



**UNIVERSIDAD DEL AZUAY**

**FACULTAD DE ADMINISTRACIÓN DE EMPRESAS**

**ESCUELA DE INGENIERÍA DE SISTEMAS**

*Estudio de las Redes Bayesianas y sus Aplicaciones*

**MONOGRAFÍA PREVIA A LA OBTENCIÓN DEL TÍTULO DE**

**INGENIERO DE SISTEMAS**

**AUTORES: KARINA ELIZABETH ARIAS REYES  
MARIA ROSA FÁREZ SANCHEZ**

**DIRECTOR: ING. LUIS CALDERON**

**CUENCA, ECUADOR**

**2007**

## DEDICATORIA

Dedico esta monografía a mis padres y hermanos ya que sin su ayuda y apoyo no hubiera logrado la meta propuesta. Agradezco a Dios por haberme permitido alcanzar este objetivo, y a mis queridos padres quienes con su inquebrantable esfuerzo y confianza, estuvieron junto a mí todos estos años.

Karina Arias Reyes

En primer lugar quiero dedicar este trabajo a mi Mami, mi mayor ejemplo de lucha y dedicación, gracias Ma porque todo el tiempo que estuvimos separadas valió la pena; gracias a ti estoy donde estoy y soy lo que soy; quiero dedicar también a mi familia, a mi abuelo, mi hermana, mis tías, tíos y primos, gracias por todo el apoyo y amor, no solo durante mi carrera sino a lo largo de toda mi vida, y quiero dedicar de manera especial a mi compañero de aula, mi amigo, mi esposo, gracias Javier, por creer siempre en mí y por ser lo mejor que me llevo de esta carrera.

María Rosa Fárez Sánchez

## AGRADECIMIENTOS

Un agradecimiento especial al Ing. Luís Calderón,  
por prestarnos su amistad, ayuda y colaboración,  
durante la elaboración de este trabajo.

## ÍNDICE DE CONTENIDOS

Dedicatoria .....	ii
Agradecimientos .....	iii
Índice de Contenidos .....	iv
Índice de Ilustraciones y Cuadros .....	vii
Resumen .....	viii
Abstract .....	ix
Introducción .....	1
Capítulo 1: Redes Bayesianas.....	2
1.1 Introducción .....	2
1.2 Grafos .....	3
1.2.1 Definición .....	4
1.2.2 Conceptos Introdutorios .....	4
1.2.2.1 Grafo Dirigido y No Dirigido .....	4
1.2.2.2 Arco Dirigido .....	5
1.2.2.3 Arco No Dirigido .....	5
1.2.2.4 Camino entre dos Nodos .....	5
1.2.2.5 Camino Cerrado .....	6
1.2.2.6 Camino Dirigido.....	6
1.2.2.7 Ciclo.....	6
1.2.2.8 Grafo Cíclico .....	6
1.2.2.9 Grafo Acíclico .... ..	6
1.2.2.10 Padre.....	6
1.2.2.11 Antepasado o Ascendiente.....	6
1.2.2.12 Conjunto Ancestral.....	6
1.2.2.13 Descendiente.. ..	6
1.2.2.14 Variable Proposicional.....	6
1.2.2.15 Hallazgo.....	6
1.2.2.16 Evidencia.....	7
1.2.2.17 Probabilidad a Priori.....	7
1.2.2.17 Probabilidad a Posteriori.....	7

1.3 Definición de una Red Bayesiana.....	7
1.4 Representación Gráfica.....	8
1.5 Teorema de Bayes.....	9
1.6 Métodos Generales de Inferencia en Redes Bayesianas.....	12
1.6.1 Métodos de Eliminación de Variables.....	12
1.6.2 Métodos de Condicionamiento.....	13
1.6.3 Métodos de Simulación Estocástica.....	13
1.6.4 Métodos de Muestreo Lógico.....	14
1.6.5 Métodos de Muestreo Uniforme.....	15
1.6.6 Métodos de Ponderación por Verosimilitud.....	15
1.6.7 Método de Muestreo hacia Atrás.....	16
1.6.8 Método de Muestreo por Importancia.....	16
1.6.9 Método de Muestreo de Markov.....	17
1.7 Tipos de Redes Bayesianas.....	17
1.8 Conclusiones del Capítulo.....	19
 Capítulo 2: Funcionamiento de una Red Bayesiana.....	 20
2.1 Introducción .....	20
2.2 ¿Cómo funciona una red Bayesiana?.....	21
2.2.1 Interpretación de una Red Probabilística.....	23
2.2.2 Aprendizaje Paramétrico.....	24
2.2.2.1 Probabilidades Previas.....	25
2.2.2.2 Probabilidades Condicionales.....	25
2.2.3 Aprendizaje Estructural.....	26
2.2.3.1 Árboles.....	26
2.2.3.2 Poliárboles.....	30
2.2.3.3 Redes Multiconectadas.....	32
2.3 Estimación de la Estructura de una Red Bayesiana.....	34
2.3.1 Selección del Modelo por Búsqueda y Store.....	34
2.3.1.1 Medidas de Calidad.....	34
2.3.1.1.1 Calidad Bayesianas.....	34
2.3.1.1.2 Longitud de Codificación.....	35
2.3.1.1.3 Información Teórica.....	35

2.3.1.1.3.1	Criterio de Inf. Máxima .....	36
2.3.1.1.3.2	Criterio de Inf. Akaike .....	36
2.3.1.1.3.3	Criterio de Inf. Bayesiana.....	36
2.3.1.2	Algoritmos de búsqueda del modelo.....	36
2.3.1.2.1	Algoritmo K2.....	36
2.3.1.2.2	Algoritmo Buntine.....	37
2.3.1.2.3	Algoritmo CB.....	37
2.3.2	Selección del Modelo usando Análisis de Dependencias.....	38
2.3.2.1	Método de Chow y Lui.....	38
2.3.3	Inferencia en Redes Bayesianas.....	39
2.3.3.1	Construcción.....	39
2.3.3.2	Inicialización.....	42
2.3.3.3	Paso de Mensajes.....	43
2.4	Construcción de Redes Bayesianas.....	43
2.4.1	A partir de una Base de Datos.....	43
2.4.2	Con la ayuda de un Experto.....	44
2.4.3	Inf. Estructural para la construcción de una Red Bayesiana....	44
2.5	Ventajas y Desventajas del Modelo Probabilístico.....	45
2.5.1	Ventajas.....	45
2.5.2	Desventajas.....	45
2.6	Conclusiones del Capítulo.....	45
Capítulo 3: Aplicación de las Redes Bayesianas.....		46
3.1	Introducción .....	46
3.2	Ejemplos del trabajo de una Red Bayesiana.....	47
3.3	Aplicación Práctica de una Red Bayesiana.....	50
3.3.1	La Fórmula de Bayes .....	50
3.3.2	Ejercicio de Aplicación .....	51
3.3.2.1	Filtrando el Spam.....	53
3.4	Conclusiones del Capítulo.....	65
Bibliografía.....		67

