
**UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

ESCUELA DE MATEMÁTICA



Universidad de El Salvador

Hacia la libertad por la cultura

“INTRODUCCIÓN A LAS REDES BAYESIANAS.”

PRESENTADO POR:

ROBERTO VLADIMIR GÁMEZ MENÉNDEZ.

PARA OPTAR AL GRADO DE:

LICENCIADO EN ESTADÍSTICA

ASESOR:

Dr. JOSÉ NERY S FUNES TORRES

CIUDAD UNIVERSITARIA, MARZO DE 2007.

UNIVERSIDAD DE EL SALVADOR
UNIVERSIDAD DE EL SALVADOR.

RECTORA:

Dra. María Isabel Rodríguez

SECRETARIA GENERAL:

Licda. Alicia Margarita Rivas de Recinos

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DECANO:

MSc. José Héctor Elías Días

SECRETARIA:

Licda. Marta Noemy Martínez de Rosales

ESCUELA DE MATEMÁTICA

DIRECTOR:

Lic. Mauricio Hernán Lovo Córdoba

TRABAJO DE GRADUACIÓN APROBADO POR:

COORDINADOR:

Lic. Mauricio Hernán Lovo Córdoba

ASESOR:

Dr. José Nerys Funes Torres

DEDICATORIAS

En primer lugar le doy gracias a Dios, por lograr terminar la carrera, a mis padres por haberme brindado la oportunidad de estudiar. A mi asesor de tesis, por su paciencia y enseñanza durante el desarrollo de la tesis, a los compañeros de la SEM, que me brindaron todo su apoyo durante todo el transcurso de la carrera y durante el desarrollo de la tesis.

ÍNDICE.

INTRODUCCIÓN GENERAL.	II
------------------------------	-----------

CAPÍTULO 1. INTRODUCCIÓN A LAS REDES BAYESIANAS.

1.1 Introducción.	1
1.2 Conceptos básicos de grafos.	2
1.2.1 Propiedades de los grafos no dirigidos.	5
1.2.2 Propiedades de los grafos dirigidos.	8
1.3 Factorización de la probabilidad.	10
1.3.1 Reseña histórica.	10
1.3.2 Factorización de una función de probabilidad conjunta.	11
1.4 Redes bayesianas.	14
1.4.1 Criterios de separación gráfica.	14
1.4.2 Propiedades de Independencia condicional, modelos de dependencia y su representación.	19

CAPÍTULO 2. INFERENCIA EN REDES BAYESIANAS.

2.1 Introducción.	29
2.2 Propagación exacta en redes bayesianas.	29
2.2.1 Propagación en poliárboles.	29
2.2.2 Método de condicionamiento.	45
2.2.3 Método de agrupamiento.	66
2.2.4 Propagación Orientada a un Objetivo	75

CAPÍTULO 3. INTRODUCCIÓN A LA ESTIMACIÓN DE UNA RED BAYESIANA.

3.1 Introducción.	79
3.2 Estimación de los parámetros de una red bayesiana.	79
3.2.1 Estimando la probabilidad de una variable.	80
3.2.2 Función de densidad a priori.	83
3.2.3 Estimando los parámetros de una red bayesiana con más de una variable.	93
3.3 Aprendizaje de la estructura.	101
3.3.1 Criterio Scoring	103
3.3.2 Algoritmo k2	106
3.4 Aplicación (Enfermedad Coronaria).	108
3.4.1 Descripción del corazón.	109
3.4.2 Problemas con las arterias del corazón.	109
3.4.3 Variables en la base de datos.	113
3.4.3.1 Ordenamiento de las variables y estimación de la red bayesiana.	115

A. ANEXOS.

A.1 Estimación de la red bayesiana utilizando el software ELVIRA versión 0.162 _____ 120

A.2 Datos Utilizados para Estimar la Red Bayesiana de la enfermedad Coronaria. _____ 123

INTRODUCCIÓN GENERAL.

Las redes bayesianas son modelos gráficos basados en grafos dirigidos y distribuciones de probabilidades, estas forman parte de una disciplina conocida como aprendizaje automático¹. El mayor desarrollo que han alcanzado las redes bayesianas ha sido a partir de los años 80, aunque la primera representación gráfica de una distribución de probabilidad fue realizada en los años 20 por el genetista Sewal Wright.

La publicación más importante sobre redes bayesianas es el libro de Pearl, de 1988, *Probabilistic Reasoning in Intelligent Systems: Networks of plausible Inference*, en él se presentan los resultados obtenidos hasta ese momento en redes bayesianas y se discute además, las relaciones que éstas tienen con las redes de Markov, los sistemas expertos, el tratamiento de la incertidumbre, etc. En este mismo año se publica el algoritmo de Lauritzen y Spiegelhalter ambos especialistas en probabilidad y estadística, desarrollaron un método conocido como árboles de cliques (método de agrupamiento), el cual además de ser un método eficiente para calcular la probabilidad, fue el primer algoritmo que permitía tratar redes con bucles; en 1990 Jensen Olesen y Andersen desarrollaron el método de los universos de conocimiento que es una modificación del algoritmo de agrupamiento (árbol de cliques). El método de agrupamiento es hoy en día uno de los algoritmos estándar para el cálculo de probabilidades en redes bayesianas.

En la presente tesis se da una pequeña introducción a las redes bayesianas, en el primer capítulo se describe los conocimientos básicos de la teoría de grafos que se utilizan al final del capítulo, y en capítulos posteriores de esta tesis, en este mismo capítulo se describe como se factoriza una distribución de probabilidad y al final de este capítulo se presenta una definición de lo que es una red bayesiana, el segundo capítulo de la tesis se presentan cuatro métodos para el cálculo de las probabilidades a posteriori de las variables en una red bayesiana, cuando conocemos con seguridad que valores toman un subconjunto de ellas, y en el tercer capítulo se presenta como se estiman las probabilidades de los nodos de una red bayesiana, para lo cual se utiliza un poco de estadística bayesiana y se presenta un algoritmo que nos permite estimar la estructura de una red bayesiana y por último se presenta estima una red bayesiana de una base de datos para lo cual se utiliza el software ELVIRA versión 0.162.

¹ Cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender.

CAPÍTULO 1.
INTRODUCCIÓN A LAS REDES
BAYESIANAS.

1. Introducción a las redes bayesianas.

1.1 Introducción.

Las redes bayesianas son modelos probabilistas inspirados en las causalidad, en cuya representación gráfica, cada nodo del grafo es una variable y las aristas representan relaciones de causalidad entre las variables, estas aparecen a principios de los años 80, debido a la debilidad del método probabilista y otros métodos que se crearon para medir la incertidumbre sin hacer uso de la probabilidad en los sistemas expertos. El cálculo de probabilidades de las variables en una red bayesiana cuando se conoce con certeza que un conjunto de variables de la red han tomado un valor concreto, se le conoce como propagación de probabilidad, los primeros algoritmos para propagar probabilidades en una red bayesiana, sólo lo hacían con redes cuya representación gráfica era un árbol o poliárbol, el primer método que permitía propagar probabilidades en redes que con bucles, apareció en 1988 descubierto por Lauritzen y Spiegelhalter , el cual titularon árbol de cliques o método de agrupamiento como se le conoce en la actualidad.

Algunas aplicaciones de redes Bayesianas son:

1. ANDES un programa informático desarrollado en la Universidad de Pittsburg, el cual es un sistema inteligente que hace deducciones de cómo un estudiante de Física resuelve un determinado problema.
2. En biología Friedman desarrolló un técnica basada en redes bayesianas, para el aprendizaje de relaciones causales entre los genes.
3. Se utilizan en negocios y finanzas por ejemplo en la valoración del riesgo en la solicitud de un préstamo.
4. Mani et.al. [1997] desarrollo Mentor, que es un sistema que predice retardo mental en niños.

En general las redes bayesianas se aplican en Medicina, Biología, Negocios y finanzas, Visión por Computadora, Hardware de computadoras, software de computadoras, minería de datos, procesamiento del lenguaje natural, Psicología, reconocimiento del idioma, control de vehículos y diagnostico de funcionamiento defectuoso., etc.

El presente capítulo consiste de tres partes en la primera de ellas se expone la teoría de grafos que utilizaremos en este capítulo y en los capítulos siguientes, en ella se expone la teoría que se utilizará tanto de grafos dirigidos como no dirigidos. La segunda parte consisten en describir como se factoriza una función de probabilidad, es decir dado un conjunto de variables, describir como se puede escribir su función de probabilidad conjunta como el producto de funciones de probabilidad definidas en menos variables. Y por último, se verá como se relacionan los grafos con la teoría de la probabilidad y las condiciones que deben cumplir los grafos, para que estos sean junto con la función de probabilidad una red bayesiana.

1.2 Conceptos básicos de grafos.

Los grafos aparecen por primera vez en el siglo XVIII cuando Leonhard Euler los creó para resolver el problema de los puentes de Königsberg y desde entonces se ha desarrollado una sólida teoría matemática sobre ellos. Los grafos son una de las herramientas más utilizadas en la actualidad, por lo que entre las representaciones más comunes que se pueden encontrar de los grafos son los organigramas que muestran la estructura formal de una empresa y los árboles genealógicos. Por lo regular se utilizan para resolver problemas de eficiencia de transporte, también se utilizan en sociología, electrónica y electricidad, en general se utilizan en aquellos campos en los que la conectividad es importante.

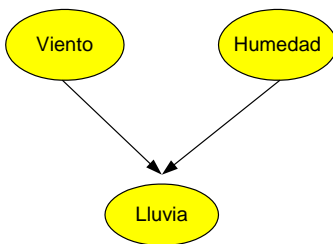


Figura 1.1

Los grafos consisten básicamente en un conjunto de objetos que se encuentran relacionados unos con otros los cuales son representados por nodos y las relaciones entre ellos por aristas, un ejemplo de grafo es el que se muestra en la figura 1.1, los nombres de las variables viento, humedad y lluvia están representadas en los nodos del grafo, mientras que las aristas que unen los nodos nos indican las relaciones que hay entre ellas.

Definición 1.2.1

Un grafo es un par de conjuntos $G = \{X, L\}$, donde $X = \{X_1, X_2, \dots, X_n\}$ es un conjunto finito de elementos (nodos), y L es un conjunto de aristas, es decir, un subconjunto de pares ordenados de elementos distintos de X .

En la definición 1.2.1 no se considera que dos nodos estén unidos por más de una arista, ni tampoco que una arista una un nodo consigo mismo, esto es debido a que en nuestro caso esto no tiene sentido ya que

utilizaremos los grafos para representar las variables de una función de probabilidad y las relaciones de independencias que hay entre ellas; por ejemplo si en la figura 1.1 se tuviera en la variable viento una arista que una ese nodo consigo mismo, se podría interpretar como que la variable viento es dependiente con ella misma, lo cual carece de sentido, ya que no aporta ninguna información sobre esa variable.

En el ejemplo de la figura 1.1 $X = \{\text{Viento, Lluvia, Humedad}\}$ y $G = \{(\text{Viento, Lluvia}), (\text{Lluvia, Humedad})\}$ observemos que no existe ninguna arista entre viento y humedad por lo que se puede decir que no están directamente relacionadas.

La notación que usaremos para representar las aristas de un grafo es $L = \{L_{ij} / X_i \text{ y } X_j \text{ estén conectados}\}$, donde L_{ij} representa la arista que une los nodos X_i y X_j .

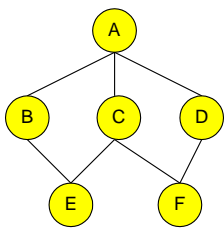


Figura 1.2

Las aristas de un grafo pueden ser dirigidas o no dirigidas, es decir, se puede tomar en cuenta o no el orden de cómo aparecen representados los nodos en el grafo, por ejemplo el grafo de la figura 1.2 se tiene el siguiente conjunto de aristas $L = \{L_{A,B}, L_{A,C}, L_{A,D}, L_{B,E}, L_{C,E}, L_{C,F}, L_{D,F}\}$, notemos que en este caso también hubiéramos podido representar el conjunto de aristas de la figura 1.2 por medio del conjunto $L = \{L_{B,A}, L_{C,A}, L_{D,A}, L_{E,B}, L_{E,C}, L_{F,C}, L_{F,D}\}$, que no es más que el

conjunto anterior con el orden de las aristas invertido, por lo que en esta representación gráfica, el orden de las aristas no es importante, sin embargo si el orden de las aristas es importante, es decir, si no es lo mismo escribir $L_{i,j}$ que $L_{j,i}$ entonces se dice que la arista es *dirigida*, y se denota por $X_i \rightarrow X_j$, lo que nos dice que la arista va del nodo X_i al nodo X_j y no en sentido inverso.

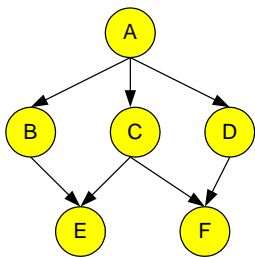


Figura 1.3

Por ejemplo en el grafo de la figura 1.3 el conjunto de aristas se puede escribir como $L = \{L_{A,B}, L_{A,C}, L_{A,D}, L_{B,E}, L_{C,E}, L_{C,F}, L_{D,F}\}$, en este caso no se puede escribir en orden inverso ya que las aristas son dirigidas. En general una arista se dice no dirigida si tanto $L_{i,j}$, como $L_{j,i}$ pertenecen a L , en caso de que sólo $L_{i,j}$ pertenece a L y $L_{j,i}$ no, se dice que la arista es dirigida. Las aristas no dirigidas se

denotan por $X_i - X_j$ o $X_j - X_i$.

Debido a esto, dependiendo de cómo sean las aristas de un grafo este será dirigido o no dirigido, por lo que se tiene la siguiente definición.

Definición 1.2.2

Un grafo en el cual todas sus aristas son dirigidas se denomina grafo dirigido y un grafo en el cual todas sus aristas son no dirigidas se denomina un grafo no dirigido.

A los grafos dirigidos también se les conoce como *dígrafos*, un ejemplo de grafo dirigido es el de la figura 1.3 en el cual todas sus aristas son dirigidas por lo que es un dígrafo y el grafo de la figura 1.2 todas sus aristas no son dirigidas por lo que es un grafo no dirigido.

En el grafo presentado en la figura 1.2, se nota que se puede ir de un nodo a otro, por ejemplo se puede ir del nodo A al nodo F, pasando por C o por D y viceversa; se puede ir del nodo F al nodo A pasando por los nodos antes mencionados, por lo que se dice que existen dos caminos entre el nodo A y el nodo F o dos caminos entre el nodo F y A.

Definición 1.2.3

Un camino entre dos nodos X_i y X_j , es una sucesión de nodos $\{X_i, \dots, X_r\}$ donde $X_i = X_i$ y $X_r = X_j$, de modo que existe una arista entre los nodos intermedios de X_i y X_j . La longitud del camino es $r-1$.

Como se mencionó anteriormente existen dos caminos entre los nodos A y F, en la figura 1.2 y que estos se puede recorrer de forma inversa, es decir, yendo del nodo F al A, mientras que en la figura 1.3 que es el mismo grafo que el de la figura 1.2, sólo que aquí las aristas son dirigidas, se tiene que existen dos caminos del nodo A al F, pero no del F al A debido a que esto sería ir en sentido contrario al de las aristas.

En la definición 1.2.3 se puede considerar que el nodo inicial en un camino coincida con el nodo terminal, por lo que en este caso se dice que se tiene *un camino cerrado*, por ejemplo en la figura 1.2 se puede tener un camino que va del nodo A al nodo A, pasando por los nodos B, E, C hasta llegar nuevamente al mismo nodo y se representa por $A-B-E-C-A$, note que en el grafo dirigido de la figura 1.3 no hay caminos cerrados. A los caminos cerrados en un grafo no dirigido se les conoce como *bucles*, y si se trata de un dígrafo al camino cerrado se le conoce como *ciclo*, por lo que un grafo dirigido se dice que es *acíclico* si no contiene ciclos, es decir si no tiene caminos cerrados.

Como podemos observar en los grafos de las figuras 1.1, 1.2, 1.3, siempre existe un nodo conectado directamente con otro nodo, por lo que si un nodo X_i está conectado directamente con el nodo X_j , entonces se dice que el nodo X_j es *adyacente* al nodo X_i o viceversa, por lo que el conjunto de nodos adyacentes a un nodo son aquellos que son directamente alcanzables desde el nodo X_i , para el caso de los dígrafos, un nodo X_j es adyacente al nodo X_i si existe la arista L_{ij} , por ejemplo en la figura 1.2 los nodos adyacentes al nodo C son {A, E, F}, mientras que en la figura 1.3 los nodos adyacente al nodo C son {E, F}. Al conjunto de nodos adyacentes de un nodo X_i , se les conoce como *vecinos del nodo X_i* .

Definición 1.2.4.

La unión de los conjuntos de vecinos de los nodos de un conjunto S, excluyendo los nodos de S, se denomina la frontera de S, la cual se denota por $frn(S)$.

Por ejemplo si consideramos el grafo de la figura 1.2 y el conjunto $S = \{C, E\}$ se tiene que los vecinos del nodo C son {A, E, F} y los vecinos del nodo E son {B, C}, por lo que la unión de los conjuntos de vecinos de los nodos C y E, excluyendo estos nodos forman el conjunto {A, B, F}, que es la frontera del conjunto S.

1.2.1 Propiedades de los grafos no dirigidos.

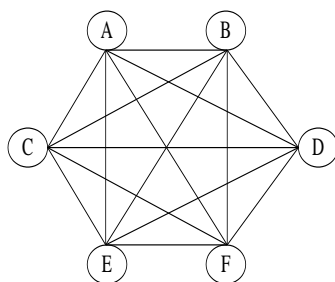


figura 1.4

Entre los grafos no dirigidos se pueden encontrar grafos en los cuales haya una arista entre cada par de nodos a este tipo de grafo se les conoce como *grafos completos* y se denotan por K_n , por ejemplo el grafo de la figura 1.4 es un grafo completo K_6 , observemos que cada nodo está conectado con todos los demás nodos del grafo. De esta definición de grafo completo se tiene que en un grafo que no necesariamente sea completo, existan

subconjuntos de nodos S que si sea completo, es decir que para todo subconjunto S de un grafo G, existe una arista en G para cada par de nodos de S. Por ejemplo cada par de nodos adyacentes forman un conjunto completo en un grafo, en la figura 1.4 cualquier subconjunto de nodos del grafo forma un subconjunto completo de nodos.

Los subconjuntos de nodos completos en un grafo son importantes en las redes probabilísticas cuando estos no son subconjuntos propios de otro subconjunto completo.

Definición 1.2.5

Un subconjunto completo de nodos C se denomina un conglomerado si no es subconjunto propio de otro conjunto completo, es decir, si es maximal.

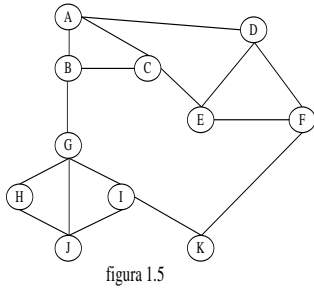


figura 1.5

Por ejemplo algunos de los conglomerados de la figura 1.5 son $C_1 = \{A, B, C\}$, $C_2 = \{D, E, F\}$, $C_3 = \{G, H, J\}$ y $C_4 = \{G, I, J\}$, como se observa en la figura cada uno de estos conglomerados forman un conjunto completo maximal.

Los grafos en los cuales existe un camino entre cada par de nodos se denominan *grafos conexos*, esto nos indica que no existen en el grafo nodos desconectados, por lo que si no hay caminos ente dos nodos se denomina *inconexo*, por ejemplo si en la figura 1.5 se eliminarán las aristas $L_{B,G}$ y $L_{F,K}$ el grafo se dividiría en dos subgrafos totalmente inconexos, si un grafo está dividido en dos grafos, es decir se trata de un grafo dividido en dos grafos, a estos grafos se les denomina *componentes conexas del grafo*.

Si en un grafo conexo existe a lo sumo un camino entre cada par de nodos entonces el grafo se denomina un *árbol*, en caso de que exista más de un camino entre cada par de nodos el grafo se denomina *múltiplemente conexo*.

Los nodos de un grafo se pueden numerar, siempre y cuando el número que se le asigna a un nodo es mayor que el número que se les asigna a sus predecesores y la *numeración de los nodos de un grafo se denomina perfecta* si para todo nodo X_i del grafo, el conjunto frontera del nodo interceptado con los demás nodos, cuya numeración sea inferior a la del nodo X_i es un conjunto completo.

Un tipo de grafo que se utiliza para propagar probabilidades en redes bayesianas y poder escribir de forma sencilla la distribución de probabilidad de una red Markoviana, son los denominados grafos triangulados, antes de definir que es un grafo triangulado es necesaria la siguiente definición.

Definición 1.2.6

Una cuerda es una arista que une dos nodos de un bucle y que no pertenecen al bucle.

Por ejemplo si en la figura 1.5 uniéramos los nodos E y K , esta arista sería una cuerda y no pertenece al bucle formado por $B-C-E-F-K-I-G-B$, notemos que los bucles de longitud tres son los únicos que no se les puede trazar una cuerda, por definición son subconjuntos completos.

Definición 1.2.7

Un grafo no dirigido es triangulado o cordal, si cada bucle de longitud mayor o igual a cuatro posee por lo menos una cuerda

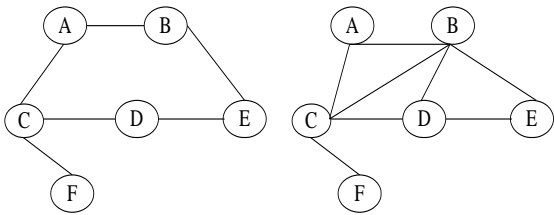


Figura 1.6

El grafo de la figura 1.6 (a) posee el bucle $A-B-E-D-C-A$ de longitud 5, como en este bucle no hay ninguna cuerda, se tiene que el grafo no es triangulado, mientras que en el grafo de la figura 1.6 (b) es un grafo triangulado que posee las cuerdas $L_{A,D}$ y

$L_{B,D}$, que no es más que el grafo de la parte (a) de la figura 1.6 triangulado, por lo que si se tiene un grafo no triangulado, es posible triangularlo añadiendo el mínimo número posible de aristas por debajo del cual el grafo no sería triangulado. El triangular un grafo no significa dividirlo en triángulos, por ejemplo en la figura 1.6 (b) no se ha añadido una arista que una los nodos F y D , que le daría la apariencia al grafo de estar dividido en triángulos, puesto que no forman ningún bucle de longitud mayor o igual a 4 con otros nodos. Una propiedad importante que cumplen los grafos no dirigidos triangulados y que se utilizará más adelante en el método de agrupamiento para la propagación de probabilidades, es la *propiedad de intercepción dinámica*, la cual se basa en que los nodos del grafo tienen una numeración perfecta, para el caso de los grafos triangulados estos siempre admiten una numeración perfecta.

Definición 1.2.8

Una numeración de los conglomerados de un grafo no dirigido $\{C_1, \dots, C_m\}$, se dice que satisface la propiedad de intercepción dinámica, si el conjunto $C_i \cap \{C_1 \cup \dots \cup C_{i-1}\}$ está contenido en al menos uno de los conglomerados $\{C_1, \dots, C_{i-1}\}$, para todo $i = 1, \dots, m$.

Una sucesión de conglomerados que satisface la propiedad de intercepción dinámica se dice que es una *cadena de conglomerados*, hay un teorema que establece que un grafo no dirigido tiene asociada una cadena de conglomerados si y sólo si el grafo es triangulado, por lo que si no se dispone de un grafo

triangulado, se puede triangular y ordenar los conglomerados de modo que cumpla la propiedad de intercepción dinámica, esto es importante ya que en la práctica las redes bayesianas con bucles no están trianguladas.

Un grafo de conglomerados de un grafo G es un grafo G' en el que los nodos del grafo son los conglomerados de G , y un grafo de conglomerados asociado a un grafo G se denomina un grafo de unión, si contiene todas la aristas posibles que unan conglomerados con algún nodo en común y un grafo de conglomerados se denomina un *árbol de unión* si es un árbol y que todo nodo que pertenezca a dos conglomerados, pertenezca también a todos los conglomerados contenidos en el camino que los une.

1.2.2 Propiedades de los grafos dirigidos.

Definición 1.2.9

Cuando una arista sale de un nodo X_i hacia el nodo X_j , se dice que el nodo X_i es padre del nodo X_j o que X_j es hijo de X_i .

La familia de un nodo está formada por un nodo y sus padres, el conjunto de padres de un nodo X_i , se representara por π_i . Un nodo X_i se dice que es un *ascendiente* de un nodo X_j , si existe un camino de X_i a X_j , y se dice que un nodo X_j es un *descendiente* del nodo X_i , si existe el camino de X_i a X_j , y un conjunto de nodos S se dice que es un *conjunto ancestral* del nodo X_i , si contiene todos sus ascendientes.

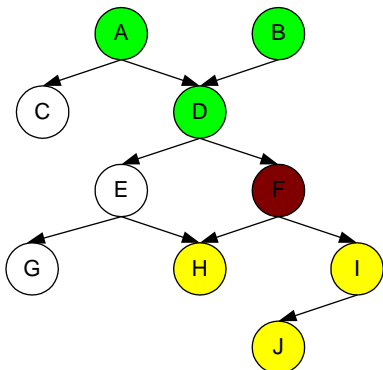


Figura 1.7

En la figura 1.7 se observa que los padres del nodo H son los nodos E y F, los ascendientes del nodo F son los nodos A, B y D, mientras que los descendientes del nodo F son H, I y J, por lo que el conjunto $\{A, B, D\}$ es el conjunto ancestral del nodo F, la familias de nodos que se pueden observar en el grafo son: $\{A, B, D\}$, $\{A, C\}$, $\{E, D, F\}$, $\{G, E\}$, $\{H, E, F\}$, $\{I, F\}$ y $\{J, I\}$.

Cuando un nodo no tiene padres se dice que es un nodo raíz, mientras que los nodos que no tienen hijos se dice que son nodos hojas, por

ejemplo en la figura 1.7, los nodos raíz son A y B, mientras que los nodos hojas son C, G, H, J.

Los grafos dirigidos pueden tener asociado un grafo no dirigido sin más que quitar la direccionalidad de las aristas en el grafo dirigido, por ejemplo el grafo dirigido de la figura 1.3 tiene asociado un grafo no dirigido,

el cual se forma cambiando las aristas dirigidas por aristas no dirigidas, este se muestra en la figura 1.2, el obtener un grafo dirigido a partir de un grafo no dirigido el problema es más complicado debido a la direccionalidad de las aristas.

Definición 1.2.10.

El grafo no dirigido asociado a un grafo dirigido que se obtiene uniendo cada par de nodos que poseen un hijo en común en el grafo no dirigido se llama grafo moral asociado al grafo dirigido.

Por ejemplo supongamos que al grafo de la figura 1.7, construimos el grafo no dirigido asociado y además unimos los nodos A y B, E y F, con una arista no dirigida, entonces el grafo no dirigido que se obtiene se llama grafo moral asociado al dígrafo de la figura 1.7.

La mayor parte de las propiedades vistas anteriormente para grafos no dirigidos se cumplen también para grafos dirigidos, por lo que se resumen en las siguientes definiciones.

Definición 1.2.11

Un grafo dirigido es conexo, si el grafo no dirigido asociado es conexo.

Definición 1.2.12

Un grafo conexo dirigido se denomina un árbol, si el grafo no dirigido asociado es un árbol, en caso contrario es un grafo múltiplemente conexo.

Definición 1.2.13

Un grafo dirigido se denomina un árbol simple, si cada nodo tiene a lo sumo un padre; en caso contrario se denomina poliárbol.

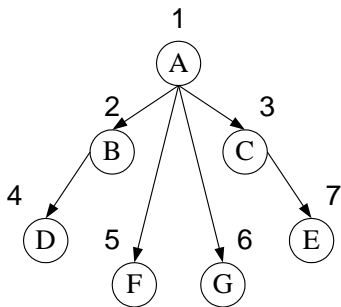


Figura 1.8

Definición 1.2.14

Una numeración de los nodos se denomina ancestral, si el número asignado a un nodo es menor que el número asignado a sus hijos.

Por ejemplo la numeración de los nodos del grafo de la figura 1.8 es una numeración ancestral, en él se muestra que el número que se le asigna a un

nodo es menor que el número asignado a su hijo.

Como podemos observar en la figura 1.8, el grafo cumple con las definiciones 1.2.11 a la 1.2.14, ya que el grafo no dirigido asociado cumple con lo que se describe en estas definiciones.

1.3 Factorización de la probabilidad.

1.3.1 Reseña histórica.

La probabilidad se empezó a desarrollar en el año de 1654, cuando el caballero de Mere quien era una persona apasionada por el juego le propuso un problema de juegos de dados a Blas Pascal (1623-1622), este caballero creía que había encontrado un error en los números al analizar el juego de los dados, observando que el comportamiento de los dados era diferente cuando se utilizaba un dado que cuando se utilizaban dos. Pascal compartía los problemas que le proponía el caballero de Mere con Pierre Fermat por medio de cartas, siendo esto el punto de partida para el estudio de la probabilidad. El primer libro de probabilidad apareció en el año de 1563 llamado "Liber de Lulo de Alae", libro sobre los juegos de azar, escrito y publicado por el italiano Girolamo Cardano (1501-1576). Desde entonces se ha venido desarrollando la teoría de la probabilidad hasta nuestros días, algunos campos en los que se puede aplicar la probabilidad son: Física, Genética, Medicina, Veterinaria, Zootecnia, Cálculo actuarial, Ciencias sociales, Astronomía, etc.

Antes de hablar de factorización de una función de probabilidad se dará la definición de independencia condicional, que es una de las definiciones más usadas en las redes bayesianas.

Definición 1.3.1.

Sea $X = \{ X_1, X_2, \dots, X_n \}$ un conjunto finito de variables y sea $P(\bullet)$ la distribución conjunta de las variables X . Sean V, Y, Z tres subconjuntos cualesquiera de las variables X . Los subconjuntos V y Y se dicen condicionalmente independientes dado Z si se cumple:

$$P(v / y, z) = p(v / z) \text{ Siempre que } p(y, z) > 0 \quad (\text{I})$$

En otras palabras la expresión anterior nos dice que el valor de Y , no proporciona ninguna información adicional acerca de V , cuando se conoce Z .

La expresión (I) se cumplirá si y sólo si, se cumple:

$$P(V, Y / Z) = P(V / Z)P(Y / Z) \quad (\text{II})$$

Esto se verifica como sigue:

- Supongamos que se cumple II, por lo que se tiene:

$$P(V / Y, Z) = \frac{P(V, Y, Z)}{P(Y, Z)} = \frac{\frac{P(V, Y, Z)}{P(Z)}}{\frac{P(Y, Z)}{P(Z)}} = \frac{P(V, Y / Z)}{P(Y / Z)}, \text{ por lo que utilizando la expresión (II) se tiene:}$$

$$P(V / Y, Z) = \frac{P(V / Z)P(Y / Z)}{P(Y / Z)} = P(V / Z) \text{ Por lo que de acuerdo a esta expresión se verifica (I).}$$

- Supongamos ahora que se cumple (I) la verificación de (II) es como sigue:

$$P(V, Y / Z) = \frac{P(V, Y, Z)}{P(Z)} = \frac{P(V, Y, Z)P(Y, Z)}{P(Z)P(Y, Z)} = P(V / Y, Z)P(Y / Z).$$

Aplicando (I) a la expresión anterior se llega al resultado que se quiere verificar.

La notación que usaremos para representar que los conjuntos X y Y son condicionalmente independientes dado un conjunto Z es: $I(X, Y / Z) \equiv P(X / Y, Z) = P(X / Z)$, y para representar que son condicionalmente dependientes se usará la notación $D(X, Y / Z) \equiv P(X / Y, Z)$. De acuerdo a esta notación para representar que dos variables son incondicionalmente independientes (dependientes), el conjunto Z será el conjunto vacío².

1.3.2 Factorización de una función de probabilidad conjunta.

La Factorización de una función de probabilidad significa, escribir la función de probabilidad conjunta de las variables, como producto de funciones de probabilidad definidas en menos variables, por ejemplo supongamos que tenemos la siguiente función de probabilidad $P(X_1, X_2, X_3, X_4)$, esta función la podemos escribir como el producto de las funciones $P(X_1, X_2)$ y $p(X_3, X_4 / X_1, X_2)$, como se puede

² $I(X, Y / \phi) \equiv p(x)p(y)$, $D(X, Y / \phi) \equiv p(x, y)$

observar el número de probabilidades a estimar no disminuye a menos que supongamos $I(X_3, X_4 / X_1, X_2)$.

La siguiente definición conocida como regla de la cadena, nos indica como se debe factorizar una función de probabilidad cualquiera, para ello consideraremos que las variables $\{Y_1, Y_2, \dots, Y_m\}$, representa cada una, un subconjunto de las variables $X = \{X_1, X_2, \dots, X_n\}$, tal que la unión de ellas nos da X .

Definición 1.3.2 Regla de la cadena.

Cualquier función de probabilidad de un conjunto de variables $\{X_1, X_2, \dots, X_n\}$, puede ser expresada como el producto de m funciones de probabilidad condicionada de la forma:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^m P(Y_i = y_i / B_i = b_i)$$

O de modo equivalente

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^m p(Y_i = y_i / A_i = a_i)$$

Donde:

- $B_i = \{Y_1, \dots, Y_{i-1}\}$, es el conjunto de variables anteriores a Y_i
- $A_i = \{Y_{i+1}, \dots, Y_m\}$, es el conjunto de variables posteriores a Y_i

Ahora supongamos que el conjunto $X = \{X_1, X_2, X_3, X_4, X_5\}$ con la siguiente partición $Y_1 = \{X_1\}$, $Y_2 = \{X_2, X_3\}$, $Y_3 = \{X_4, X_5\}$, entonces las dos formas de factorizar la función de probabilidad conjunta de X según la regla de la cadena son:

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2, X_3 / X_1)P(X_4, X_5 / X_1, X_2, X_3) \tag{1}$$

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1 / X_2, X_3, X_4, X_5)P(X_2, X_3 / X_4, X_5)P(X_4, X_5) \tag{2}$$

Si introducimos ciertas relaciones de independencia entre las variables en las funciones anteriores se disminuirá el número de probabilidades a estimar, por ejemplo si suponemos que $I(X_4, X_5 / X_1, X_2, X_3)$, la factorización de la función de probabilidad anterior se podrá escribir como:

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2, X_3 / X_1)P(X_4 / X_1, X_2, X_3)$$

Notemos que sólo la factorización (1) de la función conjunta de probabilidad está escrita en forma conveniente para introducir las relaciones de independencia de las variables, por lo que si se dispone de un conjunto de relaciones de independencia entre las variables, la factorización de la función de probabilidad debe de escribirse de forma conveniente a fin de que no resulte dificultoso el introducir dichas relaciones.

Si en la regla de la cadena suponemos que cada subconjunto $Y_i = \{Y_1, Y_2, \dots, Y_m\}$, tiene un elemento, entonces la factorización que se obtiene aplicando la regla de cadena se le conoce como *regla canónica de la cadena* y a cada una de las funciones de probabilidad condicional que intervienen en el producto, se les conoce como *función de probabilidad condicional canónica*.

Por ejemplo en la siguiente factorización:

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1 / X_2, X_3, X_5)P(X_2 / X_4, X_5)P(X_4, X_5). \quad (3)$$

Se tiene que cada una de las funciones de probabilidad condicionada del lado derecho están escritas en forma canónica, la factorización (3) no es más que la factorización (2) con las relaciones de independencia $I(X_1, X_4 / X_2, X_3, X_5)$ y $I(X_2, X_3 / X_4, X_5)$.

En el caso de que se tenga funciones de probabilidad condicional, es posible escribirlas en forma canónica, si se considera un conjunto de variables $X = \{X_1, X_2, \dots, X_n\}$ y suponga las funciones de probabilidad marginales y condicionales $P = \{P(U_1 / V_1), \dots, P(U_m / V_m)\}$ donde U_i y V_i son subconjuntos disjuntos de X , tal que $U_i \neq \emptyset$ y V_i puede ser vacío. Entonces a partir de P puede obtenerse un conjunto equivalente en el que los nuevos conjuntos U_i contienen una única variable de X .

Por ejemplo consideremos la siguiente función de probabilidad condicionada $P(X_1, X_2, X_3 / X_4, X_5)$ en este caso se tiene que $U_1 = \{X_1, X_2, X_3\}$ y $V_1 = \{X_4, X_5\}$, para convertir la función de probabilidad en canónica, lo que se hace es formar el conjunto $S = U_1 \cup V_1$ y aplicar la regla de la cadena a este conjunto, en el ejemplo $S = \{X_1, X_2, X_3, X_4, X_5\}$, por lo que la regla de la cadena nos da:

$$P(X_1, X_2, X_3 / X_4, X_5) = P(X_1 / X_1, X_2, X_3, X_4)P(X_2 / X_3, X_4, X_5)P(X_3 / X_4, X_5) \quad (4)$$

Definición 1.3.3.

