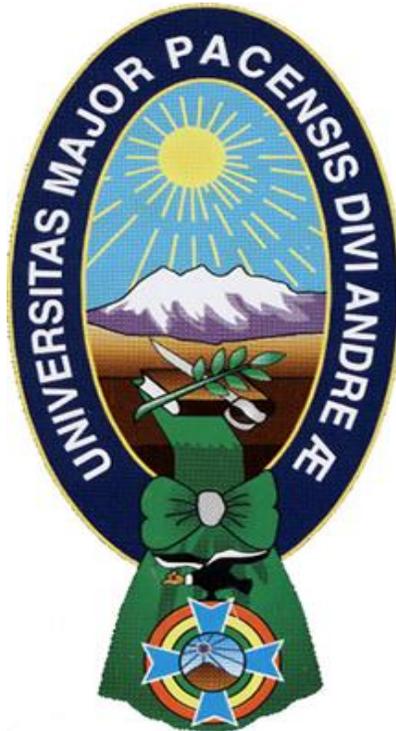




UNIVERSIDAD MAYOR DE SAN ANDRES  
FACULTAD DE CIENCIAS PURAS Y NATURALES  
CARRERA DE ESTADISTICA



## **TESIS DE GRADO**

REDES BAYESIANAS DISCRETAS APLICADAS AL ESTUDIO DE  
ENFERMEDADES EN LA PAZ

POSTULANTE: MARIELA JUANA TRUJILLO ESCOBAR

TUTOR: LIC. DINDO VALDEZ BLANCO

LA PAZ –BOLIVIA

2013

A mis padres y hermanos por toda su  
paciencia, cariño y apoyo.

# **AGRADECIMIENTOS**

A Dios por permitirme existir.

A mis padres y hermanos, porque ellos constituyen el pilar fundamental en mi vida, sobre el que se ha apoyado todo mi esfuerzo.

Al Lic. Juan Carlos flores, al Lic. Raúl Delgado por su amistad, por sus consejos y por animarme a culminar el trabajo de Tesis de Grado.

Al Lic. Dindo Valdez Blanco por darme la oportunidad de trabajar y culminar el presente trabajo.

A mis amigos por su cariño y ánimo constante.

Son numerosas las personas que directa e indirectamente han contribuido en la realización de este trabajo y a las que quiero agradecer su ayuda y comprensión.

Gracias.

# Índice General

<b>1</b>	<b><i>Introducción</i></b> .....	<b>1</b>
1.1.	Introducción. ....	1
1.2.	Antecedentes. ....	2
1.3.	Planteamiento del problema. ....	2
1.4.	Justificación de la Investigación. ....	2
1.5.	Planteamiento del Objetivo. ....	3
1.5.1.	Objetivo General. ....	3
1.5.2.	Objetivos Específicos ....	3
1.6.	Hipotesis. ....	4
1.7.	Alcances y limitaciones. ....	4
<b>2</b>	<b>Estructura de una Red Bayesiana Discreta</b> .....	<b>6</b>
2.1.	Axiomas de Probabilidad y Notaciones Básicas .....	6
2.1.1.	Distribución de Probabilidad sobre $\mathcal{X}$ .....	8
2.1.2.	Notación .....	8
2.1.3.	Variables Aleatorias y Vectores Aleatorios .....	8
2.1.4.	Distribución Marginal .....	11
2.1.5.	Independencia .....	12