## ÍNDICE DE ILUSTRACIONES

Figura 1.2	Estructura de un Grafo.....	4
Figura 1.2	Grafo Dirigido y Grafo No Dirigido.....	5
Figura 1.3	Independencia en el Grafo.....	9
Figura 1.4	Casos Básicos.....	9
Figura 1.5	Thomas Bayes.....	9
Figura 2.1	Grafos camino Dirigido y no Dirigido.....	20
Figura 2.2	Grafo Acíclico Dirigido.....	22
Figura 2.3	Estructura de Árbol.....	26
Figura 2.4	Estructura de Poliárboles.....	30
Figura 2.5	Estructura de Redes Multiconectadas .....	32
Figura 2.6	Red Bayesiana.....	39
Figura 2.7	Paso 1 del algoritmo.....	40
Figura 2.8	Paso 2 y 3 del algoritmo Grafo Triangulado.....	40
Figura 2.9	Paso 4 Identificación de los Cliques.....	41
Figura 2.10	Paso 5 Grafo del Clique.....	41
Figura 2.11	Paso 6 Eliminación de Ciclos.....	42
Tabla 2.1	Tabla del Clique.....	42
Tabla 2.2	Tabla del Separador.....	42
Figura 3.1	Ejemplo 1.....	47
Figura 3.2	Ejemplo 2.....	48
Figura 3.3	Ejemplo 3 .....	50
Tabla 3.1	Mensaje particionado en fichas.....	56
Tabla 3.2	Calculo de valores en el Filtro.....	58
Tabla 3.3	Filtro Bayesiano.....	63

## RESUMEN

Las redes Bayesianas permiten modelizar un dominio con incertidumbre, y razonar a partir de él. Los mecanismos de inferencia utilizados se basan en la topología de la red, el uso de probabilidades condicionales y en el Teorema de Bayes para su propagación. Dependiendo del dominio, las redes bayesianas pueden capturar el conocimiento incierto de manera natural y pueden ser eficientes.

Las redes bayesianas son utilizadas en diversas áreas de aplicación como el diagnóstico médico, ciencias de la computación, matemática aplicada, estadística, ingeniería, economía; las mismas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento capaces de predecir el valor de variables no observadas y explicar las observadas.

Entre las características que poseen las redes bayesianas, se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos, evitan el sobre-ajuste de los datos y pueden manejar bases de datos incompletas.

Una red bayesiana es un grafo dirigido acíclico, donde los nodos representan las variables del problema que se desea resolver. El conocimiento del problema se representa mediante la instanciación de aquellos nodos cuyo valor es conocido, propagándose tal conocimiento a través de la red mediante ciertas reglas probabilísticas.

## **ABSTRACT**

The Bayesian network allows modelling a domain uncertainly, and to reason from it. The mechanisms of inferences used are based in the topology of the network, the use of conditional probabilities and in the Bayes theorem in order to propagation. Depending to the dominium, the Bayesian network could capture the uncertain knowledge in a natural way being efficient.

These are used in different application areas such as Medical diagnosis, Computer Science, Applied Mathematics, Statistics, Engineer, Economics; these provide a compact way of representing the knowledge and flexible methods of reasoning able to predict the value of the non observed variables and to explain the observed ones.

Among the characteristics of Bayesian network it is possible to emphasize that it allows learning about dependence and casual relations. Also it allows combining knowledge and data, avoid the non adjustment of data, and could be handled from incomplete baselines.

A Bayesian network is an acyclic directed graph, where the nodes represent the variables of the problem that is desired to solve. The knowledge of the problem is represented by the installation of the nodes which value is known, propagating the knowledge through the net by certain probable rules.

## INTRODUCCIÓN

El método más antiguo para el tratamiento de la incertidumbre es la probabilidad. Dentro del campo de la inteligencia artificial, surgieron críticas contra el uso de métodos probabilistas en sistemas expertos, especialmente porque las hipótesis necesarias para hacer tratable el método bayesiano clásico eran incorrectas en la mayor parte de los problemas del mundo real. Esto motivó el desarrollo de otros métodos, como los factores de certeza o la lógica difusa, en que se introducen implícitamente hipótesis y aproximaciones aún más exigentes.

Afortunadamente, el desarrollo de las [Redes Bayesianas](#) en la década de los 80 permitió refutar las objeciones anteriores contra el uso de la probabilidad, construyendo un modelo de razonamiento causal con un sólido fundamento teórico.

Las Redes Bayesianas se encuentran entre los modelos gráficos más populares. La principal diferencia, con respecto a otros modelos, está en que sus arcos son dirigidos y representan dependencia condicional entre las variables. El nombre proviene del hecho que gran parte de la teoría relevante con este tipo de redes se basa en la estadística Bayesiana, por lo que se presentan algunos temas relacionados con el uso y la estimación de las redes Bayesianas.

## CAPITULO 1

### REDES BAYESIANAS

#### 1.1 Introducción

Una red Bayesiana es un modelo gráfico que permite representar las relaciones de dependencia entre un conjunto de variables. Las redes Bayesianas se usan para codificar el conocimiento en los Sistemas Expertos y existen numerosos algoritmos desarrollados para estimar este tipo de redes. Las redes Bayesianas, junto con los métodos Bayesianos, facilitan la combinación del conocimiento experto con la información contenida en un conjunto de datos.

En definitiva una red bayesiana, es un modelo [probabilístico](#) multivariado que relaciona un [conjunto](#) de variables aleatorias mediante un grafo dirigido, el cual indica explícitamente influencia causal. Gracias a su motor de actualización de probabilidades, las redes bayesianas permiten modelizar un dominio con incertidumbre, y razonar a partir de él. Los mecanismos de inferencia utilizados se basan en la topología de la red, el uso de probabilidades condicionales y en el Teorema de Bayes para su propagación.