Sea $\{Y_1, Y_2, \dots, Y_n\}$ una permutación del conjunto $X = \{X_1, X_2, \dots, X_n\}$. Una función de probabilidad condicionada $P(Y_i / S_i)$ se dice que es una función de probabilidad condicionada en forma canónica estándar, si Y_i está formado por una única variable y S_i contiene todas las variables anteriores a Y_i , o todas las variables posteriores a Y_i .

Por ejemplo, si tenemos la función de probabilidad conjunta de 4 variables, ésta se puede escribir como productos de funciones de probabilidad condicionada canónica estándar como se muestra a continuación:

$$P(X_1, X_2, X_3, X_4) = P(X_1 / X_2, X_3, X_4)P(X_2 / X_3, X_4)P(X_3 / X_4)P(X_4) \quad (5)$$

Donde:

$$Y_i = \{X_i\} \quad i = 1, 2, 3, 4$$

$$S_1 = \{Y_2, Y_3, Y_4\}$$

$$S_2 = \{Y_3, Y_4\}$$

$$S_3 = \{Y_4\}$$

$$S_4 = \{\phi\}$$

Cada una de las funciones de distribución de probabilidad condicionada incluyendo la marginal, son funciones de probabilidad condicionada en forma canónica estándar.

A la representación de la función de probabilidad conjunta de las variables dada en (5) se le conoce como representación canónica estándar. Por lo que cualquier función de probabilidad se puede representar en forma canónica estándar, siempre y cuando cada una de las funciones que intervienen en la factorización de la distribución conjunta de las variables, es una función de probabilidad condicionada en forma canónica estándar.

1.4 Redes bayesianas.**1.4.1 Criterios de separación gráfica.**

Cuando tenemos la distribución de probabilidad conjunta de un conjunto de variables, y se quiere conocer si dos variables son independientes lo que se hace es calcular la probabilidad conjunta de las dos variables

y las probabilidades marginales de cada variable, para luego realizar el producto de las probabilidades marginales y comparar el resultado con la probabilidad conjunta de las dos variables, para determinar si las dos variables son independientes o no. Un procedimiento similar se sigue si se quiere determinar si las variables son condicionalmente dependientes. Es obvio que el hacer estos cálculos numéricos para un gran número de variables resulta demasiado trabajo, una forma alternativa de determinar si un conjunto de variables son independientes es hacer una representación gráfica de las mismas, por ejemplo en el grafo de la figura 1.9³ se representan las variables: *Visita Asia*, *Tuberculosis*, *fumador*, *Cáncer de pulmón*, *Bronquitis*, *Cáncer de pulmón o tuberculosis*, *rayos X positivos*, *Disnea*. En la que cada variable se supone que es binaria.

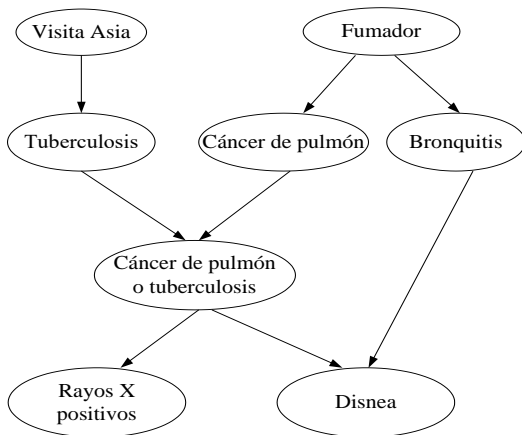


Figura 1.9

Algunas relaciones entre las variables que se pueden obtener de la figura 1.9, son las siguientes: veamos primero el subgrafo $fumador \rightarrow Bronquitis \rightarrow Disnea$, como podemos ver si una persona es fumadora podemos decir que esto le puede provocar Bronquitis y por lo tanto Disnea, sin embargo, si sabemos que la persona tiene Bronquitis, el conocer que es fumador o no, no aumenta nuestro conocimiento de que la persona padecerá Disnea, en otras palabras se tiene que la variable fumador es condicionalmente independiente de la variable Disnea, si conocemos la variable Bronquitis ($I(fumador, Disnea / Bronquitis)$). Otra relación que se

puede obtener de la figura 1.9 es la siguiente: Como sabemos, el tener cáncer de pulmón es independiente de tener Tuberculosis, está independenciam se representa en la figura 1.9 al no haber ningún arco dirigido del nodo *Tuberculosis* al nodo *Cáncer de pulmón*, pero si conocemos el valor que toma el nodo *Cáncer de pulmón o tuberculosis*, este nos hace aumentar o disminuir la creencia en uno de los nodos *Tuberculosis*, *cáncer de pulmón*; por lo que en este caso se tiene que la variable *Tuberculosis* es dependiente de la variable *cáncer de pulmón* ($D(Tuberculosis, cáncer de pulmón/tuberculosis o cáncer de pulmón)$) cuando se conoce el estado del nodo *cáncer de pulmón o tuberculosis*.

³ Esta red fue presentada por Lauritzen y Spiegelhalter como modo de ilustración de su método de propagación de probabilidades.

Antes de definir los criterios de separación gráfica para grafos dirigidos se dará la definición de separación gráfica en grafos no dirigidos a fin de mostrar con esto que la representación de las variables en grafos dirigidos es mejor que en los no dirigidos.

Definición 1.4.1 U-Separación.

Sean X , Y y Z tres conjuntos disjuntos de nodos de un grafo no dirigido G . se dice que Z separa X de Y si y sólo si cada camino entre nodos de X y nodos de Y , contiene algún nodo de Z .

A los nodos de Z se les llama conjunto de corte, ya que el criterio establece que independientemente del camino que se tome (en caso de haber más de uno) para ir del conjunto de nodos X al conjunto de nodos Y siempre habrá por lo menos un nodo de Z , en caso contrario en que existiera un camino alternativo que no contenga nodos de Z , entonces los conjuntos X e Y son condicionalmente dependientes dado Z , ya que se está condicionando a las variables a que se comuniquen por caminos que tienen nodos de Z .

De acuerdo al criterio de U-separación se tiene que las variables *viento* y *humedad* del gráfico 1.1 son marginalmente dependientes porque existe un camino entre ambos nodos, pero se vuelven condicionalmente independientes cuando se conoce el estado de la variable lluvia, es decir se cumple $I(\text{viento}, \text{humedad} / \text{lluvia})$, de acuerdo al criterio de separación U, el grafo no dirigido no representa en este caso correctamente la realidad, ya que a priori se conoce que las variables *viento* y *humedad* son marginalmente independientes, lo cual no está reflejado en el grafo.

Definición 1.4.2. Nodo de aristas convergentes.

Dado un grafo dirigido y un camino no dirigido (... — U — A — V — ...), el nodo A se denomina un nodo de aristas convergentes en este camino si las dos aristas del camino convergen a este nodo en el grafo dirigido, es decir, si el grafo dirigido contiene las aristas $U \rightarrow A \leftarrow V$.

Por ejemplo en el grafo de la figura 1.9 se tiene el camino *visita Asia — tuberculosis — cáncer de pulmón o tuberculosis — Disnea*, en este camino no hay ningún nodo de aristas convergentes, aunque el nodo *cáncer de pulmón o tuberculosis* tiene dos aristas convergentes en él, este no es un nodo de aristas convergentes en el camino considerado, ya que este camino no contiene el nodo *cáncer de pulmón*. Si consideramos el camino *visita Asia — tuberculosis — cáncer de pulmón o tuberculosis — cáncer de pulmón*

— fumador, en este camino se tiene que el nodo *Cáncer de pulmón o tuberculosis* es un nodo de aristas convergentes.

Si en la definición 1.4.2 no existe ninguna arista entre los nodos U y V , entonces a la terna de nodos U, A , y V se le conoce como v -estructura.

Definición 1.4.3 d-separación⁴.

Sean X, Y y Z tres subconjuntos disjuntos de nodos en un grafo dirigido acíclico D ; entonces se dice que Z d -separa X e Y si y sólo si a lo largo de todo camino no dirigido entre cualquier nodo de X y cualquier nodo de Y existe un nodo intermedio A tal que, o bien

1. A es un nodo de aristas convergentes en el camino y ni A ni sus descendientes están en Z , o bien
2. A no es un nodo de aristas convergentes en el camino y A está en Z .

Otra forma de escribir la definición 1.4.3 es la siguiente:

Definición 1.4.4 d-separación.

Un camino es d -separado (o bloqueado) por un conjunto de nodos Z si y solamente si:

1. El camino contiene una cadena $i \rightarrow m \rightarrow j$ o una bifurcación $i \leftarrow m \rightarrow j$ tal que el nodo m está en Z .
2. El camino contiene una bifurcación invertida $i \rightarrow m \leftarrow j$ tal que el nodo m ni los descendientes de m no están en Z .

Un conjunto de nodos Z d -separa X de Y si y solamente si Z bloquea todos los caminos de un nodo de X para un nodo de Y .

⁴ El criterio de d -separación sólo es valido para el caso de grafos dirigidos aciclicos.

Por ejemplo, en el grafo de la figura 1.9, de acuerdo al criterio de d-separación se tiene que se cumple $I(tuberculosis, cáncer de pulmón / \phi)$, ya que el único nodo de aristas convergentes *Cáncer de pulmón o tuberculosis* en el camino no dirigido que une los *nodos Tuberculosis y cáncer de pulmón* no está contenido en ϕ .

Otro criterio de d-separación equivalente a los anteriores y que se basa en un conjunto ancestral de los nodos es el siguiente.

Definición 1.4.5 d-separación.

Sean X, Y y Z tres subconjuntos disjuntos de nodos en un grafo dirigido acíclico D , entonces se dice que Z d-separa a X e Y si y sólo si Z separa X e Y en el grafo moral del menor subconjunto ancestral que contenga a los nodos de X, Y y Z .

El criterio de la definición 1.4.5 nos dice que para comprobar si dos subconjuntos de variables son condicionalmente independientes dado un tercer subconjunto, tenemos que construir el grafo no dirigido asociado al conjunto ancestral que contenga a los nodos, para los cuales se desea verificar si se cumple la relación de independencia, después se debe moralizar dicho grafo y aplicar el criterio de U-separación.

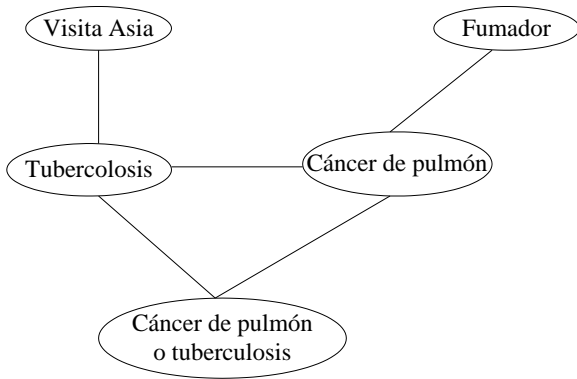


Figura 1.10

Si se quiere verificar $I(tuberculosis, cáncer de pulmón / cáncer de pulmón o tuberculosis)$, se construye el grafo moral asociado al grafo de la figura 1.9 y de él extraemos el menor conjunto ancestral que contenga a los nodos para los cuales queremos saber si se cumple la relación de independencia, este grafo se muestra en la figura 1.10, aplicando el criterio de u-separación, se tiene el camino *tuberculosis – cáncer de pulmón* este, no contiene el nodo *cáncer de pulmón o tuberculosis*,

por lo que se tiene que las variables *tuberculosis* y *cáncer de pulmón* son dependientes cuando se conoce el estado de la variable *cáncer de pulmón o tuberculosis*.

1.4.2 Propiedades de Independencia condicional, modelos de dependencia y su representación.

En ocasiones se pueden conocer algunas relaciones de independencia entre las variables, ya sea por el conocimiento que se tenga sobre el área en la cual se está investigando o por cualquier otro medio del cual dispongamos y estemos interesados en conocer o generar más relaciones de independencia entre ellas a partir de las relaciones conocidas, con el fin de simplificar el modelo probabilístico que se quiera estimar; por lo que aquí se presenta algunas propiedades a partir de las cuales se puedan obtener nuevas relaciones de independencia, dado que disponemos de un conjunto inicial de ellas.

En las siguientes propiedades suponemos que X , Y , W y Z son conjuntos de variables.

1. **Simetría:** $I(X, Y/Z) = I(Y, X/Z)$
2. **Descomposición:** $I(X, Y \cup W/Z) \Rightarrow I(X, Y/Z) \ \& \ I(X, W/Z)$
3. **Unión débil:** $I(X, Y \cup W/Z) \Rightarrow I(X, W/Z \cup Y) \ \& \ I(X, Y/Z \cup W)$
4. **Contracción:** $I(X, W/Z \cup Y) \ \& \ I(X, Y/Z) \Rightarrow I(X, Y \cup W/Z)$
5. **Intersección:** $I(X, W/Z \cup Y) \ \& \ I(X, Y/Z \cup W) \Rightarrow I(X, Y \cup W/Z)$

Las primeras cuatro propiedades las cumple cualquier función de probabilidad, y la propiedad de intersección sólo es satisfecha cuando la función de probabilidad es no extrema⁵.

A continuación se verifica que las funciones de probabilidad satisfacen las propiedades mencionadas:

- **Simetría.**

La propiedad de simetría establece que si $I(X, Y/Z)$, entonces se cumple $I(Y, X/Z)$, lo cual se verifica como sigue:

$$I(Y, X/Z) \equiv P(Y/X, Z) = \frac{P(X, Y, Z)}{P(X, Z)} = \frac{\frac{P(X, Y, Z)}{P(Z)}}{\frac{P(X, Z)}{P(Z)}} = \frac{P(X, Y/Z)}{P(X/Z)}$$

⁵ Se dice que la distribución de probabilidad es no extrema, si sus probabilidades no son valores cercanos a cero o uno.

Puesto que se cumple:

$I(X, Y / Z) \equiv P(X / Y, Z) = P(X / Z)$ Se puede aplicar II por lo que se tiene:

$$I(Y, X / Z) \equiv P(Y / X, Z) = \frac{P(X / Z)P(Y / Z)}{P(X / Z)} = P(Y / Z).$$

Lo cual verifica la propiedad de simetría.

- **Descomposición.**

La propiedad establece que se cumple

$$I(X, Y \cup W / Z) \equiv P(X / Y, W, Z) = P(X / Z) \quad \text{P2.1}$$

Verificamos que la primera implicación del lado derecho se cumple

$I(X, Y / Z) \equiv P(X / Y, Z) = P(X / Z)$:

$$I(X, Y / Z) \equiv P(X / Y, Z) = \sum_w P(x, w / y, z) = \sum_w \frac{P(x, y, w, z)}{P(y, z)} = \sum_w \frac{P(x, y, w, z)P(y, w, z)}{P(y, z)P(y, w, z)}$$

$P(X / Y, Z) = \sum_w P(x / y, w, z)P(w / y, z)$ Por P2.1 Se tiene:

$$P(X / Y, Z) = \sum_w P(x / z)P(w / y, z) = P(X, Z) \sum_w P(w, / y, z) = P(X / Z), \text{ ya que } \sum_w P(w, / y, z) = 1,$$

lo cual verifica $I(X, Y / Z)$.

Ahora se verifica la segunda implicación de la propiedad 2. $I(X, W / Z) \equiv P(X / W, Z) = P(X / Z)$.

$$P(X / W, Z) = \sum_y P(x, y / w, z) = \sum_y \frac{P(x, y, w, z)}{P(w, z)} = \sum_y \frac{P(x, y, w, z)P(y, w, z)}{P(w, z)P(y, w, z)}$$

$P(X / W, Z) = \sum_y P(x / y, w, z)P(y / w, z)$ Aplicando P2.1. Se tiene:

$$P(X / W, Z) = \sum_y P(x / z)P(y / w, z) = P(X / Z) \sum_y P(y / w, z) = P(X / Z). \text{ Donde } \sum_y P(y / w, z) = 1$$

Por lo que se verifica la propiedad de Descomposición.

- **Unión Débil**

La propiedad de Unión débil nos dice que se cumple

$$P(X / Y, W, Z) = P(X / Z) \quad \text{P3.1}$$

Por lo que se cumple cada una de las independencias que están después del signo de implicación.

$$1. \text{ Verificando } I(X, Y/W \cup Z) \equiv P(X/Y, W, Z) = P(X/Z, W). \quad \text{P3.2}$$

Aplicando la propiedad de descomposición se tiene:

$$I(X, Y \cup W/Z) = I(X, W/Z) \equiv P(X/W, Z) = P(X/Z). \quad \text{P3.3}$$

Se tiene: $I(X, Y/W \cup Z) \equiv P(X/Y, W, Z)$, por P3.1 esta equivalencia es igual a $P(X/Z)$ y por P3.3 esta última expresión es igual $P(X/W, Z)$, por lo que se verifica P3.2.

$$2. \text{ Verificando } I(X, W/Y \cup Z) \equiv P(X/W, Y, Z) \quad \text{P3.4}$$

Aplicando de nuevo la propiedad de descomposición se tiene:

$$I(X, Y \cup W/Z) = I(X, Y/Z) \equiv P(X/Y, Z) = P(X/Z) \quad \text{P3.5}$$

Por P3.1 se tiene que P3.4 es igual a $P(X/Z)$ y por P3.5 se tiene que esta última expresión es igual a $P(X/Y, Z)$, por lo que se cumple $I(X, W/Y \cup Z)$.

- **Contracción.**

La propiedad de contracción establece que se cumple las siguientes propiedades

$$1. \quad I(X, W/Z \cup Y) \equiv P(X/Y, W, Z) = P(X/Y, Z)$$

$$2. \quad I(X, Y/Z) \equiv P(X/Y, Z) = P(X/Z)$$

De los literales 1 y 2 se tiene que $P(X/Y, W, Z) = P(X/Y, Z) = P(X/Z)$, por lo que se cumple el lado derecho de la implicación $I(X, Y \cup W/Z) \equiv P(X/W, Y, Z) = P(X/Z)$ y se verifica la propiedad de contracción.

- **Intersección.**

-

La propiedad establece que si se cumple:

$$1. \quad I(X, W/Y \cup Z) \equiv P(X/W, Y, Z) = P(X/Y, Z)$$

$$2. \quad I(X, Y/W \cup Z) \equiv P(X/W, Y, Z) = P(X/W, Z)$$

Entonces se cumple el lado derecho de la implicación por lo que da:

$$I(X, Y \cup W / Z) \equiv P(X / Y, W, Z) = P(X / Z) \quad \text{P5.1}$$

Esto sólo es válido si la función de probabilidad es no extrema.

La propiedad de simetría nos dice que el conocer Y (dado que conocemos Z) no proporciona información sobre X , por lo que el conocer X tampoco proporcionará información acerca de Y .

La propiedad de descomposición establece que si se conoce dos tipos de información ($Y \cup W$) que no dice nada acerca del valor que pueda tomar X entonces cada una por separado tampoco dirá nada sobre el valor de X .

La propiedad de unión débil nos muestra que si dos tipos de información son en un principio irrelevantes para X entonces cada una de ellas por separado no puede hacer que la otra sea relevante para X .

La propiedad de contracción nos dice que una información irrelevante Y no hace que otra información irrelevante W sea relevante para X cuando se conoce Z , por lo que ambas variables W e Y , son irrelevantes para X .

La propiedad de intersección refleja el hecho de que si la variable Y no hace que otra variable W sea relevante para X y esta variable W no hace tampoco que Y sea relevante para X entonces desde un principio ambas variables eran irrelevantes para X .

Las relaciones de independencia que se deriven de un conjunto inicial de relaciones de independencia deberán ser verificadas después, para conocer si estas no contradicen la realidad del problema.

Definición 1.4.6

Dado un conjunto de variables $V = \{V_1, V_2, \dots, V_n\}$ y un conjunto de relaciones de independencia M , entonces, se dice que M es modelo de dependencia del conjunto de variables V , si se puede establecer si es o no cierta la relación de independencia $I(X, Y / Z)$, para cualesquiera tres conjuntos disjuntos de V .

Por ejemplo consideremos el siguiente modelo de dependencia:

$$M = \{I(X_1, X_2 / \phi), I(X_1, X_3 / X_2), I(X_1, X_4 / X_2, X_3)\}$$

De acuerdo a las propiedades de independencia condicional, se pueden derivar de M nuevas relaciones de independencia de las variables como sigue:

$$\text{Contracción: } I(X_1, X_2 / \phi) \text{ y } I(X_1, X_3 / X_2 \cup \phi) \Rightarrow I(X_1, X_2 \cup X_3 / \phi)$$

$$\text{Contracción: } I(X_1, X_4 / X_2, X_3) \text{ y } I(X_1, X_3 / X_2) \Rightarrow I(X_1, X_3 \cup X_4 / X_2)$$

$$\text{Descomposición: } I(X_1, X_3 \cup X_4 / X_2) \Rightarrow I(X_1, X_3 / X_2) \text{ y } I(X_1, X_4 / X_2)$$

$$\text{Unión débil: } I(X_1, X_3 \cup X_4 / X_2) \Rightarrow I(X_1, X_3 / X_2 \cup X_4) \text{ y } I(X_1, X_4 / X_2 \cup X_3)$$

Las relaciones de independencia derivadas del modelo M se muestran a continuación:

$$\{I(X_1, X_2 \cup X_3 / \phi), I(X_1, X_3 \cup X_4 / X_2), I(X_1, X_4 / X_2), I(X_1, X_4 / X_2, X_4)\}$$

Si las relaciones de independencia incluidas en el modelo de dependencia M, provienen de una función de probabilidad entonces el modelo de dependencia M se denomina probabilístico. A los modelos de dependencia que cumplen las primeras cuatro propiedades de independencia condicional mencionadas anteriormente se denominan *semigrafoide* y si cumple las cinco propiedades se denomina *grafoide*.

Un modelo de dependencia se puede representar gráficamente, si los nodos del grafo son las variables que intervienen en el modelo de dependencia y en el grafo están contenidas algunas o todas las relaciones de independencia de las variables que hay en M, las cuales pueden ser verificadas por medio de algún criterio de separación gráfica (dependiendo de si el grafo es dirigido o no). Al tipo de grafos en el cual están representadas todas las relaciones de independencia del modelo M y las relaciones de independencia adicionales que se puedan obtener del grafo también se pueden obtener de M (y viceversa) entonces a estos grafos se les denomina *mapa perfecto* del modelo de dependencia M. No todas las relaciones entre las variable de un modelo de dependencia se pueden representar por medio de un grafo, si no que sólo un conjunto de ellas, por lo que no todo modelo de dependencia posee un mapa perfecto, lo cual da lugar a las siguientes definiciones.

Definición 1.4.7 I-mapa

Un grafo G se dice que es un mapa de independencia (I-mapa) de un modelo de dependencia M si $I(X, Y / Z)_G \Rightarrow I(X, Y / Z)_M$ es decir, las independencias del grafo reflejan independencias del

modelo. O de forma equivalente $D(X, Y / Z)_M \Rightarrow D(X, Y / Z)_G$ es decir todas las dependencias del modelo quedan reflejadas en el grafo.

La definición anterior nos indica que sólo unas relaciones de independencia del modelo quedan reflejadas en el grafo, por lo que puede haber independencias en el modelo representadas como dependencias en el grafo. Un ejemplo de I-mapas son los grafos completos.

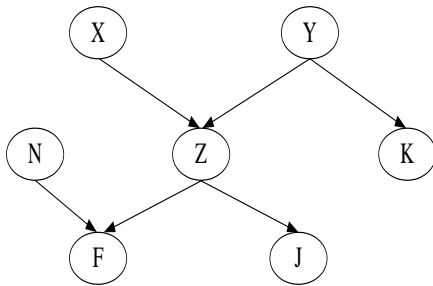


Figura 1.11

El grafo G de la figura 1.11 es un I-mapa del modelo de dependencia:

$$M = I(X, Y / \phi), I(X, K / Y), I(X, F, U, J / Z), I(X, N / F)\}.$$

Ya que todas las relaciones de independencias que están en el grafo están reflejadas en M, excepto la relación $I(X, N / F)$, que de acuerdo al criterio de d-separación estas variables son dependientes.

Definición 1.4.8 d-mapa.

Un grafo G es un mapa de dependencia (D-mapa) del modelo M si $I(X, Y / Z)_M \Rightarrow I(X, Y / Z)_G$, las relaciones de independencias del modelo quedan reflejadas en el grafo, o equivalentemente $D(X, Y / Z)_G \Rightarrow D(X, Y / Z)_M$, es decir las dependencias del grafo, son las dependencias del modelo.

La definición anterior nos dice que en un d-mapa las relaciones de independencia del modelo están reflejadas en el grafo, un ejemplo de d-mapa son los grafos con todos los nodos desconectados, por otra parte, si en el modelo de dependencia anterior se elimina la relación $I(X, N / F)$ entonces el grafo de la figura 1.11 sería un d-mapa del modelo de dependencia resultante, ya que todas sus relaciones de independencia quedarían reflejadas en el grafo.

Definición 1.4.9 I-mapa minimal

Un grafo es un I-mapa minimal cuando al eliminar una cualquiera de sus aristas este deja de ser un I-mapa.

Hasta el momento se ha definido que es un modelo de dependencia y cuando un grafo es un I-mapa y D-mapa de un modelo de dependencia, puesto que sólo nos interesa los modelos de dependencia probabilísticos, es decir aquellos en que las relaciones de independencia provienen de una distribución de probabilidad, supondremos que M es un modelo de dependencia probabilístico. Hay un teorema que nos indica como construir un I-mapa minimal de un modelo probabilístico.

En él se considera que dada una permutación $Y = \{Y_1, Y_2, \dots, Y_n\}$ de un conjunto de variables $X = \{X_1, X_2, \dots, X_n\}$ y una función de probabilidad $P(X)$ de X , el grafo dirigido acíclico creado asignando como padres de cada nodo Y_i cualquier conjunto de ascendientes Π_i , que cumpla $P(Y_i = y_i / B_i = b_i) = P(Y_i = y_i / \Pi_i = \pi_i)$ para todos los padres π_i de las variables $\Pi_i \subset B_i = \{Y_1, Y_2, \dots, Y_{i-1}\}$, es un I-mapa minimal dirigido de $p(x)$ ⁶.

El grafo construido de la función de probabilidad resulta ser siempre un I-mapa minimal de la función, pueden haber distintos I-mapas minimales de una misma función de probabilidad que correspondería a distintas ordenaciones de las variables, además se tiene que sólo las funciones de probabilidad no extremas poseen I-mapas minimales únicos.

Por ejemplo consideremos la siguiente función de probabilidad de cinco variables:

$$P(X) = P(X_1)P(X_2)P(X_3 / X_1, X_2)P(X_4 / X_2)P(X_5 / X_3, X_4) \quad (6)$$

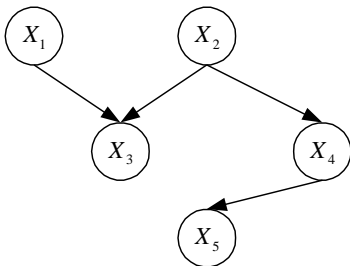


Figura 1.12

De acuerdo a lo expuesto anteriormente el grafo se construye designando como conjunto de padres de cada variable a aquellas variables de las cuales depende directamente en la factorización, el grafo de la función de probabilidad (6) se muestra en la figura 1.12 que corresponde a la ordenación de las variables $\{X_1, X_2, X_3, X_4, X_5\}$,

Dado un grafo dirigido también se puede encontrar una función de probabilidad de tal modo que el grafo sea un I-mapa minimal dirigido de la función de probabilidad que resulta. Por lo que se tiene la siguiente definición.

⁶ Ver el teorema 6.10 página 248 del libro “Sistemas Expertos y Modelos de Redes Probabilísticas”, Enrique Castillo, José Manuel Gutierrez y Ali S. Adi.

Definición 1.4.10

Dada una función de probabilidad se dice que admite una factorización recursiva según un grafo dirigido acíclico, si la función de probabilidad se puede expresar como:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(x_i / \pi_i) \tag{7}$$

Donde $P(x_i / \pi_i)$ es la función de probabilidad de X_i condicionada a sus padres Π_i .

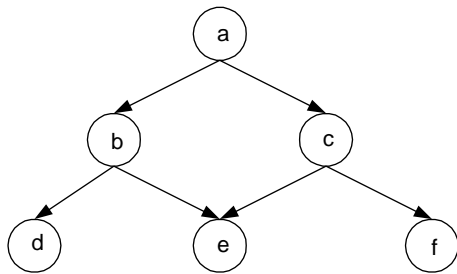


Figura 1.13

La definición anterior nos dice que es posible obtener una función de probabilidad de cualquier grafo dirigido siempre y cuando este sea acíclico. Por ejemplo consideremos el grafo de la figura 1.13

Para encontrar la función de probabilidad conjunta, para el cual el grafo de la figura 1.13 es un I-mapa minimal, lo que se hace es crear funciones de probabilidad condicionales, es decir, se crean funciones de probabilidad condicionadas de cada nodo del grafo a sus padres, por ejemplo la función de probabilidad condicionada del nodo C es: $P(c/a)$ de forma similar se hace con los demás nodos del grafo y cuando un nodo no tiene padres la función de probabilidad es su función de probabilidad marginal. La factorización de la función de probabilidad para la cual el grafo de la figura 1.13 es un I-mapa minimal, según la expresión (7) es:

$$p(a,b,c,d,e,f) = p(a)p(b/a)p(c/a)p(d/b)p(e/b,c)p(f/c)$$

Hasta el momento se ha hablado como se representa gráficamente una función de probabilidad y como se obtiene una factorización de una función probabilidad según un grafo dirigido, por lo que en la siguiente definición se da lo que es una red bayesiana.

Definición 1.4.11 red bayesiana.

Una red bayesiana es un par (D, P) , donde D es un grafo dirigido acíclico, y P es un conjunto de funciones de probabilidad condicionada, una para cada variable condicionada a sus padres en

el grafo, en su representación gráfica D . El conjunto P define una función de probabilidad asociada mediante la factorización

$$p(x) = \prod_{i=1}^n p(x_i / \pi_i)$$

El grafo dirigido acíclico D es un I-mapa minimal de $p(x)$.

Dado que el grafo dirigido acíclico D es un I-mapa minimal, se tiene que todas las relaciones de independencia entre las variables que se encuentran en el grafo, también se encuentran contenidas en la función de probabilidad.

Las redes bayesianas pueden ser continuas o discretas dependiendo, de si todos sus nodos se distribuyen según una función discreta de probabilidad o continua, o pueden ser redes bayesianas mixtas aquellas en las que algunos nodos son discretos y otros continuos.

A las redes bayesianas cuya representación gráfica es un árbol o un poliárbol se les conoce como redes

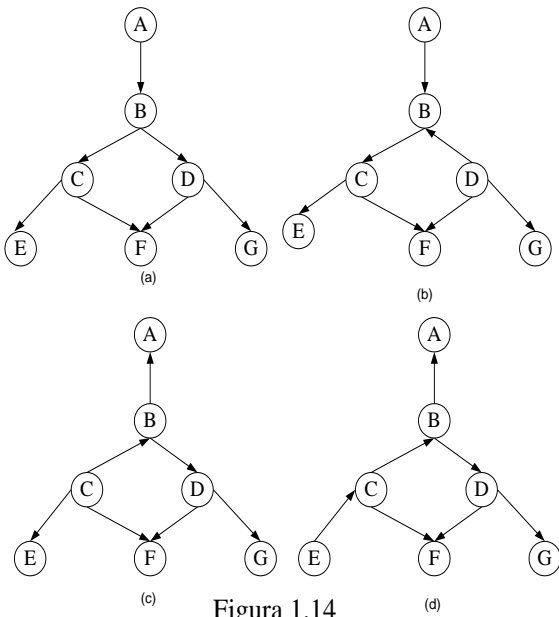


Figura 1.14

simplemente conectadas, y si la representación gráfica de una red bayesiana es más compleja se les conoce como redes multiconectadas. Las redes bayesianas también se pueden representar relaciones temporales entre los nodos, a este tipo de redes se les conoce como redes bayesianas dinámicas. Dado que varios grafos dirigidos pueden representar el mismo modelo de dependencia, se tiene que pueden haber por lo menos dos redes bayesianas que sean equivalentes. Por ejemplo los grafos de la figura 1.14, contienen las mismas relaciones de independencia que contiene el siguiente modelo:

$$M = \{I(A, F / C, D), I(A, E / C), I(A, D / B), I(B, G / D)\}$$

Por lo que se dice que los modelos gráficos de la figura 1.14 son equivalentes.

Definición 1.4.12

Dos modelos gráficos son equivalentes, si representan el mismo modelo de dependencia.

Definición 1.4.13

Dos redes bayesianas son equivalentes si tiene asociados los mismos grafos no dirigidos, las mismas v-estructuras y las distribuciones de probabilidad de ambas redes son iguales.

La definición 1.4.12 puede no ser válida cuando los nodos del grafo tengan interpretación causal, por ejemplo si suponemos que la red (a) de la figura 1.14 tiene interpretación causal es decir la causa precede el efecto, entonces aunque las redes de la (b) a la (d) sean equivalentes a la (a) no serían válidas debido a la interpretación de los nodos del grafo, ya que el cambio de sentido de una de las aristas distorsionaría la realidad del problema que se está modelando.

CAPÍTULO 2.

INFERENCIA EN REDES BAYESIANAS.

2. Inferencia en Redes Bayesianas.

2.1 Introducción.

La propagación de probabilidad en una red bayesiana significa el cálculo de las probabilidades de los nodos de la red condicionados a los valores que toman con certeza un subconjunto de nodos de la red (a este subconjunto de nodos se les conoce como evidencia), los métodos que se utilizan para propagar probabilidades en una red bayesiana se dividen en tres clases, métodos exactos que calculan las probabilidades de los nodos de la red sin otro error que el cometido por el redondeo de la computadora, los métodos aproximados que se utilizan cuando hay demasiados nodos de la red y lo que hacen es muestrear los nodos de la red y a partir de las muestras obtenidas calcular las probabilidades aproximadas de los nodos dada la evidencia y los métodos de propagación simbólica que se utilizan cuando no se dispone de los valores numéricos de las probabilidades de algunos nodos, o cuando se dispone de intervalos de valores de los nodos, también se utilizan para realizar análisis de sensibilidad de las probabilidades de los nodos de la red. En este capítulo sólo se describirán los métodos de propagación exacta.

En el presente capítulo se describen los siguientes métodos: propagación en poliárboles, métodos de condicionamiento, método de agrupamiento y propagación orientada a un objetivo.

La característica principal de estos métodos es que se basan en la estructura topológica de la red. Así por ejemplo el método de propagación para poliárboles sólo es aplicable a redes simplemente conectadas, mientras que los métodos de agrupamiento y condicionamiento es aplicable a redes más generales, es decir, con bucles, pero modifican la estructura topológica de la red, por ejemplo el método de agrupamiento construye un grafo triangulado a partir del grafo no dirigido asociado al grafo dirigido de la red.

2.2 Propagación exacta en redes bayesianas.

2.2.1 Propagación en poliárboles.

El método que se presenta a continuación sólo es aplicable a árboles y a poliárboles, ya que aprovecha la estructura gráfica de este tipo de red (En cada par de nodos de un poliárbol existe sólo un camino, por lo que al eliminar un enlace de la red ésta se divide en dos redes), por lo que su principal limitación, es que no se puede aplicar a redes con bucles, que son las que con más frecuencia se presenta en la práctica.

Este método consiste en el paso de mensajes entre los nodos para calcular las probabilidades condicionales de las variables dada la evidencia⁷ de la cual se dispone, este método fue desarrollado por Kim (1981) a partir del método desarrollado por Pearl (1980) para árboles.

Descripción del método

Puesto que en un poliárbol hay un único camino entre cada par de nodos se pueden dividir la evidencia (la cual la denotaremos con e), en dos subconjuntos disjuntos de nodos tales como $e = e_{X_i}^+ \cup e_{X_i}^-$ donde:

$e_{X_i}^+$: denota la evidencia que es accesible desde los padres del nodo X_i .

$e_{X_i}^-$: denota la evidencia que es accesible desde los hijos del nodo X_i .

Por ejemplo consideremos la figura 2.1 en que cada variable es binaria, se denota con X_i^+ , si la variable toma el valor de uno, y con X_i^- , si la variable toma el valor de cero.

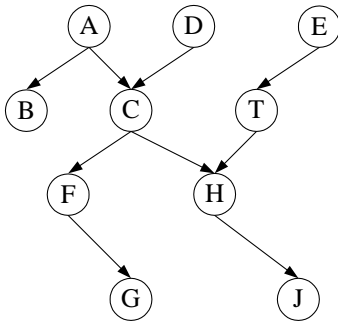


Figura 2.1

Si tenemos el siguiente conjunto evidencial $e = \{f^+, t^+, a^-, j^-\}$, entonces se puede partir este conjunto, tomando como referencia el nodo H de la siguiente manera:

$$e_H^+ = \{f^+, t^+, a^-\}$$

$$e_H^- = \{j^-\}$$

Si tomamos como referencia el nodo C se tiene:

$$e_C^+ = \{a^-\}$$

$$e_C^- = \{f^+, t^+, j^-\}$$

El interés está en calcular: $P(x_i / e)$, la cual se puede descomponer como sigue:

$$P(x_i / e) = \frac{P(x_i, e)}{P(e)}$$

$$P(x_i / e) = \frac{P(x_i)P(e / x_i)}{P(e)}$$

⁷ Entenderemos por evidencia, cuando conocemos con certeza que un conjunto de variables de la red, ha tomando un valor de su conjunto de valores.