2.1.6. Independencia Condicional .....	12
2.2. Actualización de la Probabilidad de Bayes .....	13
2.3. Factor de Bayes .....	14
2.4. Regla de Bayes aplicado a las Variables Aleatorias .....	15
2.5. Regla de Bayes .....	15
2.6. Regla de Jeffrey.....	16
2.7. Grafo .....	16
2.8. Grafos Acíclicos Dirigidos y Distribuciones de Probabilidad .....	21
2.8.1. Factorización de una Distribución de Probabilidad .....	21
2.8.2. Tipos de Conexiones .....	22
2.9. Potencial.....	25
2.10. Red Bayesiana.....	26
<b>3 Inferencia de una Red Bayesiana Discreta .....</b>	<b>28</b>
3.1. Evidencia.....	28
3.2. Evidencia Difícil.....	29
3.3. Evidencias Suaves y Evidencias Virtuales.....	30
3.4. Construcción de una Unión de Árboles. ....	30
3.4.1. Grafo Completo, Subconjunto Completo .....	30
3.4.2. Clique.....	31
3.4.3. Nodo Simplicial. ....	31
3.4.4. Conexión y Componentes Fuertes .....	31
3.4.5. Separador. ....	32

3.4.6. Separador Mínimo. ....	32
3.4.7. Árbol con raíz. ....	32
3.4.8. Diámetro. ....	32
3.4.9. Grafo Moral. ....	32
3.4.10. Acorde. ....	32
3.4.11. Triangulado.....	32
3.4.12. Descomposición.....	33
3.4.13. Grafo Descompuesto. ....	33
3.4.14. Grafo Descompuesto y Triangulación del Grafo.....	33
3.4.15. Eliminación de nodos de Secuencia Perfecta. ....	34
3.4.16. Eliminación de un nodo. ....	35
3.4.17. Eliminación de Secuencia.....	36
3.4.18. Eliminación de Dominio.....	36
3.4.19. Unión de Árboles.....	36
3.5. Potencial y Propiedad distributiva. ....	39
3.6. Potencial y Carga. ....	39
3.7. Contracción. ....	40
3.8. Marginalización.....	40
3.8.1. Marginalización y Propiedad Distributiva. ....	40
3.9. Eliminación y Dominio del Grafo.....	41
3.10. Dominio del Grafo. ....	42
3.11. Eliminando un Nodo. ....	42
3.12. Unión de Árboles ....	43

3.13. Factorización a lo largo de un Grafo no Dirigido. ....	44
3.14. Propiedad de Markov. ....	47
3.15. Factorizar a lo largo de una Unión de Árboles. ....	47
3.16. Factorización a lo largo de una Unión de Árboles y Carga Marginal.....	48
3.17. Introduciendo Evidencia. ....	49
3.18. Flujo de mensajes. ....	51
3.19. Cálculo local de la Unión de Árboles. ....	53
3.20. Contracción de una Carga sobre una Unión de Àrbol. ....	53
3.21. Transmisión del Mensaje Local. ....	53
3.22. Programa. ....	56
3.23. Sub-arbol,Vecino clique.....	56
3.24. Propagación Floja.....	56
3.25. Base de un Sub-árbol y Restricción de una Carga. ....	56
3.26. Unión de Àrboles Delgado.....	58
3.27. Usando una Unión de Árboles con Evidencia Virtual y Evidencia Suave. ....	58
<b>4 Árbol de Chow Liu .....</b>	<b>60</b>
4.1. Introducción .....	60
4.2. Asignación de una Distribución de Probabilidad a la Estructura de Grafo.....	60
4.3. Distribución apriori de la Estructura del Grafo.....	61
4.4. Maximización de la estructura del grafo .....	61
4.5. Predicción del problema.....	63

4.6. NP-Hard .....	6
4.7. Árbol de Chow-Liu .....	64
4.8. Divergencia de Kullback-Leibler .....	64
4.9. Información Mutua.....	65
<b>5 Aplicación de Redes Bayesianas Discretas .....</b>	<b>71</b>
5.1 Introducción .....	71
5.2 Descripción de las variables .....	72
5.3 Resultados con el software Hugin Lite .....	80
5.4 Relación dependencia entre variables .....	82
<b>6 Conclusiones y Recomendaciones.....</b>	<b>89</b>
6.1. Conclusiones.....	89
6.2. Recomendaciones .....	90
<b>A Redes bayesianas con R .....</b>	<b>90</b>
<b>B Software Hugin Lite .....</b>	<b>96</b>
Bibliografía	