Dependiendo del dominio, las redes bayesianas pueden capturar el conocimiento incierto de manera natural y pueden ser eficientes, además que son una herramienta extremadamente útil en la estimación de probabilidades ante nuevas evidencias.

Las redes bayesianas constituyen un método para la representación de conocimiento incierto, que permite establecer razonamientos basados en la teoría de la probabilidad.

Las redes Bayesianas proporcionan una representación gráfica para un conjunto de variables aleatorias y para las relaciones existentes entre ellas. La estructura de la red permite especificar la función de probabilidad conjunta de estas variables como el producto de funciones de probabilidad condicionadas, por lo general, más sencillas. En esta sección se definen formalmente las redes Bayesianas y se mencionan algunos métodos relacionados con su estimación.

Para el desarrollo del presente capítulo se usó la siguiente bibliografía:

<http://www.conocimientosweb.net> Laboratorio de Sistemas Inteligentes

<http://www.ia.uned.es> Guía de la Inteligencia Artificial

Heckerman, D. (1995). *A tutorial on learning bayesian networks. Technical report MSR-TR-95-06, Microsoft research, Redmond, WA.*

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems. Capítulo 3*

Morgan Kaufmann, San Mateo, CA.

## **1.2 Grafos.**

Un modelo gráfico puede usarse para representar las relaciones de dependencia entre un conjunto de variables. Los nodos corresponden a las variables aleatorias y las relaciones de dependencia existentes pueden representarse usando arcos entre los nodos. La interpretación gráfica de estas relaciones de dependencia en los modelos gráficos constituye una de sus características principales.

Un modelo probabilístico puede definirse usando un grafo que describa las relaciones existentes entre las variables. Supóngase un conjunto de variables  $\{X_1, \dots, X_n\} = X$  que pueden relacionarse entre sí. El conjunto anterior puede representarse gráficamente por una colección de nodos o vértices, asociando un nodo a cada elemento de  $X$ . Estos nodos pueden conectarse por arcos, indicando las relaciones

existentes entre los mismos. Un arco entre los nodos  $X_i$  y  $X_j$  se denotará mediante  $L_{ij}$ . Así mismo, el conjunto de todos los arcos se denotará por  $L = \{L_{ij} / X_i \text{ y } X_j \text{ están conectados}\}$ . Por tanto, un grafo puede definirse mediante el conjunto de nodos,  $X$ , y las relaciones entre los mismos,  $L$ .

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

La Figura 1.1 nos muestra un ejemplo de un grafo compuesto por siete nodos:  $X = \{A, B, \dots, G\}$  y de un conjunto de seis aristas (arcos),  $L = \{LAB, LAC, LBD, LCE, LDF, LDG\}$ .

Los nodos están representados por círculos y las aristas por líneas que unen los nodos correspondientes.

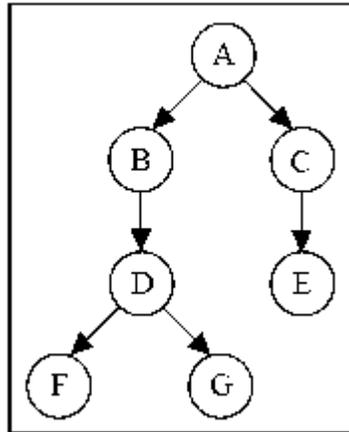


Figura 1.1 Estructura de un Grafo

## 1.2.2 Conceptos Introductorios.

**1.2.2.1 Grafo dirigido y no dirigido.** Un grafo en el cual todas las aristas (arcos) son dirigidas se denomina grafo dirigido, y un grafo en el que todas sus aristas son no dirigidas se denomina no dirigido.

Por tanto, en un grafo dirigido es importante el orden del par de nodos que definen cada arista, mientras que en un grafo no dirigido, el orden carece de importancia.

En la Figura 1.2(a), observamos un grafo dirigido definido por:

$$X = \{A, B, C, D, E, F\},$$

$$L = \{A \rightarrow D, B \rightarrow C, D \rightarrow B, F \rightarrow D, D \rightarrow E, E \rightarrow F\},$$

Mientras que para el grafo no dirigido de la Figura 1.2(b) se tiene:

$$X = \{A, B, C, D, E, F, G, H\},$$

$$L = \{A - B, B - C, C - D, D - E, E - A, E - F, F - G, G - D, D - H\}.$$

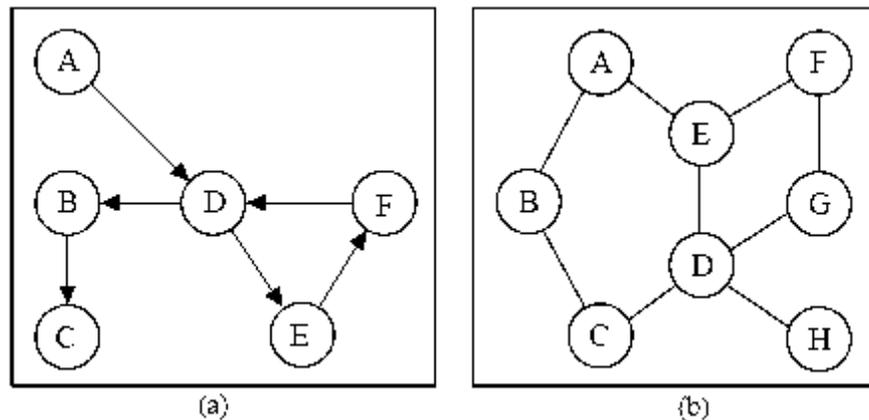


Figura 1.2 Grafo Dirigido y Grafo No Dirigido

**1.2.2.2 Arco dirigido.** Dado un grafo  $G = (X, L)$  si  $L_{ij} \in L$  y  $L_{ji} \notin L$ , el arco  $L_{ij}$  entre los nodos  $X_i$  y  $X_j$  se denomina dirigido y se denota mediante  $X_i \rightarrow X_j$ .

**1.2.2.3 Arco no dirigido.** Dado un grafo  $G = (X, L)$  si  $L_{ij} \in L$  y  $L_{ji} \in L$ , el arco  $L_{ij}$  se denomina no dirigido y se denota mediante  $X_i - X_j$  o  $X_j - X_i$ .