Si suponemos que $e = \{X_{r_1} = x_{r_1}, \dots, X_{r_k} = x_{r_k}, \dots, X_{r_j} = x_{r_j}\}$ y ordenamos este conjunto de tal forma que las primeras k variables corresponde a la evidencia que es accesible a X_i a través de sus padres y las restantes $j - k$ corresponde a la evidencia que es accesible a X_i a través de sus hijos, entonces tenemos:

$$P(x_i / e) = \frac{P(x_i)P(X_{r_1} = x_{r_1}, \dots, X_{r_k} = x_{r_k}, X_{r_{k+1}} = x_{r_{k+1}}, \dots, X_{r_j} = x_{r_j} / x_i)}{P(e)}$$

Dado que $e_{X_i}^+ = \{X_{r_1} = x_{r_1}, \dots, X_{r_k} = x_{r_k}\}$ y $e_{X_i}^- = \{X_{r_{k+1}} = x_{r_{k+1}}, \dots, X_{r_j} = x_{r_j}\}$ se tiene:

$$P(x_i / e) = \frac{P(x_i)P(e_{X_i}^+, e_{X_i}^- / x_i)}{P(e)}$$

Puesto que el nodo X_i d-separa a la evidencia que se encuentra por arriba de él, con la evidencia que se encuentra por debajo de él, se tiene:

$$P(x_i / e) = \frac{P(x_i)P(e_{X_i}^+ / x_i)P(e_{X_i}^- / x_i)}{P(e)}$$

$$P(x_i / e) = \frac{P(e_{X_i}^+, x_i)P(e_{X_i}^- / x_i)}{P(e)} \tag{2.1}$$

Llamando a:

$$\lambda_i(x_i) = P(e_{X_i}^- / x_i)$$

$$\rho_i(x_i) = P(e_{X_i}^+, x_i)$$

$$\alpha = 1 / P(e)$$

La expresión (2.1) queda como:

$$P(x_i / e) = \alpha \rho_i(x_i) \lambda_i(x_i) \tag{2.2}$$

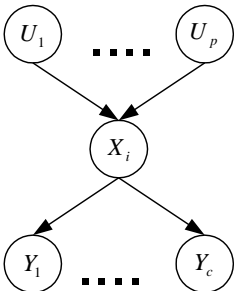


Figura 2.2

Para calcular las expresiones $\rho_i(x_i)$, $\lambda_i(x_i)$ se utilizará la siguiente partición de la evidencia:

$$e_i^+ = \{e_{u_1 x_i}^+, e_{u_2 x_i}^+, \dots, e_{u_p x_i}^+\} \tag{2.3}$$

$$e_i^- = \{e_{x_i y_1}^-, e_{x_i y_2}^-, \dots, e_{x_i y_c}^-\} \tag{2.4}$$

Donde:

e_{u_j, x_i}^+ : Representa la evidencia contenida en el subgrafo que contiene a su padre U_j cuando se elimina el enlace $u_j \rightarrow x_i$

e_{x_i, y_j}^- : Representa la evidencia contenida en el subgrafo que contiene a su hijo Y_j cuando se elimina el enlace $x_i \rightarrow y_j$.

Primeramente se ilustrará el cálculo de la función $\rho_i(x_i)$, como sigue:

$$\rho_i(x_i) = P(x_i, e_i^+) = \sum_{u_1} \sum_{u_2} \sum_{u_3} \dots \sum_{u_p} P(x_i, \{u_1, \dots, u_p\} \cup e_i^+) = \sum_u P(x_i / u \cup e_i^+) P(u \cup e_i^+)^8$$

Donde:

$u = \{u_1, \dots, u_p\}$: representa el conjunto de P padres del nodo x_i .

Utilizando (2.3) se tiene:

$$\rho_i(x_i) = \sum_u P(x_i / u \cup e_i^+) P(u \cup e_{u_1, x_i}^+, e_{u_2, x_i}^+ \cup \dots \cup e_{u_p, x_i}^+)$$

Puesto que e_{u_j, x_i}^+ esta d-separado de e_{u_k, x_i}^+ (Son marginalmente independientes) se tiene:

$$\rho_i(x_i) = \sum_u P(x_i / u \cup e_i^+) P(u_1 \cup e_{u_1, x_i}^+) \dots P(u_p \cup e_{u_p, x_i}^+)$$

$$\rho_i(x_i) = \sum_u P(x_i / u \cup e_i^+) \prod_{j=1}^P P(u_j \cup e_{u_j, x_i}^+)$$

$$\rho_i(x_i) = \sum_u P(x_i / u \cup e_i^+) \prod_{j=1}^P \rho_{u_j, x_i}(u_j)$$

Puesto que X_i esta d-separado de e_i^+ dado el conjunto de padres del nodo X_i , se tiene:

$$\rho_i(x_i) = \sum_u P(x_i / u) \prod_{j=1}^P \rho_{u_j, x_i}(u_j) \tag{2.5}$$

Donde:

$$\rho_{u_j, x_i}(u_j) = P(u_j \cup e_{u_j, x_i}^+) \tag{2.6}$$

$\rho_{u_j, x_i}(u_j)$: Representa el mensaje que el nodo U_j envía a su hijo X_i .

⁸ Del cálculo de probabilidades se tiene: $P(x_1, x_2) = \sum_{x_3} \sum_{x_4} \sum_{x_5} \dots \sum_{x_n} P(x_1, x_2, \dots, x_n)$.

Si el nodo U_j es un nodo evidencial es decir $u_j = e_j$ para un valor particular de U_j entonces se tiene:

$\rho_{u_j, x_i}(u_j) = 1$ y los demás valores del nodo U_j se le asigna $\rho_{u_j, x_i}(u_j) = 0$.

Si el nodo X_i no tiene padres y no es un nodo evidencial se tiene: $\rho_i(X_i) = P(X_i)$ para todos los valores de X_i .

Para el cálculo de la función $\lambda_i(X_i)$ se utiliza (2.4) como sigue:

$$\lambda_i(x_i) = P(e_i^- / x_i) = P(e_{x_i, y_1}^-, e_{x_i, y_2}^-, \dots, e_{x_i, y_c}^- / x_i)$$

Dado que X_i d-separa a e_{x_i, y_j}^- de e_{x_i, y_k}^- se tiene:

$$\lambda_i(x_i) = P(e_{x_i, y_1}^- / x_i) P(e_{x_i, y_2}^- / x_i) \dots P(e_{x_i, y_c}^- / x_i)$$

$$\lambda_i(x_i) = \prod_{j=1}^c P(e_{x_i, y_j}^- / x_i)$$

$$\lambda_i(x_i) = \prod_{j=1}^c \lambda_{y_j, x_i}(x_i) \tag{2.7}$$

$$\lambda_{y_j, x_i}(x_i) = P(e_{x_i, y_j}^- / x_i) \tag{2.8}$$

Donde $\lambda_{y_j, x_i}(x_i)$ representa el mensaje que el nodo Y_j envía a su padre X_i .

Por lo que (2.2) se puede escribir como:

$$P(x_i / e) = \alpha \left(\sum_u P(x_i / u) \prod_{j=1}^p \rho_{u_j, x_i}(u_j) \right) \left(\prod_{j=1}^c \lambda_{y_j, x_i}(x_i) \right) \tag{2.9}$$

2.2.1 Cálculo de cada uno de los mensajes.

❖ Cálculo del mensaje $\rho_{x_i, y_j}(x_i)$

Para el cálculo de este mensaje se considera la evidencia que se encuentra por arriba de X_i y la que esta por el subgrafo que se obtiene al eliminar el enlace $X_i \rightarrow Y_j$ para los demás efectos de X_i distintos de Y_j como sigue:

$$e_{x_i, y_j}^+ = e_i^+ \bigcup_{k \neq j} e_{x_i, y_k}^-$$

Por lo que se tiene:

$$\rho_{x_i y_j}(x_i) = P(x_i \cup e_{x_i y_j}^+)$$

$$\rho_{x_i y_j}(x_i) = P\left(x_i \cup \left\{ e_i^+ \cup \bigcup_{k \neq j} e_{x_i y_k}^- \right\}\right)$$

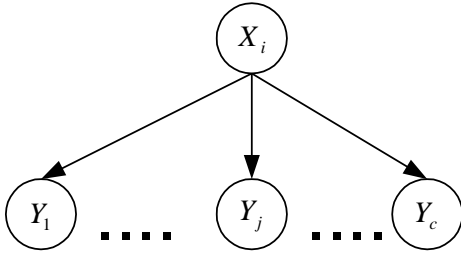


Figura 2.3

$$\rho_{x_i y_j}(x_i) = P(x_i, e_i^+) P\left(\bigcup_{k \neq j} e_{x_i y_k}^- / x_i \cup e_i^+\right)$$

Como se puede observar en la figura 2.3 cada $e_{x_i y_k}^-$ esta d-separado de e_i^+ por el nodo x_i , por lo que la expresión anterior queda como:

$$\rho_{x_i y_j}(x_i) = P(x_i, e_i^+) \prod_{k \neq j} P(e_{x_i y_k}^- / x_i)$$

$$\rho_{x_i y_j}(x_i) = \rho_i(x_i) \prod_{k \neq j} \lambda_{y_k x_i}(x_i) \tag{2.10}$$

Por lo que para enviar el mensaje del nodo X_i a su hijo Y_j , este debe de haber recibido el mensaje de todos sus demás hijos distintos de Y_j y haber calculado su función $\rho_i(x_i)$.

❖ **Cálculo del mensaje $\lambda_{y_j x_i}(x_i)$**

En este caso se divide la evidencia que se encuentra por debajo del enlace $X_i \rightarrow Y_j$, en la que se encuentra por debajo del nodo Y_j y la evidencia que es accesible a través de sus padres distintos de X_i por lo que nos queda:

$$e_{x_i y_j}^- = e_{y_j}^- \cup e_{v y_j}^+ \tag{2.11}$$

El cálculo del mensaje $\lambda_{y_j x_i}(x_i)$ es como sigue:

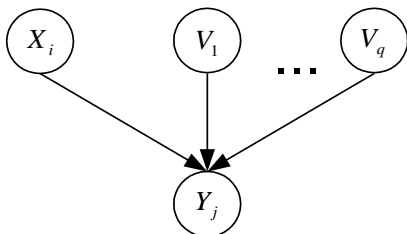


Figura 2.4

$$\lambda_{y_j x_i}(x_i) = P(e_{x_i y_j}^- / x_i) = \frac{P(e_{x_i y_j}^-, x_i)}{P(x_i)} = \frac{\sum_{y_j} \sum_v P(y_j, v, e_{x_i y_j}^-, x_i)}{P(x_i)}$$

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j} \sum_v P(y_j, v, e_{x_i y_j}^- / x_i)$$

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j, v} P(y_j, v, e_{y_j}^-, e_{v y_j}^+ / x_i)$$

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j, v} P(e_{y_j}^- / y_j, v, e_{vy_j}^+ x_i) P(y_j / v, e_{vy_j}^+, x_i) P(v, e_{vy_j}^+ / x_i)$$

Como se observa en la figura 2.4 $e_{y_j}^-$ esta d-separado de v , $e_{vy_j}^+$, x_i por y_j , por lo que se tiene:

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j, v} P(e_{y_j}^- / y_j) P(y_j / v, e_{vy_j}^+, x_i) P(v, e_{vy_j}^+ / x_i)$$

y_j Esta d-separado de $e_{vy_j}^+$ por v y x_i , por lo que se obtiene lo siguiente:

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j, v} P(e_{y_j}^- / y_j) P(y_j / v, x_i) P(v, e_{vy_j}^+ / x_i)$$

Finalmente v , $e_{vy_j}^+$ son marginalmente independientes por d-separación con X_i , por lo que nos queda:

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j} P(e_{y_j}^- / y_j) \sum_v P(y_j / v, x_i) P(v, e_{vy_j}^+)$$

$$\lambda_{y_j x_i}(x_i) = \sum_{y_j} \lambda_j(y_j) \sum_{v_1, \dots, v_q} P(y_j / \pi_j) \prod_{k=1}^q \rho_{v_k y_j}(v_k) \quad (2.12)$$

Con esto se termina el cálculo de los mensajes.

El valor de α que aparece en la ecuación (2.2) se calcula de la siguiente manera:

$$\sum_{x_i} P(x_i) = \sum_{x_i} \alpha \lambda_i(x_i) \rho_i(x_i) = 1$$

$$\alpha = \left(\sum_{x_i} \lambda_i(x_i) \rho_i(x_i) \right)^{-1} \quad (2.13)$$

Para aplicar el método de propagación en poliárboles se hace lo siguiente:

1. Primero se hace: $\left\{ \begin{array}{l} \rho_i(x_i) = 1; \text{ si } x_i \in e \\ \rho_i(x_i) = 0; \text{ si } x_i \notin e \end{array} \right\}$ y $\left\{ \begin{array}{l} \lambda_i(x_i) = 1; \text{ si } x_i \in e \\ \lambda_i(x_i) = 0; \text{ si } x_i \notin e \end{array} \right\}$
2. Para los nodos sin padre se hace: $\rho_i(x_i) = P(x_i)$
3. Para los nodos que no tienen hijos se hace: $\lambda_i(x_i) = 1$.

Para empezar a propagar la evidencia se puede empezar calculando los mensajes que se enviarán de aquellos nodos que no tienen padre y tienen sólo un hijo utilizando (2.10), también se puede empezar enviando los mensajes de los nodos que no tienen hijos y que sólo tienen un padre utilizando (2.12).

En el caso de no disponer de evidencia el método de propagación en poliarboles calcula las probabilidades marginales de los nodos de la red.

Ejemplo 2.1: Consideremos el poliarbol de la figura 2.5, en él aparecen las probabilidades condicionales de cada nodo, aquí consideramos que cada variable es binaria.

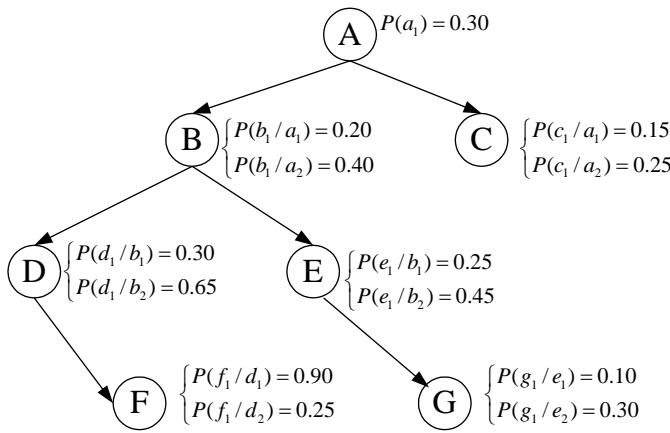


Figura 2.5

Se ilustrará el método de propagación de probabilidades en el caso en que se dispone con la siguiente evidencia:

$$e = \{D = d_1, C = c_2\}. \quad (\text{E 2.1})$$

Lo que se quiere es calcular es:

$$P(X_i = x_i / e) = P(X_i = x_i / D = d_1, C = c_1)$$

Para todos los nodos no evidenciales de la red 2.5.

Recordemos primero que $\rho_D(d_1) = P(d_1, e_D^+)$, puesto que el estado d_1 ocurre con seguridad (forma parte de la evidencia), tenemos que la probabilidad $P(d_1, e_D^+)$, es igual a uno.

Para el nodo D:

$$\begin{cases} \rho_D(d_1) = 1; & d_1 \in e \\ \rho_D(d_2) = 0 & d_2 \notin e \end{cases} \text{ y } \begin{cases} \lambda_D(d_1) = 1; & d_1 \in e \\ \lambda_D(d_2) = 0 & d_2 \notin e \end{cases}$$

De igual forma para el nodo C:

$$\begin{cases} \rho_C(c_1) = 0; & c_1 \in e \\ \rho_C(c_2) = 1; & c_2 \notin e \end{cases} \text{ y } \begin{cases} \lambda_C(c_1) = 0; & c_1 \in e \\ \lambda_C(c_2) = 1; & c_2 \notin e \end{cases}$$

Luego para los nodos no evidenciales.

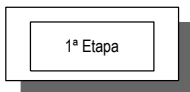
Para los nodos sin padre se tiene:

$$\begin{cases} \rho_A(a_1) = P(a_1) \\ \rho_A(a_2) = P(a_2) \end{cases} \Rightarrow \begin{cases} \rho_A(a_1) = 0.3 \\ \rho_A(a_2) = 0.7 \end{cases}$$

Para los nodos sin hijos excluyendo el nodo evidencial C se tiene:

$$\begin{cases} \lambda_F(f_1) = 1 \\ \lambda_F(f_2) = 1 \end{cases} \text{ y } \begin{cases} \lambda_G(g_1) = 1 \\ \lambda_G(g_2) = 1 \end{cases}$$

En lo que sigue se dividirá el paso de mensajes entre nodos en tres etapas.



1. El nodo A

Su función $\rho_A(a)$ ha sido calculada, pero no puede enviar ningún mensaje a sus dos hijos B y C, ya que no ha recibido ningún mensaje de alguno de ellos.

2. El nodo B.

Este nodo no ha recibido ningún mensaje de su padre A, ni de sus hijos D, E, por lo que no se puede enviar ningún mensaje a su padre e hijos.

3. El nodo C.

La función $\lambda_C(c)$ ya ha sido calculada, dado que sólo tiene un padre que es A puede enviarle el mensaje

$\lambda_{CA}(a)$, utilizando (2.12) se tiene:

$$\lambda_{CA}(a) = \sum_c \lambda_C(c) p(c/a)$$

Para $A = a_1$ se tiene:

$$\lambda_{c_A}(a_1) = \lambda_c(c_1)P(c_1/a_1) + \lambda_c(c_2)P(c_2/a_1) \Rightarrow \lambda_{c_A}(a_1) = 0(0.15) + 1(0.85)$$

$$\lambda_{c_A}(a_1) = 0.85$$

Para $A = a_2$ se tiene:

$$\lambda_{c_A}(a_2) = \lambda_c(c_1)P(c_1/a_2) + \lambda_c(c_2)P(c_2/a_2) \Rightarrow \lambda_{c_A}(a_2) = 0(0.25) + 1(0.75)$$

$$\lambda_{c_A}(a_2) = 0.75$$

4. El nodo D.

El nodo D es un nodo evidencial su función $\rho_D(d)$ ha sido calculada, puesto que sólo tiene un hijo F, puede enviarle el mensaje $\rho_{DF}(d)$ utilizando (2.10) como sigue:

$$\rho_{DF}(d) = \rho_D(d). \text{ Por lo que para } D = d_1 \text{ y } D = d_2 \text{ se tiene: } \begin{cases} \rho_{DF}(d_1) = 1 \\ \rho_{DF}(d_2) = 0 \end{cases}$$

Dado que también su función $\lambda_D(d)$ ha sido calculada puede enviarle el mensaje $\lambda_{DB}(b)$ a su único padre B, utilizando (2.12) como sigue:

$$\lambda_{DB}(b) = \sum_d \lambda_D(d)P(d/b)$$

Para $B = b_1$ se tiene:

$$\lambda_{DB}(b_1) = \lambda_D(d_1)P(d_1/b_1) + \lambda_D(d_2)P(d_2/b_1) \Rightarrow \lambda_{DB}(b_1) = 1(0.3) + 0(0.75)$$

$$\lambda_{DB}(b_1) = 0.3$$

Para $B = b_2$ se tiene:

$$\lambda_{DB}(b_2) = \lambda_D(d_1)P(d_1/b_2) + \lambda_D(d_2)P(d_2/b_2) \Rightarrow \lambda_{DB}(b_2) = 1(0.65) + 0(0.35)$$

$$\lambda_{DB}(b_2) = 0.65$$

5. El nodo E.

En este nodo no se ha recibido ningún mensaje de su padre B, ni de su hijo G, y ninguna de sus funciones $\lambda_E(e)$ y $\rho_E(e)$ ha sido calculada por lo que no se puede enviar ningún mensaje.

6. El nodo G.

Su función $\lambda_G(g)$ ha sido calculada, puesto que sólo tiene un padre puede enviarle el mensaje $\lambda_{GE}(e)$ utilizando (2.12) como sigue:

$$\lambda_{GE}(e) = \sum_g \lambda_G(g)P(g/e)$$

Para $E = e_1$ se tiene:

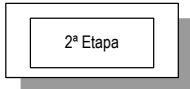
$$\lambda_{GE}(e_1) = \lambda_G(g_1)P(g_1/e_1) + \lambda_G(g_2)P(g_2/e_1) \Rightarrow \lambda_{GE}(e_1) = 1(0.10) + 1(0.9)$$

$$\lambda_{GE}(e_1) = 1$$

Para $E = e_2$ se tiene:

$$\lambda_{GE}(e_2) = \lambda_G(g_1)P(g_1/e_2) + \lambda_G(g_2)P(g_2/e_2) \Rightarrow \lambda_{GE}(e_2) = 1(0.3) + 1(0.7)$$

$$\lambda_{GE}(e_2) = 1$$



1. El nodo A.

Ha recibido el mensaje $\lambda_{CA}(a)$ de su hijo C y su función $\rho_A(a)$ ha sido calculada, por lo que puede enviar el mensaje $\rho_{AB}(a)$ a su hijo B utilizando (2.12) como sigue:

$$\rho_{AB}(a) = \rho_A(a)\lambda_{CA}(a)$$

Para $A = a_1$ se tiene:

$$\rho_{AB}(a_1) = \rho_A(a_1)\lambda_{CA}(a_1) \Rightarrow \rho_{AB}(a_1) = 0.3(0.85)$$

$$\rho_{AB}(a_1) = 0.255$$

Para $A = a_2$ se tiene:

$$\rho_{AB}(a_2) = \rho_A(a_2)\lambda_{CA}(a_2) \Rightarrow \rho_{AB}(a_2) = 0.7(0.75) \Rightarrow \rho_{AB}(a_2) = 0.525$$

2. El nodo B.

El nodo B ha recibido el mensaje $\rho_{AB}(a)$ de su padre A, por lo que se puede calcular su función $\rho_B(b)$ utilizando (2.5) como sigue:

$$\rho_B(b) = \sum_a P(b/a)\rho_{AB}(a)$$

Para $B = b_1$ se tiene:

$$\rho_B(b_1) = P(b_1/a_1)\rho_{AB}(a_1) + P(b_1/a_2)\rho_{AB}(a_2) \Rightarrow \rho_B(b_1) = 0.2(0.255) + 0.4(0.525)$$

$$\rho_B(b_1) = 0.261$$

Para $B = b_2$ se tiene:

$$\rho_B(b_2) = P(b_2/a_1)\rho_{AB}(a_1) + P(b_2/a_2)\rho_{AB}(a_2) \Rightarrow \rho_B(b_2) = 0.8(0.255) + 0.6(0.525)$$

$$\rho_B(b_2) = 0.519$$

Dado que el nodo B ha recibido el mensaje $\lambda_{DB}(b)$ de su hijo B, y se ha calculado la función $\rho_B(b)$ se puede calcular el mensaje $\rho_{BE}(b)$, utilizando (2.10) como sigue:

$$\rho_{BE}(b) = \rho_B(b)\lambda_{DB}(b)$$

Para $B = b_1$ se tiene:

$$\rho_{BE}(b_1) = \rho_B(b_1)\lambda_{DB}(b_1) \Rightarrow \rho_{BE}(b_1) = 0.261(0.3)$$

$$\rho_{BE}(b_1) = 0.0783$$

Para $B = b_2$ se tiene:

$$\rho_{BE}(b_2) = \rho_B(b_2)\lambda_{DB}(b_2) \Rightarrow \rho_{BE}(b_2) = 0.519(0.65)$$

$$\rho_{BE}(b_2) = 0.33735$$

3. El nodo C.

Este nodo es evidencial, y como ya se envió el mensaje a su padre, ya no se realiza ningún cálculo.

4. El nodo D.

Este nodo también es evidencial y dado que ya mando todos los mensajes a su padre e hijos, ya no se realiza ningún cálculo con él.

5. El nodo E.

Este nodo ha recibido el mensaje $\rho_{BE}(b)$ de su único padre B por lo que se puede calcular $\rho_E(e)$ utilizando 2.5 como sigue:

$$\rho_E(e) = \sum_b P(e/b)\rho_{BE}(b)$$

Para $E = e_1$ se tiene:

$$\rho_E(e_1) = P(e_1/b_1)\rho_{BE}(b_1) + P(e_1/b_2)\rho_{BE}(b_2) \Rightarrow \rho_E(e_1) = 0.25(0.0783) + 0.45(0.33755)$$

$$\rho_E(e_1) = 0.1714725$$

Para $E = e_2$ se tiene:

$$\rho_E(e_2) = P(e_2/b_1)\rho_{BE}(b_1) + P(e_2/b_2)\rho_{BE}(b_2) \Rightarrow \rho_E(e_2) = 0.75(0.0783) + 0.55(0.33755)$$

$$\rho_E(e_2) = 0.2443775$$

Puesto que el nodo E, tiene sólo un hijo y se ha calculado la función $\rho_E(e)$, se puede calcular el mensaje

$\rho_{EG}(e)$ como sigue:

$$\rho_{EG}(e) = \rho_E(e)$$

Para $E = e_1$ se tiene: $\rho_{EG}(e_1) = \rho_E(e_1) \Rightarrow \rho_{EG}(e_1) = 0.1714725$

Para $E = e_2$ se tiene: $\rho_{EG}(e_2) = \rho_E(e_2) \Rightarrow \rho_{EG}(e_2) = 0.2442675$

Dado que el nodo E también ha recibido el mensaje $\lambda_{GE}(e)$ se puede calcular su función $\lambda_E(e)$ utilizando

(2.7) como sigue: $\lambda_E(e) = \lambda_{GE}(e)$

Para $E = e_1$ se tiene: $\lambda_E(e_1) = \lambda_{GE}(e_1) \Rightarrow \lambda_E(e_1) = 1$

Para $E = e_2$ se tiene: $\lambda_E(e_2) = \lambda_{GE}(e_2) \Rightarrow \lambda_E(e_2) = 1$

Dado que para el nodo E ya se ha calculado la función $\lambda_E(e)$ se puede enviar el mensaje $\lambda_{EB}(b)$ utilizando (2.12) se tiene:

$$\lambda_{EB}(b) = \sum_e \lambda_E(e)P(e/b)$$

Para $E = e_1$ se tiene:

$$\lambda_{EB}(b_1) = \lambda_E(e_1)P(e_1/b_1) + \lambda_E(e_2)P(e_2/b_1) = 1(0.25) + 1(0.75) \Rightarrow \lambda_{EB}(b_1) = 1$$

Para $E = e_2$ se tiene:

$$\lambda_{EB}(b_2) = \lambda_E(e_1)P(e_1/b_2) + \lambda_E(e_2)P(e_2/b_2) = 1(0.45) + 1(0.55) \Rightarrow \lambda_{EB}(b_2) = 1$$

6. El nodo F.

Este nodo ha recibido el mensaje $\rho_{DF}(d)$ de su padre D, por lo que se puede calcular la función $\rho_F(f)$ utilizando 2.5, como sigue:

$$\rho_F(f) = \sum_d P(f/d)\rho_{DF}(d)$$

Para $F = f_1$ se tiene:

$$\rho_F(f_1) = P(f_1/d_1)\rho_{DF}(d_1) + P(f_1/d_2)\rho_{DF}(d_2) \Rightarrow \rho_F(f_1) = 0.9(1) + 0.25(0)$$

$$\rho_F(f_1) = 0.9$$

Para $F = f_2$ se tiene:

$$\rho_F(f_2) = P(f_2/d_1)\rho_{DF}(d_1) + P(f_2/d_2)\rho_{DF}(d_2) \Rightarrow \rho_F(f_2) = 0.1(1) + 0.75(0)$$

$$\rho_F(f_2) = 0.10$$

7. El nodo G.

Este nodo ha recibido el mensaje $\rho_{EG}(e)$ de su padre E, por lo que se puede calcular su función $\rho_G(g)$ utilizando (2.5).

Para $G = g_1$ se tiene:

$$\rho_G(g_1) = P(g_1/e_1)\rho_{EG}(e_1) + P(g_1/e_2)\rho_{EG}(e_2) = 0.1(0.1714725) + 0.3(0.2442675)$$

$$\rho_G(g_1) = 0.0904275$$

Para $G = g_2$ se tiene:

$$\rho_G(g_1) = P(g_2 / e_1)\rho_{EG}(e_1) + P(g_2 / e_2)\rho_{EG}(e_2) = 0.9(0.1714725) + 0.7(0.2442675)$$

$$\rho_G(g_1) = 0.3253125$$



El único mensaje que queda por enviarse es el del nodo B al A, por lo que en esta etapa se empezará con el nodo B.

1. Nodo B.

Este nodo ya ha recibido los mensajes de sus dos hijos D y E, por lo que se calcula su función $\lambda_B(b)$ utilizando (2.7) como sigue:

$$\text{Para } B = b_1 \text{ se tiene: } \lambda_B(b_1) = \lambda_{DB}(b_1)\lambda_{EB}(b_1) = 0.3(1) \Rightarrow \lambda_B(b_1) = 0.3$$

$$\text{Para } B = b_2 \text{ se tiene: } \lambda_B(b_2) = \lambda_{DB}(b_2)\lambda_{EB}(b_2) = 0.65(1) \Rightarrow \lambda_B(b_2) = 0.65$$

Puesto que el nodo B ha calculado su función $\lambda_B(b)$ se puede enviar el mensaje $\lambda_{BA}(a)$ a su padre A, utilizando (2.12) se tiene:

$$\lambda_{BA}(a) = \sum_b \lambda_B(b)P(b/a)$$

Para $A = a_1$ se tiene:

$$\lambda_{BA}(a_1) = \lambda_B(b_1)P(b_1/a_1) + \lambda_B(b_2)P(b_2/a_1) = 0.3(0.2) + 0.65(0.8)$$

$$\lambda_{BA}(a_1) = 0.58$$

Para $A = a_2$ se tiene:

$$\lambda_{BA}(a_2) = \lambda_B(b_1)P(b_1/a_2) + \lambda_B(b_2)P(b_2/a_2) = 0.3(0.4) + 0.65(0.6)$$

$$\lambda_{BA}(a_2) = 0.51$$

2. Nodo A.

Este nodo ya ha recibido los mensajes de todos sus hijos por lo que se puede calcular la función $\lambda_A(a)$ como sigue:

$$\lambda_A(a) = \lambda_{BA}(a)\lambda_{CA}(a)$$

Para $A = a_1$ se tiene: $\lambda_A(a_1) = \lambda_{BA}(a_1)\lambda_{CA}(a_1) = 0.58(0.85) \Rightarrow \lambda_A(a_1) = 0.493$

Para $A = a_2$ se tiene: $\lambda_A(a_2) = \lambda_{BA}(a_2)\lambda_{CA}(a_2) = 0.51(0.75) \Rightarrow \lambda_A(a_2) = 0.3825$

En la tabla 2.1 se encuentran los valores de las funciones $\rho_X(x)$ y $\lambda_X(x)$ por cada uno de los nodos y sus probabilidades condicionadas a la evidencia (E 2.1).

Tabla 2.1 Funciones $\rho_X(X = x)$ y $\lambda_X(X = x)$

Nodo A				Nodo B			
A	$\lambda_A(a)$	$\rho_A(a)$	$P(A = a / C = c_2, D = d_1)$	B	$\lambda_B(b)$	$\rho_B(b)$	$P(B = b / C = c_2, D = d_1)$
a_1	0.493	0.3	0.35582822	b_1	0.3	0.261	0.188379646
a_2	0.3825	0.7	0.644171779	b_2	0.65	0.519	0.811620353

Nodo C				Nodo D			
C	$\lambda_C(c)$	$\rho_C(c)$	$P(C = c)$	D	$\lambda_D(d)$	$\rho_D(d)$	$P(D = d)$
c_1	0	0	0	d_1	1	1	1
c_2	1	1	1	d_2	0	0	0

Nodo E				Nodo F			
E	$\lambda_E(e)$	$\rho_E(e)$	$P(E = e / C = c_2, D = d_1)$	F	$\lambda_F(f)$	$\rho_F(f)$	$P(F = f / C = c_2, D = d_1)$
e_1	1	0.1714725	0.412342686	f_1	1	0.9	0.9
e_2	1	0.2443775	0.587658515	f_2	1	0.1	0.1

Nodo G			
G	$\lambda_G(g)$	$\rho_G(g)$	$P(G = g / C = c_2, D = d_1)$
g_1	1	0.0904275	0.217509741
g_2	1	0.3253125	0.782490258

Comentario:

Observemos que para calcular las probabilidades condicionales que aparecen en la tabla 2.1, cada una de las variables X debe haber recibido todos los mensajes λ_{hX} de parte de sus hijos h y todos los mensajes ρ_{PX} de parte de sus padres P , para con ellos calcular los valores λ_X y ρ_X que se utilizarán para el cálculo de la probabilidad condicional que se desea y también se utilizarán para que la variable X pueda

calcular los mensajes que deberá enviar tanto a sus hijos como a sus padres para que estos puedan calcular las probabilidades condicionales.

2.2.2 Método de condicionamiento.

Como se menciona al principio el método descrito en la sección anterior, sólo funciona para redes simplemente conectadas, ya que como hemos visto la evidencia observada se divide en dos subconjuntos, tomando como referencia un nodo X_i que hace que la evidencia esté d-separada dado ese nodo, esto es lo que no ocurre si tuviéramos una red con bucle, por ejemplo en la figura 2.6 si consideramos el conjunto evidencial $e = \{X_1, X_5\}$, el nodo X_2 no d-separa la evidencia observada ya que el nodo X_3 no es un nodo de aristas convergentes y no esta contenido en el conjunto formado por X_2 ⁹, por lo que habría que agregar el nodo X_3 al conjunto formado por sólo X_2 para d-separar a la evidencia en este caso particular.

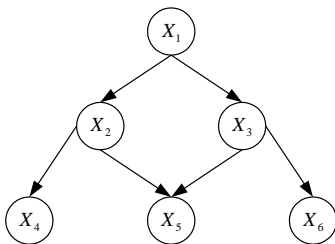


Figura 2.6

Supongamos que no sabemos que el método para poliárboles descrito anteriormente se aplica sólo a redes simplemente conectadas, y lo aplicamos al grafo de la figura 2.6 (supondremos que no disponemos de evidencia), los primeros mensajes que enviaríamos serían $\lambda_{X_4 X_2}(X_2)$ y $\lambda_{X_6 X_3}(X_3)$ ya que los nodos X_4 y X_6 sólo tienen un padre.

El nodo X_1 no puede enviar ningún mensaje a ninguno de sus hijos, ya que ninguno de estos le ha enviado uno, el nodo X_2 no puede enviar ningún mensaje a sus hijos ya que éste no ha recibido ningún mensaje de su padre y tampoco le puede enviar un mensaje a su único padre, ya que no ha recibido todos los mensajes de parte de sus hijos. La situación del nodo X_3 es similar a la del nodo X_2 , el nodo X_5 no puede enviar ningún mensaje a ninguno de sus padres, ya que no ha recibido ningún mensaje de parte de sus padres. Por lo que como podemos observar el método de poliárboles

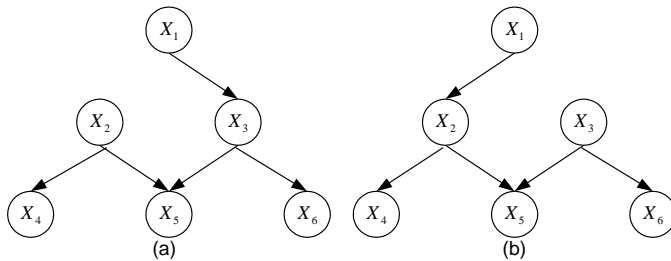


Figura 2.7

queda estancado, ya que no se pueden enviar los mensajes necesarios para calcular las distribuciones de probabilidad marginal de los nodos.