# Índice de Figuras

Figura 5.1 Red apriori con el programa R .....	76
Figura 5.2 Grafo Moralizado con el programa R.....	77
Figura 5.3 Grafo Triangulado con el programa R .....	78
Figura 5.4 Cliques o grupo de nodos .....	79
Figura 5.5 Árbol de Chow Liu con R .....	80
Figura 5.6 Árbol de Chow Liu con Hugin Lite.....	81

## Índice de Tablas

Tabla 5.1.1 Desequilibrio Electrolítico según Hiponatremia.....	83
Tabla 5.1.2 Consistencia según Convulsiones .....	83
Tabla 5.1.3 Acidosis Metabólica según Desequilibrio Electrolítico .....	84
Tabla 5.1.4 Plan de Hidratación según el Estado de Hidratación.....	84
Tabla 5.1.5 Vómitos según Plan de Hidratación .....	85
Tabla 5.1.6 Convulsiones según Estado de Hidratación.....	86
Tabla 5.1.7 Método de medición según Fiebre .....	86
Tabla 5.1.8 Fiebre según el Plan de Hidratación .....	87
Tabla 5.1.9 Hiponatremia según Desequilibrio Electrolítico .....	87
Tabla 5.1.10 Estado de Hidratación según Acidosis Metabólica .....	88
Tabla 5.1.11 Estado de Hidratación.....	88

# RESUMEN

Las Redes Bayesianas son un tipo de modelos gráficos probabilísticos, que se caracterizan por modelizar relaciones de dependencia e independencia entre nodos o variables. Dentro de las Redes Bayesianas, se describen las Redes Bayesianas Discretas y las Redes Bayesianas Gaussianas, siendo el enfoque principal de este trabajo las Redes Bayesianas Discretas, como su nombre indica este tipo de análisis se enfoca únicamente en las variables discretas.

**Capítulo 1.** Se presenta la introducción a las Redes Bayesianas y principalmente el objetivo general y objetivos específicos del presente trabajo.

**Capítulo 2.** Se describen las Redes Bayesianas y sus principales características (definiciones de la teoría de la probabilidad y grafos), una red bayesiana es la interacción entre la teoría de la probabilidad y la teoría de grafos.

**Capítulo 3.** Se describe la actualización de la información que se tiene acerca de las variables del problema, este proceso se denomina actualización de la evidencia y se realiza para conocer y describir las variables de interés del problema cuando se tiene información o evidencia acerca de la ocurrencia de otra u otras variables del mismo.

**Capítulo 4.** Se describe el algoritmo de Chow Liu para obtener la red Bayesiana Discreta con el mejor peso y por supuesto el más preciso, el algoritmo de Chow Liu es otra forma de realizar la propagación de la red.

**Capítulo 5.** Se realiza la aplicación de la Teoría de Redes Bayesianas y el algoritmo de Chow Liu y así obtener un resultado preciso con las herramientas desarrolladas en los capítulos anteriores.

# *Capítulo 1*

## *Introducción*

### **1.1 Introducción.**

Hoy en día la estadística clásica ha sido difundida ampliamente en toda nuestra sociedad por esta razón se utiliza en numerosas aplicaciones prácticas y diversas disciplinas. Sin embargo, en los últimos años se ha producido un notable avance en la estadística bayesiana como es el caso de las Redes Bayesianas.

La redes bayesianas es una técnica para tratar el razonamiento con incertidumbre la misma trabaja con una interacción entre la teoría de la probabilidad y la teoría de grafos, el conjunto de variables aleatorias conforman los nodos de la red y la dependencia entre variables está representada por las aristas dirigidas (o arcos) que conectan pares de nodos o variables y cada nodo tiene una distribución de probabilidad condicional. Las redes bayesianas se dividen en dos partes una es la parte cualitativa y la otra es la parte cuantitativa, en la parte cualitativa se puede entender como la representación gráfica (Grafo), la ventaja de representar el modelo gráficamente es que podemos indicar relaciones de dependencia condicional entre variables, en la parte cuantitativa cada nodo está asociado a la distribución de probabilidad condicional así cada nodo lleva una probabilidad, esta probabilidad es susceptible de ser modificado cuando se conoce alguna evidencia de las variables del modelo, esta información se propaga en la red y se re-calculan los valores asociados a otras variables.