**1.2.2.4 Camino entre dos nodos.** Un camino del nodo  $X_i$  al nodo  $X_j$  es una sucesión de nodos  $(X_{i_1}, \dots, X_{i_r})$ , comenzando en  $X_{i_1} = X_i$  y finalizando en  $X_{i_r} = X_j$ , de forma

que existe un arco del nodo  $X_{i_k}$  al nodo  $X_{i_{k+1}}$ ,  $k = 1, \dots, r - 1$ . La longitud del camino,  $(r-1)$ , se define como el número de arcos que contiene.

**1.2.2.5 Camino cerrado.** Un camino  $(X_{i_1}, \dots, X_{i_r})$ , se dice que es cerrado si el nodo inicial coincide con el final, es decir,  $X_{i_1} = X_{i_r}$ .

**1.2.2.6 Camino dirigido.** Es una secuencia ordenada de nodos  $(X_{i_1}, \dots, X_{i_r})$ .

**1.2.2.7 Ciclo.** Es un camino no dirigido que empieza y termina en el mismo nodo  $X$ .

**1.2.2.8 Grafo cíclico.** Es un grafo que contiene al menos un ciclo.

**1.2.2.9 Grafo acíclico.** Es un grafo que no contiene ciclos.

**1.2.2.10 Padre.**  $X$  es un padre de  $Y$  si y sólo si existe un arco  $X \rightarrow Y$ . Se dice también que  $Y$  es hijo de  $X$ . Al conjunto de los padres de  $X$  se representa como  $pa(X)$ , y al de los hijos de  $X$  por  $S(X)$ .

**1.2.2.11 Antepasado o ascendiente.**  $X$  es un *antepasado* o ascendiente de  $Z$  si existe un camino dirigido de  $X$  a  $Z$ .

**1.2.2.12 Conjunto ancestral.** De un nodo  $X$  es un conjunto que contiene a  $X$  y a todos sus antepasados.

**1.2.2.13 Descendiente.**  $Z$  es un *descendiente* de  $X$  si existe un camino de  $X$  a  $Z$ .

**1.2.2.14 Variable proposicional.** Es una variable aleatoria que toma un conjunto exhaustivo y excluyente de valores. La denotaremos con letras mayúsculas, por ejemplo  $X$ , y a un valor cualquiera de la variable con la misma letra en minúscula,  $x$ .

**1.2.2.15 Hallazgo.** Determinación del valor de una variable, a partir de un dato (una observación, una medida...).

**1.2.2.16 Evidencia.** Conjunto de todos los hallazgos disponibles en un determinado momento.

**1.2.2.17 Probabilidad a priori.** Es la probabilidad de una variable o subconjunto de variables cuando no hay ningún hallazgo, es decir antes de conocer los síntomas.

**1.2.2.18 Probabilidad a posteriori.** Es la probabilidad de una variable o subconjunto de variables dada la evidencia, es decir después de conocer los síntomas.

### 1.3 Definición de Red Bayesiana

Existen muchas definiciones del que puede ser una red bayesiana, se han recopilado algunas de ellas, las más importantes para poder llegar a una idea concreta.

- Es un grafo dirigido acíclico conexo más una distribución de probabilidad sobre sus variables.
  
- Una Red Probabilística (RP) es un Grafo Acíclico Dirigido (DAG) en la cual cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres.
  
- Una red bayesiana es un Grafo Dirigido Acíclico que consta de:
  - Un conjunto de nodos, uno por cada variable aleatoria del “mundo”
  - Un conjunto de arcos dirigidos que conectan los nodos; si hay un arco de X a Y decimos que X es un padre de Y (padres(X) denota el conjunto de v.a. que son padres de X)
  - Cada nodo  $X_i$  contiene la distribución de probabilidad condicional  $P(X_i | \text{padres}(X_i))$

- Una red bayesiana es:
  - Un conjunto de variables proposicionales,  $V$ .
  - Un conjunto de relaciones binarias definida sobre las variables de  $V$ ,  $E$ .
  - Una distribución de probabilidad conjunta sobre las variables de  $V$ . tales que:
    - $(V, E)$  forman un grafo acíclico, conexo y dirigido  $G$ .
    - $(G, P)$  cumplen las *hipótesis de independencia condicional*, también llamadas de *separación direccional*, que se enuncian a continuación.

#### 1.4 Representación Gráfica

Las redes bayesianas son una representación gráfica de dependencias para razonamiento probabilístico, en la cual los nodos y arcos representan:

- Nodo: Variable proposicional.
- Arcos: Dependencia probabilística.

La variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste.

Una Red Bayesiana representa en forma gráfica las dependencias e independencias entre variables aleatorias, en particular las independencias condicionales.

- Independencia en la distribución

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

**Notación:**

$$I(X, Z, Y)$$

- Independencia en el grafo

X “separada” de Y por Z

**Notación:**

$$\langle X | Z | Y \rangle$$



Figura 1.3 Independencia en el Grafo

### Separación “D”

Existen tres casos básicos:

- Arcos Divergentes
- Arcos en Secuencia
- Arcos Convergentes

caso 1: Secuencia:



caso 2: Divergentes:



caso 3: Convergentes:



Figura 1.4 Casos Básicos

## 1.5 Teorema de Bayes



Figura 1.5 Thomas Bayes

El teorema de Bayes, descubierto por [Thomas Bayes](#), en la [teoría de la probabilidad](#), es el resultado que da la [distribución de probabilidad condicional](#) de una [variable aleatoria](#)  $A$  dada  $B$  en términos de la distribución de probabilidad condicional de la variable  $B$  dada  $A$  y la [distribución de probabilidad marginal](#) de sólo  $A$ .

La fórmula del Teorema de Bayes es:

$$P(A_i/B) = \frac{P(A_i) * P(B/A_i)}{\sum P(A_i) * P(B/A_i)}$$

Tratar de explicar esta fórmula con palabras es muy complejo, así que vamos a intentar explicarla con un ejemplo. De todos modos, antes de entrar en el ejercicio, recordar que este teorema también exige que el suceso A forme un sistema completo.

El parte meteorológico ha anunciado tres posibilidades para el fin de semana:

- a) **Que llueva:** probabilidad del 50%.
- b) **Que nieve:** probabilidad del 30%
- c) **Que haya niebla:** probabilidad del 20%.

