida de robustez, siendo dichas distribuciones normales multivariantes.

Como caso particular, se enuncia la medida de robustez cuando sólo hay una variable de interés en el problema  $X_i$  y una variable evidencial  $X_e$ . En este caso, los parámetros que afectan a la variable de interés, para la cual se obtiene su distribución final como la salida de interés de la red, son la media de interés y la evidencial,  $\delta_i$  y  $\delta_e$  y las varianzas y covarianzas entre ambas variables,  $\delta_{ii}$ ,  $\delta_{ee}$ ,  $\delta_{ie}$ . De esta forma, la medida de robustez definida por las expresiones (3.19) y (3.20) queda reducida al siguiente resultado

$$R(f, f^\delta) = \frac{1}{2} \left[ \ln \left( \frac{\sigma_{ii}^{Y|E,\delta}}{\sigma_{ii}^{Y|E}} \right) + \frac{\frac{(\sigma_{ie} + \delta_{ie})^2}{\sigma_{ee} + \delta_{ee}} - \frac{\sigma_{ie}^2}{\sigma_{ee}} - \delta_{ii} + \left( \delta_i + \frac{\sigma_{ie} + \delta_{ie}}{\sigma_{ee} + \delta_{ee}} (e - \mu_e - \delta_e) - \frac{\sigma_{ie}(e - \mu_e)}{\sigma_{ee}} \right)^2}{\sigma_{ii}^{Y|E,\delta}} \right] \quad (3.21)$$

siendo  $\sigma_{ii}^{Y|E,\delta}$  la varianza final de interés calculada tras propagar la evidencia en el modelo perturbado y  $\sigma_{ii}^{Y|E}$  esa misma varianza pero para el modelo original. Las perturbaciones vienen dadas por  $\delta_i$ ,  $\delta_e$ , asociadas a la media de interés y evidencial y por  $\delta_{ii}$ ,  $\delta_{ee}$  y  $\delta_{ie}$  asociadas a las varianzas de interés y evidencial y

a la covarianza entre ambas. El resto de parámetros, son los que definen la red inicialmente.

A continuación, se presentan dos ejemplos donde se estudia la robustez de las Redes Bayesianas Gaussianas del Ejemplo 3.1 y del Ejemplo 3.4. En la primera red se trabaja con la expresión (3.21) para redes con una variable de interés y una variable evidencial, y en la segunda red se emplea la expresión recogida en (3.20) para un red con un conjunto de variables de interés y otro de variables evidenciales.

### Ejemplo 3.5

*Estudiar la robustez de la Red Bayesiana Gaussiana presentada en el Ejemplo 3.1, teniendo en cuenta las perturbaciones propuestas para el estudio de la sensibilidad de una vía propuesto.*

La red del Ejemplo 3.1, presenta como variable de interés a  $X_5$  y como variable evidencial  $X_2$ , siendo  $E = \{X_2 = 4\}$ . Además, las perturbaciones propuestas indican que el vector de perturbación de medias  $\delta$  y la matriz de perturbación de varianzas  $\Delta$ , son

$$\delta = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 0 \\ 2 \end{pmatrix} \quad \Delta = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 \end{pmatrix}$$

Lo primero que hay que hacer es estudiar si las perturbaciones propuestas hacen que la matriz de covarianzas perturbada sea una matriz definida positiva y si la matriz de covarianzas que describe la distribución final es también definida positiva. Tras dicha comprobación se obtiene que  $\Sigma^\Delta = \Sigma + \Delta$  y  $\Sigma^{\mathbf{Y}|\mathbf{E},\Delta}$  son matrices definidas positivas.

Los parámetros finales que describen las salidas de la red, tras la propagación de la evidencia son.