Las redes bayesianas es una técnica bastante útil y precisa para resolver problemas que contiene incertidumbre y complejidad, por esta razón las Redes Bayesianas son utilizadas principalmente en el análisis de grandes cantidades de datos.

## **1.2 Antecedentes.**

A principios de los años 80 aparece la Red Bayesiana como un un modelo probabilista, que lleva asociado un modelo gráfico que está inspirado en la causalidad. El extraordinario desarrollo experimentado por las Redes Bayesianas en las últimas décadas ha permitido construir modelos gráficos para problemas de tamaño considerable. Las Redes Bayesianas surgieron como alternativa a los sistemas expertos clásicos orientados a la toma de decisiones y a la predicción bajo incertidumbre en términos probabilísticos.

En los primeros sistemas expertos, se utilizaba la probabilidad como una medida para tratar la incertidumbre pero, se encontraron problemas, debido a que le dieron un uso incorrecto a algunas hipótesis de independencia, en los primeros resultados de los sistemas expertos, la probabilidad fue considerada como una medida de incertidumbre poco práctica, esto se debió a que se manejaba una gran cantidad de parámetros que imposibilitó la estimación precisa de los mismos.

Posteriormente, surgieron medidas alternativas a la probabilidad, como los factores de certeza, las credibilidades, las plausibilidades, las necesidades o las posibilidades, para tratar la incertidumbre. Sin embargo, con la aparición de las redes probabilísticas, principalmente las Redes Bayesianas y Markovianas, la probabilidad ha resurgido de forma espectacular y a la fecha es una de las medidas más aceptadas. De hecho es un campo actual de investigación que permanece abierto.

## **1.3 Planteamiento del problema.**

Una de las ventajas de las Redes Bayesianas es que con ella se puede manejar grandes cantidades de datos pero esto no queda ahí ya que esto se complica más si hay muchas variables en estudio, pues el número de redes crece exponencialmente al aumentar el número de variables. Es por esta razón que es necesario buscar la mejor estructura de la red y visualizar las relaciones entre los nodos o variables.

## **1.4 Justificación de la investigación.**

Se pretende mostrar un enfoque diferente tanto teórico como práctico ya que las investigaciones realizadas con un amplio número de variables donde existen relaciones complejas, requieren un análisis complejo, además Redes Bayesianas nos ayuda a descubrir relaciones ocultas entre las variables. La Red Bayesiana puede ser aplicada en diferentes disciplinas o áreas de nuestro entorno social, como: la incertidumbre de las fluctuaciones del mercado financiero, en los diagnósticos médicos, en las redes sociales, entre otras. En nuestro país normalmente aplicamos la estadística clásica pero sería una excelente herramienta utilizar la estadística bayesiana como es el caso de las redes bayesianas que por supuesto resulta ser un método efectivo en la toma de decisiones.

Entre las ventajas que tienen las redes bayesianas, tenemos:

- ✓ Nos permite presentar la dimensión cualitativa y cuantitativa de un problema en un entorno gráfico.
- ✓ Nos permite trabajar con datos perdidos de una manera eficiente.
- ✓ Nos permite construir un modelo gráfico probabilístico a partir de una base de datos que contenga un conjunto de observaciones sobre un conjunto de variables.

## **1.5 Planteamiento del objetivo.**

### **1.5.1 Objetivo general.**

Desarrollar la teoría de Redes Bayesianas discretas y mostrar las ventajas que tiene al construir un modelo gráfico probabilístico, en base al cual se diseñará la red bayesiana para el análisis de enfermedades en ciudad de La Paz.