Según estos posibles estados meteorológicos, la posibilidad de que ocurra un accidente es la siguiente:

- a) **Si llueve:** probabilidad de accidente del 20%.
- b) **Si nieva:** probabilidad de accidente del 10%
- c) **Si hay niebla:** probabilidad de accidente del 5%.

Resulta que efectivamente ocurre un accidente y como no estábamos en la ciudad no sabemos que tiempo hizo (llovió, nevó o hubo niebla). El teorema de Bayes nos permite calcular estas probabilidades:

Las probabilidades que manejamos antes de conocer que ha ocurrido un accidente se denominan "**probabilidades a priori**" (lluvia con el 50%, nieve con el 30% y niebla con el 20%).

Una vez que incorporamos la información de que ha ocurrido un accidente, las probabilidades del suceso A cambian: son probabilidades condicionadas  $P(A/B)$ , que se denominan "**probabilidades a posteriori**".

Vamos a aplicar la fórmula:

$$P(A_i/B) = \frac{P(A_i) * P(B/A_i)}{\sum P(A_i) * P(B/A_i)}$$

**a) Probabilidad de que estuviera lloviendo:**

$$P(A_i/B) = \frac{0,50 * 0,20}{(0,50 * 0,20) + (0,30 * 0,10) + (0,20 * 0,05)} = 0,714$$

La probabilidad de que efectivamente estuviera lloviendo el día del accidente (probabilidad a posteriori) es del 71,4%.

**b) Probabilidad de que estuviera nevando:**

$$P(A_i/B) = \frac{0,30 * 0,10}{(0,50 * 0,20) + (0,30 * 0,10) + (0,20 * 0,05)} = 0,214$$

La probabilidad de que estuviera nevando es del 21,4%.

**c) Probabilidad de que hubiera niebla:**

$$P(A_i/B) = \frac{0,20 * 0,05}{(0,50 * 0,20) + (0,30 * 0,10) + (0,20 * 0,05)} = 0,071$$

La probabilidad de que hubiera niebla es del 7,1%

El teorema de Bayes es válido en todas las aplicaciones de la teoría de la probabilidad. Sin embargo, hay una controversia sobre el tipo de probabilidades que emplea. En esencia, los seguidores de la [estadística](#) tradicional sólo admiten probabilidades basadas en experimentos repetibles y que tengan una confirmación empírica mientras que los llamados estadísticos bayesianos permiten probabilidades subjetivas.

El teorema puede servir entonces para indicar cómo debemos modificar nuestras probabilidades subjetivas cuando recibimos información adicional de un experimento. La estadística bayesiana está demostrando su utilidad en ciertas estimaciones basadas en el conocimiento subjetivo a priori y permitir revisar esas estimaciones en función de la evidencia es lo que está abriendo nuevas formas de hacer conocimiento.

## **1.6 Métodos Generales de Inferencia en Redes Bayesianas**

La inferencia probabilística en una red bayesiana consiste en el cálculo de las probabilidades a posteriori de las variables no observadas, dados los valores que toman ciertas variables observadas de la red.

A partir de las leyes de la teoría de la probabilidad, cualquier probabilidad condicionada o marginal se puede obtener a partir de la probabilidad conjunta.

La probabilidad conjunta de las variables de una red bayesiana es el producto de todas las probabilidades condicionales incluidas en la red. Desafortunadamente, aunque este método para el cálculo de las probabilidades a posteriori de las variables parece el más inmediato, su complejidad crece exponencialmente con el número de nodos de la red.

### **1.6.1 Métodos de Eliminación de Variables y Métodos de Agrupamiento**

Una forma de evitar la complejidad propia del método mencionado en el párrafo anterior consiste en sumar primero sobre ciertas variables antes de multiplicar todos los potenciales. (Cada tabla de probabilidad es un potencial y el resultado de multiplicar dos potenciales es un nuevo potencial)

### 1.6.2 Métodos de Condicionamiento

Existen algoritmos para la propagación de la evidencia en redes bayesianas simplemente conexas (aquellas en las que entre dos cualesquiera de sus nodos hay exactamente un camino), que están basados en el paso de mensajes probabilísticos  $\pi$  y  $\lambda$  entre nodos vecinos. En concreto, Pearl propuso un algoritmo de este tipo para redes con estructura de árbol. Posteriormente, Kim generalizó dicho algoritmo para poliárboles (grafos dirigidos simplemente conexos). Estos algoritmos poseen una complejidad proporcional al número de nodos y se prestan a una implementación distribuida.

Los métodos que acabamos de mencionar, basados en el paso de mensajes, no permiten tratar redes con bucles, por lo que en la práctica no resultan de gran utilidad. Los algoritmos de condicionamiento surgieron ante la necesidad de desarrollar métodos que pudieran tratar los bucles y que admitieran una implementación distribuida.

Díez creó un método de condicionamiento local basado en condicionar sólo dentro de cada bucle. De este modo se reduce drásticamente la complejidad del algoritmo de condicionamiento global, la cual varía exponencialmente con el número de nodos del conjunto de corte.

### 1.6.3 Métodos de Simulación Estocástica

Algunos de los métodos exactos que hemos citado sólo permiten realizar inferencia en redes con estructura de árbol o poliárbol. A su vez, aquellos métodos exactos que sí permitían tratar con redes de estructura general, resultan cada vez más ineficientes a medida que aumenta el número de nodos de la red o la complejidad estructural de la misma.

Existen métodos aproximados para la propagación de la evidencia, aplicables a cualquier estructura de red. Estos métodos obtienen las probabilidades a posteriori de cada variable de forma aproximada. Básicamente, generan muestras de tamaño  $N$  (número de nodos o variables de la red) a partir de las cuales se aproximan los valores de probabilidad buscados, teniendo en cuenta la frecuencia de aparición de cada valor en la muestra y el tamaño de la misma.

Este esquema básico se puede ampliar asignando un peso o probabilidad a cada muestra. Al final del proceso de simulación, los pesos son normalizados y utilizados en la estimación de las probabilidades a posteriori.

En general, en lo que resta de sección supondremos la existencia de un conjunto de variables  $X = (X_1, \dots, X_N)$  con función de probabilidad conjunta  $P(x)$  y un subconjunto  $E$  de variables de  $X$  que toma valores conocidos. Dada una función de probabilidad conjunta  $P(x)$  asociada a una red bayesiana, cada método aproximado particular define un modo específico de generar muestras y calcular sus pesos.