Pearl [1986] propuso una solución a este problema, la cual consiste en instanciar un nodo

⁹ Consideramos que el nodo X_2 forma parte de un conjunto de un sólo elemento, por lo que el nodo X_3 no esta en este conjunto.

de cada ciclo¹⁰ de la red y luego borrar uno de los enlaces con sus hijos, de tal forma de convertirla en un poliárbol y sustituir sus distribuciones de probabilidad condicional, por distribuciones de probabilidad marginal que en realidad serán distribuciones de probabilidad condicional condicionadas a un valor particular del nodo de corte, por ejemplo si en la red de la figura 2.6¹¹ instanciamos el nodo $X_1 = 0$ tenemos dos casos los cuales se muestran en la figura 2.7, en la parte (a) de esta figura se ha eliminado el enlace $X_1 \rightarrow X_2$ y en la parte (b) se ha eliminado el enlace $X_1 \rightarrow X_3$, observemos que cada una de estas redes es un poliárbol a las cuales se les puede aplicar el método descrito en la sección 2.2.1, por ejemplo si escogemos la red de la parte (a) de la figura 2.7 para realizar los cálculos de las probabilidades, lo que seguiría sería sustituir $P(X_2 / X_1)$ por $P_2(X_2) = P(X_2 / X_1 = x_1)$.

Para la obtención del conjunto de corte se utilizará la siguiente regla:

Regla para la obtención del conjunto de corte: El conjunto de corte de ciclo debe de contener al menos un nodo de cada ciclo de la red bayesiana y tal que el nodo elegido, para un ciclo sea hijo, a lo sumo de un nodo del ciclo.

En general el método de condicionamiento consiste en que una vez seleccionado el conjunto de corte $C = \{C_1, C_2, \dots, C_k\}$, se sustituye la distribución de probabilidad condicional $P(X_i / \pi_i)$ de uno de los hijos de cada nodo del conjunto de corte, por la distribución $P_1(X_i / \pi_i \setminus C_j, C_j = c_j)$ ¹², con el objetivo de romper el bucle, luego para cada realización del conjunto de corte y la evidencia disponible $e = \{e_1, \dots, e_v\}$ se utiliza la siguiente ecuación:

$$P(x_i / e) = \frac{P(x_i, e)}{P(e)} = \sum_{c_1} \sum_{c_2} \dots \sum_{c_k} \frac{P(x_i, e, c_1, \dots, c_k)}{P(e)} = \sum_c \frac{P(x_i, e, c)}{P(e)} = \sum_c \frac{P(x_i, e, c)P(e, c)}{P(e, c)P(e)}$$

¹⁰ Al conjunto de nodos que son instanciados en una red con bucles, para romper los bucles se les conoce como conjunto de corte, que en nuestro caso lo denotaremos como $C = \{C_1, C_2, \dots, C_k\}$.

¹¹ Consideremos que las variables de la red son binarias con los valores: $X_i = 1$ o $X_i = 0$

¹² Suponiendo que uno de los padres de nodo X_i forma parte del conjunto de corte y se elimina el enlace $C_j \rightarrow X_i$ entonces esta probabilidad representa la probabilidad condicional del nodo X_i dados sus padres que no forman parte del conjunto de corte, para un valor particular del padre C_j que forma parte del conjunto de corte.

$$P(x_i/e) = \sum_c P(x_i/e,c)P(c/e) \quad (2.14)$$

Como se puede observar en la ecuación (2.14) hay que calcular $P(x_i/e,c)$, esta expresión se puede calcular utilizando el método de propagación en poliárboles visto en la sección 2.2.1, para cada realización del conjunto de corte.

La expresión $P(c/e)$ se calcula suponiendo que el conjunto evidencial esta ordenado $\sigma_E = \{e_1, \dots, e_v\}$, $\sigma_E(i) = e_i$ y $E_{i-1} = (e_1, \dots, e_{i-1})$ utilizando la siguiente expresión:

$$P(c/E_i) = \frac{P(c, E_i)}{P(E_i)}$$

$$P(c/E_i) = \frac{P(c, e_1, \dots, e_i)}{P(e_1, \dots, e_i)}$$

$$P(c/E_i) = \frac{P(c, e_1, \dots, e_{i-1}, \sigma_E(i))}{P(e_1, \dots, e_{i-1})P(e_i/e_1, \dots, e_{i-1})}$$

$$P(c/E_i) = \frac{P(c, E_{i-1}, \sigma_E(i))}{P(E_{i-1})P(\sigma_E(i)/E_{i-1})}$$

$$P(c/E_i) = \frac{P(\sigma_E(i)/c, E_{i-1})P(c, E_{i-1})}{P(\sigma_E(i)/E_{i-1})P(E_{i-1})}$$

$$P(c/E_i) = \alpha P(\sigma_E(i)/c, E_{i-1})P(c/E_{i-1}) \quad (2.15)$$

Donde:

$$\alpha = \frac{1}{P(\sigma_E(i)/E_{i-1})}$$

En la expresión (2.15) se debe aplicar el método de propagación para poliárboles, para el cálculo de $P(\sigma_E(i)/c, E_{i-1})$ y para $i=1$ se tiene que calcular $P(c)$, ésta se calcula utilizando el siguiente algoritmo de Suermondt y Cooper¹³:

Cálculo de la probabilidad Marginal para las variables de corte.

Se denotará por C ¹⁴ al conjunto de variables de corte en una red causal.

1. Realizar una ordenación σ_n ¹⁵ de todos los nodos del grafo.

Esta ordenación se basa en el siguiente criterio: dado un nodo i , entonces ninguno de los ascendientes de i , en el grafo es posterior a i en ese orden. Es decir $i < j \Leftrightarrow i$ no es descendiente de j .

2. Cálculo de $P(C_{\sigma_k(i)} / C_{i-1})$

Para simplificar y sin pérdida de generalidad supongamos que la ordenación obtenida sobre el conjunto de variables es: $X = \{X_1, X_2, \dots, X_n\}$. Establecido dicho orden, este establece una ordenación en el conjunto C , y supongamos que es $C_{\sigma_k} = \{C_1, C_2, \dots, C_k\}$. Denotemos por S_i al subconjunto de X formado por los nodos comprendidos entre C_i y C_{i+1} , es decir:

$$S_0 = \{X_1, \dots, X_p\} \quad p = \text{Nodo anterior a } C_1 \text{ según } \sigma_n$$

$$S_i = \{C_i, \dots, C_{i+1}\} \quad \forall \quad 1 \leq i \leq k-1$$

$$S_k = \{X_t, X_{t+1}, \dots, X_{n-1}, X_n\} \quad X_t \text{ Nodo siguiente a } C_k \text{ según } \sigma_n$$

Definidos estos dos conjuntos realizamos los siguientes dos pasos:

a. Cálculo de las probabilidades marginales de los nodos de S_0

Para cada nodo de $i \in S_0$ en el orden que establece la ordenación σ_n hacer:

¹³ Tomado de la tesis doctoral “Diseño y validación de nuevos algoritmos para el tratamiento de grafos de dependencia”, autor Luís Daniel Hernández Molinero, paginas 78-79.

¹⁴ $C_{\sigma_k} = \{C_1, C_2, \dots, C_k\}$

¹⁵ $\sigma_n(i)$ denotará a la variable i-ésima en la ordenación es decir: $\sigma_n(i) = X_i$

$$P(x_i) = \sum_{\pi_i} P(x_i / \pi_i) \left(\prod_{j \in \pi_i} P(x_j) \right) \quad (2.16)$$

Nótese que esta expresión sólo es válida si los elementos de π_i son marginalmente independientes, que se podrá garantizar si el nodo i no tiene predecesores que sean miembros del conjunto de corte, ya que estos entre si podrían formar un bucle y la fórmula no sería aplicable. Como S_0 no tiene elementos de conjunto de corte, puede aplicarse dicha expresión. En el caso particular de que el nodo i no tenga nodos padres la expresión anterior se reduce a la distribución a priori $P(x_i)$.

b. Cálculo de la probabilidad condicional $P(X_{\sigma_{ck}(i)} / C_{i-1})$ ¹⁶ para cada uno de los nodos de corte.

Si siguiendo el orden de σ_n llegaremos al primer nodo del conjunto corte C_1 . Llegado a éste, si instanciamos C_1 a un valor, los sucesores de C_1 llegan a ser marginalmente independientes y podremos conocer las creencias a priori de los nodos comprendidos entre C_1 y C_2 . En efecto, puesto que estamos realizando una instanciación, podemos interpretar ésta como la captación de una evidencia $e = C_1 = c_1$ sobre el grafo dirigido acíclico (G.D.A.) definido por el subgrafo $S_0 \cup S_1 = \{X_1, X_2, \dots, C_1, \dots, C_2\}$. En este subgrafo podemos aplicar el método para poliárboles, con lo que obtendremos las distribuciones de probabilidad condicionadas de las variables asociadas a $S_0 \cup S_1$ al valor instanciado de C_1 ; y en particular conoceremos la creencia de C_2 ($P(C_2 = c_2 / C_1 = c_1)$). Al llegar en el orden a C_2 realizamos el mismo proceso ya que sus hijos son marginalmente independientes y al ser el subgrafo $S_0 \cup S_1 \cup S_2 = \{X_1, X_2, \dots, C_3\}$ un G.D.A. podremos calcular $P(C_3 = c_3 / C_1 = c_1, C_2 = c_2)$. Este proceso se repite hasta terminar de recorrer los nodos del grafo en el orden marcado por σ_n .

¹⁶ Donde $C_{i-1} = \{C_{\sigma_n(i)}, \dots, C_{\sigma_n(i-1)}\}$ y $\sigma_{ck}(i) = c_i$ si $i = 0$ entonces $P(X_{\sigma_{ck}(i)} / C_{i-1}) = P(X_{\sigma_{ck}(i)})$

3. Calcular la probabilidad marginal $P(C)$

Para el cálculo de esta probabilidad se utiliza la ecuación:

$$\begin{aligned}
 P(C) &= P(C_1, \dots, C_k) = P(C_k / C_1, \dots, C_{k-1}) P(C_{k-1} / C_1, \dots, C_{k-2}) \dots P(C_3 / C_1, C_2) P(C_2 / C_1) P(C_1) \\
 P(C) &= P(C_1, \dots, C_k) = P(C_k / C_1, \dots, C_{k-1}) P(C_{k-1} / C_1, \dots, C_{k-2}) \dots P(C_3 / C_1, C_2) \prod_{i=1}^2 P(C_{\sigma_k(i)} / C_{i-1}) \\
 P(C) &= P(C_1, \dots, C_k) = P(C_k / C_1, \dots, C_{k-1}) P(C_{k-1} / C_1, \dots, C_{k-2}) \dots \prod_{i=1}^3 P(C_{\sigma_k(i)} / C_{i-1}) \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad \cdot \\
 P(C) &= P(C_1, \dots, C_k) = P(C_k / C_1, \dots, C_{k-1}) \prod_{i=1}^{k-1} P(C_{\sigma_k(i)} / C_{i-1}) \\
 P(C) &= \prod_{i=1}^k P(C_{\sigma_{ck}(i)} / C_{i-1}) \tag{2.17}
 \end{aligned}$$

En este algoritmo en la última etapa del literal 2 calcularemos $P(x_i / e, c)$ ya que sólo el cálculo de $P(x_i / c)$ no nos sirve de nada¹⁷, por lo que después de aplicar el algoritmo anterior lo único que queda por calcular es $P(\sigma_E(i) / c, E_{i-1})$, la cual se obtiene aplicando el algoritmo para poliárboles, para cada valor de la evidencia y del conjunto de corte.

Ejemplo 2.2: En la figura 2.8(a) se presenta una red bayesiana de 11 nodos binarios con 3 bucles, suponemos que conocemos con certeza que $X_1 = 0$ y $X_6 = 1$ que será nuestro conjunto e . Lo que queremos es conocer $P(X_i = x_i / e) = P(X_i = x_i / X_1 = 0, X_6 = 1)$ para todos los demás nodos de la red. Puesto que se trata de una red con bucles lo primero que hacemos es determinar el conjunto de corte aplicando la regla ya mencionada como sigue: Para el bucle $X_2 - X_4 - X_7 - X_{10} - X_6 - X_3 - X_2$ según la regla se puede seleccionar cualquier nodo del bucle para formar parte del conjunto de corte excepto el

¹⁷ Si no disponemos de evidencia se calculará $P(x_i / c)$ para obtener las probabilidades marginales de los nodos de la red.

nodo X_{10} el cual además de tener dos padres, si tuviera descendientes estos no pertenecerían al bucle, por lo que no se rompería el bucle si condicionamos la red a ese nodo, en este caso elegiremos arbitrariamente el nodo X_7 para formar parte del conjunto de corte.

En este bucle $X_3 - X_6 - X_9 - X_5 - X_3$ se puede elegir cualquier nodo para formar parte del conjunto de corte, excepto X_9 por lo que escogemos el nodo X_3 . Finalmente en el bucle $X_4 - X_8 - X_{11} - X_7 - X_4$ se puede elegir cualquier nodo para formar parte del conjunto de corte excepto el nodo X_{11} por lo que escogemos el nodo X_4 .

Por lo que el conjunto de corte elegido es: $C_{\sigma_3} = \{X_3, X_4, X_7\}$

La factorización de la probabilidad conjunta de las variables según la figura 2.8 (a) es:

$$P(X) = P(x_1)P(x_2/x_1)P(x_3/x_2)P(x_4/x_2)P(x_5/x_3)P(x_6/x_3)P(x_7/x_4)P(x_8/x_4)P(x_9/x_5, x_6)P(x_{10}/x_6, x_7)P(x_{11}/x_7, x_8)$$

Instanciando las variables del conjunto de corte, por un valor particular y eliminando los enlaces $X_3 \rightarrow X_5$, $X_4 \rightarrow X_8$, $X_7 \rightarrow X_{10}$ para convertir la red en un poliárbol, esto se ilustra en la parte (b) de la figura 2.8 y sustituimos las probabilidades condicionales de cada nodo del conjunto $\{X_5, X_{10}, X_8\}$, por:

$$P_1(X_5) = P(X_5/X_3 = x_3), \quad P_1(X_{10}/X_6) = P(X_{10}/X_6, X_7 = x_7), \quad P_1(X_8) = P(X_8/X_4 = x_4).$$

Por lo que la factorización de probabilidad conjunta de las variables de la red queda como:

$$P(X) = P(x_1)P(x_2/x_1)P(x_3/x_2)P(x_4/x_2)P_1(X_5), P(x_6/x_3)P(x_7/x_4)P_1(X_8)P(x_9/x_5, x_6)P_1(X_{10}/X_6), P(x_{11}/x_7, x_8)$$

Figura 2.8: Red bayesiana de 11 nodos, con tres bucles.

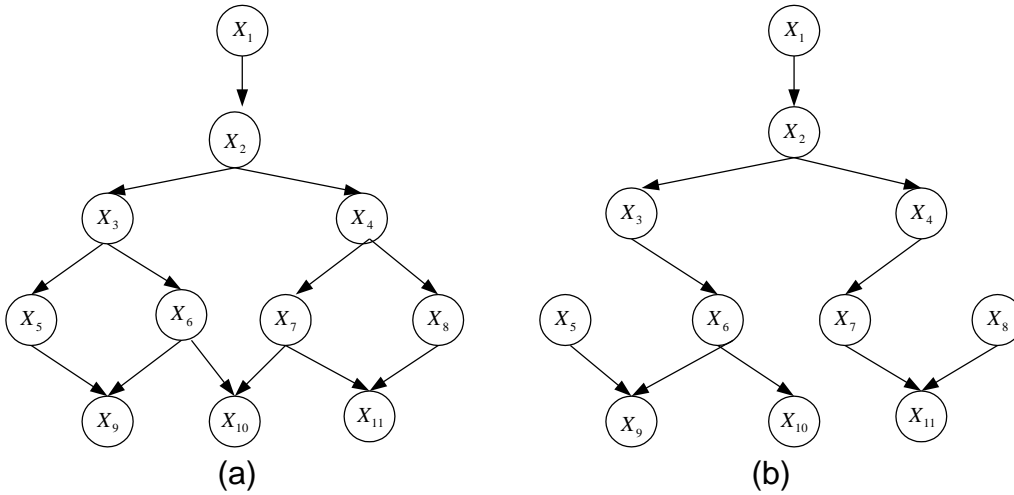


Tabla 2.2: Probabilidades de los nodos de la red 2.8 (a)

X_1	$P(X_1)$
0	0.35
1	0.65

X_1	X_2	$P(X_2 / X_1)$
0	0	0.3
0	1	0.7
1	0	0.4
1	1	0.6

X_2	X_3	$P(X_3 / X_2)$
0	0	0.13
0	1	0.87
1	0	0.2
1	1	0.8

X_2	X_4	$P(X_4 / X_2)$
0	0	0.35
0	1	0.65
1	0	0.25
1	1	0.75

X_3	X_5	$P(X_5 / X_3)$
0	0	0.5
0	1	0.5
1	0	0.75
1	1	0.25

X_3	X_6	$P(X_6 / X_3)$
0	0	0.6
0	1	0.4
1	0	0.8
1	1	0.2

X_4	X_7	$P(X_7 / X_4)$
0	0	0.4
0	1	0.6
1	0	0.3
1	1	0.7

X_4	X_8	$P(X_8 / X_4)$
0	0	0.9
0	1	0.1
1	0	0.65
1	1	0.35

X_5	X_6	$P(X_9 / X_5, X_6)$
0	0	0.55
0	0	0.45
0	1	0.15
0	1	0.85
1	0	0.45
1	0	0.55
1	1	0.56
1	1	0.44

X_6	X_7	X_{10}	$P(X_{10} / X_6, X_7)$
0	0	0	0.85
0	0	1	0.15
0	1	0	0.10
0	1	1	0.9
1	0	0	0.08
1	0	1	0.92
1	1	0	0.01
1	1	1	0.99

X_7	X_8	X_{11}	$P(X_{11} / X_7, X_8)$
0	0	0	0.95
0	0	1	0.05
0	1	0	0.5
0	1	1	0.5
1	0	0	0.5
1	0	1	0.5
1	1	0	0
1	1	1	1

Se empezará calculando la probabilidad conjunta del conjunto de corte aplicando el algoritmo descrito anteriormente, como sigue:

1. Puesto que las variables del grafo así como el conjunto de corte ya están ordenados esta etapa esta completa.
2. Se realizará el cálculo de $P(C_{\sigma_3(i)} / C_{i-1})$

De acuerdo al algoritmo se forman los siguientes conjuntos:

$S_0 = \{X_1, X_2\}$ En este conjunto están los nodos que no pertenecen a ningún bucle de la red

$S_1 = \{X_3, X_4\}$ En este conjunto están los dos primeros nodos del conjunto de corte.

$S_2 = \{X_4, X_5, X_6, X_7\}$ En este conjunto están el segundo y tercer nodo del conjunto de corte y los nodos comprendidos entre ellos que no forman parte del conjunto de corte.

$S_3 = \{X_8, X_9, X_{10}, X_{11}\}$ En este conjunto están comprendidos los nodos que están después del último nodo del conjunto de corte según la ordenación.

Estos conjuntos se forman con el objetivo de ir obteniendo las probabilidades condicionales de un nodo del conjunto de corte condicionado a uno o más nodos del conjunto de corte, por ejemplo con S_1 se calcula $P(X_4 = x_4 / X_3 = x_3)$, al agregar S_2 se calcula $P(X_7 / X_3 = x_3, X_4 = x_4)$ y finalmente al agregar S_3 se obtiene $P(X_i / C = c, X_1 = 0, X_6 = 1)$

- a. Cálculo de las probabilidades marginales del conjunto S_0

Las cuales se calculan utilizando (2.16) se obtiene: $(P(X_1 = 0), P(X_1 = 1)) = (0.35, 0.65)$ estas probabilidades se obtiene directamente de la tabla 2.2.

Para $P(X_2 = x_2)$ se realiza los siguientes cálculos:

$$P(X_2 = 0) = \sum_{x_1=0}^{x_1=1} P(X_2 = 0 / X_1) (P(X_1))$$

$$P(X_2 = 0) = P(X_2 = 0 / X_1 = 0)P(X_1 = 0) + P(X_2 = 0 / X_1 = 1)P(X_1 = 1)$$

$$P(X_2 = 0) = 0.3(0.35) + 0.4(0.65)$$

$$P(X_2 = 0) = 0.365$$

Similarmente se calcula para $X_2 = 1$ obteniéndose:

$$(P(X_2 = 0), P(X_2 = 1)) = (0.365, 0.635)$$

- b. Cálculo de la probabilidad condicional $P(X_{\sigma_{ck}(i)} / C_{i-1})$ para cada uno de los nodos del conjunto de corte.

Empezando con el grafo formado por $S_1 \cup S_2 = \{X_1, X_2, X_3, X_4\}$, se obtiene el poliárbol mostrado en la figura 2.9, tomando como evidencia a X_3 para cada valor que toma y aplicando el algoritmo para poliárboles se obtiene¹⁸:

Para $X_3 = 0$:

Para el nodo evidencial se tiene:

$$\rho_{X_3}(X_3) = \begin{cases} 1; & \text{si } X_3 = 0 \\ 0; & \text{si } X_3 = 1 \end{cases} \quad \text{y} \quad \lambda_{X_3}(X_3) = \begin{cases} 1; & \text{si } X_3 = 0 \\ 0; & \text{si } X_3 = 1 \end{cases}$$

Para el nodo sin padre:

$$\rho_{X_1}(X_1) = P(X_1 = x_1) = \begin{cases} 0.35; & \text{si } X_1 = 0 \\ 0.65; & \text{si } X_1 = 1 \end{cases}$$

Para el nodo sin hijo se tiene:

$$\lambda_{X_3}(X_3) = \begin{cases} 1; & \text{si } X_4 = 0 \\ 1; & \text{si } X_4 = 1 \end{cases}$$

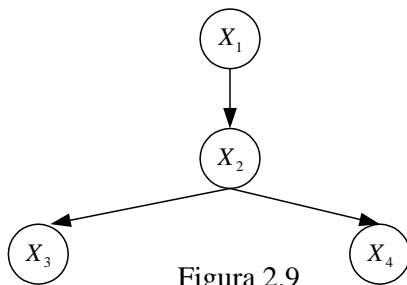


Figura 2.9

Enviando el mensaje de X_3 a X_2 como sigue:

$$\lambda_{X_3, X_2}(X_2) = \sum_{X_3=0}^1 \lambda_{X_3}(X_2) P(X_3 / X_2)$$

¹⁸ Sólo se mostraran las probabilidades $P(C_i = c_i / C_k = c_k)$ que son las que por el momento nos interesan, y al final del algoritmo se mostraran las probabilidades $P(X_i = x_i / C = c, e)$.

$$\lambda_{X_3X_2}(X_2 = 0) = 1(0.13) + 0(0.87) \Rightarrow \lambda_{X_3X_2}(X_2 = 0) = 0.13$$

$$\lambda_{X_3X_2}(X_2 = 1) = 1(0.20) + 0(0.80) \Rightarrow \lambda_{X_3X_2}(X_2 = 1) = 0.2$$

Enviando el mensaje de X_1 a X_2 como sigue:

$$\rho_{X_1X_2}(X_1) = \rho_{X_1}(X_1)$$

$$\rho_{X_1X_2}(X_1 = 0) = 0.35$$

$$\rho_{X_1X_2}(X_1 = 1) = 0.65$$

Enviando el mensaje de X_1 a X_2 como sigue:

$$\rho_{X_1X_2}(X_1) = \rho_{X_1}(X_1)$$

$$\rho_{X_1X_2}(X_1 = 0) = 0.35$$

$$\rho_{X_1X_2}(X_1 = 1) = 0.65$$

Calculando $\rho_{X_2}(X_2)$ como sigue:

$$\rho_{X_2}(X_2) = \sum_{X_1=0}^1 P(X_2 / X_1) \rho_{X_1X_2}(X_1)$$

$$\rho_{X_2}(X_2 = 0) = 0.3(0.35) + 0.4(0.65) \Rightarrow \rho_{X_2}(X_2 = 0) = 0.365$$

$$\rho_{X_2}(X_2 = 1) = 0.7(0.35) + 0.6(0.65) \Rightarrow \rho_{X_2}(X_2 = 1) = 0.635$$

Enviando el mensaje de X_2 a X_4 como sigue:

$$\rho_{X_2}(X_2) = \rho_{X_2}(X_2) \lambda_{X_3X_2}(X_3)$$

$$\rho_{X_2}(X_2 = 0) = 0.365(0.13) \Rightarrow \rho_{X_2}(X_2 = 0) = 0.0475$$

$$\rho_{X_2}(X_2 = 1) = 0.365(0.20) \Rightarrow \rho_{X_2}(X_2 = 1) = 0.127$$

Cálculo de $\rho_{X_4}(X_4)$ como sigue:

$$\rho_{X_4}(X_4) = \sum_{X_2=0}^1 P(X_4 / X_2) \rho_{X_2X_4}(X_2)$$

$$\rho_{X_4}(X_4 = 0) = 0.35(0.0475) + 0.25(0.127) \Rightarrow \rho_{X_4}(X_4 = 0) = 0.048375$$

$$\rho_{X_4}(X_4 = 1) = 0.65(0.048375) + 0.75(0.127) \Rightarrow \rho_{X_4}(X_4 = 1) = 0.1236125$$

Calculado $P(X_4 / X_3 = 0)$ como sigue:

$$P(X_4 / X_3 = 0) = \frac{\lambda_{X_4}(X_4)\rho_{X_4}(X_4)}{\sum_{X_4=0}^1 \lambda_{X_4}(X_4)\rho_{X_4}(X_4)}$$

$$P(X_4 = 0 / X_3 = 0) = \frac{1(0.048375)}{1(0.048375) + 1(0.126125)} \Rightarrow P(X_4 = 0 / X_3 = 0) = 0.2772$$

$$P(X_4 = 1 / X_3 = 0) = \frac{1(0.126125)}{1(0.048375) + 1(0.126125)} \Rightarrow P(X_4 = 1 / X_3 = 0) = 0.7228$$

$$(P(X_4 = 0 / X_3 = 0), P(X_4 = 1 / X_3 = 0)) = (0.2772, 0.7228)$$

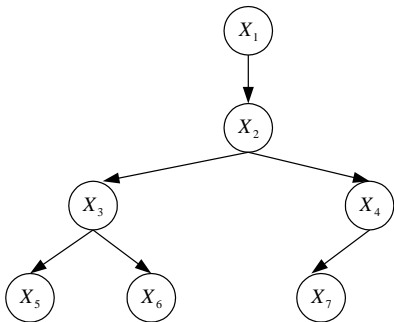


Figura 2.10

De forma similar se calcula para $X_3 = 1$ por lo que se obtiene:

$$(P(X_4 = 0 / X_3 = 1), P(X_4 = 1 / X_3 = 1)) = (0.3481, 0.6519)$$

Cálculo de las probabilidades de los nodos del grafo $S_0 \cup S_1 \cup S_2$:

En este caso se instancia para cada valor que tome el par de nodos X_3 y

X_4 , para calcular $P(X_7 / X_3 = x_3, X_4 = x_4)$ como sigue:

En este caso aplicaríamos el método para poliárboles tomando como evidencia cada realización de los nodos X_3 y X_4 , sin embargo si aplicamos el criterio de d-separación, observamos que se cumple $I(X_7, X_3 / X_4)$ es decir $P(X_7 / X_3, X_4) = P(X_7 / X_4)$ por lo que las probabilidades para cada realización de X_3 y X_4 , se obtienen directamente de la tabla 2.2. Por lo que se obtiene:

Para $X_3 = 0$ y $X_4 = 0$ se tiene:

$$(P(X_7 = 0 / X_3 = 0, X_4 = 0), P(X_7 = 1 / X_3 = 0, X_4 = 0)) = (0.4, 0.6)$$

Para $X_3 = 0$ y $X_4 = 1$ se tiene:

$$(P(X_7 = 0 / X_3 = 0, X_4 = 1), P(X_7 = 1 / X_3 = 0, X_4 = 1)) = (0.3, 0.7)$$

Para $X_3 = 1$ y $X_4 = 0$:

$$(P(X_7 = 0 / X_3 = 1, X_4 = 0), P(X_7 = 1 / X_3 = 1, X_4 = 0)) = (0.4, 0.6)$$

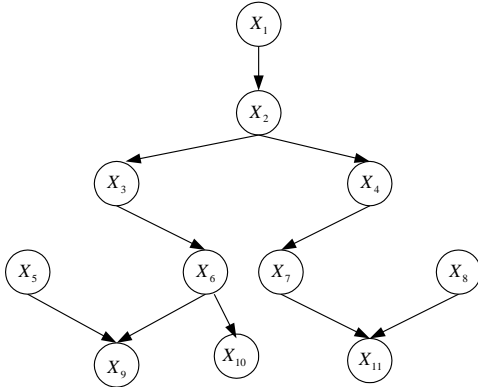


Figura 2.11

Para $X_3 = 1$ y $X_4 = 1$:

$$(P(X_7 = 0 / X_3 = 1, X_4 = 1), P(X_7 = 1 / X_3 = 1, X_4 = 1)) = (0.3, 0.7)$$

Cálculo de las probabilidades de todos los nodos de la red dado el conjunto de corte y la evidencia ($P(X_i / C, e)$):

En esta parte se obtiene $P(X_i / X_1 = 0, X_6 = 1, C)$ por lo que el conjunto evidencial para la red de la figura 2.11 es $C'_i = C \cup e$ se ilustrara el cálculo para el conjunto

$C'_1 = \{X_3 = 0, X_4 = 0, X_7 = 0\} \cup \{X_1 = 0, X_6 = 1\}$ y el nodo X_2 como sigue:

En este caso como podemos observar en la figura 2.11 la familia del nodo X_2 y sus hijos forman parte del conjunto evidencial por lo que para los cálculos sólo se necesita el subgrafo formado por los nodos $\{X_1, X_2, X_3, X_4\}$ aplicamos el algoritmo de propagación para poliárbol como sigue:

Para los nodos evidenciales se tiene:

$$\rho_{X_1}(X_1) = \begin{cases} 1; & X_1 = 0 \\ 0; & X_1 = 1 \end{cases} \text{ y } \lambda_{X_1}(X_1) = \begin{cases} 1; & X_1 = 1 \\ 0; & X_1 = 0 \end{cases}$$

$$\rho_{X_3}(X_3) = \begin{cases} 1; & X_3 = 0 \\ 0; & X_3 = 1 \end{cases} \text{ y } \lambda_{X_3}(X_3) = \begin{cases} 1; & X_3 = 1 \\ 0; & X_3 = 0 \end{cases}$$

$$\rho_{X_4}(X_4) = \begin{cases} 1; & X_4 = 0 \\ 0; & X_4 = 1 \end{cases} \text{ y } \lambda_{X_4}(X_4) = \begin{cases} 1; & X_4 = 1 \\ 0; & X_4 = 0 \end{cases}$$

Enviando el mensaje de X_1 a X_2 como sigue:

$$\rho_{X_1 X_2}(X_1) = \rho_{X_1}(X_1)$$

$$\rho_{X_1 X_2}(X_1) = \begin{cases} 1; & X_1 = 0 \\ 0; & X_1 = 1 \end{cases}$$

Enviando el mensaje de X_3 a X_2 como sigue:

$$\lambda_{X_3 X_2}(X_2) = \sum_{X_3=0}^1 \lambda_{X_3}(X_3)P(X_3 / X_2)$$

$$\lambda_{X_3 X_2}(X_2 = 0) = 1(0.15) + 0(0.87) \Rightarrow \lambda_{X_3 X_2}(X_2 = 0) = 0.13$$

$$\lambda_{X_3 X_2}(X_2 = 1) = 1(0.2) + 0(0.8) \Rightarrow \lambda_{X_3 X_2}(X_2 = 1) = 0.2$$

Enviando el mensaje de X_4 a X_2 como sigue:

$$\lambda_{X_4 X_2}(X_2) = \sum_{X_4=0}^1 \lambda_{X_4}(X_4)P(X_4 / X_2)$$

$$\lambda_{X_4 X_2}(X_2 = 0) = 1(0.35) + 0(0.65) \Rightarrow \lambda_{X_4 X_2}(X_2 = 0) = 0.35$$

$$\lambda_{X_4 X_2}(X_2 = 1) = 1(0.25) + 0(0.75) \Rightarrow \lambda_{X_4 X_2}(X_2 = 1) = 0.25$$

Cálculo de la función $\lambda_{X_2}(X_2)$ como sigue:

$$\lambda_{X_2}(X_2) = \lambda_{X_3 X_2}(X_2)\lambda_{X_4 X_2}(X_2)$$

$$\lambda_{X_2}(X_2 = 0) = (0.13)(0.35) \Rightarrow \lambda_{X_2}(X_2 = 0) = 0.0455$$

$$\lambda_{X_2}(X_2 = 1) = 0.2(0.25) \Rightarrow \lambda_{X_2}(X_2 = 1) = 0.05$$

Cálculo de la función $\rho_{X_2}(X_2)$ como sigue:

$$\rho_{X_2}(X_2) = \sum_{X_1=0}^1 P(X_2 / X_1)\rho_{X_1 X_2}(X_1)$$

$$\rho_{X_2}(X_2 = 0) = 0.3(1) + 0.4(0) \Rightarrow \rho_{X_2}(X_2 = 0) = 0.3$$

$$\rho_{X_2}(X_2 = 1) = 0.7(1) + 0.6(0) \Rightarrow \rho_{X_2}(X_2 = 1) = 0.7$$

Cálculo de la probabilidad $P(X_2 / C'_1)$ como sigue:

$$P(X_2 / C'_1) = \frac{\lambda_{X_2}(X_2)\rho_{X_2}(X_2)}{\sum_{X_2=0}^1 \lambda_{X_2}(X_2)\rho_{X_2}(X_2)}$$

$$P(X_2 = 0 / C'_1) = \frac{0.3(0.0455)}{0.3(0.0455) + 0.05(0.7)} \Rightarrow P(X_2 = 0 / C'_1) = 0.2806$$

$$P(X_2 = 1 / C'_1) = \frac{0.05(0.7)}{0.3(0.0455) + 0.05(0.7)} \Rightarrow P(X_2 = 1 / C'_1) = 0.7194$$

Para el cálculo de la probabilidad $P(X_5 / C'_1)$ puede observarse en la figura 2.11 que se cumple

$$I(X_5, \{X_1, X_2, X_3, X_4, X_6, X_7, X_8, X_{10}, X_{11}\} / \phi)$$

Es decir $P(X_5 / X_1, X_2, X_3, X_4, X_6, X_7, X_8, X_{10}, X_{11}) = P_1(X_5)$ ¹⁹ la cual se obtiene directamente de la tabla 2.2 por lo que da:

$$(P(X_5 = 0 / C'_1), P(X_5 = 1 / C'_1)) = (0.5, 0.5)$$

Similar a las probabilidades anteriores se obtienen las siguientes probabilidades para todos los conjuntos C'_i

$$(P(X_8 = 0 / C'_1), P(X_8 = 1 / C'_1)) = (0.9, 0.1)$$

$$(P(X_9 = 0 / C'_1), P(X_9 = 1 / C'_1)) = (0.355, 0.645)$$

$$(P(X_{10} = 0 / C'_1), P(X_{10} = 1 / C'_1)) = (0.08, 0.92)$$

$$(P(X_{11} = 0 / C'_1), P(X_{11} = 1 / C'_1)) = (0.905, 0.095)$$

¹⁹ recordemos que $P_1(X_5) = P(X_5 / X_3 = x_3)$ en este caso $X_3 = 0$

Tomando como referencia las probabilidades recién calculadas, aquellas que se repitan con respecto a éstas no se escribirán, en las demás realizaciones del conjunto de corte:

Para $C'_2 = \{X_3 = 0, X_4 = 0, X_7 = 1\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_{10} = 0/C'_2), P(X_{10} = 1/C'_2)) = (0.01, 0.99)$$

$$(P(X_{11} = 0/C'_2), P(X_{11} = 1/C'_2)) = (0.45, 0.55)$$

Para $C'_3 = \{X_3 = 0, X_4 = 1, X_7 = 0\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_3), P(X_2 = 1/C'_3)) = (0.1945, 0.8055)$$

$$(P(X_8 = 0/C'_3), P(X_8 = 1/C'_3)) = (0.65, 0.35)$$

$$(P(X_{11} = 0/C'_3), P(X_{11} = 1/C'_3)) = (0.7925, 0.2075)$$

Para $C'_4 = \{X_3 = 1, X_4 = 0, X_7 = 0\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_4), P(X_2 = 1/C'_4)) = (0.3949, 0.6051)$$

$$(P(X_5 = 0/C'_4), P(X_5 = 1/C'_4)) = (0.75, 0.25)$$

$$(P(X_9 = 0/C'_4), P(X_9 = 1/C'_4)) = (0.2525, 0.7475)$$

$$(P(X_{11} = 0/C'_4), P(X_{11} = 1/C'_4)) = (0.905, 0.095)$$

Para $C'_5 = \{X_3 = 0, X_4 = 1, X_7 = 1\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_5), P(X_2 = 1/C'_5)) = (0.1945, 0.8055)$$

$$(P(X_8 = 0/C'_5), P(X_8 = 1/C'_5)) = (0.65, 0.35)$$

$$(P(X_{10} = 0/C'_5), P(X_{10} = 1/C'_5)) = (0.01, 0.99)$$

$$(P(X_{11} = 0/C'_5), P(X_{11} = 1/C'_5)) = (0.325, 0.675)$$

Para $C'_6 = \{X_3 = 1, X_4 = 1, X_7 = 0\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_6), P(X_2 = 1/C'_6)) = (0.2877, 0.7123)$$

$$(P(X_5 = 0/C'_6), P(X_5 = 1/C'_6)) = (0.75, 0.25)$$

$$(P(X_8 = 0/C'_6), P(X_8 = 1/C'_6)) = (0.65, 0.35)$$

$$(P(X_9 = 0/C'_6), P(X_9 = 1/C'_6)) = (0.2525, 0.7475)$$

$$(P(X_{11} = 0/C'_6), P(X_{11} = 1/C'_6)) = (0.7925, 0.2075)$$

Para $C'_7 = \{X_3 = 1, X_4 = 0, X_7 = 1\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_7), P(X_2 = 1/C'_7)) = (0.3949, 0.6051)$$

$$(P(X_5 = 0/C'_7), P(X_5 = 1/C'_7)) = (0.75, 0.25)$$

$$(P(X_9 = 0/C'_7), P(X_9 = 1/C'_7)) = (0.2525, 0.7475)$$

$$(P(X_{10} = 0/C'_7), P(X_{10} = 1/C'_7)) = (0.01, 0.99)$$

$$(P(X_{11} = 0/C'_7), P(X_{11} = 1/C'_7)) = (0.45, 0.55)$$

Para $C'_8 = \{X_3 = 1, X_4 = 1, X_7 = 1\} \cup \{X_1 = 0, X_6 = 1\}$ se tiene:

$$(P(X_2 = 0/C'_8), P(X_2 = 1/C'_8)) = (0.2877, 0.7123)$$

$$(P(X_5 = 0/C'_8), P(X_5 = 1/C'_8)) = (0.75, 0.25)$$

$$(P(X_8 = 0/C'_8), P(X_8 = 1/C'_8)) = (0.65, 0.35)$$

$$(P(X_9 = 0/C'_8), P(X_9 = 1/C'_8)) = (0.2525, 0.7475)$$

$$(P(X_{10} = 0/C'_8), P(X_{10} = 1/C'_8)) = (0.01, 0.99)$$

$$(P(X_{11} = 0/C'_8), P(X_{11} = 1/C'_8)) = (0.325, 0.675)$$

Con las probabilidades obtenidas anteriormente, calcularemos la probabilidad conjunta del conjunto de corte utilizando (2.17), como sigue:

$$P(C) = \prod_{i=1}^3 P(C_{\sigma_{C_i}(i)} / C_{i-1}) \Rightarrow P(C) = P(X_3)P(X_4 / X_3)P(X_7 / X_3, X_4)$$

Por lo que se obtiene la siguiente tabla:

X_3	X_4	X_7	$P(X_3, X_4, X_7)$
0	0	0	0.0193
0	0	1	0.0290
0	1	0	0.0378
1	0	0	0.1149
0	1	1	0.0883
1	1	0	0.1615
1	0	1	0.1724
1	1	1	0.4112

Lo que falta es calcular $P(\sigma_E(i)/c, E_{i-1})$ para $i > 1$ ya que para $i = 1$ se tiene $E_{i-1} = \phi$ por lo que da $P(e_1 / C) = P(X_1 = 0 / C)$. Las probabilidades $P(\sigma_E(i)/c, E_{i-1})$ se calculan utilizando el algoritmo para poliárboles.