### **1.5.2 Objetivos específicos.**

- a) Entender la topología o estructura de la red y la inferencia aplicado a la misma.
- b) Construir un modelo gráfico probabilístico para el análisis de enfermedades en la ciudad de La Paz.
- c) Desarrollar el algoritmo de Chow Liu que nos permita obtener la mejor estructura de la red.

- d) Aplicar el algoritmo de Chow liu en el diseño de la red bayesiana discreta al estudio de enfermedades diarreicas en La Paz.

**1.6 Hipótesis.** Las Redes Bayesianas aplicadas al estudio de los modelos de relación probabilísticos permiten estimar las probabilidades condicionales con una gran cantidad de parámetros sin considerar el supuesto de independencia entre los mismos.

### **1.7 Alcances y limitaciones.**

El presente trabajo tiene los siguientes alcances y limitaciones:

- En el presente tema, nuestra atención se centrara en los modelos gráficos yprobabilísticos dirigidos. Los modelos gráficos probabilísticos mucho depende del tipo de grafo que se utilice al definir el problema, es decir dependiendo del tipo de arista que se emplea en la representación gráfica del mismo, podemos distinguir tres tipos de modelos gráficos probabilísticos:
  - ✓ Modelos gráficos probabilísticos no dirigidos. Redes de Markov.
  - ✓ Modelos gráficos probabilísticos dirigidos. Redes Bayesianas.
  - ✓ Modelos gráficos probabilísticos dirigidos y no dirigidos, o mixtos. Redes Cadena.
- Una vez a aclarada el tipo de modelos probabilísticos que se desarrollara a lo largo del tema el primer enfoque se basara en construir la Red Bayesiana es decir la topología o estructura de una Red Bayesiana y en un segundo enfoque tenemos como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de nuestra estructura dada.
- Este trabajo tiene la limitación de construir redes bayesianas discretas es decir el manejo de las variables serán discretas, una vez construido la estructura de la red se aplica el algoritmo de Chow Liu que reduce la estructura de la red a una mejor estructura para después realizar la propagación de probabilidades y así obtener resultados eficientes. Cabe mencionar también que las Redes Bayesianas se subdividen en Redes Bayesianas Discretas, Redes Bayesianas Gaussianas,

Redes Bayesianas Mixtas pero nuestro caso se centrara solo en Redes Bayesianas Discretas. Aunque hubiera sido interesante desarrollar los diferentes tipos de Redes Bayesianas mencionadas anteriormente, el tema está abierto a mucha más investigación ya que existen diversos tipos de redes probabilistas que por supuesto son temas de investigacion bastante importante en el campo de la estadística, como en todo trabajo disponemos de una limitación y en nuestro caso, nos concentraremos principalmente en la Redes Bayesianas Discretas dejamos a consideración los temas que no será posibles desarrollar en el presente documento.

## Capítulo 2

### *Estructura de una Red Bayesiana Discreta*

En este capítulo se desarrollan las herramientas necesarias para la construcción de una red bayesiana, por tal motivo es importante definir los axiomas de probabilidad y la definición de grafo, en las definiciones se mencionan las variables discretas y continuas como también se menciona a las aristas dirigidas y no dirigidas, cabe recordar que en este trabajo se desarrollan Redes Bayesianas Discretas por tal motivo se consideran variables discretas y por supuesto solo aristas dirigidas es decir que la estructura de la red es un grafo acíclico dirigido lo que significa que la arista si y solo si tiene dirección de una variable a otra, lo que significa que tenemos dos variables  $\alpha_i$  y  $\alpha_j$  donde la arista dirigida se representa de esta manera  $\alpha_i \rightarrow \alpha_j$ , en el caso de las aristas no dirigidas la representación es de la siguiente manera  $\alpha_i - \alpha_j$ .