#### **1.6.4 Método de Muestreo Lógico.**

El método de muestreo lógico asigna valores secuencialmente a todas las variables, incluidas las observadas, muestreando primero los padres y luego los hijos. Por tanto, sólo se muestrea una variable cuando todos sus padres ya han sido simulados y tienen valores asignados. Cada variable  $X_i$  es simulada a partir de su TPC,  $P(x_i | pa(x_i))$ , donde  $pa(x_i)$  representa una configuración de valores para los padres de  $X_i$ .

En el momento en que el valor simulado para un nodo observado difiera del valor realmente observado, se rechaza la muestra. Rechazar una muestra sería equivalente a asignarle un peso nulo. El caso contrario supone aceptar la muestra asignándole un peso unitario.

Cualquier probabilidad a posteriori o condicional que queramos obtener se puede aproximar hallando el cociente entre el número de casos no rechazados que satisfacen nuestra consulta y el número total de casos no rechazados.

Con más frecuencia de la deseada, el presente método puede conducir a un alto porcentaje de rechazos, lo cual implica aumentar el número de simulaciones necesario para obtener resultados fiables. Esto es así cuando la probabilidad a priori de la evidencia,  $P(e)$ , es pequeña.

El resto de los métodos aproximados revisados en este capítulo intenta solucionar el problema de las muestras rechazadas que aparecen en el muestreo lógico.

#### **1.6.5 Método de Muestreo Uniforme**

En este método, las variables no observadas pueden ser simuladas en cualquier orden. Cuando el conjunto de posibles muestras no esté distribuido uniformemente, existe el inconveniente de que se pueden llegar a generar muchas muestras poco representativas y, por tanto, los valores de probabilidad obtenidos pueden ser erróneos.

#### **1.6.6 Método de Ponderación por Verosimilitud**

El método de ponderación por verosimilitud intenta ofrecer una solución al problema de los pesos descompensados que se producía en el método de muestreo uniforme.

El orden de simulación de las variables debe ser descendente, ya que los padres deben ser muestreados antes de hacer lo propio con su hijo. El peso asociado a una muestra  $X = (X_1, \dots, X_N)$  es  $\prod_{i \in E} P(e_i | pa(x_i))$ . Si la probabilidad a priori de la evidencia no está muy próxima a cero, este método se comporta satisfactoriamente.

### 1.6.7 Método de Muestreo hacia atrás

En el método de muestreo hacia atrás no es necesario muestrear los padres antes que los hijos, ya que combina dos tipos de muestreo: uno hacia adelante y otro hacia atrás.

En concreto, utiliza muestreo hacia adelante en los nodos cuyos descendientes no han sido observados. En cambio, se realiza muestreo hacia atrás una vez asignado un valor al nodo en nodos que tienen algún descendiente con evidencia.

En el muestreo que se realiza hacia atrás, únicamente se podrán simular los padres de un nodo  $X_i$  si éste es un nodo evidencia o si ya ha sido previamente muestreado.

Los valores de  $pa(x_i)$  se generan en función de los parámetros probabilísticas  $p(x_i | pa(x_i))$  normalizados.

### 1.6.8 Método de Muestreo por Importancia

El método de muestreo por importancia simula el valor de cada variable no observada a partir de una distribución de probabilidad  $f$  denominada *función de importancia*, definida sobre cada variable de la red. Existen distintos métodos de este tipo, que se diferencian por la forma de cálculo de la función de importancia.

La convergencia hacia los resultados correctos será tanto más rápida cuanto más parecida sea la función de importancia utilizada a la distribución de probabilidades a posteriori que se pretende calcular.

En este método, el peso asociado a una muestra  $x$  viene dado por la probabilidad a priori de  $x$ ,  $P(x)$ , dividida por la probabilidad de seleccionar  $x$  a partir de  $f$ .

### 1.6.9 Método de Muestreo de Markov

El muestreo de Markov genera inicialmente una muestra de forma aleatoria, se simulan secuencialmente las variables no observadas siguiendo un orden arbitrario hasta obtener otra muestra.

La simulación de cada variable no observada  $X_i$  se realiza a partir de una función de probabilidad asociada, que depende únicamente del propio nodo  $X_i$  y de su manto de Markov, que está formado por sus padres, sus hijos y los demás padres de sus hijos.

En este método, los valores obtenidos en la última muestra sirven de base para hallar la próxima muestra. Además, no se requiere una ordenación de los nodos, ya que para las variables a las que no se les ha asignado todavía un nuevo valor en la simulación actual se utilizan los valores de la simulación previa. En este método, los pesos asociados a cada muestra son siempre la unidad.

### 1.7 Tipos de Redes Bayesianas

Existen distintos tipos de Redes Bayesianas:

- Naive Bayes = bayes “ingenuo” o Idiot's Bayes  
Forma de “**V**” =>  $2^n$  estados en el nodo inferior
- DBNs = Redes Bayesianas Dinámicas  
Cambian con el tiempo (t, t+1, t+2...)  
Lo pasado en t, tiene relación con lo que suceda en t+1  
Se obtienen redes muy complejas estructuralmente  
Cada variable debería representar un evento irreversible (sólo ocurre una vez a lo largo del tiempo) y no un evento que puede producirse múltiples veces.

Nuestro dominio no es markoviano ya que puede haber diferentes retardos entre dos nodos evento

- Redes Gaussianas = distribución gaussiana  
Para nodos con variables continuas
- Cadenas de Markov = Subconjunto de las Redes Bayesianas  
Un proceso de Markov es un proceso que se va moviendo de estado en estado dependiendo exclusivamente de los N estados anteriores.

En un proceso de Markov aparte de los estados, está la matriz de transiciones entre estados, que indica la probabilidad de pasar de un estado a otro o a sí mismo.

Además hace falta tener una matriz que indique como se encuentra el sistema al inicio, vector de probabilidades iniciales.

Cadenas de Markov Ocultas(HMM): tenemos dos partes diferenciadas y no vale para modelar simplemente con una cadena de Markov anterior.

Por ejemplo, en el caso del reconocimiento de palabras.

Sonido que oímos = cuerdas vocales+lengua+labios...

Este tipo de sistemas trabajan considerando que la producción interna de fonemas es una secuencia de estados ocultos (no observables) y que los sonidos resultantes son una secuencia observable generada por los estados ocultos. El número de estados ocultos puede ser diferente del número de estados observables.