Se ilustrara el cálculo de $P(X_1 = 0 / X_3 = 0, X_4 = 0, X_7 = 0)$ como sigue:

Podemos observar en la figura (2.10) que se cumple $I(X_1, X_7 / X_4)$ por lo que se emplea en este caso el método de propagación para poliárboles al subgrafo formado por los nodos $\{X_1, X_2, X_3, X_4\}$ como sigue:

- Las funciones de los nodos evidenciales son:

$$\rho_{X_3}(X_3) = \begin{cases} 1; & X_3 = 0 \\ 0; & X_3 = 1 \end{cases} \text{ y } \lambda_{X_3}(X_3) = \begin{cases} 1; & X_3 = 0 \\ 0; & X_3 = 1 \end{cases}$$

$$\rho_{X_4}(X_4) = \begin{cases} 1; & X_4 = 0 \\ 0; & X_4 = 1 \end{cases} \text{ y } \lambda_{X_4}(X_4) = \begin{cases} 1; & X_4 = 0 \\ 0; & X_4 = 1 \end{cases}$$

- La función para el nodo sin padre es:

$$\rho_{X_1}(X_1) = \begin{cases} 0.35; & X_1 = 0 \\ 0.65; & X_1 = 1 \end{cases}$$

Enviando el mensaje de X_3 a X_2 como sigue:

$$\lambda_{X_3X_2}(X_2 = 0) = 1(0.13) + 0(0.87) \Rightarrow \lambda_{X_3X_2}(X_2 = 0) = 0.13$$

$$\lambda_{X_3X_2}(X_2 = 1) = 1(0.2) + 0(0.8) \Rightarrow \lambda_{X_3X_2}(X_2 = 1) = 0.2$$

Enviando el mensaje de X_4 a X_2 como sigue:

$$\lambda_{X_4X_2}(X_2 = 0) = 1(0.35) + 0(0.65) \Rightarrow \lambda_{X_4X_2}(X_2 = 0) = 0.35$$

$$\lambda_{X_4X_2}(X_2 = 1) = 1(0.25) + 0(0.75) \Rightarrow \lambda_{X_4X_2}(X_2 = 1) = 0.25$$

Calculando $\lambda_{X_2}(X_2)$ como sigue:

$$\lambda_{X_2}(X_2 = 0) = 0.13(0.35) \Rightarrow \lambda_{X_2}(X_2 = 0) = 0.0455$$

$$\lambda_{X_2}(X_2 = 1) = 0.2(0.25) \Rightarrow \lambda_{X_2}(X_2 = 1) = 0.05$$

Enviando el mensaje de X_2 a X_1 como sigue:

$$\lambda_{X_2X_1}(X_1 = 0) = 0.0455(0.3) + 0.05(0.7) \Rightarrow \lambda_{X_2X_1}(X_1 = 0) = 0.04865$$

$$\lambda_{X_2X_1}(X_1 = 1) = 0.0455(0.4) + 0.05(0.6) \Rightarrow \lambda_{X_2X_1}(X_1 = 1) = 0.0482$$

Calculando $\lambda_{X_1}(X_1)$ como sigue:

$$\lambda_{X_1}(X_1) = \lambda_{X_2X_1}(X_1)$$

$$\lambda_{X_1}(X_1 = 0) = 0.04865$$

$$\lambda_{X_1}(X_1 = 1) = 0.0482$$

Calculando $P(X_1 = 0 / X_3 = 0, X_4 = 0, X_7 = 0)$ como sigue:

$$P(X_1 = 0 / X_3 = 0, X_4 = 0, X_7 = 0) = \frac{0.04865(0.35)}{0.04865(0.35) + 0.0482(0.65)} = 0.3521$$

Las demás probabilidades para las restantes 7 realizaciones del conjunto de corte se calculan de forma similar, los cálculos se presentan en la siguiente tabla:

X_3	X_4	X_7	$P(X_1 = 0 / X_3, X_4, X_7)$
0	0	0	0.3521
0	0	1	0.3521
0	1	0	0.3618
0	1	1	0.3618
1	0	1	0.3400
1	1	0	0.3513
1	0	0	0.3400
1	1	1	0.3513

En la siguiente tabla se presenta el cálculo de la probabilidad $P(X_6 = 1 / X_3, X_4, X_7, X_1 = 0)$ estas se han calculado de forma similar a las de la tabla anterior considerando, con un conjunto evidencial $e = \{X_3, X_4, X_7, X_1 = 0\}$ para cada una de las ocho realizaciones del conjunto de corte.

X_3	X_4	X_7	$P(X_6 = 1 / X_3, X_4, X_7, X_1 = 0)$
0	0	0	0.4
0	0	1	0.4
0	1	0	0.4
0	1	1	0.4
1	0	1	0.2
1	1	0	0.2
1	0	0	0.2
1	1	1	0.2

La probabilidad $P(c/e)$ se calcula utilizando (2.15) como sigue:

$$P(X_3, X_4, X_7 / X_1 = 0, X_6 = 1) = \alpha P(X_6 = 1 / X_3, X_4, X_7, X_1 = 0) P(X_1 = 0 / X_3, X_4, X_7) P(X_3, X_4, X_7)$$

Donde:

$$\alpha = \left(\sum_{X_3=0}^1 \sum_{X_4=0}^1 \sum_{X_7=0}^1 P(X_6 = 1 / X_3, X_4, X_7, X_1 = 0) P(X_1 = 0 / X_3, X_4, X_7) P(X_3, X_4, X_7) \right)^{-1}$$

Los resultados para todas las realizaciones del conjunto de corte se dan en la siguiente tabla:

X_3	X_4	X_7	$P(X_3, X_4, X_7 / X_1 = 0, X_6 = 1)$
0	0	0	0.0328
0	0	1	0.0492
0	1	0	0.0659
0	1	1	0.1659
1	0	1	0.0723
1	1	0	0.1491
1	0	0	0.1170
1	1	1	0.3478

En la tabla 2.4, se muestra el cálculo de las probabilidades condicionales de cada nodo dada la evidencia, los cálculos se ha realizado utilizando (2.14):

$$P(X_i = x_i / X_1 = 0, X_6 = 1) = \sum_{X_3=0}^1 \sum_{X_4=0}^1 \sum_{X_7=0}^1 P(X_i = x_i / X_1 = 0, X_6 = 1, X_3, X_4, X_7) P(X_3, X_4, X_7 / X_1 = 0, X_6 = 1)$$

Tabla 2.4: Probabilidades condicionales dada la evidencia obtenida por el método de agrupamiento ($P(X_i / X_1 = 0, X_6 = 1)$)

X_2	$P(X_2 / X_1 = 0, X_6 = 1)$	X_3	$P(X_3 / X_1 = 0, X_6 = 1)$	X_4	$P(X_4 / X_1 = 0, X_6 = 1)$
0	0.28578565	0	0.31386275	0	0.27124032
1	0.71421435	1	0.68613718	1	0.72875968
X_5	$P(X_5 / X_1 = 0, X_6 = 1)$	X_7	$P(X_7 / X_1 = 0, X_6 = 1)$	X_8	$P(X_8 / X_1 = 0, X_6 = 1)$
0	0.67153432	0	0.3647436	0	0.71781009
1	0.32846568	1	0.6352564	1	0.28218991
X_9	$P(X_9 / X_1 = 0, X_6 = 1)$	X_{10}	$P(X_{10} / X_1 = 0, X_6 = 1)$	X_{11}	$P(X_{11} / X_1 = 0, X_6 = 1)$
0	0.267659621	0	0.03553205	0	0.527550546
1	0.732340379	1	0.96446795	1	0.472449454

2.2.3 Método de agrupamiento.

Otro de los métodos que se utiliza para calcular $P(x_i / e)$ en redes bayesianas con bucles, es el método de agrupamiento, el cual consiste en moralizar el grafo dirigido y después triangular el grafo obtenido en caso de que no lo sea. Una forma de triangular el grafo moral es primeramente numerar los nodos de tal forma que si X_i es descendiente de X_j , entonces el número asignado a X_j es menor que el asignado a X_i , posteriormente a la numeración de los nodos, se comprueba si la numeración es perfecta, si durante el proceso en que se esta comprobando, si la numeración es perfecta se llega a un nodo del grafo en que la numeración no es perfecta, se agregan las aristas necesarias para que esta sea perfecta y se vuelve a iniciar el proceso para comprobar si la numeración de los nodos es perfecta, el proceso termina hasta que se ha verificado o logrado que el grafo tenga una numeración perfecta.

Una vez triangulado el grafo moral, identificar los conglomerados (los nodos y sus padres deben de estar dentro de un mismo conglomerado) de éste y numerarlos²⁰. A continuación se aplica la propiedad de intersección dinámica, para obtener una cadena de los conglomerados.

Puesto que se tiene que la distribución conjunta de las variables del grafo esta escrita como:

$$P(X) = \prod_{i=1}^n P(x_i / \pi_i)$$

Se puede obtener una nueva representación de la probabilidad conjunta de las variables, como producto de funciones que dependen de los nodos contenidos en los conglomerados, para hacer esto supondremos que disponemos de m conglomerados y que cada conglomerado C_i tiene N_i nodos por lo que la función del conglomerado C_i se puede escribir como:

$$\psi_i(C_i) = \prod_{x_i \in N_i} P(x_i / \pi_i)$$

²⁰ Puesto que los nodos dentro de los conglomerados se encuentran enumerados, los conglomerados se ordenan de acuerdo al número máximo que se le ha asignado a uno de sus nodos dentro del conglomerado.

A esta función se le denomina función potencial del conglomerado C_i , y sus valores no necesariamente suma uno, como en una distribución de probabilidad; por lo que la distribución de probabilidad conjunta de las variables se puede escribir como:

$$P(X) = \frac{\psi_1(C_1)\psi_2(C_2)\dots\psi_m(C_m)}{\sum_x \prod_{i=1}^m \psi_i(C_i)} \quad (2.18)$$

A la representación de la probabilidad conjunta de las variables dada en (2.18) se le conoce como representación potencial.

En lo que sigue supondremos que no disponemos de evidencia, lo que se quiere es obtener la probabilidad conjunta de cada conglomerado, para luego poder obtener la probabilidad marginal de cada nodo.

Puesto que los conglomerados cumplen la propiedad de intersección dinámica se pueden definir los siguientes conjuntos:

$$S_i = C_i \cap \{C_1 \cup \dots \cup C_{i-1}\} \quad (2.19)$$

$$R_i = C_i \setminus S_i \quad (2.20)$$

Donde:

S_i : Se le conoce como conjunto separador.

R_i : Se le conoce como conjunto residual y contiene las variables que se encuentran en el conglomerado i y que no están en los demás conglomerados.

En los conjuntos definidos anteriormente se busca escribir la función de probabilidad conjunta de las variables como:

$$P(X) = \prod_{i=1}^m P(R_i / S_i) \quad (2.21)$$

Empezaremos calculando $P(R_m / S_m)$ como sigue.

La propiedad de intersección dinámica establece que los conjuntos separadores S_m separan los conjuntos residuales R_m de los conjuntos $(C_1 \cup C_2 \cup \dots \cup C_{m-1} \setminus S_m)$, lo que quiere decir que se cumple:

$$P(R_m / C_1 \cup \dots \cup C_{m-1}) = P(R_m / S_m) \quad (2.22)$$

Por lo que utilizando (2.22) se tiene:

$$P(R_m / S_m) = P(R_m / C_1, \dots, C_{m-1}) = \frac{P(C_1 \cup C_2 \cup \dots \cup C_{m-1}, R_m)}{P(C_1 \cup C_2 \cup \dots \cup C_{m-1})} \quad (2.23)$$

Primero calcularemos el numerador de (2.23) como sigue:

$$P(C_1 \cup C_2 \cup \dots \cup C_{m-1}, R_m) = P(C_1 \cup C_2 \cup \dots \cup C_m) = K^{-1} \prod_{i=1}^m \psi_i(C_i)^{21} \quad (2.24)$$

Calculando el denominador de (2.23) como sigue:

$$P(C_1 \cup C_2 \cup \dots \cup C_{m-1}) = k^{-1} \sum_{R_m} \prod_{i=1}^m \psi_i(C_i) = k^{-1} \prod_{i=1}^{m-1} \psi_i(C_i) \sum_{R_m} \psi_m(C_m) \quad (2.25)$$

Puesto que $\sum_{R_m} \psi_m(C_m)$ sólo dependen del conjunto separador S_m se multiplica el resultado de esta suma, por la función potencial del conglomerado que lo contiene diferente de C_m , por lo que de (2.24) y (2.25) se obtiene:

²¹ $K = \sum_x \prod_{i=1}^m \psi_i(C_i)$

$$P(R_m / S_m) = \frac{k^{-1} \prod_{i=1}^m \psi_i(C_i)}{k^{-1} \prod_{i=1}^{m-1} \psi_i(C_i) \sum_{r_m} \psi_m(C_m)} = \frac{\psi_m(C_m)}{\sum_{r_m} \psi_m(C_m)} \quad (2.26)$$

Similarmente se obtiene $P(R_{m-1} / S_{m-1})$, lo que da:

$$P(R_{m-1} / S_{m-1}) = \frac{\psi_{m-1}(C_{m-1})}{\sum_{r_{m-1}} \psi_{m-1}(C_{m-1})} \quad (2.27)$$

Se sigue así hasta llegar a $P(R_1 / S_1) = P(R_1) = P(C_1)$.

Lo que se va haciendo una vez que se calcula la suma $\sum_{R_i} \psi_i(C_i)$ es multiplicar esta suma por la función potencial del conglomerado que contiene S_i , y a partir de esta nueva función potencial se calcula $P(R_{i-1} / S_{i-1})$ y así sucesivamente hasta llegar S_1 donde $S_1 = \phi$, por lo que la probabilidad del conglomerado C_1 se calcula utilizando la siguiente expresión:

$$P(C_1) = \frac{\psi_1(C_1)}{\sum_{x_i \in N_1} \psi_1(C_1)}$$

Una vez que hemos calculado cada una de las funciones de 2.21 se puede obtener $P(C_i)$ como sigue:

$$P(C_i) = p(r_i, s_i) = P(r_i / s_i) P(s_i) \quad (2.28)$$

El cálculo de la probabilidad conjunta de cada conglomerado se comienza con el segundo conglomerado por ejemplo: $P(C_2) = p(r_2, s_2) = P(r_2 / s_2) P(s_2)$, la función $P(s_2)$ se calcula marginalizando la función de probabilidad conjunta de primer conglomerado, ya que este conglomerado contiene las variables

de s_2 . En general la probabilidad conjunta del conglomerado i se calcula una vez que han sido calculadas las probabilidades conjuntas de los conglomerados anteriores a i , ya que uno de estos contiene las variables s_i y por lo tanto sólo hay que marginalizar su función de probabilidad conjunta para el cálculo de $P(s_i)$.

Una vez estimadas la probabilidad conjunta de cada conglomerado, es decir una vez estimado $P(C_i)$, la probabilidad marginal de cada nodo se calcula, marginalizando la función de probabilidad del conglomerado que lo contiene²².

En el caso de disponer de evidencia, se sustituye en cada función potencial que contiene a los nodos evidenciales por la siguiente función.

$$\psi_i(C_i) = \begin{cases} 0, & \text{si el valor de } c_i \text{ es diferente de } e \\ \psi_i(C_i), & \text{en otro caso} \end{cases} \quad (2.29)$$

Y se repite el procedimiento descrito anteriormente y lo que se obtiene es: $P(x_i / e)$.

Ejemplo 2.3.

Se ilustrara el método utilizando la red que se muestra en la parte (a) de la figura 2.11, las tablas de probabilidades se muestran en la tabla 2.3.

Lo primero que se hace es moralizar el grafo, el cual se muestra en la figura 2.11 (b), los conglomerados de este grafo moralizado son:

$$C_1 = \{X_1, X_2\} \quad C_2 = \{X_2, X_3, X_4\}$$

$C_3 = \{X_3, X_4, X_5\}$. Observemos que el grafo de la figura 2.11 (b) es triangulado y el orden de los conglomerados mencionados anteriormente cumple la propiedad de intercepción dinámica.

²² Puesto que un nodo puede pertenecer a varios conglomerados se escoge aquel que contiene el menor número de nodos para realizar el cálculo.

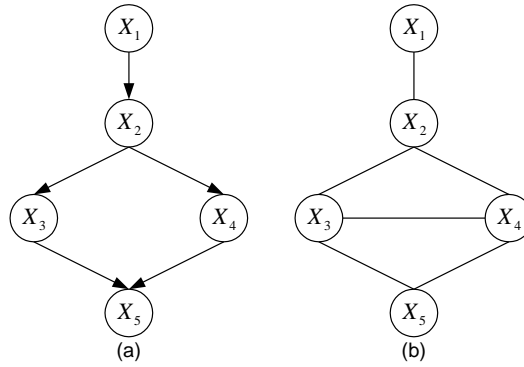


Figura 2.11

Tabla 2.3 Probabilidades de los nodos de la red de la figura 2.11 (a)

X_1	$P(X_1)$
0	0.35
1	0.65

X_1	X_2	$P(X_2 / X_1)$
0	0	0.3
0	1	0.7
1	0	0.4
1	1	0.6

X_2	X_3	$P(X_3 / X_2)$
0	0	0.13
0	1	0.87
1	0	0.2
1	1	0.8

X_2	X_4	$P(X_4 / X_2)$
0	0	0.35
0	1	0.65
1	0	0.25
1	1	0.75

X_3	X_4	X_5	$P(X_5 / X_3, X_4)$
0	0	0	0.55
0	0	1	0.45
0	1	0	0.15
0	1	1	0.85
1	0	0	0.45
1	0	1	0.55
1	1	0	0.56
1	1	1	0.44

Ahora calculamos las funciones potenciales de cada uno de los conglomerados, por ejemplo para el conglomerado C_1 se utiliza la expresión: $\psi_1(X_1, X_2) = P(X_1)P(X_2 / X_1)$.

Para el conglomerado C_2 se utiliza: $\psi_2(X_2, X_3, X_4) = P(X_3 / X_2)P(X_4 / X_2)$, para el conglomerado C_3 se utiliza: $\psi_3(X_3, X_4, X_5) = P(X_5 / X_3, X_4)$, en la tabla 2.6 se muestran las funciones potenciales.

Tabla 2.6: Funciones potenciales de los conglomerados

X_1	X_2	$\psi_1(X_1, X_2)$
0	0	0.105
0	1	0.245
1	0	0.26
1	1	0.39

X_2	X_3	X_4	$\psi_2(X_2, X_3, X_4)$
0	0	0	0.0455
0	0	1	0.0845
0	1	0	0.3045
0	1	1	0.5655
1	0	0	0.05
1	0	1	0.15
1	1	0	0.2
1	1	1	0.6

X_3	X_4	X_5	$\psi_3(X_3, X_4, X_5)$
0	0	0	0.55
0	0	1	0.45
0	1	0	0.15
0	1	1	0.85
1	0	0	0.45
1	0	1	0.55
1	1	0	0.56
1	1	1	0.44

El método de agrupamiento lo aplicaremos suponiendo que disponemos de la evidencia $e = \{X_3 = 1\}$, por lo que modificaremos las funciones potenciales que contengan al nodo X_3 utilizando (2.29), los cálculos se muestran en la tabla 2.7.

Tabla 2.7: Funciones potenciales correspondiente a la evidencia observada

X_2	X_3	X_4	$\psi_2(X_2, X_3, X_4)$	X_3	X_4	X_5	$\psi_3(X_3, X_4, X_5)$
0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0
0	1	0	0.3045	0	1	0	0
0	1	1	0.5655	0	1	1	0
1	0	0	0	1	0	0	0.45
1	0	1	0	1	0	1	0.55
1	1	0	0.2	1	1	0	0.56
1	1	1	0.6	1	1	1	0.44

Cálculo de las funciones $P(R_i / S_i)$

Para $i = 3$ se tiene: $r_3 = \{X_5\}$

$$M_{X_5}(X_3 = 1, X_4 = 0) = \sum_{X_5} \psi_3(X_3 = 1, X_4 = 0, X_5) \Rightarrow M_{X_5}(X_3 = 1, X_4 = 0) = 1$$

$$M_{X_5}(X_3 = 1, X_4 = 1) = \sum_{X_5} \psi_3(X_3 = 1, X_4 = 1, X_5) \Rightarrow M_{X_5}(X_3 = 1, X_4 = 1) = 1$$

Por lo que se tiene:

$$P(X_5 = 0 / X_3 = 1, X_4 = 0) = \frac{\psi_3(X_3 = 1, X_4 = 1, X_5 = 0)}{M_{X_5}(X_3 = 1, X_4 = 0)} = \frac{0.45}{1} = 0.45$$

Similarmente se obtienen las demás probabilidades, para los nodos del conglomerado C_3 , los cálculos se ilustran en la tabla 2.8.

Figura 2.8: Funciones $P(R_i = r_i / S_i = s_i)$

X_3	X_4	X_5	$P(X_5 / X_3 = 1, X_4)$
1	0	0	0.45
1	0	1	0.55
1	1	0	0.56
1	1	1	0.44

X_2	X_3	X_4	$P(X_3, X_4 / X_2)$
0	1	0	0.35
0	1	1	0.65
1	1	0	0.25
1	1	1	0.75

X_1	X_2	$P(X_1, X_2)$
0	0	0.110653503
0	1	0.237417479
1	0	0.273999152
1	1	0.377929864

Puesto que el conglomerado C_2 contiene el conjunto $S_3 = \{X_3, X_4\}$, se multiplica la función potencial $\psi_2(X_2, X_3, X_4)$ por $M_{X_5}(X_3 = 1, X_4 = 0) = 1$ y $M_{X_5}(X_3 = 1, X_4 = 1) = 1$ el resultado produce la misma tabla 2.7, por lo que utilizando esta nueva función potencial actualizada se calcula la siguiente función de probabilidad.

Para $i = 2$ se tiene: $r_2 = \{X_3, X_4\}$

$$M_{X_3X_4}(X_2 = 0) = \sum_{X_4} \psi_2(X_2 = 0, X_3 = 1, X_4) \Rightarrow M_{X_3X_4}(X_2 = 0) = 0.87$$

$$M_{X_3X_4}(X_2 = 1) = \sum_{X_4} \psi_2(X_2 = 1, X_3 = 1, X_4) \Rightarrow M_{X_3X_4}(X_2 = 1) = 0.8$$

Por lo que se tiene:

$$P(X_3 = 1, X_4 = 0 / X_2 = 0) = \frac{\psi_3(X_2 = 0, X_3 = 1, X_4 = 0)}{M_{X_3X_4}(X_2 = 0)} = \frac{0.3045}{0.87} = 0.35$$

Los cálculos de las demás probabilidades se encuentran en la tabla 2.8.

Puesto que $S_2 = \{X_2\}$ se encuentran en el primer conglomerado se multiplica la función potencial de conglomerado C_1 por $M_{X_3X_4}(X_2 = 0)$ y $M_{X_3X_4}(X_2 = 1)$, los cálculos se muestran en la tabla 2.9.

Tabla 2.9: Función potencial del conglomerado C_1 actualizada

X_1	X_2	$\psi_1(X_1, X_2)$
0	0	0.09135
0	1	0.196
1	0	0.2262
1	1	0.312

Para $i = 1$ se tiene: $r_1 = \{X_1, X_2\}$ y $S_1 = \{\phi\}$

$$M_{X_1 X_2}(\phi) = \sum_{X_1, X_2} \psi_1(X_1, X_2) \Rightarrow M_{X_3 X_4}(\phi) = 0.822555$$

Por lo que la probabilidad conjunta de las variables X_1 y X_2 se calcula como sigue:

$$P(X_1 = 0, X_2 = 0) = \frac{\psi_1(X_1 = 0, X_2 = 0)}{M_{X_1 X_2}(\phi)} = \frac{0.09135}{0.822555} = 0.110653503$$

Los cálculos de las demás probabilidades se muestran en la tabla 2.8.

Cálculo de la probabilidad conjunta de los conglomerados.

Para el conglomerado C_1 no se realiza ningún cálculo, ya que esta ha sido calculada, para el conglomerado C_2 se utiliza la expresión $P(X_2, X_3, X_4) = P(X_3 = 1, X_4 / X_2)P(X_2)$ donde $P(X_2)$ se obtiene marginalizando la función de probabilidad del conglomerado C_1 , para el nodo X_2 , el cálculo de la función de probabilidad conjunta del conglomerado C_3 se obtiene por la expresión $P(X_3, X_4, X_5) = P(X_5 / X_3 = 1, X_4)P(X_3 = 1, X_4)$, la probabilidad $P(X_3 = 1, X_4)$ se obtiene marginalizando la función de probabilidad conjunta del conglomerado C_2 para las variables $X_3 = 1$, y X_4 . Los cálculos se muestran en la tabla 2.10 para los conglomerados C_2 y C_3 .

Tabla 2.10: Funciones de probabilidad para los conglomerados dos y tres

X_2	X_3	X_4	$P(X_2, X_3, X_4)$
0	1	0	0.134628429
0	1	1	0.250024225
1	1	0	0.153836836
1	1	1	0.461510508

X_3	X_4	X_5	$P(X_3, X_4, X_5)$
1	0	0	0.129809369
1	0	1	0.158655895
1	1	0	0.398459451
1	1	1	0.313075285

En la tabla 2.11 se presenta $P(X_i / e = X_3 = 1)$ para cada nodo no evidencial de la red (2.11) (b), estas se han obtenido marginalizando las funciones de probabilidad de los conglomerados dadas en la tabla 2.8 y la del conglomerado C_1 dado en la tabla 2.10.

Tabla 2.11: Probabilidades Condicionales de los nodos de la red dada la evidencia Observada.

X_1	$P(X_1)$	X_2	$P(X_2)$	X_3	$P(X_3)$	X_4	$P(X_4)$	X_5	$P(X_5)$
0	0.348070982	0	0.384652655	0	0	0	0.288465265	0	0.52826882
1	0.651929016	1	0.615347345	1	1	1	0.711534735	1	0.47173118

2.2.4 Propagación Orientada a un Objetivo

En muchas ocasiones, se puede tener una red bayesiana con un gran número de nodos y estar interesados en la probabilidad de unos pocos nodos dada la evidencia, los métodos descritos en las secciones anteriores nos proporcionan estas probabilidades pero trabajan sobre toda la red, por lo que se obtienen las probabilidades $P(X_i = x_i / e)$ de todos los nodos de la red.

El método de calcular las probabilidades de un conjunto de nodos de la red, que llamaremos conjunto objetivo dada la evidencia se le denomina propagación orientada a un objetivo.

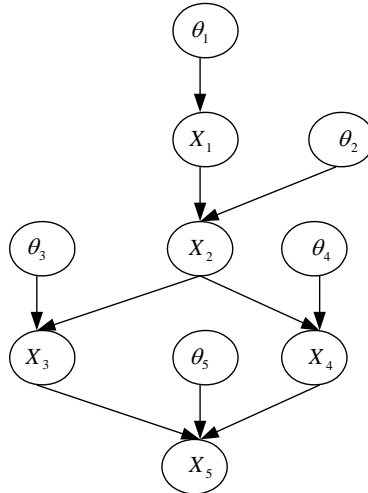
El procedimiento consiste en que dado un conjunto objetivo de variables $O = \{X_i, \dots, X_k\}$ se desea calcular $P(X_j / e)$, $\forall X_j \in O$, lo primero que se hace es construir una nueva presentación gráfica, con nuevos nodos $\theta = \{\theta_1, \dots, \theta_n\}$ uno para cada nodo de la red y agregar el enlace $\theta_i \rightarrow X_i$, a continuación se aplica el criterio de d-separación para obtener un subconjunto de θ que no este d-separado de O por e , a este subconjunto de nodos de θ , se le conoce como nodos relevantes.

Una vez identificado el conjunto de nodos relevantes, se eliminan del grafo los nodos irrelevantes y con el nuevo grafo obtenido se aplica algún método de propagación descrito, para calcular las probabilidades condicionales del conjunto objetivo.

Ejemplo 2.4:

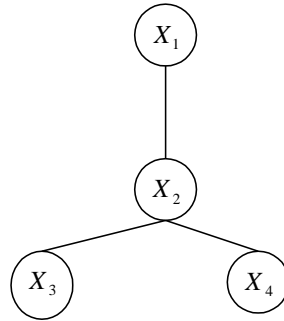
Consideremos la red de la figura 2.11(a) y supongamos que sólo estamos interesados en calcular $P(X_1 / X_3 = 1)$ por lo que en este caso el conjunto objetivo es solamente la variable X_1 , para encontrar el conjunto de nodos que no esta d-separado del nodo X_1 , por la evidencia X_3 se crea un grafo con un nodo auxiliar por cada nodo de la red de la figura 2.11 (a), esta se ilustra en la figura 2.12.

Figura 2.12: Nodos Auxiliares por cada nodo de la red 2.11 (a)



Como podemos observar en la figura 2.12 sólo el nodo θ_5 es independiente de X_1 dado el nodo evidencial X_3 debido a que el camino $X_1 - X_2 - X_3 - X_5 - \theta_5$ contiene al nodo evidencial, por lo que se elimina el nodo X_5 de la figura 2.12, lo cual se muestra en la figura 2.13, los demás nodos de la red no están d-separados de X_1 por X_3 por lo que ya no se eliminan más nodos.

Figura 2.13: Red Simplificada de la figura 2.12.



Como podemos observar la red obtenida (figura 2.13) es un poliárbol por lo que se puede aplicar el método descrito en la sección 2.2.1 para obtener la probabilidad buscada como sigue:

Funciones para el nodo Evidencial:

$$\rho_{X_3}(X_3) = \begin{cases} 0; & X_3 = 0 \\ 1; & X_3 = 1 \end{cases}$$

$$\lambda_{X_3}(X_3) = \begin{cases} 0; & X_3 = 0 \\ 1; & X_3 = 1 \end{cases}$$

Función para el nodo sin padre:

$$\rho_{X_1}(X_1) = \begin{cases} 0.35; & X_1 = 0 \\ 0.65; & X_1 = 1 \end{cases}$$

Función para el nodo sin hijo

$$\lambda_{X_4}(X_4) = \begin{cases} 1; & X_4 = 0 \\ 1; & X_4 = 1 \end{cases}$$

Enviando el mensaje de X_3 a X_2

$$\lambda_{X_3X_2}(X_2 = 0) = 0(0.13) + 1(0.87) \Rightarrow \lambda_{X_3X_2}(X_2 = 0) = 0.87$$

$$\lambda_{X_3X_2}(X_2 = 1) = 0(0.2) + 1(0.8) \Rightarrow \lambda_{X_3X_2}(X_2 = 1) = 0.8$$

Enviando el mensaje de X_4 a X_2

$$\lambda_{X_4 X_2}(X_2 = 0) = 1(0.35) + 1(0.65) \Rightarrow \lambda_{X_4 X_2}(X_2 = 0) = 1$$

$$\lambda_{X_4 X_2}(X_2 = 1) = 1(0.25) + 1(0.75) \Rightarrow \lambda_{X_4 X_2}(X_2 = 1) = 1$$

Calculando $\lambda_{X_2}(X_2)$

$$\lambda_{X_2}(X_2 = 0) = 0.87(1) \Rightarrow \lambda_{X_2}(X_2 = 0) = 0.87$$

$$\lambda_{X_2}(X_2 = 1) = 0.8(1) \Rightarrow \lambda_{X_2}(X_2 = 1) = 0.8$$

Enviando el mensaje de X_2 a X_1

$$\lambda_{X_2 X_1}(X_1 = 0) = 0.87(0.3) + 0.8(0.7) \Rightarrow \lambda_{X_2 X_1}(X_1 = 0) = 0.821$$

$$\lambda_{X_2 X_1}(X_1 = 1) = 0.87(0.4) + 0.8(0.6) \Rightarrow \lambda_{X_2 X_1}(X_1 = 1) = 0.828$$

Calculando $\lambda_{X_1}(X_1)$

$$\lambda_{X_1}(X_1 = 0) = 0.821$$

$$\lambda_{X_1}(X_1 = 1) = 0.828$$

Calculando $P(X_1 / X_3 = 1)$

$$P(X_1 = 0 / X_3 = 1) = \frac{0.821(0.35)}{0.821(0.35) + 0.828(0.65)} \Rightarrow P(X_1 = 0 / X_3 = 1) = 0.34807$$

$$P(X_1 = 1 / X_3 = 1) = \frac{0.828(0.65)}{0.821(0.35) + 0.828(0.65)} \Rightarrow P(X_1 = 1 / X_3 = 1) = 0.65192$$

Por lo que la probabilidad buscada es:

X_1	$P(X_1 / X_3 = 1)$
0	0.34807
1	0.65192

CAPÍTULO 3.
INTRODUCCIÓN A LA ESTIMACIÓN
DE UNA RED BAYESIANA.

3. Introducción a la estimación de una red bayesiana.

3.1 Introducción.

La estimación de una red bayesiana se divide en dos partes, las cuales consisten en la estimación de las probabilidades de las variables, que se le conoce como *estimación de parámetros*, y la representación gráfica de una red bayesiana (es decir el grafo dirigido acíclico (GDA)) que se le conoce como *estructura*, este capítulo tratará la estimación de una red bayesiana en el caso de variables discretas multinomiales, primero veremos como estimar las probabilidades de la red, para lo cual suponemos que conocemos el GDA de la red y después veremos un algoritmo para estimar la estructura de la red, es decir, como se relacionan las variables gráficamente.

En general existen dos formas de estimar una red bayesiana, ya sea por medio de un experto que conozca la realidad que se esta modelando o por medio de una base de datos que se disponga para estimar tanto los parámetros como la estructura de la red bayesiana, este capítulo trata la segunda forma. Una de las razones por las que han surgido técnicas para estimar una red bayesiana de una base de datos es debido a que no en todas las realidades que se quieran modelar con una red bayesiana se dispone de un experto, para la estimación de la red y si se dispone de un experto puede ser dificultoso para la persona dar estimaciones subjetivas de las probabilidades de las variables cuando el número de variables es grande y personas diferentes darán estimaciones diferentes de la probabilidad de una misma variable.