- Para el modelo original, la distribución final es  $\mathbf{Y}|\mathbf{E} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , donde

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 2 \\ 4 \\ 4 \\ 6 \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 3 & 6 & 0 & 6 \\ 6 & 13 & 0 & 13 \\ 0 & 0 & 2 & 4 \\ 6 & 13 & 4 & 24 \end{pmatrix}$$

- Para el modelo perturbado, la distribución final de la red viene dada por  $\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}, \Delta \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}, \Delta}, \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta})$ , siendo

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}, \boldsymbol{\delta}, \Delta} = \begin{pmatrix} 2 \\ \frac{17}{5} \\ 4 \\ \frac{38}{5} \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}, \Delta} = \begin{pmatrix} 3 & 6 & 0 & 6 \\ 6 & \frac{71}{5} & 0 & \frac{69}{5} \\ 0 & 0 & 2 & 4 \\ 6 & \frac{69}{5} & 4 & \frac{111}{5} \end{pmatrix}$$

Con los parámetros finales de los modelos original y perturbado, se calcula la medida de robustez, aplicando la expresión (3.21). De esta forma, se obtiene que la medida de robustez para las perturbaciones propuestas es

$$R^{\boldsymbol{\delta}, \Delta}(f, f^{\boldsymbol{\delta}, \Delta}) = 0.0025$$

Como se puede observar, la medida de robustez es pequeña, por tanto se podría concluir que la Red Bayesiana Gaussiana del Ejemplo 3.1 es robusta, frente a las perturbaciones propuestas.

### Ejemplo 3.6

*Considérese la Red Bayesiana Gaussiana descrita en el Ejemplo 3.4. Se desea estudiar, para las perturbaciones propuestas, la robustez de la misma.*

Para los datos del Ejemplo 3.4, se tiene que el vector de perturbación de medias  $\boldsymbol{\delta}$  y la matriz de perturbación de covarianzas  $\Delta$ , vienen dados por

$$\boldsymbol{\delta} = \begin{pmatrix} 0 \\ 2 \\ -2 \\ -1 \\ 0 \end{pmatrix} \quad \Delta = \begin{pmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 \\ -1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 2 & 2 \end{pmatrix}$$

Se está trabajando con el conjunto de variables de interés  $\mathbf{Y} = \{\mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5\}$  y el conjunto de variables evidenciales, dadas por  $\mathbf{E} = \{\mathbf{X}_1 = 2, \mathbf{X}_2 = 4\}$ .

Se estudian las matrices  $\Sigma^\Delta = \Sigma + \Delta$  y  $\Sigma^{\mathbf{Y}|\mathbf{E},\Delta}$  y se concluye que son definidas positivas.

Posteriormente, tras la propagación de la evidencia se obtienen las siguientes salidas de interés.

- Para el modelo original, la distribución final de interés es normal multivariante dada por  $\mathbf{Y}|\mathbf{E} \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}}, \Sigma^{\mathbf{Y}|\mathbf{E}})$ , donde

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 4 \\ 4 \\ 6 \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 4 \\ 1 & 4 & 12 \end{pmatrix}$$

- Para el modelo perturbado, la salida de interés final de la red viene dada por  $\mathbf{Y}|\mathbf{E}, \delta, \Delta \sim \mathbf{N}(\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E},\delta,\Delta}, \Sigma^{\mathbf{Y}|\mathbf{E},\Delta})$ , siendo

$$\boldsymbol{\mu}^{\mathbf{Y}|\mathbf{E},\delta,\Delta} = \begin{pmatrix} \frac{17}{5} \\ 3 \\ \frac{28}{5} \end{pmatrix} \quad \Sigma^{\mathbf{Y}|\mathbf{E},\Delta} = \begin{pmatrix} \frac{118}{15} & 0 & \frac{19}{5} \\ 0 & 3 & 6 \\ \frac{19}{5} & 6 & \frac{71}{5} \end{pmatrix}$$

En función de las salidas presentadas se calcula la medida de robustez, aplicando la expresión (3.20). De esta forma, se obtiene que

$$R^{\delta,\Delta}(f, f^{\delta,\Delta}) = 15.347$$

Teniendo en cuenta los valores de las medidas de sensibilidad calculadas para el Ejemplo 3.4 y el valor de la medida de robustez obtenida en este ejemplo, no se puede concluir que esta Red Bayesiana Gaussiana sea robusta para las perturbaciones propuestas. Por tanto, es muy importante que se definan con la mayor precisión posible los parámetros que describen esta Red Bayesiana Gaussiana para que los resultados que se obtienen, tras introducir nueva evidencia, sean los adecuados.