La conexión entre los estados ocultos y las variables observables, símbolos, representa la probabilidad de generar un símbolo en particular desde un determinado estado oculto.

## **1.8 Conclusiones del Capítulo**

Al finalizar este capítulo tenemos ya una idea clara sobre las ventajas que nos brinda la Teoría de Grafos para la implementación de una Red Bayesiana además todos los conceptos básicos, como el Teorema de Bayes que nos sirve para conocer los principios esenciales del funcionamiento de este Sistema Experto.

## CAPITULO 2

### FUNCIONAMIENTO DE UNA RED BAYESIANA

#### 2.1. Introducción.

Las redes bayesianas son un tipo de herramienta estadística que representan un conjunto de incertidumbres relacionadas. Se basan en la teoría de la probabilidad y en la Teoría de grafos (Figura 2.1) para representar modelos de la realidad. Las redes bayesianas se han utilizado, entre otras, para tratar la incertidumbre que caracteriza las fluctuaciones de los mercados financieros, en la toma de decisiones implícita en el diagnóstico médico o en el estudio de ecosistemas naturales.

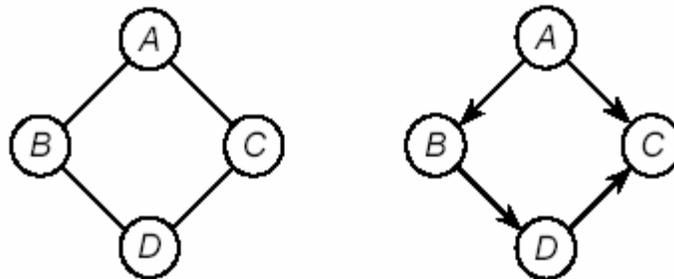


Figura 2.1 Grafos, Camino Dirigidos y No Dirigidos.

Las redes bayesianas tienen una dimensión cualitativa y otra cuantitativa. En la parte cualitativa, las redes bayesianas se podrían entender como una representación gráfica (como un grafo) que representa un conjunto de relaciones de dependencia entre variables; lo que se ha venido en denominar como un grafo dirigido acíclico. Esta forma de representar la información de un modelo tiene la ventaja de codificar relaciones de dependencia e independencia condicional entre variables, lo que facilita la interpretación y los cálculos sobre el modelo.

La incertidumbre asociada a cada variable de una red bayesiana es tratada bajo la óptica de la Teoría de la probabilidad. Así, cada variable de un modelo tiene un

número limitado de posibles estados (o niveles) y cada uno de ellos lleva asociado un valor, una probabilidad, para la ocurrencia de ese estado. Estas probabilidades son susceptibles de modificación cuando conocemos el valor que toma una determinada variable del modelo, esa información se propaga por la red y se recalculan los valores asociados a otras variables utilizando como operador básico el teorema de Bayes. Gracias a los principios de dependencia e independencia condicional codificados por la estructura gráfica y a los algoritmos desarrollados para operar sobre ellos, estos cálculos son realizados en un corto espacio de tiempo frente a lo que habría que esperar si tuviésemos que manejar todo el conjunto de variables que componen el modelo.

En las redes bayesianas, cada nodo está asociado a una variable aleatoria, que puede tomar valores dentro de un rango discreto o continuo. Dichos valores son exclusivos y exhaustivos para cada variable de la red.

Referencias utilizadas en este capítulo:

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Capítulo 4,5.

[Castillo] Castillo, E. y otros. *Sistemas expertos y modelos de redes probabilísticas*. Monografías de la academia de ingeniería. Madrid.

<http://www.fi.uba.ar/laboratorios/lsi/cacic2003-optimizacionredesbayesianas.pdf>

<http://www.sc.ehu.es/ccwbytes/post-script/lalaguna260702.pdf>

<http://www.grad.uprmedu/tesis/lopezdecastilla.pdf>

## 2.2 ¿Cómo funciona una Red Bayesiana?

Una red bayesiana es un grafo acíclico dirigido (Figura 2.2) en el que cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres; la variable a

la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste.

La topología o estructura de la red nos da información sobre las dependencias probabilísticas entre las variables pero también sobre las independencias condicionales de una variable (o conjunto de variables) dada otra u otras variables; dichas independencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades).

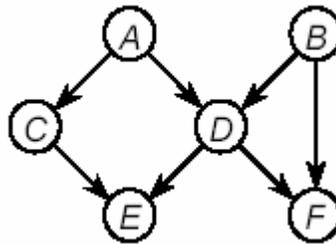


Figura 2.2 Grafo Acíclico Dirigido.

El obtener una red bayesiana a partir de datos, es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico.

La primera de ellas, consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas.

La segunda etapa, tiene como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de una estructura dada.

Las redes bayesianas son utilizadas en diversas áreas de aplicación como por ejemplo el diagnóstico médico, minería de datos, métodos de aprendizaje, seguridad en correo electrónico, etc. Las mismas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento basados en las teorías

probabilísticas capaces de predecir el valor de variables no observadas y explicar las observadas; además son una representación gráfica de dependencias para razonamiento probabilístico en sistemas expertos, en la cual los nodos representan las Variables proposicionales y arcos representan Dependencia probabilística

Entre las características que poseen las redes bayesianas, se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos, evitan el sobre-ajuste de los datos y pueden manejar bases de datos incompletas.

### 2.2.1 Interpretación de una Red Probabilística

Una Red Probabilística puede ser interpretada de dos formas:

Distribución de probabilidad: Representa la distribución de la probabilidad conjunta de las variables representadas en la red. Por ejemplo:

$$P(A, B, C, D, E, F, G) = P(G|D)P(F|C, D)P(E|B) \\ P(D|A, B)P(C|A)P(B)P(A)$$

Base de reglas: Cada arco representa un conjunto de reglas que asocian las variables involucradas, Por ejemplo:

*Si C, D entonces F*

Dichas reglas están cuantificadas por las probabilidades respectivas.

La topología o estructura de la red nos da información sobre las dependencias probabilísticas entre las variables.

La red también representa las independencias condicionales de una variable (o conjunto de variables) dada otra variable(s).

Ej.:  $\{E\}$  es cond. indep. de  $\{A,C,D,F,G\}$  dado  $\{B\}$

Esto es:  $P(E | A,C,D,F,G,B) = P(E | B)$

Esto se representa gráficamente por el nodo  $B$  separando al nodo  $E$  del resto de las variables.