3.2 Estimación de los parámetros de una red bayesiana.

En esta parte trataremos la estimación de las probabilidades de las variables de la red de acuerdo a como éstas estén representadas en el GDA, supondremos en todo momento que las variables son discretas multinomiales²³, que no hay datos faltantes y que conocemos todos los valores que pueden tomar cada variable aleatoria.

De acuerdo al cálculo de probabilidades existen tres métodos para estimar las probabilidades, a saber:

1. *El Método clásico*: Consiste en que si conocemos el total de resultados posibles en que ocurre cada valor que tome una variable aleatoria, entonces la probabilidad de cada resultado se define

²³ En esta tesis no se tratará la estimación de parámetros de una red bayesiana en el caso de que variables de la red sean continuas, ni tampoco el caso en que la red tiene variables continuas y discretas.

como el cociente del número de veces en que ocurre el resultado por la suma del número de veces en que ocurre cada uno de los resultados de la variable aleatoria.

2. *Método de frecuencia relativa:* Si no conocemos el número de veces en que ocurre cada resultado de una variable aleatoria, pero podemos obtener por medio de algún experimento, un cierto número de veces en que ocurre cada resultado, entonces nuestra estimación de la probabilidad de cada variable aleatoria sería la frecuencia relativa con la que ocurre cada resultado en el experimento.
3. *Método Subjetivo:* Éste consiste en que las probabilidades de cada uno de los valores que tome una variable aleatoria las puede dar la persona que realice el estudio, ya sea que la haya obtenido por medio de la experiencia personal, por medio de la literatura especializada de algún campo del conocimiento o por un experto, a este tipo de probabilidades se le llama *probabilidades subjetivas*, ya que representa la creencia personal de un individuo, sobre cada uno de los valores que tomen las variables aleatorias.

El método que utilizaremos para estimar las probabilidades de una red bayesiana será el método frecuentista, en ciertos casos utilizaremos cualquiera de los tres métodos para asignar probabilidades a priori de los nodos de la red.

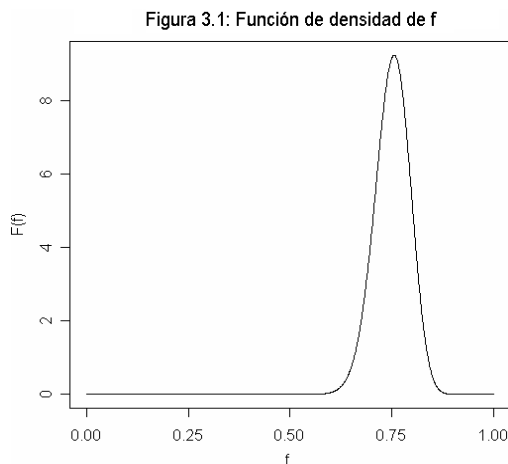
3.2.1 Estimando la probabilidad de una variable.

Supongamos que tenemos una variable aleatoria X cuyos posibles valores son $X = \{1, 2, \dots, r\}$, por ejemplo supongamos que nuestra variable aleatoria representa el color de los ojos de las personas, y que nos encontramos en un lugar extraño en que solo dos colores son posibles, ojos cafés y negros, por lo que si X representa el color de los ojos de una persona escogida aleatoriamente, entonces nuestra variable toma su valor del conjunto $X = \{1 = \text{cafés}, 2 = \text{negros}\}$, si conociésemos el número de personas que viven en ese lugar y que tienen sus ojos cafés podríamos dar la probabilidad de que al escoger una persona aleatoriamente sus ojos sean cafés o negros, sin embargo antes de realizar el estudio no tenemos conocimiento acerca de la cantidad de personas que tienen ojos cafés, si contáramos con un grupo de personas y determináramos el color de sus ojos, nuestra estimación de probabilidad de color de los ojos cafés, sería la frecuencia relativa del conjunto de personas que tienen sus ojos cafés, pero puede suceder que esta estimación de probabilidad no nos satisfaga debido a que en un principio podríamos creer que la

frecuencia relativa de personas que tienen sus ojos cafés estén alrededor de un valor particular y que los que proporcionan los datos obtenidos refleja un valor muy alto o uno muy bajo de lo que pensábamos, por lo que mejor preferiríamos dar una estimación de la probabilidad de las personas que tienen ojos cafés de acuerdo a nuestro criterio y que luego de muestrear un conjunto de personas y determinar el color de sus ojos actualizar nuestra probabilidad previa, por ejemplo conocemos que en nuestro medio la mayoría de personas tienen sus ojos cafés, por lo que podríamos creer que alrededor del 75% de las personas tienen sus ojos cafés y por lo tanto un 25% tienen ojos negros, estos valores representarían nuestras probabilidades a priori, es decir:

$$\begin{aligned} P(X = 1) &= 0.75 \\ P(X = 2) &= 0.25 \end{aligned} \tag{3.1}$$

Al principio del estudio en que no disponemos de información de la población estas probabilidades serían las que utilizaríamos, hasta que obtengamos información y así actualizamos las probabilidades dadas en (3.1), sin embargo conocemos que si tomamos varias muestras aleatoriamente del mismo tamaño obtendremos distintas actualizaciones de las probabilidades dadas en (3.1), por lo que en lugar de obtener



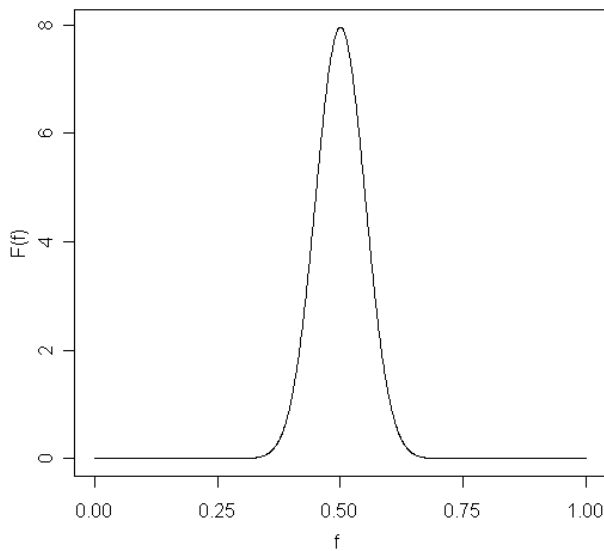
una probabilidad actualizada, podríamos tener una distribución de proporciones que nos indicarán el valor de nuestra creencia actualizada, pero debido a que no podemos tomar varias muestras de la población de estudio, lo mejor sería tener una variable aleatoria que represente nuestro conocimiento a priori acerca del color de los ojos de las personas y así en lugar de actualizar una probabilidad como la dada en (3.1) actualizaríamos una distribución de proporciones.

Más concretamente las probabilidades dadas en (3.1) representan a nuestro criterio que alrededor de esos valores es que puede andar la proporción de personas que tienen ojos cafés y negros, por lo que nos gustaría tener una variable aleatoria que represente nuestra creencia a priori en lugar de un valor puntual, es decir preferimos una variable aleatoria f cuyos valores se encuentren entre 0 y 1, y que represente nuestra creencia con respecto a la proporción de personas que tienen ojos cafés de tal forma que la mayoría de sus valores se encuentren alrededor de 0.75, en otras palabras deseamos una variable aleatoria cuya distribución de valores sea

como la que se muestra en la figura 3.1., como observamos en la figura los valores más frecuentes de la variable están alrededor de 0.75 siendo este último el de mayor frecuencia, por lo que cuando observemos el color de los ojos de una muestra de personas, actualizaremos no un valor particular si no que una distribución de valores que representa nuestro conocimiento a priori o creencia personal.

Consideremos ahora el lanzamiento de una moneda, conocemos que al lanzar una moneda la probabilidad de observar una cara es 0.5 (si la moneda esta equilibrada), por lo que antes de realizar una serie de lanzamientos y observar la frecuencia relativa de cara, creemos que ésta si bien no será 0.5, sabemos que el valor de la frecuencia relativa que se observe estará alrededor de 0.5, por lo que en este caso tendremos que nuestra variable aleatoria f que representa nuestro conocimiento a priori tendrá una distribución de valores en que la mayoría de ellos se concentre alrededor de 0.5, es decir deseamos que nuestra variable aleatoria tenga la distribución mostrada en la figura (3.2).

Figura 3.2:Función de densidad de f

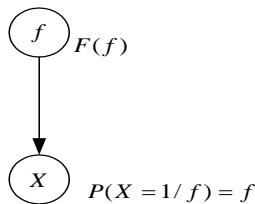


Como se muestra en la figura, la mayor área bajo la curva se tiene alrededor de 0.5, siendo este valor donde la función toma su máximo, por lo que cuando lanzamos la moneda n veces, esperamos que la frecuencia relativa de que la moneda aterrice con el lado cara hacia arriba esté cerca de 0.5.

En general cuando una variable aleatoria es binaria y consideramos que sus valores son igualmente probables, deseáramos una variable cuya distribución sea igual a la de la figura 3.2, para estimar la frecuencia relativa con que se observará

el resultado de la variable.

Figura 3.3: Probabilidad a priori de la variable X



De aquí en adelante supondremos que f es una variable aleatoria continua entre 0 y 1, que representa nuestra probabilidad o conocimiento a priori y que utilizaremos para estimar la frecuencia relativa con que ocurre el resultado de la variable que estemos observando, de tal forma que los

valores más probables de esta variable sean aquellos números que creemos son los mejores candidatos para ser la probabilidad de la variable aleatoria, por lo que antes de recoger los datos tenemos:

$$P(X = 1 / f) = f \tag{3.2}$$

Tanto nuestro conocimiento a priori como a la variable cuya probabilidad estamos estimando se pueden representar por medio de una red bayesiana la cual se muestra en la figura 3.3.

En la expresión (3.2) tenemos representada nuestra probabilidad a priori, de que la variable aleatoria X tome el valor de 1, si conociéramos la función de probabilidad de f podríamos dar una estimación de la frecuencia relativa con que la variable X sea igual a 1 de la siguiente forma:

$$P(X = 1) = \int_{-\infty}^{\infty} P(X = 1 / f) \rho(f) df \tag{3.3}$$

Cuando obtenemos la muestra actualizaríamos la probabilidad (3.3) y esta sería nuestra estimación para $X = 1$ y la que ocuparíamos en la red.

3.2.2 Función de densidad a priori.

Si tenemos una variable aleatoria con dos resultados posibles, en que el primer resultado ($X = 1$) puede ocurrir a veces y el segundo resultado ($X = 2$) ocurre b veces entonces se dice que el primer resultado es equivalente a observarlo a veces de un total de $N = a + b$ veces, por ejemplo si lanzamos una moneda 2 veces al aire esperaríamos observar una cara una vez en las dos veces que es lanzada.

En general deseáramos una variable aleatoria f en que la mayor cantidad del área bajo la curva de su función de densidad se encuentre alrededor del valor $a / (a + b)$, como hemos observado en las figuras 3.1 y 3.2 las funciones de densidad que se ha graficado cumple los requisitos que acabamos de mencionar, por lo que de aquí en adelante supondremos que la función de densidad que tiene la variable

aleatoria f será una beta, cuando represente nuestro conocimiento a priori de una variable aleatoria binaria²⁴, la cual definimos como sigue:

Definición 3.1 si tenemos una variable aleatoria f cuyo rango de valores es $[0,1]$ y que tiene la siguiente función de densidad:

$$\rho(f) = \frac{\Gamma(N)}{\Gamma(a)\Gamma(b)} f^{a-1} (1-f)^{b-1}$$

Se dice que la variable aleatoria f tiene una distribución beta con parámetros a , b y $N = a + b$

Por lo que ahora podemos dar una estimación de la frecuencia relativa de los ejemplos anteriores usando 3.3, puesto que f tiene una distribución beta la expresión 3.3, no es más que el valor esperado de la variable aleatoria f por lo que la estimación de la frecuencia relativa sería:

$$P(X = 1) = E(f) = \frac{a}{N} \quad (3.4)$$

En el ejemplo del color de los ojos de las personas, tenemos que nuestra creencia es observar 75 personas con color de ojos cafés de un total de 100, por lo que nuestra estimación de la frecuencia relativa con que observaríamos las personas con ojos cafés ($X = 1$) es la dada en 3.1 con $a = 75$ y $b = 25$, equivalentemente se obtiene la probabilidad de la moneda, en ella suponemos que de un total de 100 lanzamientos que observemos esperamos ver 50 caras y aplicando (3.4) obtenemos el resultado.

En general si tenemos una variable aleatoria que puede tomar sus valores del conjunto $(1, 2, 3, \dots, r)$ nuestro conocimiento a priori a cerca de los valores que puede tomar la variable aleatoria se puede representar por medio de un conjunto de variables $\{f_1, \dots, f_r\}$ donde f_i representa nuestro conocimiento a priori con que ocurre el i -ésimo valor de la variable X , donde la distribución conjunta de las variables f_i es una distribución Dirichlet la cual se define a continuación.

²⁴ La razón por lo que se usa una distribución beta para representar nuestro conocimiento a priori acerca de una variable binomial es debido a que esta distribución es a priori conjugado de una distribución binomial, en el caso de variables multinomiales su a priori conjugado es una distribución Dirichlet.

Definición 3.2 La función de densidad Dirichlet con parámetros enteros a_1, \dots, a_r , mayores o iguales a 1 y

$N = \sum_{k=1}^r a_k$, se define como

$$\rho(f_1, \dots, f_{r-1}) = \frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} f_1^{a_1-1} f_2^{a_2-1} f_3^{a_3-1} \dots f_{r-1}^{a_{r-1}-1} f_r^{a_r-1} \quad 0 \leq f \leq 1 \quad \sum_{i=1}^r f_i = 1$$

Las variables aleatorias f_1, \dots, f_r que tiene esta función de densidad se dice que tienen una distribución Dirichlet la cual la representaremos como $dir(f_1, \dots, f_r; a_1, a_2, \dots, a_r)$.²⁵

Por ejemplo si tenemos una variable aleatoria X que nos representa el resultado del lanzamiento de un dado entonces su espacio de valores es el conjunto $\{1, 2, 3, 4, 5, 6\}$ que representa cada una de las caras de un dado equilibrado, podemos representar la frecuencia relativa de cada una de las caras por las variables aleatorias $\{f_1, f_2, f_3, \dots, f_6\}$ donde f_i representa la frecuencia relativa con que el dado aterriza con el lado i hacia arriba. De acuerdo a nuestro conocimiento a priori, podemos decir que al lanzar el dado 12 veces se espera observar cada cara dos veces en doce lanzamientos, por lo que se tiene la siguiente función de densidad conjunta de las variables f_i :

$$\rho(f_1, \dots, f_{r-1}) = \frac{\Gamma(12)}{\Gamma(2)\Gamma(2)\Gamma(2)\Gamma(2)\Gamma(2)\Gamma(2)} f_1^{2-1} f_2^{2-1} f_3^{2-1} \dots f_r^{2-1}$$

Donde $a_i = 2 \quad i = 1, 2, \dots, 6$

Por lo que la probabilidad a priori de obtener el i -ésimo resultado está dada por:

$$P(X = i / f_1, \dots, f_i, \dots, f_r) = f_i$$

²⁵ $f_r = 1 - \sum_{i=1}^{r-1} f_i$

La estimación de la frecuencia relativa con que la variable aleatoria X toma su i -ésimo valor se calcula como sigue:

$$P(X = i) = \int_0^1 \int_0^1 \dots \int_0^1 P(X = i / f_1, \dots, f_i, \dots, f_r) \rho(f_1, \dots, f_r) df_1, \dots, df_r$$

$$P(X = i) = \int_0^1 \int_0^1 \dots \int_0^1 f_i \rho(f_1, \dots, f_r) df_1, \dots, df_r \quad (3.5)$$

Puesto que la distribución conjunta de las F_i es una dirichlet se tiene que la expresión (3.5) no es más que el valor esperado de una distribución dirichlet de su i -ésima variable aleatoria por lo que se tiene:

$$P(X = i) = \frac{a_i}{N} \quad (3.6)$$

Por ejemplo, en el caso del dado tenemos que nuestra estimación de la frecuencia relativa a priori antes de observar el resultado de un conjunto de lanzamientos es por (3.6):

$$P(X = i) = \frac{1}{6}, \text{ para } i = 1, 2, 3, \dots, 6$$

En general si tenemos una variable aleatoria multinomial y obtenemos una muestra en la cual observamos el resultado que toma la variable aleatoria en cada individuo, a este tipo de muestra recibe un nombre especial el cual se define a continuación.

Definición 3.3 supongamos que tenemos una muestra de tamaño M tal que:

1. Para cada X^h su espacio de valores es $\{1, 2, \dots, r\}$
2. Si $F = \{f_1, \dots, f_r\}$ y para $1 \leq h \leq M$ y $1 \leq k \leq r$ se tiene:

$$P(X^h = k / f_1, \dots, f_k, \dots, f_r) = f_k.$$

Entonces $D = \{X^1, \dots, X^M\}$ es llamada una muestra multinomial de tamaño M con parámetro F .

La definición 3.3 nos dice que si tenemos una variable aleatoria multinomial X y tomamos una muestra de tamaño M , entonces cada observación X^h de la muestra constituye una realización de X , por lo que la frecuencia relativa a priori con que $X^h = k$ en la muestra es f_k debido a que este valor representa nuestro conocimiento a priori con respecto al k -ésimo resultado que puede tomar la variable X y el literal 2 nos indica que observar un resultado en el h -ésimo individuo, este no influye en el resultado de la variable en los demás individuos, es decir, las observaciones de la variable de un individuo a otro son independientes. Para actualizar nuestro conocimiento acerca de la frecuencia relativa con que la variable aleatoria X toma sus valores necesitamos calcular la probabilidad de la muestra que hemos obtenido. Esta se calcula como sigue:

Teorema 3.1 Supongamos que D es una muestra multinomial de tamaño M con parámetro F y hemos observado una realización de las variables en D la cual la representaremos como:

$$d = \{x^1, \dots, x^M\}$$

Denotando por s_k el número de variables en d iguales a K . Entonces se tiene:

$$P(d) = \frac{\Gamma(N)}{\Gamma(N+M)} \prod_{k=1}^r \frac{\Gamma(a_k + s_k)}{\Gamma(a_k)} \quad (3.7)$$

$$M = \sum_{k=1}^M s_k$$

Prueba:

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 P(d / f_1, \dots, f_r) \rho(f_1, \dots, f_r) df_1 df_2 \dots df_r$$

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 P(x^1, x^2, \dots, x^M / f_1, \dots, f_r) \rho(f_1, \dots, f_r) df_1 df_2 \dots df_r$$

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 P(x^1 / f_1, \dots, f_r) \dots P(x^M / f_1, \dots, f_r) \rho(f_1, \dots, f_r) df_1 df_2 \dots df_r$$

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 \prod_{h=1}^M P(x^h / f_1, \dots, f_r) \rho(f_1, \dots, f_r) df_1 df_2 \dots df_r$$

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 \prod_{h=1}^r f_h^{s_h} \rho(f_1, \dots, f_r) df_1 df_2 \dots df_r$$

$$P(d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 f_1^{s_1} f_2^{s_2} \dots f_r^{s_r} \frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} f_1^{a_1-1} f_2^{a_2-1} \dots f_r^{a_r-1} df_1 df_2 \dots df_r$$

$$P(d) = \frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 f_1^{a_1+s_1-1} f_2^{a_2+s_2-1} \dots f_r^{a_r+s_r-1} df_1 df_2 \dots df_r$$

$$P(d) = \frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} \frac{\Gamma(a_1 + s_1) \Gamma(a_2 + s_2) \dots \Gamma(a_r + s_r)}{\Gamma(a_1 + a_2 + a_3 + \dots + a_r + s_1 + s_2 + \dots + s_r)}$$

$$P(d) = \frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} \frac{\prod_{i=1}^r \Gamma(a_i + s_i)}{\Gamma(M + N)}$$

$$P(d) = \frac{\Gamma(N)}{\Gamma(M + N)} \prod_{i=1}^r \frac{\Gamma(a_i + s_i)}{\Gamma(a_i)}$$

La tercera igualdad se debe a que las observaciones en la muestra d son independientes, la quinta igualdad se obtiene de la definición 3.3, la sexta igualdad se obtiene de la definición de la distribución Dirichlet, la octava igualdad se obtiene al integrar la séptima igualdad. Y las demás igualdades se obtienen al agrupar y ordenar términos. ■

Ejemplo 3.1 Consideremos que hemos lanzado un dado equilibrado 10 veces y obtenemos el siguiente resultado:

$$d = \{2, 3, 4, 5, 5, 1, 3, 2, 2, 4\}$$

Entonces $a_1 = a_2 = a_3 = \dots = a_6 = 2$, $s_1 = 1$, $s_2 = 3$, $s_3 = 2$, $s_4 = 2$, $s_5 = 2$, $s_6 = 0$, $N = 10$, y $M = 10$

Por el teorema 3.1 se tiene:

$$P(d) = \frac{\Gamma(12)}{\Gamma(12+10)} \frac{\Gamma(2+1)}{\Gamma(2)} \frac{\Gamma(2+3)}{\Gamma(2)} \frac{\Gamma(2+2)}{\Gamma(2)} \frac{\Gamma(2+2)}{\Gamma(2)} \frac{\Gamma(2+2)}{\Gamma(2)} \frac{\Gamma(2+0)}{\Gamma(2)}$$

$$P(d) = 1.620081^{-7}$$

El siguiente teorema nos indica como calcular la función de densidad condicionada de las variables f_1, \dots, f_r dada la muestra d como sigue.

Teorema 3.2 supongamos que D es una muestra multinomial de tamaño M con parámetro F cuyas variables tienen una distribución dirichlet, tenemos un conjunto de datos $d = \{x^1, \dots, x^M\}$ y s_k es el número de variables en d iguales a k , entonces la función de densidad condicionada de F a los datos d es:

$$dir(f_1, \dots, f_r; a_1 + s_1, a_2 + s_2, \dots, a_r + s_r).$$

Prueba:

Aplicando la definición de probabilidad condicional se tiene:

$$\rho(f_1, f_2, \dots, f_r / d) = \frac{\rho(f_1, f_2, \dots, f_r, d)}{P(d)}$$

Aplicando el teorema de bayes se tiene:

$$\rho(f_1, f_2, \dots, f_r / d) = \frac{\rho(f_1, f_2, \dots, f_r) P(d / f_1, f_2, \dots, f_r)}{P(d)}$$

Puesto que la probabilidad conjunta $\rho(f_1, \dots, f_r)$ es una dirichlet y de la definición (3.3) y del hecho que las observaciones son independientes entre ellas y del teorema (3.1) se tiene:

$$\rho(f_1, f_2, \dots, f_r / d) = \frac{\left(\frac{\Gamma(N)}{\prod_{i=1}^r \Gamma(a_i)} f_1^{a_1-1} f_2^{a_2-1} \dots f_r^{a_r-1} \right) \left(\prod_{i=1}^M P(x^i / f_1, \dots, f_r) \right)}{\frac{\Gamma(N)}{\Gamma(N+M)} \prod_{i=1}^r \frac{\Gamma(a_i + s_i)}{\Gamma(a_i)}}$$

$$\rho(f_1, f_2, \dots, f_r / d) = \frac{(\Gamma(N)\Gamma(N+M) f_1^{a_1-1} f_2^{a_2-1} \dots f_r^{a_r-1}) \left(\prod_{i=1}^M f_i^{s_i} \right)}{\Gamma(N) \prod_{i=1}^r \Gamma(a_i) \frac{\prod_{i=1}^r \Gamma(a_i + s_i)}{\prod_{i=1}^r \Gamma(a_i)}}$$

Simplificando la igualdad anterior se obtiene la siguiente función de densidad dirichlet:

$$\rho(f_1, f_2, \dots, f_r / d) = \frac{\Gamma(N+M)}{\prod_{i=1}^r \Gamma(a_i + s_i)} f_1^{a_1+s_1-1} f_2^{a_2+s_2-1} \dots f_r^{a_r+s_r-1}.$$

■

Hasta ahora hemos visto que en una muestra multinomial de tamaño M las observaciones en ella son independientes, y hemos representado la frecuencia relativa con que la i -ésima observación tome el valor de k sea f_k , que no es más que nuestro conocimiento a priori de que la variable aleatoria X tome su k -ésimo valor en los datos antes de que se observen, lo que nos interesa es que una vez observada la muestra y determinados los valores de la variable X que en ella aparecen, es calcular la probabilidad a posteriori en base a los datos observados de cada uno de los valores que puede tomar la variable X , es decir nos interesa conocer la probabilidad de que al observar el resultado de la $(m+1)$ -ésima observación este sea igual a k para todo $k \in \{1, 2, \dots, r\}$.

Teorema 3.3 Si tenemos una muestra multinomial D de tamaño M con parámetro F y un conjunto de datos $d = \{x^1, \dots, x^M\}$ y definimos s_k como el número de variables en d iguales a k y creamos una muestra

multinomial de tamaño $M + 1$ agregando a D la variable X^{m+1} entonces la probabilidad de que esta variable tome el k -ésimo valor de los posibles valores de X es:

$$P(X^{M+1} = k / d) = \frac{a_k + s_k}{N + M}. \quad (3.8)$$

Prueba:

$$P(X^{M+1} = k / d) = \frac{P(X^{M+1} = k, d)}{P(d)}$$

$$P(X^{M+1} = k / d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 \frac{P(X^{M+1} = k, d, f_1, \dots, f_r)}{P(d)} df_1 \dots df_r$$

$$P(X^{M+1} = k / d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 \frac{P(X^{M+1} = k / d, f_1, \dots, f_r) P(d, f_1, \dots, f_r)}{P(d)} df_1 \dots df_r$$

$$P(X^{M+1} = k / d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 \frac{f_k \rho(f_1, \dots, f_r / d) P(d)}{P(d)} df_1 \dots df_r$$

$$P(X^{M+1} = k / d) = \int_0^1 \int_0^1 \int_0^1 \dots \int_0^1 f_k \rho(f_1, \dots, f_r / d) df_1 \dots df_r$$

$$P(X^{M+1} = k / d) = E(F_k / d)$$

Puesto que la función condicionada de las variables aleatorias $F = \{f_1, \dots, f_r\}$ es una dirichlet con parámetros $a_1 + s_1, \dots, a_r + s_r$ se tiene:

$$P(X^{M+1} = k / d) = \frac{a_k + s_k}{N + M} \quad (3.9)$$

■

Ejemplo 3.2 Consideremos que tenemos la muestra del lanzamiento del dado equilibrado que aparece en el ejemplo 3.1, como vimos en ese ejemplo los datos que proporciona la muestra de tamaño 10 son:

$$s_1 = 1, \quad s_2 = 3, \quad s_3 = 2, \quad s_4 = 2, \quad s_5 = 2, \quad s_6 = 0$$

Como hemos supuesto nuestro conocimiento a priori acerca de los valores que puede tomar nuestra variable aleatoria X (Resultado del lanzamiento del dado) lo representamos por medio de la distribución Dirichlet con parámetros $a_i = 2$, $i = 1, \dots, 6$ y $N = 12$, por lo que nuestra estimación de probabilidad de que la variable aleatoria tome su k -ésimo valor en el 11º lanzamiento es debido al teorema 3.3:

$$P(X^{11} = 1) = \frac{2+1}{12+10} \Rightarrow P(X^{11} = 1) = \frac{3}{22}$$

$$P(X^{11} = 2) = \frac{2+3}{12+10} \Rightarrow P(X^{11} = 2) = \frac{5}{22}$$

$$P(X^{11} = 3) = \frac{2+2}{12+10} \Rightarrow P(X^{11} = 3) = \frac{2}{11}$$

$$P(X^{11} = 4) = \frac{2+2}{12+10} \Rightarrow P(X^{11} = 4) = \frac{2}{11}$$

$$P(X^{11} = 5) = \frac{2+2}{12+10} \Rightarrow P(X^{11} = 5) = \frac{2}{11}$$

$$P(X^{11} = 6) = \frac{2+0}{12+10} \Rightarrow P(X^{11} = 6) = \frac{1}{11}$$

Por lo que las probabilidades dadas anteriormente serían las que ocuparíamos para realizar inferencias en una red bayesiana.

En los ejemplos anteriores hemos supuesto que tenemos conocimiento a priori acerca de los valores que puede tomar la variable aleatoria y de acuerdo a eso hemos podido determinar los valores a_i de la distribución Dirichlet, pero en el caso de que no tengamos conocimiento a priori de cuales de los valores son los más probables que se pueden observar, podemos suponer que todos los valores de la variable aleatoria son igualmente probables y por lo tanto los parámetros a_i de la distribución Dirichlet deben ser iguales a 1 y las variables aleatorias que representen nuestro conocimiento a priori darán igual probabilidad a priori a cada resultado de la variable aleatoria que se este estimando su probabilidad.

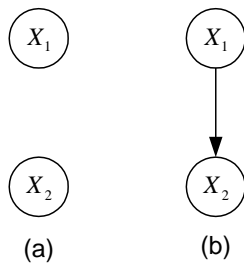
Si la variable a la cual estamos estimando las probabilidades de sus valores, consideramos que el k -ésimo valor que toma la variable aleatoria es el más probable entonces los valores a_i de la distribución Dirichlet debe de ser mayor a 1, de tal forma que la función de densidad dirichlet concentre la mayoría de sus valores alrededor de a_k / N .

3.2.3 Estimando los parámetros de una red bayesiana con más de una variable.

Hasta ahora hemos visto como estimar la probabilidad de una variable aleatoria multinomial y hemos utilizado un conjunto de variables aleatorias $F = \{f_1, \dots, f_r\}$ que nos representa nuestro conocimiento a priori acerca de las probabilidades de los valores de nuestra variable aleatoria, y hemos supuesto de acuerdo a la estadística bayesiana, sobre proporciones que el conjunto F tiene una distribución Dirichlet, supuesto que mantendremos en el resto del capítulo.

En el caso de tener dos variables aleatorias, estas pueden estar representadas como se muestra en la figura 3.4, en la parte (a) de la figura se tiene que las variables X_1 y X_2 son independientes, por lo que la estimación de la probabilidad de las variables de esa red se hace por separado y para el cálculo de la probabilidad conjunta de las variables de la red se realiza el producto de las probabilidades de las variables de la red, mientras que en la red bayesiana de la figura 3.4 (b) las probabilidades de la variable X_1 se estiman de acuerdo a la sección anterior, sin embargo, las estimaciones de las probabilidades de la variable X_2 se realizan tomando en cuenta los valores que puede tomar la variable X_1 debido a que se trata de una probabilidad condicional.

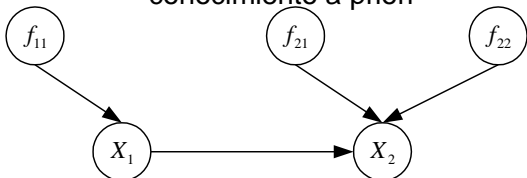
Figura 3.4: Red bayesiana



Por ejemplo, consideremos que las variables de la figura 3.4 son binarias y que pueden tomar los valores $\{1, 2\}$, entonces representaremos nuestro conocimiento a priori de la primera variable, por f_{11} que nos representa la probabilidad a priori de que la primera variable sea igual a 1, denotaremos por f_{21} nuestra probabilidad a priori de que X_2 tome el valor de 1 dado que la variable $X_1 = 1$ y f_{22} representa nuestro

conocimiento a priori de que la variable X_2 tome el valor de 1 dado que la variable $X_1 = 2$ al igual como

Figura 3.5: Red Bayesiana representa nuestro conocimiento a priori



representamos nuestro conocimiento a priori en el caso de una variable estas probabilidades a priori también se pueden representar por medio de una red bayesiana como se muestra en la figura 3.5, por lo que nuestras probabilidades a priori se

pueden escribir como:

$$P(X_1 = 1 / f_{11}) = f_{11}$$

$$P(X_2 = 1 / X_1 = 1, f_{21}) = f_{21}$$

$$P(X_2 = 1 / X_1 = 2, f_{22}) = f_{22}$$

Como podemos observar en la figura 3.5 cada variable f_{ij} es representada por un nodo de la red y tiene solamente un hijo, que es el que corresponde al nodo al cual le esta sirviendo como representación de su conocimiento a priori, al tipo de red bayesiana de la figura 3.5 se le conoce como *red bayesiana aumentada*.

Definición 3.2 una *red bayesiana aumentada multinomial* (G, F, ρ) es una red bayesiana aumentada con las siguientes propiedades:

1. Una grafo dirigido acíclico $G = (X, L)$ donde $X = \{X_1, X_2, \dots, X_n\}$ y cada X_i es una variable aleatoria con espacio $\{1, 2, \dots, r_i\}$.
2. Para cada variable X_i se denotará el número total de instanciaciones²⁶ de todos sus padres por q_i , por lo que la variable que representa nuestro conocimiento a priori se define como el conjunto:

$$F_i = F_{i1} \cup F_{i2} \cup \dots \cup F_{iq_i}$$

Y cada una de estos subconjuntos se definen como:

$$F_{ij} = \{F_{ij1}, F_{ij2}, \dots, F_{ijr_i}\}$$

²⁶ Suponemos que las realizaciones de los padres del nodo X_i están ordenadas, según el orden lexicográfico y representaremos por π_{ij} al conjunto de padres del nodo, cuando se encuentran en su j -ésima realización.

Donde cada F_{ij} tienen un único hijo y es padre de la variable X_i y representa nuestro conocimiento a priori acerca del valor que puede tomar la variable X_i cuando sus padres están en su j -ésima instanciación y f_{ijk} representa nuestro conocimiento a priori de que la variable X_i tome su k -ésimo valor cuando sus padres están en su j -ésima realización. Cada variable F_{ij} tiene una función de densidad Dirichlet tal que:

$$\sum_{k=1}^{r_i} f_{ijk} = 1$$

3. Para cada $i, j,$ y k se tiene:

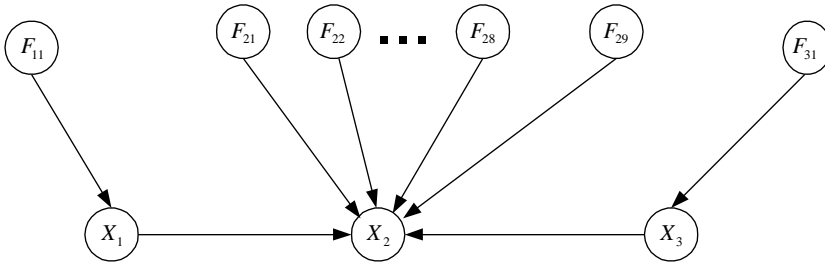
$$P(X_i = k / \pi_i, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) = f_{ijk}$$

Por ejemplo consideremos la red bayesiana de la figura 3.6, en donde cada variable aleatoria tiene espacio

$\{1, 2, 3\}$, por lo que el número de

instanciaciones de los padres de la primera variable es $q_1 = 1$ debido a que el único padre que tiene es el que representa su conocimiento a priori, acerca de cada uno de los valores que puede tomar, y el

Figura 3.6: Red Bayesiana Aumentada Multinomial



conjunto está definido como:

$$F_{11} = \{f_{111}, f_{112}, f_{113}\}$$

Mientras que el número de instanciaciones de los padres de X_2 es $q_2 = 9$ debido a que tiene dos padres, y cada uno puede tomar 3 valores, por lo que el conjunto F_2 es:

$$F_{22} = F_{21} \cup F_{22} \cup F_{23} \cup F_{24} \cup F_{25} \cup F_{26} \cup F_{27} \cup F_{28} \cup F_{29}$$

Donde cada F_{2j} está definido como:

$$F_{2j} = \{f_{2j1}, f_{2j2}, f_{2j3}\}$$

El caso de la variable X_3 es similar al caso de la variable X_1 .