### 3.7. Conclusiones y futuras líneas de investigación

En esta Memoria se han desarrollado dos métodos para estudiar la sensibilidad de una Red Bayesiana Gaussiana y una aproximación para evaluar la robustez de la misma.

Los análisis presentados se apoyan en el cálculo de la divergencia de Kullback-Leibler, para estudiar la diferencia existente entre dos distribuciones de interés. Se trabaja con esta medida por ser una medida de divergencia usual en términos de la Teoría de la Información cuando se quiere establecer una discrepancia estructural entre distribuciones.

La metodología seguida para estudiar la sensibilidad y la robustez de una Red Bayesiana Gaussiana concreta, consiste en comparar el modelo original, que describe la parte cuantitativa de la Red Bayesiana Gaussiana, con un modelo perturbado, que se obtiene tras añadir una o un conjunto de perturbaciones a los parámetros del modelo original. Ambos modelos se definen inicialmente, es decir, antes de propagar la evidencia por la red. Con la medida definida se comparan dichos modelos finales, esto es, una vez que se ha propagado la evidencia.

En función del análisis de sensibilidad, se introducen uno a uno los parámetros perturbados o en grupo. Las medidas de sensibilidad presentadas se calculan para cada parámetro o conjunto de parámetros inciertos. De esta forma, se obtienen distintas medidas para cada tipo de perturbación, que una vez comparadas permiten determinar que parámetro o conjunto de parámetros afecta más a la salida de interés de la red.

Con el valor de la medida calculada se puede concluir hasta que punto la Red Bayesiana Gaussiana es sensible o no a las perturbaciones propuestas.

En el análisis de sensibilidad de una vía desarrollado, se trabaja con una variable de interés  $X_i$  y una variable evidencial  $X_e$ , de forma que en cada paso del mismo se introduce una única perturbación asociada a un parámetro incierto de la red, y se calcula una medida de sensibilidad para cada perturbación.

Es importante señalar, que en este caso, sólo los parámetros que describen inicialmente dicha variable de interés y los que describen la variable evidencial, influyen sobre la salida de la red. Esto es debido a que se trabaja con la distribución conjunta de la red y no con las distribuciones condicionadas de cada

variable, dada la ocurrencia de los padres en el DAG.

Este resultado es interesante porque evidencia que se ha de precisar al máximo la descripción inicial de  $X_i$ , al igual que la descripción inicialmente de aquellas variables que puedan ser evidenciales. Para el resto de variables de la red, la inexactitud al definir los parámetros que las describen no afecta a la distribución final de interés de la red.

Para este análisis de una vía, se estudia además el comportamiento de la medida de sensibilidad descrita cuando la perturbación, que cuantifica la incertidumbre acerca de un parámetro, es extrema. Se observa como la medida de sensibilidad refleja el comportamiento extremo del parámetro, para todos los casos, salvo cuando es incierta la varianza evidencial. Para este caso, se obtiene como límite de la medida de sensibilidad un valor finito. Por tanto, cuando las incertidumbres acerca de los parámetros sean grandes, es conveniente estudiar dicho límite, para poder obtener algún resultado acerca del valor que toma la medida de sensibilidad cuando la varianza evidencial es incierta.

También, se estudian las medidas de sensibilidad cuando la variable de interés y la evidencial son independientes o dependientes linealmente. Dichas relaciones de dependencia se asocian de forma directa con la conexión entre las mismas en el DAG. Con este estudio, se simplifican mucho las medidas de sensibilidad, de forma que si son linealmente dependientes, con coeficiente de correlación lineal  $\rho_{ie}^2 = 1$ , la medida de sensibilidad vale infinito, para cualquier perturbación posible; y si son independientes, con  $\rho_{ie}^2 = 0$ , la salida de interés de la red, dada por la distribución final de la variable  $X_i$ , no se ve influida por la evidencia introducida. Ambos resultados, son los esperados.

El análisis de sensibilidad de  $n$  vías, se presenta como una generalización del análisis de una vía desarrollado. De esta forma, se trabaja con una Red Bayesiana Gaussiana, con un conjunto de variables evidenciales y un conjunto de variables de interés. Además, las perturbaciones se introducen de forma conjunta en el modelo perturbado, en función de si los parámetros inexactos, a los cuales se suman las perturbaciones, describen variables de interés o evidenciales y si se presentan en el vector de medias o en la matriz de covarianzas.