En general, el conjunto de variables  $A$  es independiente del conjunto  $B$  dado  $C$  si al remover  $C$  hace que  $A$  y  $B$  se desconecten. Es decir, NO existe una trayectoria entre  $A$  y  $B$  en que las siguientes condiciones sean verdaderas.

Todos los nodos con flechas convergentes están o tiene descendientes en  $C$ .

Todos los demás nodos están fuera de  $C$ .

Esto se conoce como Separación- $D$ .

En una Red Probabilística todas las relaciones de independencia condicional representadas en el grafo corresponden a relaciones de independencia en la distribución de probabilidad.

Dichas independencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades).

### 2.2.2 Aprendizaje Paramétrico

El aprendizaje paramétrico consiste en encontrar los parámetros asociados a una estructura dada de una red bayesiana. Dichos parámetros consisten en las probabilidades *a priori* de los nodos raíz y las probabilidades condicionales de las demás variables, dados sus padres.

Si se conocen todas las variables, es fácil obtener las probabilidades requeridas. Las probabilidades previas corresponden a las marginales de los nodos raíz, y las condicionales se obtienen de las conjuntas de cada nodo con su(s) padre(s).

Para que se actualicen las probabilidades con cada caso observado, éstas se pueden representar como razones enteras, y actualizarse con cada observación. En el caso de un árbol, las fórmulas para modificar las probabilidades correspondientes son:

### 2.2.2.1 Probabilidades previas:

$$i = k \quad P(A_i) = (a_i+1) / (s+1)$$

$$i \neq k \quad P(A_i) = a_i / (s+1)$$

### 2.2.2.2 Probabilidades condicionales:

$$i = k \text{ y } j = l \quad P(B_j | A_i) = (b_j + 1) / (a_i + 1)$$

$$i = k \text{ y } j \neq l \quad P(B_j | A_i) = b_j / (a_i + 1)$$

$$i \neq k \quad P(B_j | A_i) = b_j / a_i$$

Donde **s** corresponde al número de casos totales,

**i, j** los índices de las variables,

**k, l** los índices de las variables observadas.

El algoritmo anterior corresponde a una simplificación, suponiendo que las probabilidades tienen un valor preciso, es decir que no hay incertidumbre en las probabilidades. Un enfoque más adecuado, pero un poco más complejo, es utilizar una distribución de probabilidad para las probabilidades.

Normalmente se utiliza, para el caso de variables binarias, la distribución *Beta* y para variables multivaluadas, su generalización que es la distribución.

Para fines prácticos, se utiliza como estimado de la probabilidad el valor medio de las distribuciones; el cual corresponde, aproximadamente, al que se obtiene en el algoritmo anterior.

### 2.2.3 Aprendizaje Estructural

El razonamiento probabilístico o aprendizaje estructural consiste en propagar la evidencia a través de la red para conocer la probabilidad *a posteriori* de las variables. La propagación consiste en darle valores a ciertas variables (evidencia), y obtener la probabilidad posterior de las demás variables dadas las variables conocidas (instanciadas).

Los algoritmos de propagación dependen de la estructura de la red:

#### 2.2.3.1 Árboles



Figura 2.3 Estructura de Árbol.

El método para aprendizaje estructural de árboles (ver figura 2.3) se basa en el algoritmo desarrollado por Chow y Liu para aproximar una distribución de probabilidad por un producto de probabilidades de segundo orden, lo que corresponde a un árbol. La probabilidad conjunta de  $n$  variables se puede representar como:

$$A = \{A_1, A_2, \dots, A_n\}$$

Cada nodo corresponde a una variable discreta, , con su respectiva matriz de probabilidad condicional,  $P(B|A)=P(B_j|A_i)$

Dada cierta evidencia  $E$  --representada por la instanciación de ciertas variables-- la probabilidad posterior de cualquier variable  $B$ , por el teorema de Bayes:

$$P(B_i|E) = P(B_i)P(E|B_i)/P(E)$$

Ya que la estructura de la red es un árbol, el Nodo  $B$  la separa en dos subárboles, por lo que podemos dividir la evidencia en dos grupos:

$E^-$ : Datos en el árbol que cuya raíz es  $B$

$E^+$ : Datos en el resto del árbol

Entonces:

$$P(B_i|E) = P(B_i)P(E, E^+|B_i)/P(E)$$

Pero dado que ambos son independientes y aplicando nuevamente Bayes:

$$P(B_i|E) = \alpha P(B_i|E^+)P(E^-|B_i)$$

Donde  $\alpha$  es una constante de normalización.

Esto separa la evidencia para actualizar la probabilidad de  $B$ . Además vemos que no requerimos de la probabilidad *a priori*, excepto en el caso de la raíz donde:

$$P(A_i|E^+) = P(A_i)$$

Si definimos los siguientes términos:

$$\lambda(B_i) = P(E^-|B_i)$$

$$\pi(B_i) = P(B_i|E^+)$$

Entonces:

$$P(B_i|E) = \alpha\pi(B_i)\lambda(B_i)$$

Dado que los hijos son condicionalmente independientes dado el padre:

$$\lambda(B_i) = \prod_k P(E_k^-|B_i) = \prod_k \lambda_k(B_i)$$

Donde  $E_k^-$  corresponde a la evidencia que proviene del hijo  $k$  de  $B$ , denotado por  $S_k$ .

Condicionando cada término en la ecuación anterior respecto a todos los posibles valores de cada nodo hijo, obtenemos:

$$\lambda(B_i) = \prod_k [\sum_j P(E_k^-|B_i, S_j^k) P(S_j^k|B_i)]$$

Dado que  $B$  es condicionalmente de la evidencia bajo cada hijo dado éste y usando la definición de  $\lambda$ :

$$\lambda(B_i) = \prod_k [\sum_j P(S_j^k|B_i)\lambda(S_j^k)]$$

En forma análoga obtenemos una ecuación para  $\pi$ . Primero la condicionamos sobre todos los posibles valores del padre:

$$\pi(B_i) = \sum_j P(B_i|E^+, A_j)P(A_j|E^+)$$

Podemos eliminar  $E^+$  del primer termino dada independencia condicional. El segundo término representa la probabilidad posterior de  $A$  sin contar la evidencia de subárbol de  $B$ , por lo que podemos expresarla usando la ecuación para  $P(B_i|E)$  y la descomposición de  $\lambda$ .