Teorema 3.4 supongamos que tenemos una red bayesiana multinomial, si cada F_{ij} tiene una distribución

Dirichlet con parámetros $a_{ij1}, a_{ij2}, a_{ij3}, \dots, a_{ijr_i}$ y $N_{ij} = \sum_{k=1}^{r_i} a_{ijk}$ entonces la probabilidad de que la variable

X_i sea igual a K dada una instanciación de sus padres es:

$$P(X_i = k / \pi_{ij}) = \frac{a_{ijk}}{N_{ij}} \quad (3.10)$$

Prueba:

$$P(X_i = k / \pi_{ij}) = \frac{P(X_i = k, \pi_{ij})}{P(\pi_{ij})}$$

$$P(X_i = k / \pi_{ij}) = \int_0^1 \int_0^1 \dots \int_0^1 \frac{P(X_i = k, \pi_{ij}, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i})}{P(\pi_{ij})} df_{i1} df_{i2} \dots df_{iq_i}$$

$$P(X_i = k / \pi_{ij}) = \int_0^1 \int_0^1 \dots \int_0^1 \frac{P(X_i = k / \pi_{ij}, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) P(\pi_{ij}, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i})}{P(\pi_{ij})} df_{i1} df_{i2} \dots df_{iq_i}$$

$$P(X_i = k / \pi_{ij}) = \int_0^1 \int_0^1 \dots \int_0^1 \frac{P(X_i = k / \pi_{ij}, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) \rho(f_{i1}, \dots, f_{ij}, \dots, f_{iq_i} / \pi_{ij}) P(\pi_{ij})}{P(\pi_{ij})} df_{i1} df_{i2} \dots df_{iq_i}$$

$$P(X_i = k / \pi_{ij}) = \int_0^1 \int_0^1 \dots \int_0^1 P(X_i = k / \pi_{ij}, f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) \rho(f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) df_{i1} df_{i2} \dots df_{iq_i}$$

$$P(X_i = k / \pi_{ij}) = \int_0^1 \int_0^1 \dots \int_0^1 f_{ijk} \rho(f_{i1}, \dots, f_{ij}, \dots, f_{iq_i}) df_{i1} df_{i2} \dots df_{iq_i}$$

$$P(X_i = k / \pi_{ij}) = E(F_{ijk})$$

$$P(X_i = k / \pi_{ij}) = \frac{a_{ijk}}{N_{ijk}}$$

■

Si disponemos de una muestra multinomial D de tamaño M con parámetro (G, F) de n variables aleatorias tenemos los datos:

$$x^1 = \begin{pmatrix} x_1^1 \\ \vdots \\ x_n^1 \end{pmatrix}, \dots, x^M = \begin{pmatrix} x_1^M \\ \vdots \\ x_n^M \end{pmatrix}$$

Donde representamos por $d = \{x^1, \dots, x^M\}$ a la realización de la muestra D y s_{ijk} el número de veces en que la i -ésima variable toma el valor de k cuando sus padres están en su j -ésima realización, entonces se puede demostrar que la probabilidad de la muestra obtenida es:

$$P(d) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \quad (3.11)$$

Donde:

$$M_{ij} = \sum_{k=1}^{r_i} s_{ijk}$$

Para más detalles acerca de la expresión (3.11) lo puede encontrar en Neapolitan (2003).

Puesto que las probabilidades a priori de la red bayesiana se representa por variables aleatorias que tienen distribución Dirichlet, entonces al observar una muestra de las realizaciones de nuestras variables aleatorias para actualizar las probabilidades de la red dada la muestra, sería lo mismo que actualizar las variables aleatorias que representan nuestro conocimiento a priori y a partir de estas variables actualizadas calcular las probabilidades a posteriori de las variables que serían usadas en la red para realizar inferencias.

Se puede demostrar que si s_{ijk} representa el número de casos en la muestra en los que los padres de X_i están en su j -ésima instanciación, cuando $X_i = k$ entonces la distribución de probabilidad a posteriori de las variables que representan nuestro conocimiento a priori de la red bayesiana tienen una distribución Dirichlet, la cual se representa como sigue:

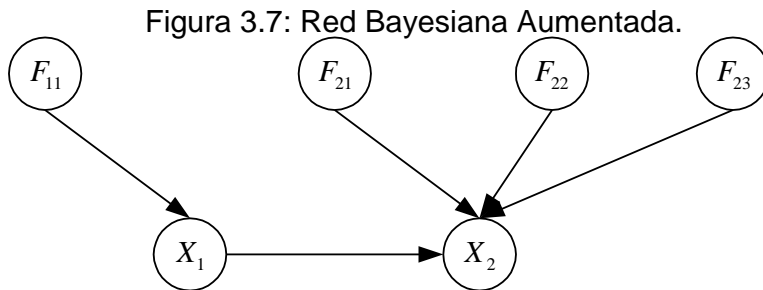
$$\rho(f_{ij1}, f_{ij2}, \dots, f_{ijr_i} / d) = Dir(f_{ij1}, f_{ij2}, \dots, f_{ijr_i}; a_{ij1} + s_{ij1}, a_{ij2} + s_{ij2}, \dots, a_{ijr_i} + s_{ijr_i}) \quad (3.12)$$

A partir de la expresión (3.12) se puede obtener, que la probabilidad a posteriori para el $(M + 1)$ -ésimo caso es:

$$P(X_i = k / \pi_{ij}) = \frac{a_{ijk} + s_{ijk}}{N_{ij}} \quad (3.13)$$

Para ver las expresiones (3.12) y (3.13) de una forma más formal ver Neapolitan (2003), con estas

expresiones se calculan las probabilidades a posteriori de cada variable de una red bayesiana.



Ejemplo 3.2.

En la figura 3.7 se muestra una red bayesiana aumentada con tres variables, donde la variable X_1 toma sus valores del conjunto $\{1, 2, 3\}$ y la variable X_2 toma sus valores del conjunto $\{1, 2, 3, 4\}$.

El conjunto de variables F_{11} que representa nuestro conocimiento a priori acerca de los valores que puede tomar la variable X_1 está definido como sigue:

$$F_{11} = \{f_{111}, f_{112}, f_{113}\}$$

Donde la variable f_{ijk} representa nuestro conocimiento a priori de que la variable X_i tome el valor k de su espacio de valores, en este ejemplo supondremos que la distribución de probabilidad a priori de la variable F_{11} es:

$$Dir(f_{111}, f_{112}, f_{113}; 4, 8, 10)$$

Los padres F_{ij} de la variable X_2 están definidos como sigue:

$$F_{2j} = \{f_{2j1}, f_{2j2}, f_{2j3}\} \quad (3.14)$$

Para $j = 1, 2, 3$ supondremos que la distribución de probabilidades a priori para cada valor de j en la expresión (3.14) es:

$$\text{Para } j = 1 \text{ dir}(f_{211}, f_{212}, f_{213}, f_{214}; 1, 1, 1, 1)$$

$$\text{Para } j = 2 \text{ dir}(f_{221}, f_{222}, f_{223}, f_{224}; 2, 4, 1, 1)$$

$$\text{Para } j = 3 \text{ dir}(f_{231}, f_{232}, f_{233}, f_{234}; 1, 3, 4, 2)$$

Por lo que de acuerdo al teorema 3.4 expresión (3.10) la estimación de la frecuencia relativa antes de recoger la muestra es:

Para la variable X_1 :

$$P(X_1 = 1) = \frac{a_{111}}{N_{11}} = \frac{2}{11} \quad P(X_1 = 2) = \frac{a_{112}}{N_{11}} = \frac{4}{11} \quad P(X_1 = 3) = \frac{a_{113}}{N_{11}} = \frac{5}{11}$$

Para la variable X_2 es:

$$P(X_2 = 1 / X_1 = 1) = \frac{a_{211}}{N_{21}} = \frac{1}{4} \quad P(X_2 = 2 / X_1 = 1) = \frac{a_{212}}{N_{21}} = \frac{1}{4} \quad P(X_2 = 3 / X_1 = 1) = \frac{a_{213}}{N_{21}} = \frac{1}{4}$$

$$P(X_2 = 4 / X_1 = 1) = \frac{a_{214}}{N_{21}} = \frac{1}{4} \quad P(X_2 = 1 / X_1 = 2) = \frac{a_{221}}{N_{22}} = \frac{1}{4} \quad P(X_2 = 2 / X_1 = 2) = \frac{a_{222}}{N_{22}} = \frac{1}{2}$$

$$P(X_2 = 3 / X_1 = 2) = \frac{a_{223}}{N_{22}} = \frac{1}{8} \quad P(X_2 = 4 / X_1 = 2) = \frac{a_{224}}{N_{22}} = \frac{1}{8} \quad P(X_2 = 1 / X_1 = 3) = \frac{a_{231}}{N_{23}} = \frac{1}{10}$$

$$P(X_2 = 2 / X_1 = 3) = \frac{a_{232}}{N_{23}} = \frac{3}{10} \quad P(X_2 = 3 / X_1 = 3) = \frac{a_{233}}{N_{23}} = \frac{2}{5} \quad P(X_2 = 4 / X_1 = 3) = \frac{a_{234}}{N_{23}} = \frac{1}{5}$$

Supongamos que obtenemos la siguiente muestra de cada una de las variables:

Tabla 1: Muestra de los nodos de la red de la figura 3.7

Observación	X_1	X_2
1	1	2
2	1	3
3	2	1
4	2	2
5	3	4
6	2	2
7	3	3
8	2	1
9	3	4
10	1	1
11	1	4
12	2	1
13	3	2
14	2	3
15	1	1

De la expresión (3.12) se tiene que s_{ijk} representa el número de casos en la muestra en la que los padres de X_i están en su j -ésima realización cuando $X_i = k$ por lo que de la tabla se tiene:

Para el nodo X_1 : $s_{111} = 5$, $s_{112} = 6$, y $s_{113} = 4$.

Para el nodo X_2 : $s_{211} = 2$, $s_{212} = 1$, $s_{213} = 1$, $s_{214} = 1$, $s_{221} = 3$, $s_{222} = 2$, $s_{223} = 1$, $s_{224} = 0$, $s_{231} = 0$, $s_{232} = 1$, $s_{233} = 1$, $s_{234} = 2$.

Por lo que de acuerdo a la expresión (3.12) la distribución de probabilidad actualizada de las variables F_{ij} es:

Para el nodo X_1 : $Dir(f_{111}, f_{112}, f_{113}; 9, 14, 14)$

Para el nodo X_2 se tiene para cada una de las realizaciones de su padre X_1 :

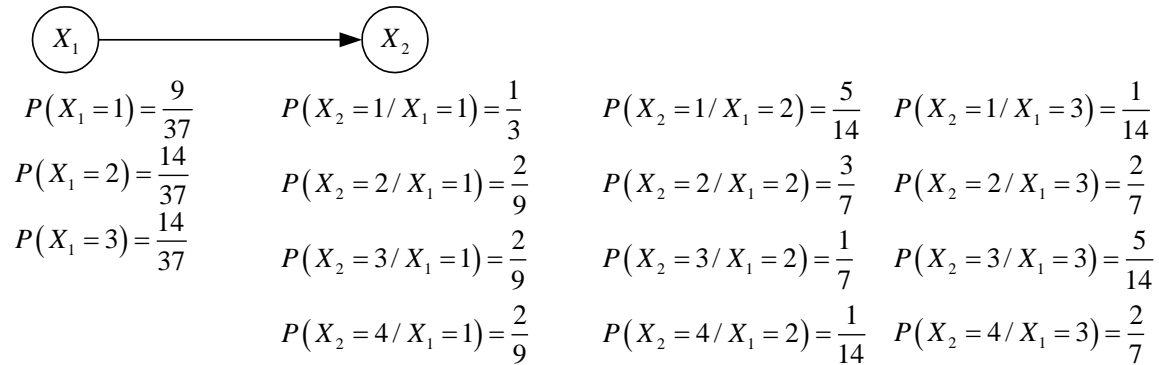
Para $j = 1$ $dir(f_{211}, f_{212}, f_{213}, f_{214}; 3, 2, 2, 2)$

Para $j = 2$ $dir(f_{221}, f_{222}, f_{223}, f_{224}; 5, 6, 2, 1)$

Para $j = 3$ $dir(f_{231}, f_{232}, f_{233}, f_{234}; 1, 4, 5, 4)$

Las probabilidades de los nodos de la red Bayesiana se actualizan usando (3.13) a partir de las distribuciones actualizadas, utilizando (3.12), las probabilidades actualizadas se muestran en la figura 3.8.

Figura 3.8: Red Bayesiana Actualizada.



La red bayesiana de la figura 3.8 es la que se utilizaría para realizar inferencias con respecto a los valores que pueden tomar las variables en la red.

3.3 Aprendizaje de la estructura.

En lo que sigue se mostrará como se estima la estructura de una red bayesiana a partir de una base de datos, es decir, como se representan las variables gráficamente, en general hay tres tipos de métodos para estimar la estructura de una red bayesiana:

1. **Métodos basados en la detección de independencias:** Estos métodos consisten en detectar relaciones de independencia (condicional o incondicional) entre las variables por medio de test estadísticos y construyen la red que represente el máximo número de independencias encontradas.

2. **Basados en funciones de evaluación y algoritmos de búsqueda:** Estos métodos consisten en que a partir de un conjunto de estructuras candidatas, se encuentra aquella que maximiza una medida de calidad, para decidir cual es la mejor de entre todas las candidatas y utiliza algoritmos de búsqueda para seleccionar un pequeño subconjunto de redes bayesianas de alta calidad de entre las cuales se seleccionará la mejor.

3. **Métodos Híbridos:** Esta clase de métodos es una combinación de los dos anteriores.

En esta parte se verá un algoritmo que se basa en una medida de calidad bayesiana para encontrar la estructura gráfica que represente a las variables. Para estimar la red bayesiana que mejor represente los datos supondremos que tenemos n variables aleatorias multinomiales, que disponemos de una muestra completa²⁷ de tamaño M de estas variables y que no hay variables ocultas.

La cantidad de grafos dirigidos acíclicos con que se pueden representar un conjunto de variables, crece exponencialmente con el número de variables, por ejemplo si tenemos 2 variables la cantidad de grafos es 3, con 3 variables el número de grafo acíclicos es 25, con cinco variables el número de grafos es 29,000 por lo que el algoritmo que se verá encuentra una configuración de red bayesiana que maximiza una medida de calidad, de tal forma que la configuración gráfica que proporciona el algoritmo si bien no es óptima se supone que esta cerca de la solución óptima.

De aquí en adelante representaremos la muestra de los datos por:

$$D = \{X^1, \dots, X^M\} \tag{3.15}$$

Donde:

$$X^j = \begin{pmatrix} X_1 = x_1^j \\ \vdots \\ X_n = x_n^j \end{pmatrix}$$

²⁷ Lo que se quiere decir es que para cualquier vector aleatorio de la muestra, este contiene una realización de cada una de las variables, por lo que no hay observaciones perdidas de una variable en la muestra.

3.3.1 Criterio Scoring

Si suponemos que tenemos un conjunto $G^g = \{G_1, \dots, G_g\}$ de grafos dirigidos acíclicos para representar las variables, lo que queremos es seleccionar aquella red que maximice la siguiente probabilidad:

$$P(G_j / D) = \frac{P(G_j)P(D/G_j)}{P(D)}$$

$$P(G_j / D) = \alpha P(G_j)P(D/G_j)^{28} \tag{3.16}$$

El término $P(D/G_j)$ que aparece en el lado derecho de la expresión (3.16) depende del grafo j -ésimo, por lo que se puede considerar como *la probabilidad marginal de los datos dada por el grafo j* . Puesto que estamos suponiendo que la distribución a priori de las variables es Dirichlet y al igual que en la sección anterior denotaremos por F_{ij} al conjunto de variables aleatorias que representa la probabilidad a priori de que la variable X_i tome sus valores dado que sus padres están en su j -ésima realización, se tiene lo siguiente:

$$P(D/G_j) = P(X^1, \dots, X^M)$$

En la expresión anterior en el lado derecho se ha omitido el término G_j , pero supondremos que los cálculos se realizan basados en el j -ésimo gráfico.

Puesto que los vectores de observaciones contenidas en (3.15) son independientes se tiene:

$$P(D/G_j) = \prod_{v=1}^M P(X^v)$$

²⁸ $\alpha = [P(D)]^{-1}$

$$P(D/G_j) = \prod_{v=1}^M P(x_1^v, \dots, x_n^v)$$

Factorizando el vector de observaciones según la distribución de probabilidad conjunta se tiene:

$$P(D/G_j) = \prod_{v=1}^M \prod_{z=1}^n P(x_z^v / \pi_z^v)$$

$$P(D/G_j) = \int_0^1 \int_0^1 \dots \int_0^1 \prod_{z=1}^n \prod_{v=1}^M P(x_z^v / \pi_z^v, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) \rho(f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) df_{11}, \dots, df_{nq_n} \quad (3.17)$$

De la expresión anterior se tiene lo siguiente:

$$\prod_{v=1}^M P(x_z^v / \pi_z^v, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) = P(x_z^1 / \pi_z^1, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) \dots P(x_z^M / \pi_z^M, f_{11}, \dots, f_{ij}, \dots, f_{nq_n})$$

Por lo que este producto se puede escribir como sigue:

$$\prod_{v=1}^M P(x_z^v / \pi_z^v, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) = \prod_{j=1}^{q_z} \prod_{k=1}^{r_z} P(X_z = k / \pi_{zj}, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) \quad (3.18)$$

Donde:

q_z : Es el número de realizaciones de los padres de la variable X_z .

r_z : Es el número de realizaciones de la variable X_z .

En la expresión (3.18) en la práctica solo aparecerán las realizaciones de las variables y de sus respectivos padres que se han observado. Por lo que sustituyendo en (3.17) se obtiene:

$$P(D/G_j) = \int_0^1 \int_0^1 \dots \int_0^1 \prod_{z=1}^n \prod_{j=1}^{q_z} \prod_{k=1}^{r_z} P(X_z = k / \pi_{zj}, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) \rho(f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) df_{11}, \dots, df_{nq_n}$$

$$P(D/G_j) = \prod_{z=1}^n \prod_{j=1}^{q_z} \int_0^1 \int_0^1 \dots \int_0^1 \prod_{k=1}^{r_z} P(X_z = k / \pi_{zj}, f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) \rho(f_{11}, \dots, f_{ij}, \dots, f_{nq_n}) df_{11}, \dots, df_{nq_n}$$

De la definición (3.2) y denotando por s_{zjk} al número de veces en que la variable X_z toma su k-ésimo valor, cuando sus padres están en su j-ésima realización es:

$$P(D/G_j) = \prod_{z=1}^n \prod_{j=1}^{q_z} \int_0^1 \int_0^1 \dots \int_0^1 \prod_{k=1}^{r_z} f_{zjk}^{s_{zjk}} \rho(f_{11}) \dots \rho(f_{zj}) \dots \rho(f_{nq_n}) df_{11}, \dots, df_{nq_n} \dots df_{nq_n}$$

$$P(D/G_j) = \prod_{z=1}^n \prod_{j=1}^{q_z} \int_0^1 \int_0^1 \dots \int_0^1 \prod_{k=1}^{r_z} f_{zjk}^{s_{zjk}} \rho(f_{zj1}, \dots, f_{zjk}, \dots, f_{zq_z r_z}) df_{zj1}, \dots, df_{zq_z r_z}$$

$$P(D/G_j) = \prod_{z=1}^n \prod_{j=1}^{q_z} E \left(\prod_{k=1}^{r_z} f_{zjk}^{s_{zjk}} \right) \tag{3.19}$$

Puesto que cada variable f_{zjk} tiene una distribución Dirichlet con parámetros a_{zj1}, \dots, a_{zjk} , $N_{zj} = \sum_{k=1}^{r_z} a_{zjk}$

por lo que (3.19) se puede escribir como sigue:

$$P(D/G_j) = \prod_{z=1}^n \prod_{j=1}^{q_z} \frac{\Gamma(N_{zj})}{\Gamma(N_{zj} + M_{zj})} \prod_{k=1}^{r_z} \frac{\Gamma(a_{zjk} + s_{zjk})}{\Gamma(a_{zjk})} \tag{3.20}$$

Donde:

$$M_{zj} = \sum_{k=1}^{r_z} s_{zjk}$$

La expresión (3.20) depende del tipo de grafo que se utilice, ya que en el cálculo del último término del lado derecho de (3.20) depende de los padres que tenga la variable que se este considerando y dado que en distintas configuraciones de red no todos los padres de las variables serán los mismos el resultado que se obtenga será diferente para cada configuración gráfica de las variables.

Para establecer la probabilidad a priori de las estructuras de una red bayesiana, se siguen los siguientes pasos:

1. Se asigna una probabilidad uniforme a todas las estructuras gráficas candidatas, cuando no se tiene conocimiento acerca de las probabilidades a priori de las estructuras.
2. Si se tiene un conjunto de redes, lo mejor sería preguntar a una persona que conozca sobre el problema que se está modelando, para que esa persona asigne las probabilidades a priori a las redes candidatas.

El cálculo de (3.16) para todas las posibles estructuras gráficas de las variables nos es conveniente, ya que el número de representaciones gráficas crece exponencialmente con el número de variables, por lo que se describirá un algoritmo llamado k_2 , que construye la estructura de una red bayesiana utilizando (3.20).

Algoritmo k_2

Este algoritmo fue propuesto por Cooper y Herskovits (1992) y se basa en un ordenamiento previo de las variables y supone que una variable no puede ser padre de sus predecesoras en el ordenamiento, la medida que se utiliza en este algoritmo para encontrar la configuración gráfica más probable es la siguiente expresión:

$$g(\pi_i) = \prod_{j=1}^{q_j} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})} \quad (3.21)$$

Al comparar (3.21) con (3.20) vemos que solo estamos tomando en cuenta los cálculos referentes a los padres de las variables, esto se debe a que el algoritmo supone un orden de las variables y que en un inicio no hay aristas entre ellas, y escoge como padres de una variable X_i aquellas variables predecesoras que dan el valor máximo de (2.21) el algoritmo es el siguiente.

Algoritmo k2.

1. Se ordenan las variables con que se estima la estructura de la red bayesiana.

Etapa iterativa

Para $i = 1$ hasta n

Repetir

Calculamos (3.21) para cada una de las variables anteriores a X_i y que no forman parte del conjunto de padres de la variable X_i y seleccionamos aquella variable Y de las anteriores a X_i en la ordenación con el mayor valor de (3.21) y que denotaremos como $g(\pi_i \cup Y)$ y calculamos la siguiente diferencia:

$$gr = g(\pi_i \cup Y) - g(\pi_i)$$

Si $gr > 0$

Agregamos la variable Y al conjunto de padres como sigue:

$$\Pi_i = \Pi_i \cup Y$$

Fin Si

Esto se repite hasta que $gr \leq 0$ o hasta que no haya más variables anteriores a X_i que puedan formar parte del conjunto de padres de X_i .

Fin para



Al utilizar el algoritmo anterior se tiene que al inicio, la primera variable de la ordenación no tiene ninguna variable anterior a ella, por lo que $gr = 0$ y se continua con la siguiente variable en la ordenación, puesto que en la segunda variable solo X_1 es anterior a ella en la ordenación se calcula $gr(\Pi_2 = \{X_1\})$ utilizando (3.21) suponiendo que esta variable es padre de X_2 y se calcula (3.21) en el caso de que la variable X_2 no tenga padres $gr(\Pi_2 = \phi)$, a continuación se calcula $gr = gr(\Pi_2 = \{X_1\}) - gr(\Pi_2 = \phi)$ y se agrega la variable X_1 al conjunto de padres de X_2 si $gr > 0$

puesto que ya no quedan más variables anteriores a X_2 , se continua con la siguiente variable en la ordenación X_3 , para esta variable se calcula $gr(\Pi_3 = \{X_1\})$ y $gr(\Pi_3 = \{X_2\})$ de estos dos cálculos se selecciona el que tiene el valor máximo, supongamos que el valor máximo es $gr(\Pi_3 = \{X_1\})$ y calculamos $gr = gr(\Pi_3 = \{X_1\}) - gr(\Pi_3 = \{\emptyset\})$ si este resultado es menor que cero pasamos a la variable X_4 que es la siguiente en la ordenación y dejamos a la variable X_3 sin padres, supongamos que el resultado es mayor a cero entonces agregamos la variable X_1 al conjunto de padres de la variable X_3 es decir $\Pi_3 = \{X_1\}$, puesto que todavía queda una variable anterior a X_3 en la ordenación se calcula $gr(\Pi_3 = \{X_1, X_2\})$ suponiendo que la variable X_2 ya forma parte del conjunto de padres y se calcula $gr(\Pi_3 = \{X_1\})$ y por último se calcula $gr = gr(\Pi_3 = \{X_1, X_2\}) - gr(\Pi_3 = \{X_1\})$, si $gr > 0$ el conjunto de padres de la variable X_3 sería $\Pi_3 = \{X_1, X_2\}$ y se continuaría con la siguiente variable en la ordenación, en caso de que $gr \leq 0$ el conjunto de padres de X_3 sería $\Pi_3 = \{X_1\}$ y así sucesivamente.

El ordenamiento de las variables se puede realizar a partir del conocimiento que se tenga de ellas, pero resulta ser un problema, cuando lo que queremos es determinar las relaciones que puedan existir entre las variables, es decir, cuando no podemos establecer a priori un ordenamiento de ellas, en la literatura existen otros algoritmos que no requieren un ordenamiento de las variables que se podrían utilizar cuando no se tiene un ordenamiento de ellas. En la siguiente aplicación se utiliza el algoritmo K2.

3.4 Aplicación (Enfermedad Coronaria).

Estimar una red bayesiana para predecir la enfermedad coronaria dado algunos síntomas que presente una persona y alguna prueba que se haya realizado. (Enfermedad coronaria: es que la persona tenga tapadas las venas en más del 50% de su diámetro²⁹), para la estimación de la red bayesiana se utilizó una base de datos que contenía 296 casos de pacientes y en cada paciente se estudiaron las variables, sexo, edad, abstinencia en grasas, colesterol, enfermedad coronaria, dolor de pecho, hipertensión, electrocardiograma y ejercicio de angina. La base de datos que se utilizó fue obtenida de un repositorio de aprendizaje de Internet, la cual fue donada por "V.A. Medical Center, Long Beach and Cleveland Clinic Foundation"

²⁹ En este ejemplo se considera un número muy reducido de variables para el diagnóstico de este tipo de enfermedad, aquí se presenta con el objetivo de conocer como se estima una red bayesiana por medio de una base de datos aplicando el algoritmo k2 y a la vez conocer un poco como se aplican las redes Bayesianas al diagnóstico médico.

3.4.1 Descripción del corazón.

El corazón es un órgano muscular situado en medio del tórax que posee, tanto en el lado derecho como en el izquierdo, una cavidad superior (aurícula), que recibe la sangre, y una cavidad inferior (ventrículo), que la expulsa. Para asegurarse de que la sangre fluya en una sola dirección, los ventrículos tienen una válvula de entrada y otra de salida.

Las funciones primarias del corazón consisten en proporcionar oxígeno a todo el organismo y, al mismo tiempo, liberarlo de los productos de desecho (anhídrido carbónico). En concreto, esta función supone recoger la sangre del organismo, pobre en oxígeno, y bombearla hacia los pulmones, donde se oxigena y libera el anhídrido carbónico; luego el corazón conduce esta sangre rica en oxígeno hacia todos los tejidos del organismo.

3.4.2 Problemas con las arterias del corazón.

Para poder llevar a cabo la tarea de bombear sangre, el músculo cardíaco necesita un aporte correcto de sangre rica en oxígeno, lo que se garantiza a través de la red que constituyen las arterias coronarias. La enfermedad coronaria es el resultado final de Aterosclerosis, una enfermedad caracterizada por un endurecimiento de las paredes de las arterias, que provoca un estrechamiento de las mismas. A este estrechamiento se le denomina Estenosis, y causa una dificultad de la llegada de la sangre oxigenada al corazón. Esta falta de sangre oxigenada daña el tejido cardíaco, provocando una situación llamada Isquemia, y si la arteria se tapona en su totalidad, el daño puede ser tan grande que produce la muerte de las células del corazón, lo que se conoce como Infarto del Miocardio.

La Aterosclerosis está provocada por la presencia en cantidades excesiva de unas partículas inestables llamadas radicales libres, que se unen y alteran a otras moléculas en un proceso conocido como oxidación. Cuando los radicales libres llegan a las paredes de las arterias reaccionan con las lipoproteínas de baja densidad, también llamadas LDL o “Colesterol malo”, oxidándolas. Las LDL crean entonces una capa de colesterol oxidado en las paredes de las arterias y este colesterol, con el paso del tiempo se va acumulando y endureciendo, creando lo que se conoce como placas de colesterol. Sin embargo, el proceso no acaba aquí, el daño a las arterias provoca una alerta en el sistema inmune, el que reacciona enviando a la zona glóbulos blancos (especialmente neutrófilos y macrófagos), iniciando una respuesta inflamatoria. Los macrófagos, literalmente se comen los cuerpos extraños, en este caso el colesterol oxidado, y se transforman en células espumosas, las que se adhieren a las células musculares de las paredes de las arterias, haciéndolas crecer. El sistema inmune, también detecta esta lesión y produce unas

sustancias llamadas citoquinas que estimulan el envío de más glóbulos blancos a la zona, perpetuando la reacción y creando, por tanto, un círculo vicioso. La lesión en las paredes internas de las arterias hace que las células no sean capaces de secretar el suficiente óxido nítrico, una sustancia de vital importancia para mantener la elasticidad de las arterias, por lo que las arterias lesionadas se calcifican y se vuelven cada vez más rígidas, enlenteciendo la velocidad de la sangre. Cuando en la placa de colesterol se desarrollan fisuras, las plaquetas se adhieren a ella para sellar esta lesión y forman un trombo en la arteria, lo que tapona por completo el paso de sangre produciendo el Infarto de Miocardio.

Los síntomas más comunes de la enfermedad coronaria son:

1. cansancio severo.
2. falta de aire.
3. Dolor en el cuello o en el maxilar inferior.

Otro síntoma de la enfermedad coronaria es:

La angina de pecho indica la presencia de un dolor a nivel del pecho causada por un trastorno en la circulación que irriga la musculatura del corazón y habitualmente suele ser el antecesor de un ataque agudo al corazón (infarto de miocardio). La angina de pecho es una opresión o dolor temporal que se inicia en el pecho y a veces se extiende hacia la parte superior del tórax y cuello. Comienza de repente y por lo general dura pocos minutos. Es causada por el cierre transitorio (espasmo) de las arterias que llevan la sangre y el oxígeno al corazón (coronarias). Puede acompañar a situaciones que exigen esfuerzo al corazón, por ejemplo, al haber tensión emocional, al hacer ejercicios o exponerse a temperaturas extremas. En los casos agudos, cuando los ataques son impredecibles y cada vez más difíciles de controlar, se la llama angina inestable y puede ser el anuncio del ataque al corazón.

Síntomas.

- Dolor torácico y sensación de opresión aguda y sofocante, generalmente detrás del esternón, y a veces extendida a uno u otro brazo. El dolor torácico suele durar entre 1 – 2 minutos y 10 – 15 minutos (a veces se percibe una sensación de pesadez u opresión en el pecho que no llega a dolor).
- Sentimiento de ansiedad o de muerte inminente.

- Sudoración profusa.
- Palidez

La angina de pecho no es una enfermedad. Es el resultado directo de la falta de sangre en el músculo cardíaco, lo que se conoce con el nombre de isquemia.

La angina del pecho se produce cuando el corazón no recibe la cantidad necesaria de oxígeno y otras sustancias. Generalmente sucede porque depósitos grasos obstruyen las arterias coronarias, que rodean y nutren el corazón (aterosclerosis). El dolor se produce por la necesidad de oxígeno del músculo cardíaco, condición que se conoce como isquemia. Algunas personas pueden sufrir un ataque de angina cuando no están esforzando el corazón, aun al dormir. Los científicos creen que este tipo de angina llamado variante o latente, se produce por una contracción de las arterias coronarias, usualmente en las partes ateroscleróticas.

Angina de pecho estable: cuando el dolor cede con el reposo.

Angina de pecho inestable: cuando el dolor se hace prolongado, no cede con el reposo, y se corre el riesgo inmediato de llegar a un infarto. Generalmente se debe a obstrucción de la luz de las arterias coronarias en más de 70% y rotura de una placa de colesterol.

Entre los factores de riesgo de los problemas coronarios están:

1. El sexo, el hombre esta más propenso a enfermedades cardiacas que la mujer antes de la menopausia.
2. Edad, entre más edad tenga la personas mayor es el riesgo.

Entre algunos de los factores que predisponen al desarrollo de una enfermedad coronaria se mencionan los siguientes:

- *Colesterol.* Los niveles normales de colesterol totales deben de ser iguales o menores de 200 mg/dl. Si su número se encuentra entre 200 y 239, usted está en riesgo de tener un ataque al corazón. Si su número es 240 o más, usted tiene un alto nivel de colesterol en la sangre y usted tiene más alto riesgo de sufrir un ataque al corazón.

- *Hipertensión.* Dos números se usan para medir la presión de la sangre. El número más alto, *la presión sistólica*, refiere a la presión dentro de la arteria cuando el corazón se contrae y está bombeando la sangre a través del cuerpo. El número más bajo, *la presión diastólica*, se refiere a la presión dentro de la arteria cuando el corazón está descansando y se está llenando de sangre. Tanto la presión sistólica como la diastólica se miden en “mm Hg” (milímetros de mercurio). Estos registros representan que tan alta la columna de mercurio es levantada por la presión de su sangre.

Se considera presión arterial alta cuando:

- La presión sistólica de 140 mm Hg o mayor.
- Presión diastólica de 90 mm Hg o mayor.

Se tiene prehipertensión cuando:

- La presión sistólica se encuentra entre 120 mm Hg – 139 mm Hg.
- La presión diastólica se encuentra entre 80 mm Hg – 89 mm Hg.

Se tiene la presión de sangre normal cuando:

- Presión sistólica de 120 mm Hg o menor.
- Presión diastólica de 80 mm Hg o menor.

Pruebas para detectar problemas cardiacos.

El electrocardiograma: es la representación gráfica de la actividad eléctrica del corazón detectada a través de una serie de electrodos colocados en la superficie corporal. Las aplicaciones de esta prueba en atención primaria son varias siendo las principales las siguientes: el estudio y la evolución de la cardiopatía isquémica, la identificación y el control del tratamiento de las arritmias, el control evolutivo de la repercusión de la hipertensión arterial y de las vasculopatías sobre el corazón y la valoración de trastornos metabólicos o iónicos.

Análisis de sangre: se hacen siempre que se quiere evaluar el estado del sistema cardiovascular. Interesa, entre otras mediciones, el nivel de colesterol total y de sus fracciones: el de baja densidad (IDL o malo) y el de alta densidad (HDL o bueno), también interesan especialmente los niveles de azúcar (la diabetes aumenta el riesgo de cardiopatía) y de hormonas tiroideas (un tiroides hiperactivo). Si en la sangre hay

menos de 100 miligramos por decilitro (mg/dl) se considera normal y la persona esta saludable, si el nivel de azúcar se encuentra entre 100 y 125 mg/dl se considera que la persona tiene pre-diabetes o diabetes “Fronteriza”, lo que ocasiona en la persona mayor riesgo de enfermedad cardiovascular y futuras diabetes. Y si su nivel de la sangre es mayor o igual de 126 mg/dl en dos días diferentes en que se mide el azúcar corre el riesgo de daño en los ojos, riñones, vasos sanguíneos, corazón y nervios, así como declinación cognitiva y demencia.

3.4.3 Variables en la base de datos.

A continuación se presenta los nombres de las variables que se encuentran en la base de datos y como éstas están codificadas de acuerdo a lo comentado anteriormente.

- *Edad.*

Edad 1: $29 \leq edad \leq 39$

Edad 2: $40 \leq edad \leq 50$

Edad 3: $51 \leq edad \leq 61$

Edad 4: $62 \leq edad \leq 72$

Edad 5: $73 \leq edad \leq 83$

- *Sexo.*

1 = Masculino.

2 = Femenino.

- *Dolor_Pecho.*

1. Angina típica.

2. Angina Atípica.

3. No hay dolor anginal.

4. Asintomático.

- *Colesterol.*

1. Normal: $colesterol < 200$.

2. Riesgo: $200 \leq colesterol \leq 239$

3. Alto _ riesgo: $colesterol \geq 240$

- *abstinencia _ de_grasas*³⁰.

Esta variable se refiere a la abstinencia de azúcar en la sangre > 120 mg/dl, y puede tomar el valor de “VERDADERO” si los niveles de azúcar es menor de 120 y “FALSO”, si es mayor de 120.

- *Electrocardiograma.*

1. Normal.
2. Anormal_1: tiene una onda ST-T anormal.
3. Anormal_2: define una hipertrofia³¹ ventricular izquierda.

- *Hipertensión.*

Esta variable se refiere a la velocidad máxima del corazón en el momento de ser ingresado en el hospital, está codificada de la siguiente manera:

1. Normal: presión_sistólica < 120.
2. Prehipertensión: $120 \leq \text{presión_sistólica} \leq 139$
3. Alta: presión_sistólica ≥ 140 .

- *Ejercicio _ angina.*

Esta variable se utiliza para conocer si se tiene angina de pecho, y toma el valor de “Si” si la persona soporta el ejercicio y el valor de “No” si no lo hace.

- *Enfermedad Coronaria*

Esta variable toma el valor de “ausente” si el porcentaje de estrechamiento de la vena es menor del 50% de su diámetro y toma el valor de presente si el estrechamiento de la cedula es mayor del 50% del diámetro.

³⁰ Cuando se tiene niveles elevados de azúcar el hígado la transforma en grasas y esta es conducida por la sangre al corazón.

³¹ En este caso se refiere al aumento del tamaño del ventrículo izquierdo.

Ordenamiento de las variables y estimación de la red bayesiana.