Para este análisis, se define la medida de sensibilidad para distribuciones normales multivariantes y se calcula para comparar la salida final de la red, del modelo original, con dicha salida, para el modelo perturbado. Para cada conjunto

de perturbaciones se tiene un modelo perturbado distinto, por tanto se calculan varias medidas de sensibilidad para distribuciones multivariantes en función del conjunto de parámetros inciertos. A pesar de que las expresiones asociadas a la medida de sensibilidad para distribuciones multivariantes son grandes y aparentemente complicadas, la aplicación de las mismas es sencilla y en la mayoría de los casos solo necesita de la especificación inicial de la red.

Tras realizarse este análisis de sensibilidad de  $n$  vías, se obtienen distintas medidas de sensibilidad y en función de su valor se determina él o los conjuntos de parámetros que más afectan a la salida final de interés.

La aproximación al estudio de la robustez de una Red Bayesiana Gaussiana surge cuando se obtienen valores pequeños asociados a las medidas de sensibilidad. Con el fin de estudiar la robustez de una red concreta, se define la medida de robustez que compara el modelo original con el modelo perturbado. En este caso, solo hay un modelo perturbado que viene dado por el original más todas las perturbaciones asociadas a los parámetros inciertos que describen la red.

Cabe señalar, que las conclusiones que se obtienen para una Red Bayesiana Gaussiana concreta, son acerca de la robustez de la red para las incertidumbre cuantificadas, es decir, si cambian las posibles perturbaciones, el comportamiento de la red también cambiará. No obstante, con el objetivo de establecer comparaciones, se pueden imponer perturbaciones dentro de unos límites y estudiar su efecto en distintas redes.

Existen muchos caminos abiertos para seguir estudiando la sensibilidad y la robustez en Redes Bayesianas con las ideas y herramientas introducidas en la Memoria.

Resultaría interesante determinar en Redes Bayesianas Gaussianas como afecta a la salida de la red la modificación de la estructura de dependencias presentada en el DAG. Así, si se elimina una arista del grafo o se introduce una nueva arista, los parámetros que describen la red cambian y cuantificar dicho cambio puede ayudar a conocer más profundamente la Red Bayesiana Gaussiana que describe un problema concreto. Además, se podría plantear la eliminación de una variable de la red, simplificándose así el modelo.

Otra línea de investigación planteada consiste en la aplicación de la metodología

propuesta cuando los parámetros perturbados corresponden a las distribuciones condicionadas de las variables por sus padres, que también determinan la Red Bayesiana Gaussiana.

Concretar los análisis de sensibilidad mostrados para Redes Bayesianas Discretas, es también una futura línea de trabajo.





# Bibliografía

- Andersen, H.H., Højbjerg, M., Sørensen, D. y Eriksen, P.S. (1995). *Linear and graphical models for the multivariate complex normal distribution*. Lecture Notes in Statistics 101. Springer-Verlag, New York.
- Andersen, S.K., Olesen, K.G., Jensen, F.V. y Jensen, F. (1989). HUGIN - A shell for building Bayesian belief universes for expert systems. En *Proceedings of the 11<sup>th</sup> International Joint Conference on Artificial Intelligence* (ed. N.S. Sridharan), San Mateo, California. Morgan Kaufmann.
- Anderson, T.W. (2003). *An introduction to multivariate statistical analysis* (3<sup>a</sup> edición). Wiley-Interscience.
- Aguelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G. y Ng, A. (2005). En *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol 2, Washington, DC, USA, pp. 169-176.
- Arnold, B., Castillo, E. y Sarabia, J.M. (1999). *Conditional Specification of Statistical models*. Springer-Verlag, New York.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal Statistics*, **12**, 171-178.
- Bednarski, M., Cholewa, W. y Frid, W. (2004). Identification of sensitivities in Bayesian networks. *Engineering Applications of Artificial Intelligence*, **17**, 327-335.
- Berge, C. (1973). *Graphs and Hypergraphs*. Amsterdam: North-Holland.

- Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, **36**, 192-236.
- Buntine, W.L. (1994) Operations for Learning with Graphical models. *Journal of Artificial Intelligence Research*, **2**, 159-225.
- Capitanio, A., Azzalini, A. y Stanghellini, E. (2003). Graphical models for skew-normal variates. *Scandinavian Journal of Statistics*, **30**, 129-144.
- Castillo, E., Gutiérrez, J.M. y Hadi, A.S. (1997a). *Expert Systems and Probabilistic Network Models*. Springer-Verlag, New York.
- Castillo, E., Gutiérrez, J.M., Hadi, A.S. y Solares, C. (1997b). Symbolic propagation and sensitivity analysis in Gaussian Bayesian networks with application to damage assessment. *Artificial Intelligence in Engineering*, **11**, 173-181.
- Castillo, E., Gutiérrez, J.M., y Hadi, A.S. (1997c). Sensitivity Analysis in Discrete Bayesian Networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, **27**, 412-423.
- Castillo, E., Kjærulff, U. y van der Gaag, L.C. (2001). Sensitivity Analysis in Gaussian Networks. En *Proceedings of Sensitivity Analysis in Model Output, SAMO*, Madrid, España.
- Castillo, E. y Kjærulff, U. (2003). Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. *Reliability Engineering and System Safety*, **79**, 139-148.
- Chan, B. y Shachter, R. (1992). Structural controllability and observability in influence diagrams. In Dubois, D., Wellman, M., D'Ambrosio, B. and Smets, P. (Eds.), (1992). En *Uncertain in Artificial Intelligence: Proceedings of the Eight Conference*, Stanford, California, pp. 25-32.
- Chan, H. y Darwiche, A. (2004). Sensitivity Analysis in Bayesian Networks: From Single to Multiple Parameters. En *Proceeding of the 20<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, Arlington, Virginia, USA, pp. 67-75. AUAI Press.
- Chan, H. y Darwiche, A. (2005). A distance Measure for Bounding Probabilistic Belief Change. *International Journal of Approximate Reasoning*, **38** (2), 149-174.

- Cooper, G. (1990). Computational complexity of probabilistic inference using Bayesian belief networks (Research note). *Artificial Intelligence*, **42**, 393-405.
- Coupé, V.M.H., van der Gaag, L.C. y Habbema, J.D.F. (2000). Sensitivity analysis: an aid for belief-network quantification. *The Knowledge Engineering Review*, **15**(3), 215-232.
- Coupé, V.M.H. y van der Gaag, L. (2002). Properties of Sensitivity Analysis of Bayesian Belief Networks. *Annals of Mathematics and Artificial Intelligence*, **36**, 323-356.
- Cowell, R. G., Dawid, A. P., Lauritzen, S. L. y Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*. Springer, Barcelona.
- Cowell, R. G. (2005). Local Propagation in Conditional Gaussian Bayesian Networks. *Journal of Machine Learning Research*, **6**, 1517-1550.
- Dagum, P. y Luby, M. (1993). Approximating probabilistic inference in bayesian belief networks is NP-hard. *Artificial Intelligence*, **60**, 141-153.
- Dagum, P., Galper, A., Horvitz, E. y Seiver, A. (1995). Uncertain reasoning and forecasting. *International Journal of Forecasting*, **11**(1), 73-87.
- D'Ambrosio, B. (1991). Local expression languages for probabilistic dependence. En *Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence*, Los Angeles, CA, pp. 95-102. Morgan Kaufmann.
- Dawid, A. P. (1979). Conditional independence in Statistical Theory. *Journal of the Royal Statistical Society, Series B*, **41**, 1-31.
- Dawid, A. P. (1992). Applications of a general propagation algorithm for probabilistic expert systems. *Statistics and Computing*, **2**, 25-36.
- Dean, T. y Wellman, M. (1991). *Planning and Control*. San Mateo, California: Morgan Kaufmann.
- Dempster, A.P. (1972). Covariance selection. *Biometrics*, **28**, 157-175.
- Dobra, A., Hans, C., Jones, B., Nevins, J.R., Guang, Y. y West, M. (2004). Sparse Graphical Models for Exploring Gene Expression Data. *Journal of multivariate analysis*, **90**(1), 196-212.