#### **2.1 Axiomas de probabilidad y notaciones básicas.**

Las notaciones usadas a lo largo del tema serán las siguientes:

- El conjunto universal se denota por  $\Omega$  la misma consiste en el conjunto de todos los resultados posibles de un experimento, junto con todos los posibles valores de los parámetros desconocidos.
- La notación  $\mathcal{X}$  se utiliza para denotar el espacio de la muestra, el conjunto de todos los posibles resultados del experimento.
- $\Theta$  se utiliza para denotar el espacio de parámetros, de modo que  $\Omega = \mathcal{X} \times \Theta$

Las siguientes notaciones se tomarán en cuenta en el momento de considerar conjuntos:

- Si  $A$  y  $B$  son dos conjuntos, entonces  $A \cup B$  denota su unión. Si  $A_1, \dots, A_n$  son una colección finita de conjuntos, entonces  $\bigcup_{j=1}^n A_j$  denota su unión.
- Si  $A$  y  $B$  son dos conjuntos, entonces  $A \cap B$  puede ser utilizado para denotar su intersección. Si  $A_1, \dots, A_n$  son una colección finita de conjuntos, entonces  $\bigcap_{i=1}^n A_i$
- $A \subset B$  denota que  $A$  es un subconjunto estricto de  $B$ .  $A \subseteq B$ , indica que  $A$  es un subconjunto de  $B$ , posiblemente igual a  $B$ .
- El conjunto vacío se denota por  $\varnothing$ .
- $A^c$  denota el complemento de  $A$ , es decir  $\Omega \setminus A$  donde  $\Omega$  denota el conjunto universal. El símbolo  $\setminus$  denota la exclusión.
- Junto con el conjunto universal  $\Omega$ , el espacio de eventos  $F$ . Esta es una colección de subconjuntos de  $\Omega$ . El espacio de eventos  $F$  es un álgebra de conjuntos. Es decir,  $F$  satisface a lo siguiente:
  1.  $\varnothing \in F$  y  $\Omega \in F$ .
  2. Cada  $A \in F$  es construido. Es decir, para cada  $A \in F$ , sea  $i_A : \Omega \rightarrow \{0,1\}$  denotan la asignación de tal manera que  $i_A(w) = 1$  si  $w \in A$  y  $i_A(w) = 0$  si  $w \in A^c$ . Entonces existe un procedimiento para determinar el valor de  $i_A(w)$  para cada  $w \in \Omega$ .
  3. Si para un conjunto finito de eventos  $(A_j)_{j=1}^n$  cada  $A_j$  cumple que  $A_j \in F$  entonces  $\bigcup_{j=1}^n A_j \in F$ .
  4. Si  $A \in F$ , entonces  $A^c \in F$ , donde  $A^c = \Omega \setminus A$  denota el complemento de  $A$ .
  5. Cada  $A \in F$  satisface  $A \subseteq \Omega$ .

Para  $\Omega$ ,  $F$  cumple las condiciones anteriores, una distribución de probabilidad  $P$  sobre  $(\Omega, F)$ , una función  $P : F \rightarrow [0,1]$  satisface la siguiente versión de los axiomas de Kolmogorov:

1.  $p(\varphi) = 0$  y  $p(\Omega) = 1$ .
2. Si  $(A_j)_{j=1}^n$  es una colección finita de tal manera que cada  $A_j \in F$  y los eventos satisface  $A_j \cap A_k = \emptyset$  Para todo  $j \neq k$  entonces  $p\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n p(A_j)$ .
3.  $0 \leq p(A) \leq 1$  para todo  $A \in F$ .

Esta es una versión reducida de los axiomas de Kolmogorov.

### 2.1.1 Distribución de probabilidad sobre $\mathcal{X}$

Si  $\mathcal{X}$  contiene un número finito de elementos, entonces,  $F_{\mathcal{X}}$  contiene un número finito de elementos. En este contexto, una distribución de probabilidad sobre  $\mathcal{X}$  debe satisfacer:

- $p(\{x\}) \geq 0$  para todo  $x \in \mathcal{X}$ .
- Para cualquier  $A \in F_{\mathcal{X}}$ ,  $\sum_{x \in A} p(\{x\}) = p(A)$ .
- En particular,  $\sum_{x \in \mathcal{X}} p(\{x\}) = 1$ .