El orden que se ha escogido de las variables es el siguiente³²:

1. edad
2. sexo
3. abstienencia_de_grasas.
4. Colesterol

Estas variables aparecen primero, ya que las primeras 2 variables constituyen factores de riesgo para desarrollar la enfermedad y al tener niveles elevados de azúcar y colesterol se puede desarrollar la enfermedad. El orden de las demás variables se muestra a continuación.

5. Enfermedad Coronaria.
6. Dolor de pecho.
7. Hipertensión.
8. Electrocardiograma.
9. Ejercicio _ angina.

Las variables “dolor de pecho” e “Hipertensión” representan síntomas de que las venas del corazón estén estrechadas y las variables “electrocardiograma” y “ejercicio _ angina” representan pruebas que se realizan para detectar problemas cardiovasculares.

La red bayesiana estimada aplicando el algoritmo k2 se muestra en la figura 3.9. Esta red bayesiana, no es exactamente la red bayesiana que estima el algoritmo k2, la red que estima este algoritmo deja desconectado el nodo “Colesterol”, pero dado que sabemos que esta variable influye en que las venas del corazón se obstruyan, se impone la restricción de que se incluya. La red de la figura 3.9, se estimó utilizando el algoritmo k2, pero imponiendo la restricción de que el nodo “Colesterol” pase a ser automáticamente el padre del nodo “enfermedad Coronaria”. En la figura 3.10 se pueden apreciar las probabilidades a priori de los nodos de la figura 3.9.

³² Para el ordenamiento de las variables se contó con la colaboración del doctor Carlos Enrique Ávila director del Centro de Investigación Docente Estudiantil (CIDE).

De acuerdo a lo comentado anteriormente se puede observar que la representación gráfica de la red (figura 3.9) concuerda con lo esperado, excepto el enlace entre los nodos “abstinencia _ azúcar” con el nodo “Ejercicio de angina”. Por lo que el algoritmo k2 genera con bastante precisión una red bayesiana que represente bastante bien el conocimiento que se quiera modelar.

Figura 3.9: Red bayesiana estimada mediante el algoritmo k2.

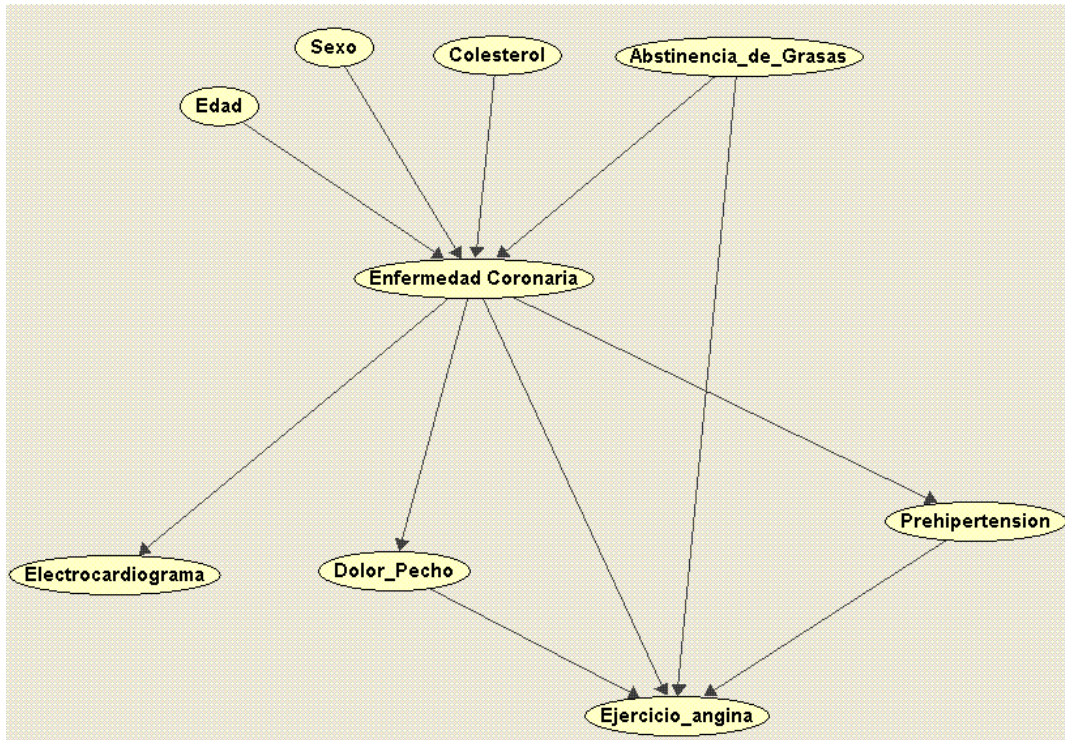
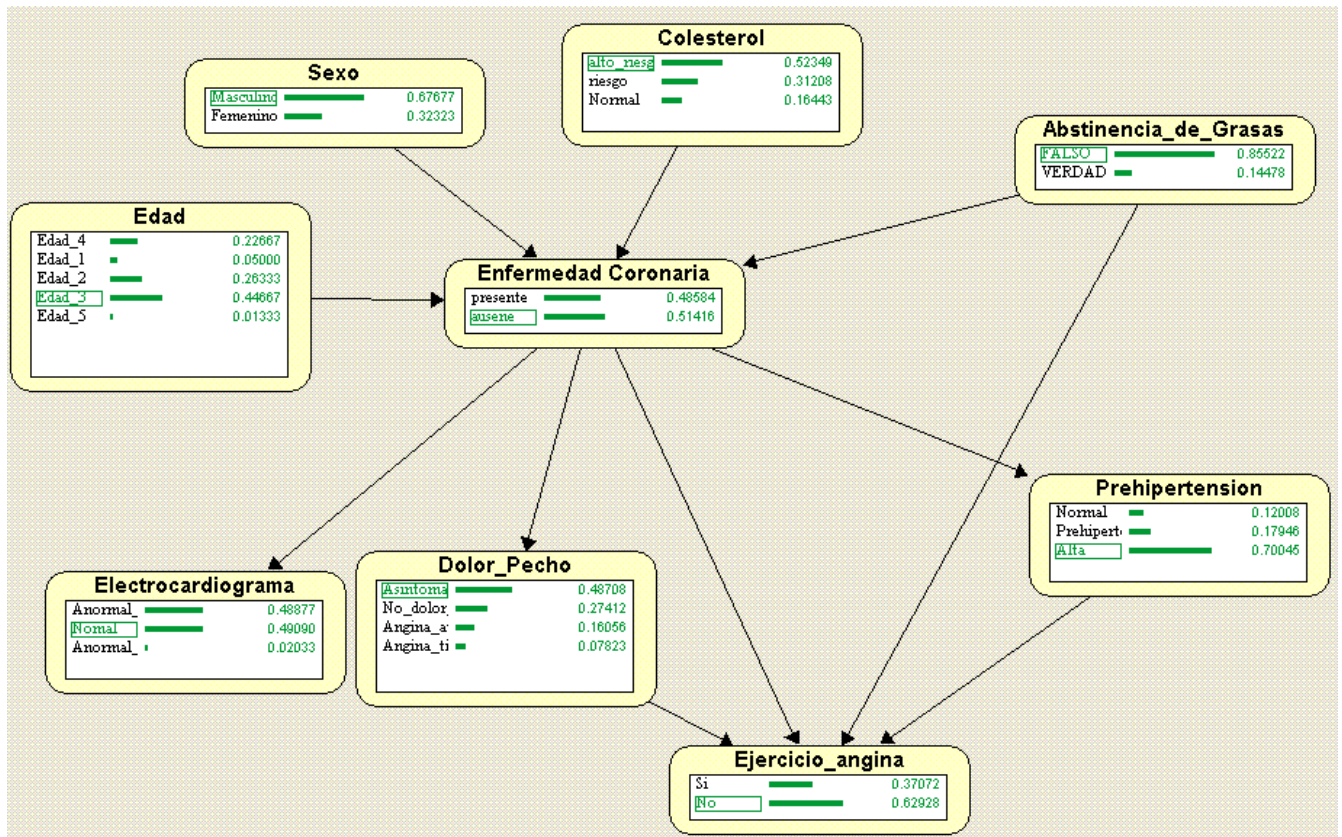


Figura 3.10: Probabilidades a priori de la red de la figura 3.9.



Con la red de la figura 3.9, se pueden hacer un poco de inferencia con respecto al nodo enfermedad Coronaria, en la tabla 2, se presenta la probabilidad de que una persona tenga las venas coronarias estrechadas en más del 50% de su diámetro dado los valores que tomen algunas variables, en las columnas de la tabla aparecen los nombres de las variables y en las filas los valores que se han observado de las variables, las celdas de la tabla que tienen un guión significa que no se ha observado ningún valor de dicha variable.

Tabla 2: Probabilidades de la red.

Edad	Sexo	Colesterol	Abstinencia Grasas	Electrocardiograma	Dolor de pecho	Prehipertensión	Ejercicio Angina	$P(EC > 50\% / e)$ ³³
Edad_1	Masculino	Normal	-	-	-	-	-	0.41448.
Edad_3	-	Riesgo	-	-	-	Prehipertensión	-	0.69394.
Edad_5	Femenino	-	-	Anormal_1	-	-	-	0.71535.
Edad_5	Masculino	-	-	Anormal_1	-	-	-	0.75896.
-	-	-	Verdadero	-	Angina típica	-	Si	0.46351.
Edad_3	Masculino	Alto Riesgo	Falso	Anormal_2	Asintomático	Alta	Si	0.92749.
Edad_1	Masculino	Alto Riesgo	Falso	Anormal_1	Angina Atípica	Alta	No	0.18925.

De la tabla se tiene lo siguiente:

- Si tenemos una persona de sexo masculino entre 51 y 61 años de edad con un nivel de colesterol mayor o igual a 240 mg/dl, un nivel de azúcar por arriba de 120 mg/dl, con hipertensión alta, es asintomático al dolor de pecho, y resiste el ejercicio de angina, entonces la probabilidad de que esta persona tenga las arterias coronarias estrechadas por arriba del 50% de su diámetro es de 0.92749, es decir, una persona en el que se observe estos valores en las variables es aproximadamente 13 veces más probable que tenga las venas coronarias estrechadas más del 50% de su diámetro.
- Uno de los casos menos probables que ocurre según la tabla 2, es cuando se tiene una persona del sexo masculino con edad entre los 29 y 39 años, con un nivel de colesterol por arriba de 240 mg/dl, con un nivel de azúcar por arriba de los 120 mg/dl, presenta dolor anginal atípico, prehipertensión alta, no resiste el ejercicio de angina, el

³³ CE: Enfermedad Coronaria.

Capítulo 3. Introducción a la estimación de una Red Bayesiana.

electrocardiograma indica una onda ST-T anormal, por lo que tiene una probabilidad de tener las venas coronarias estrechadas por arriba del 50% de su diámetro de 0.18925.

BIBLIOGRAFÍA.

1. Castillo, Enrique; Gutiérrez, José Manuel y Ali S Hadi. (1997). “Sistemas Expertos y Modelos de Redes Probabilistas”, Academia de Ingeniería, Madrid.
2. Díez Vegas, Francisco Javier (1994). “Sistema Experto para Ecocardiografía”, Universidad Autónoma de Madrid.
3. Fernández Galán, Severino (2002). “Redes Bayesianas Temporales: Aplicaciones Médicas e Industriales”, Universidad Complutense de Madrid.
4. Gutiérrez, José M; Cano, Rafael y Corfiño, Antonio S (2004). “Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas”. Ministerio del medio ambiente, Dirección General del Instituto Nacional de Metereología.
5. Hernández Molinero, Luís Daniel (1995). “Diseño y Validación de Nuevos Algoritmos para El Tratamiento de Grafos de Dependencias”. Universidad de Granada.
6. Newman, D.J. & Hettich, S. & Blake, C.L. & Merz, C.J. (1998). UCI Repository of machine learning databases [<http://www.ics.uci.edu/~mlearn/MLRepository.html>]. Irvine, CA: University of California, Department of Information and Computer Science.
7. Richard E. Neapolitan, (2003). “Learning Bayesian Networks”. Prentice Hall.
8. Rodero, Carmen Lacave, (2002). “Explicación en Redes Bayesianas Causales, Aplicaciones Médicas”, Universidad Complutense de Madrid.

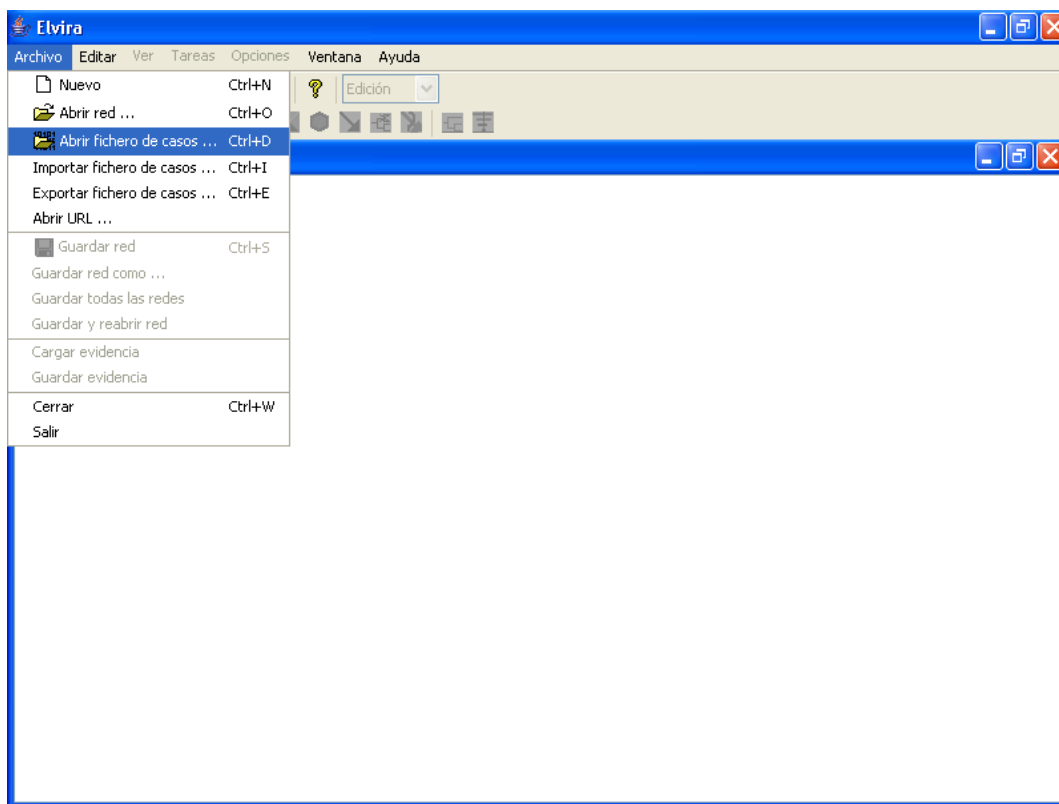
A. ANEXOS.

A.1 Estimación de la red bayesiana utilizando el software ELVIRA versión 0.162

Para estimar la red bayesiana a partir de una base de datos utilizando el programa ELVIRA, se realizó los siguientes pasos³⁴:

1. Se da clic en el menú “Archivo” y a continuación se da clic en la opción “Abrir fichero de casos” como se muestra en la figura 3.11.
2. Aparecerá una ventana como la que se muestra en la figura 3.12, presionamos el botón seleccionar y buscamos el archivos de casos y lo seleccionamos.

Figura 3.11: Entorno de ELVIRA.



3. A continuación seleccionamos la viñeta “aprendizaje automático” que se muestran en la figura 3.12, se mostrará una ventana como la que se observa en la figura 3.13, en la ventana que se

³⁴ Supondremos que la base de datos se encuentra en formato .dbc (data base cases) y que no contiene datos faltantes.

muestra en la figura 3.13, seleccionamos la sub_viñeta titulada “Factorización”, el aspecto de la ventana cambiará al mostrado en la figura 3.14.

4. En la casilla “MÉTODO DE APRENDIZAJE” aparece seleccionado automáticamente la opción “k2 Learnig”, que se refiere al algoritmo k2, con el cual estimaremos la red bayesiana, en la figura 3.14, nos muestra que el método de estimación que aparece seleccionado por defecto es el “Bayesiano”, por lo que damos clic en el botón, “Procesar”. Y el programa estima la red bayesiana y cuando haya terminado la mostrará.

Figura 3.12: Cuadro “Abrir Fichero de casos”

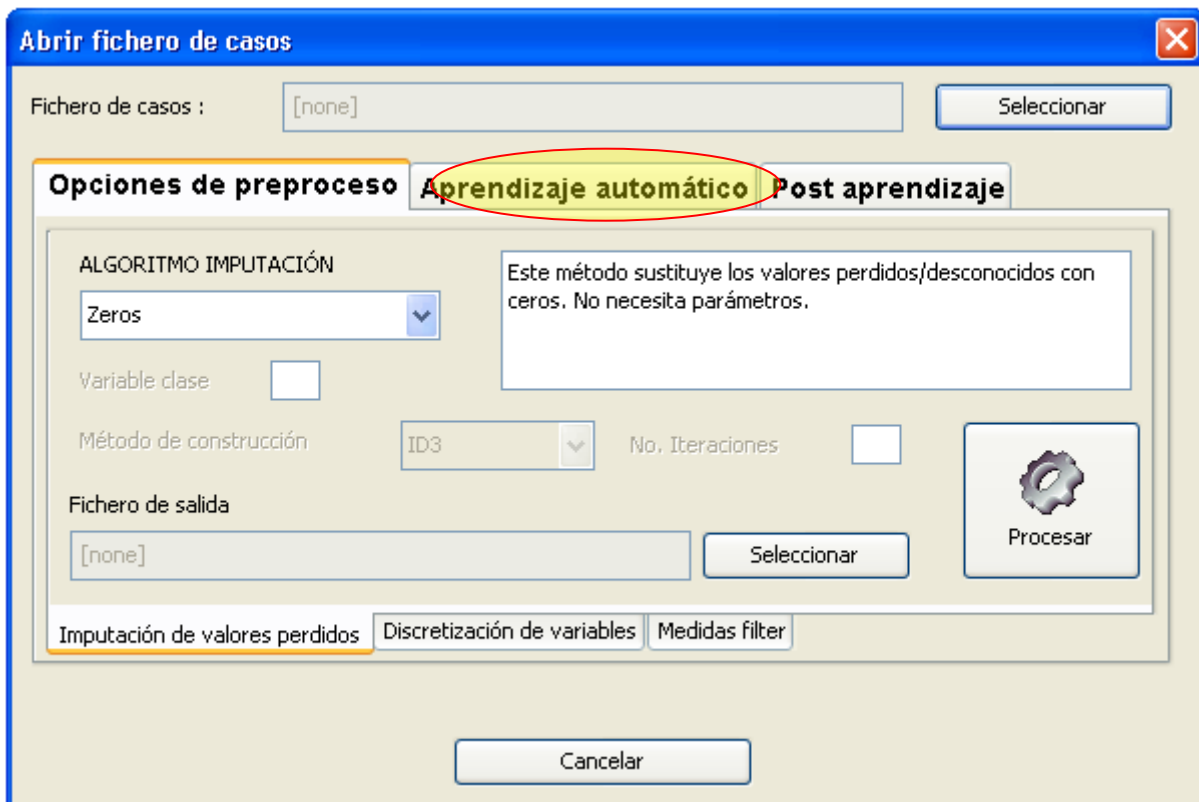


Figura 3.13: Viñeta Aprendizaje automático.

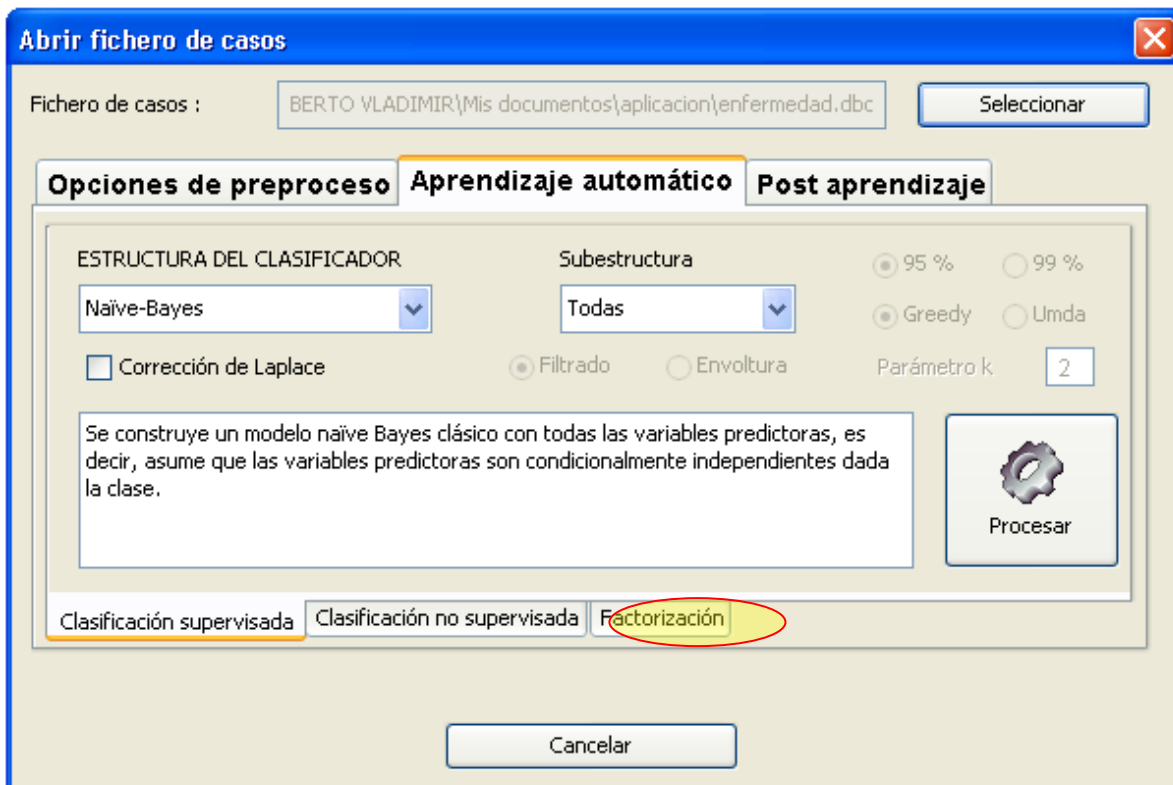
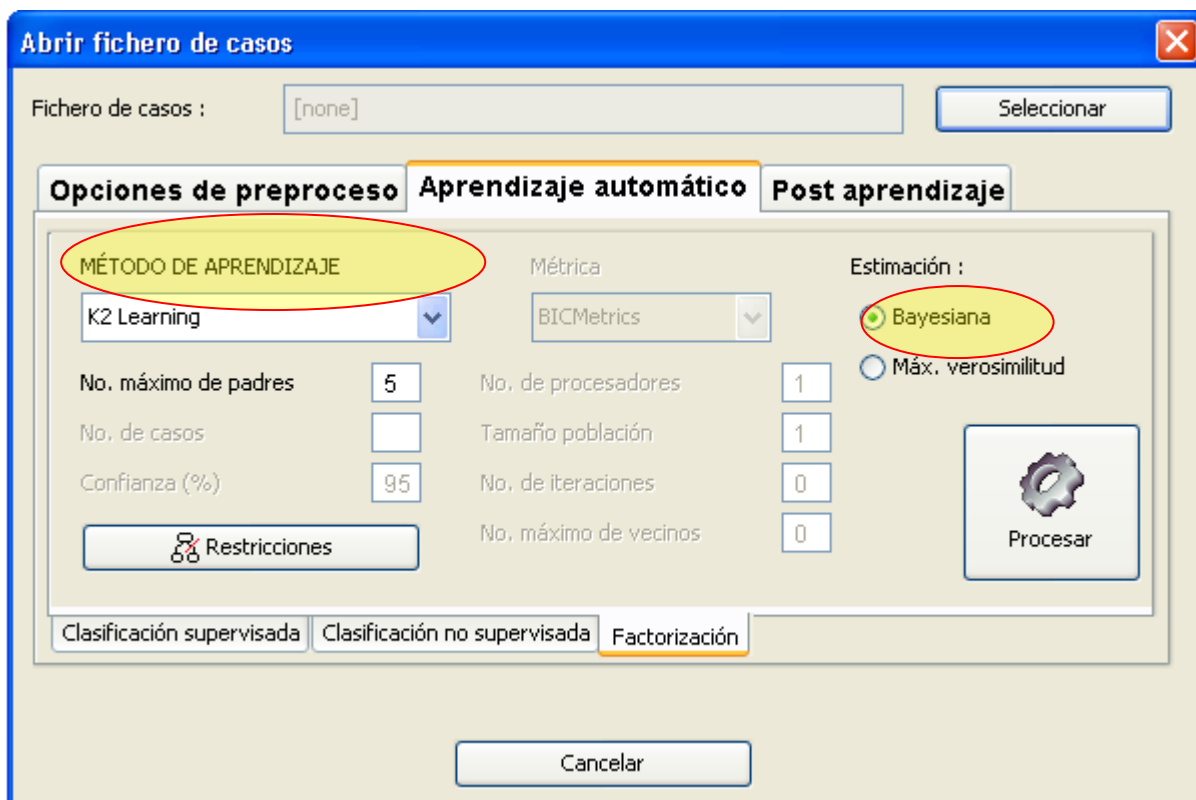


Figura 3.14: Cuadro de la sub. Viñeta "Factorización".



A.2 Datos Utilizados para Estimar la Red Bayesiana de la enfermedad Coronaria.

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_4	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_1	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Femenino	FALSO	riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Nomal	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_3	Masculino	VERDADERO	riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	Normal	ausente	Asintomático	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	VERDADERO	alto_riesgo	presente	No_dolor_anginal	Alta	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	presente	Angina_atipica	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Prehipertensión	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_4	Masculino	FALSO	riesgo	ausente	Angina_tipica	Alta	Anormal_2	Si
Edad_3	Femenino	VERDADERO	alto_riesgo	ausente	Angina_tipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_2	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Femenino	FALSO	riesgo	ausente	Angina_tipica	Normal	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	Normal	presente	Asintomatico	Normal	Anormal_2	Si

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_4	Femenino	FALSO	riesgo	ausente	Angina_típica	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	Si
Edad_2	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	Normal	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Nomal	Si
Edad_3	Masculino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Prehipertensión	Nomal	Si
Edad_4	Femenino	FALSO	riesgo	presente	Asintomático	Normal	Anormal_2	No
Edad_2	Masculino	FALSO	Normal	ausente	Angina_típica	Alta	Nomal	Si
Edad_4	Femenino	FALSO	alto_riesgo	ausente	Angina_atípica	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	Normal	ausente	No_dolor_anginal	Prehipertensión	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Anormal_2	No
Edad_4	Femenino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	VERDADERO	Normal	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Femenino	FALSO	Normal	ausente	Angina_atípica	Alta	Nomal	No
Edad_4	Masculino	FALSO	Normal	ausente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	riesgo	ausente	Angina_atípica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_2	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Masculino	FALSO	Normal	presente	Asintomático	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	Angina_típica	Prehipertensión	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_2	Femenino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Anormal_2	Si

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_3	Masculino	FALSO	riesgo	presente	Asintomatico	Prehipertención	Anormal_2	Si
Edad_3	Femenino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	presente	Asintomatico	Normal	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	Normal	presente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_2	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Masculino	VERDADERO	alto_riesgo	presente	Asintomatico	Alta	Nomal	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Normal	Nomal	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	Normal	presente	Asintomatico	Alta	Anormal_2	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Normal	Anormal_2	Si
Edad_2	Masculino	FALSO	riesgo	ausente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_1	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_4	Masculino	VERDADERO	alto_riesgo	presente	No_dolor_anginal	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_4	Masculino	FALSO	alto_riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_4	Femenino	FALSO	Normal	presente	Asintomatico	Alta	Anormal_2	No
Edad_4	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Femenino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_1	Masculino	FALSO	Normal	ausente	Angina_tipica	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Asintomatico	Alta	Anormal_2	No
Edad_4	Femenino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Prehipertensión	Anormal_2	No
Edad_2	Masculino	FALSO	Normal	presente	No_dolor_anginal	Prehipertensión	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	presente	Asintomatico	Alta	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Nomal	Si
Edad_1	Masculino	FALSO	riesgo	presente	Asintomatico	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Masculino	VERDADERO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	Normal	ausente	Angina_tipica	Alta	Anormal_2	No
Edad_2	Femenino	VERDADERO	alto_riesgo	presente	Asintomatico	Prehipertensión	Anormal_2	Si
Edad_4	Femenino	FALSO	alto_riesgo	presente	No_dolor_anginal	Normal	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	Angina_atipica	Prehipertensión	Nomal	No
Edad_3	Masculino	VERDADERO	riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_1	Femenino	FALSO	Normal	ausente	Asintomatico	Alta	Nomal	No
Edad_4	Masculino	VERDADERO	alto_riesgo	presente	Asintomatico	Prehipertensión	Anormal_2	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Prehipertensión	Anormal_2	No
Edad_2	Masculino	VERDADERO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_4	Femenino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	presente	Asintomatico	Normal	Nomal	Si
Edad_4	Masculino	VERDADERO	alto_riesgo	presente	Angina_tipica	Alta	Anormal_2	No
Edad_2	Femenino	FALSO	riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Femenino	VERDADERO	alto_riesgo	presente	Asintomatico	Prehipertensión	Anormal_2	Si

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_3	Masculino	FALSO	riesgo	presente	Asintomatico	Prehipertención	Nomal	Si
Edad_2	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_4	Femenino	FALSO	riesgo	ausente	Asintomatico	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	Si
Edad_1	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Asintomatico	Alta	Anormal_2	Si
Edad_2	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_4	Masculino	FALSO	Normal	presente	Asintomatico	Prehipertención	Nomal	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Angina_atipica	Normal	Anormal_2	No
Edad_1	Masculino	FALSO	Normal	presente	Asintomatico	Prehipertención	Nomal	Si
Edad_3	Masculino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Angina_tipica	Alta	Anormal_2	No
Edad_3	Masculino	VERDADERO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	No_dolor_anginal	Prehipertención	Nomal	Si
Edad_3	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	presente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	alto_riesgo	presente	Asintomatico	Prehipertención	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	alto_riesgo	ausente	Angina_tipica	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Asintomatico	Prehipertención	Anormal_2	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Normal	Anormal_2	Si
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Normal	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Nomal	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomatico	Alta	Anormal_2	No

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_4	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_2	Masculino	VERDADERO	Normal	ausente	Angina_atipica	Alta	Nomal	No
Edad_5	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_3	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Prehipertensión	Anormal_2	No
Edad_2	Masculino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	Si
Edad_3	Femenino	VERDADERO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	Si
Edad_1	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_2	Femenino	FALSO	Normal	ausente	Angina_atipica	Prehipertensión	Nomal	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	No_dolor_anginal	Normal	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	ausente	Asintomático	Normal	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_4	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Anormal_2	No
Edad_4	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Nomal	Si
Edad_3	Masculino	VERDADERO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	presente	Asintomático	Normal	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_típica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Angina_típica	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_4	Femenino	FALSO	Normal	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	VERDADERO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	Angina_atipica	Prehipertensión	Nomal	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Angina_atipica	Alta	Anormal_2	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Nomal	Si

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_4	Femenino	VERDADERO	alto_riesgo	presente	Asintomático	Normal	Nomal	No
Edad_4	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Normal	Anormal_2	No
Edad_4	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_4	Masculino	VERDADERO	riesgo	ausente	Angina_típica	Prehipertensión	Anormal_2	No
Edad_2	Femenino	FALSO	riesgo	ausente	Asintomático	Alta	Anormal_2	Si
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Angina_atípica	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Angina_típica	Prehipertensión	Anormal_2	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Anormal_2	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Nomal	Si
Edad_3	Masculino	VERDADERO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Prehipertensión	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_2	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Angina_atípica	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_1	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_1	Masculino	FALSO	riesgo	presente	Angina_típica	Alta	Nomal	Si
Edad_2	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_4	Femenino	VERDADERO	riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	ausente	Angina_típica	Alta	Anormal_2	No
Edad_2	Femenino	FALSO	riesgo	ausente	Angina_atípica	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Anormal_2	Si
Edad_4	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Prehipertensión	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Anormal_2	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	Si
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_1	Femenino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Nomal	Si
Edad_4	Femenino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_1	Femenino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_4	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_3	Femenino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_1	Si
Edad_2	Masculino	FALSO	Normal	presente	No_dolor_anginal	Prehipertensión	Anormal_2	No
Edad_5	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Prehipertensión	Anormal_2	Si
Edad_3	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_3	Masculino	VERDADERO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_2	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Angina_típica	Alta	Nomal	No
Edad_3	Femenino	VERDADERO	Normal	ausente	No_dolor_anginal	Normal	Nomal	No
Edad_4	Masculino	FALSO	riesgo	presente	Asintomático	Normal	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Normal	Anormal_2	Si
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Alta	Nomal	No
Edad_4	Masculino	VERDADERO	riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	riesgo	ausente	Asintomático	Prehipertensión	Nomal	Si
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Normal	Nomal	No
Edad_4	Masculino	FALSO	alto_riesgo	ausente	Asintomático	Normal	Nomal	Si
Edad_3	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	ausente	Asintomático	Alta	Nomal	No
Edad_2	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_4	Femenino	FALSO	riesgo	ausente	Asintomático	Alta	Nomal	No

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_5	Femenino	FALSO	Normal	ausente	No_dolor_anginal	Normal	Anormal_1	No
Edad_4	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Angina_atipica	Alta	Nomal	No
Edad_2	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	VERDADERO	alto_riesgo	presente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Angina_tipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Nomal	Si
Edad_3	Masculino	VERDADERO	riesgo	presente	No_dolor_anginal	Prehipertensión	Nomal	No
Edad_2	Masculino	FALSO	riesgo	presente	Asintomático	Alta	Nomal	No
Edad_2	Masculino	FALSO	Normal	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_2	Si
Edad_4	Masculino	FALSO	riesgo	ausente	Asintomático	Prehipertensión	Anormal_2	No
Edad_2	Masculino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Nomal	Si
Edad_4	Femenino	FALSO	Normal	ausente	Asintomático	Prehipertensión	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Angina_tipica	Alta	Nomal	No
Edad_4	Masculino	FALSO	riesgo	ausente	Angina_tipica	Alta	Anormal_2	No
Edad_4	Femenino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Anormal_2	No
Edad_1	Femenino	FALSO	riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Femenino	FALSO	Normal	ausente	Asintomático	Prehipertensión	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Nomal	Si
Edad_2	Masculino	FALSO	alto_riesgo	ausente	No_dolor_anginal	Alta	Nomal	No
Edad_3	Femenino	FALSO	riesgo	presente	Asintomático	Prehipertensión	Anormal_1	Si
Edad_1	Masculino	FALSO	Normal	ausente	Angina_atipica	Alta	Nomal	No
Edad_3	Masculino	FALSO	riesgo	presente	Asintomático	Alta	Nomal	No
Edad_3	Masculino	FALSO	alto_riesgo	presente	Asintomático	Alta	Anormal_1	No
Edad_3	Femenino	VERDADERO	riesgo	presente	Asintomático	Alta	Anormal_2	Si
Edad_3	Masculino	FALSO	riesgo	ausente	Angina_atipica	Alta	Anormal_2	No
Edad_3	Masculino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_4	Masculino	FALSO	riesgo	presente	No_dolor_anginal	Alta	Anormal_2	No

<i>Edad</i>	<i>Sexo</i>	<i>Abstinencia azúcar</i>	<i>Colesterol</i>	<i>Enfermedad</i>	<i>Dolor Pecho</i>	<i>Prehipertensión</i>	<i>Electrocardiograma</i>	<i>Ejercicio angina</i>
Edad_3	Femenino	FALSO	alto_riesgo	ausente	Angina_atipica	Alta	Nomal	No
Edad_2	Masculino	FALSO	Normal	presente	Asintomático	Alta	Nomal	Si
Edad_4	Masculino	FALSO	Normal	presente	Asintomático	Alta	Anormal_2	Si
Edad_4	Femenino	FALSO	Normal	presente	Asintomático	Prehipertensión	Nomal	Si
Edad_2	Masculino	FALSO	Normal	ausente	Angina_atipica	Alta	Nomal	No
Edad_3	Masculino	VERDADERO	Normal	presente	Asintomático	Normal	Anormal_2	No
Edad_3	Femenino	FALSO	alto_riesgo	presente	Asintomático	Prehipertensión	Nomal	Si
Edad_2	Masculino	FALSO	alto_riesgo	presente	Angina_tipica	Prehipertensión	Nomal	No
Edad_4	Masculino	VERDADERO	Normal	presente	Asintomático	Alta	Nomal	No
Edad_3	Masculino	FALSO	Normal	presente	Asintomático	Normal	Nomal	Si
Edad_3	Femenino	FALSO	riesgo	presente	Angina_atipica	Alta	Anormal_2	No