- Franklin, R. C. G., Spiegelhalter, D. J., Macartney, F., y Bull, K. (1991). Evaluation of an algorithm for neonates. *British Medical Journal*, **302**, 935-939.
- Frydenberg, M. (1990). The chain graph Markov property. *Scandinavian Journal of Statistics*, **17**, 333-53.
- Gilks, W., Clayton, D., Spiegelhalter, D., Best, N., McNeil, A., Sharples, L. y Kirby, A. (1993). Modelling complexitiy: applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society B*, **55**, 39-102.
- Gilks, W., Thomas, A. y Spiegelhalter, D. (1994). A language and a program for complex Bayesian modelling. *The Statistician* **43**, 169-178.
- Giudici, P. y Green, P.J. (1999). Decomposable graphical Gaussian model determination. *Biometrika*, **86**(4), 785-801.
- Golumbic, M. C. (1980). *Algorithmic Graph Theory and Perfect Graphs*. London: Academic Press.
- Gómez-Villegas, M.A., Maín, P. y Susi, R. (2006). Sensitivity analysis of extreme inaccuracies in Gaussian Bayesian Networks. En *Proceedings of the Third European Workshop on Probabilistic Graphical Models*, Prague, Czech Republic, pp. 139-146.
- Gómez-Villegas, M.A., Maín, P. y Susi, R. (2007). Sensitivity Analysis in Gaussian Bayesian Networks Using a Divergence Measure. *Communications in Statistics: Theory and Methods*, **36**(3) (En prensa).
- Heckerman, D. (1995). A Tutorial on Learning With Bayesian Networks. *Technical Report, Msr TR-95-06, Microsoft Research, Redmond, WA*. Véase <ftp://ftp.research.microsoft.com>.
- Jaakkola, T. y Jordan, M. (1996). Computing upper and lower bounds on likelihoods in intractable networks. En *Proceedings of Twelfth Conference on Uncertainty in Artificial Intelligence*, Portland, OR, pp. 340-348. Morgan Kaufmann.
- Jensen, F. V. (1988). *Junction Trees and Decomposable Hypergraphs*. JUDEX Research Report, Aalborg, Denmark.

- 
- Jensen, F.V., Lauritzen, S.L. y Olesen, K.G. (1990a). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, **4**, 269-282.
- Jensen, F.V., Olesen, K.G., and Andersen, S.K. (1990b). An Algebra of Bayesian Belief Universes for Knowledge-Based Systems. *Networks*, **20**, 637-660.
- Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*. Springer, Barcelona.
- Kim, J.H. y Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. En *Proceeding of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 190-193. Morgan Kaufmann, Los Altos, CA.
- Kjærulff, U. (1992). A computational scheme for reasoning in dynamic probabilistic networks. In Dubois, D., Wellman, M., D'Ambrosio, B. and Smets, P. (Eds.), (1992). En *Uncertain in Artificial Intelligence: Proceedings of the Eight Conference*, Standford, California, pp. 121-129.
- Kjærulff, U. y van der Gaag, L.C. (2000). Making Sensitivity Analysis Computationally Efficient. En *Proceedings of the 16<sup>th</sup> Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, pp. 315-325. Morgan Kaufmann.
- Kullback, S. y Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
- Laskey, K. B. (1995). Sensitivity Analysis for Probability Assessments in Bayesian Networks. *IEEE Transactions on Systems, Man and Cybernetics*, **25**, 901-909.
- Lauritzen, S. L. y Wermuth, N. (1984). Mixed interaction models. *Research Report R-84-8, Inst. Elec. Sys., Univ. Aalborg 1984*.
- Lauritzen, S. L. y Spiegelhalter, D. J. (1988). Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. *Journal of the Royal Statistical Society, Series B*, **50**(2), 157-224.
- Lauritzen, S. L. y Wermuth, N. (1989a). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, **17**, 31-57.

- Lauritzen, S. L. (1989b). Mixed graphical association models. *Scandinavian Journal of Statistics*, **16**, 273–306.
- Lauritzen, S. L., Dawid, A. P., Larsen, B. N. y Leimer, H. G. (1990). Independence Properties of Directed Markov Fields. *Networks*, **20**, 491–505.
- Lauritzen, S. L. (1992). Propagation of probabilities, means and variances in mixed graphical association models. *Journal of the American Statistical Association*, **87**, 1098–1108.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press.
- Lauritzen, S. L. y Jensen, F. (2001) Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, **11**, 191–203.
- Lauritzen, S. L. y Sheehan, N.A. (2003). Graphical models for Genetic Analyses. *Statistical Science*, **18**, 489–514.
- Normand, S.L. y Tritchler, D. (1992). Parameter Updating in Bayes Network. *Journal of the American Statistical Association*, **87**, 1109–1115.
- Olmsted, S. (1983). *On representing and solving decision problems*. PhD Thesis, Department of Engineering-Economic Systems, Stanford University.
- Onisko, A., Druzdzal, M.J. y Wasyluk, H. (1999). A Bayesian Network Model for Diagnosis of Liver Disorders. En *Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering*, Warsaw, Poland, pp. 842–846.
- Pearl, J. (1986a). A constraint-propagation approach to probabilistic reasoning. En *Proceedings of American Association for Artificial Intelligence National Conference on AI*, Pittsburgh, Pennsylvania, pp. 133–136.
- Pearl, J. (1986b). Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, **29**, 241–288.
- Pearl, J. y Paz, A. (1987). Graphoids: A graph based logic for reasoning about relevancy relations. *Advances in Artificial Intelligence - II*, (ed. B. D. Boulay, D. Hogg, and L. Steel). Amsterdam: North-Holland. 357–363.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. University Press, Cambridge, United Kingdom.
- Pérez, A., Larrañaga, P. y Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *International Journal of Approximate Reasoning*, **34**(1), 1-25.
- Rose, D. J., Tarjan, R. E. y Leuker, G. S. (1976). Algorithmic Aspects of Vertex Elimination on Graphs. *SIAM Journal of Computing*, **5**, 266-283.
- Saltelli, A. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. Hoboken, NJ: Wiley.
- Saul, L., Jaakkola, T. y Jordan, M. (1996). Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, **4**, 61-76.
- Shachter, R. (1988). Probabilistic inference and influence diagrams. *Operations Research*, **36**, 589-604.
- Shachter, R. y Kenley, C. (1989). Gaussian influence diagrams. *Management Science*, **35**, 527-550.
- Shafer, G. y Shenoy, P. (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence*, **2**, 327-352.
- Sham, P. (1998). *Statistics in Human Genetics*. Edward Arnold, London, United Kingdom.
- Spirtes, P., Glymour, C., y Scheines, R. (2000). *Causation, Prediction and Search*. (2ª edición). The MIT Press, Cambridge, United Kingdom.
- Studeny, M. y Bouckaert, R. R. (1998). On chain graph models for description of conditional independence structures. *Annals of Statistics*, **26**, 1434-1495.
- Stephenson, T.A., Magimai-Doss, M. y Boulard, H. (2002). Mixed Bayesian Networks with Auxiliary Variables for Automatic Speech Recognition. En *Proceedings of the 16th International Conference on Pattern Recognition*, vol 4, Washington, DC, USA.
- Susi, R. (2006). Sensitivity analysis algorithm. [www.ucm.es/info/eue/pagina/APOYO/RosarioSusiGarcia/S\\_algorithm.pdf](http://www.ucm.es/info/eue/pagina/APOYO/RosarioSusiGarcia/S_algorithm.pdf).



- Tarjan, R. E. y Yannakakis, M. (1984). Simple Linear-Time Algorithms to Test Chordality of Graphs, Test Acyclity of Hypergraphs and Selectively Reduce Acyclic Hypergraphs. *SIAM Journal of Computing*, **13**, 566-579.
- Verma, T. y Pearl, J. (1990). Causal Networks: Semantics and expressiveness. En *Uncertainty in Artificial Intelligence 4*, North Holland, Amsterdam. The Netherlands, pp. 69-76.
- Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *Journal of the American Statistical Association*, **75**, 963-972.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. John Wiley and Sons, Chichester.
- Yannakakis, M. (1981). Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic and Discrete Methods*, **2**, 77-9.
- Yule, G.U. (1907). On the Theory of Correlation for Any Number of Variables, treated by a New System of Notation. *Proc. Royal Soc. London, Serie A*, **79**, 182-193.