### 2.1.2 Notación.

Sea  $\mathcal{X}$  un espacio de estados finitos y  $A_{\mathcal{X}}$  denota el conjunto de todos los subconjuntos de  $\mathcal{X}$  (incluyendo  $\varphi$  el conjunto vacío y  $\mathcal{X}$ ). Para la distribución de probabilidad  $p : A_{\mathcal{X}} \rightarrow [0,1]$  definido anteriormente,  $P$  denota la función  $p : \mathcal{X} \rightarrow [0,1]$  tal que  $p(x) = p(\{x\})$ .

### 2.1.3 Variables aleatorias y vectores aleatorios.

Las variables aleatorias discretas, variables aleatorias continuas y vectores aleatorios satisfacen las siguientes propiedades:

Una variable aleatoria discreta  $Y$  es una función  $Y:\Omega \rightarrow \zeta$ , donde  $\zeta$  es un espacio contable, que satisface las siguientes condiciones: para cada  $x \in \zeta$ ,  $\{w | Y(w) = x\} \in F$  y hay una función  $p_Y:\zeta \rightarrow R_+$ , conocida como la función de probabilidad de  $Y$ , tal que para cada  $x \in \zeta$ ,

$$p_Y(x) = p(\{w | Y(w) = x\})$$

La probabilidad de  $Y$  es instanciado en  $x$ . Por lo tanto, para cualquier subconjunto  $C \subseteq \zeta$ ,  $p_Y$ , satisface

$$\sum_{x \in C} p_Y(x) = p(\{w | Y(w) \in C\})$$

En particular, tomando  $C = \zeta$ ,

$$\sum_{x \in \zeta} p_Y(x) = 1.$$

Una variable aleatoria continua  $\Xi$  se define como una función  $\Xi:\Omega \rightarrow R$  tal que para cualquier conjunto  $A \subseteq R$  tal que la función  $1_{A \cap [-N, N]}$  es la integral de Riemann para todo  $N \in Z^+$ , el conjunto  $\{w | \Xi(w) \in A\} \in F$ , y para los cuales hay una función de  $\pi_\Xi$ , conocida como función de densidad, de tal manera que para cualquier  $A \subset R$  con la función integral de Riemann<sup>1</sup>,

$$p(\{w | \Xi(w) \in A\}) = \int_A \pi_\Xi(x) dx.$$

En particular,

$$p(\{w | \Xi(w) \in R\}) = p(\Omega) = \int_R \pi_\Xi(x) dx = 1.$$

Un vector aleatorio  $\underline{y}$  es un vector de variables aleatorias. Será tomado como un vector fila las diferentes características de una sola observación, será tomada como un vector columna las variables aleatorias independientes e idénticamente distribuidas.

---

<sup>1</sup>Departamento de análisis matemático “calculo diferencial e integral”

Es representada como una matriz de datos. Cada columna de una matriz de datos representa un atributo diferente, cada fila representa una observación diferente.

Un vector fila aleatorio  $\underline{Y}$  es una colección de variables aleatorias que satisface lo siguiente: supongamos que  $\underline{Y} = (Y_1, \dots, Y_m)$  y para cada  $j = 1, \dots, m$ ,  $Y_j$  es una variable aleatoria discreta que toma valores en un espacio contables  $\zeta_j$  y sea  $\zeta = \zeta_1 \times \dots \times \zeta_m$ , entonces  $\underline{Y}$  es un vector aleatorio, si por cada,

$$(y_1, \dots, y_m) \in \zeta, \{w \mid (Y_1(w), \dots, Y_m(w)) \in F\} \in \mathcal{F}$$

hay una función de probabilidad conjunta  $p_{Y_1, \dots, Y_m} : \zeta \rightarrow [0, 1]$  tal que para cualquier conjunto  $C \subset \zeta$ ,

$$\sum_{(y_1, \dots, y_m) \in C} p_{Y_1, \dots, Y_m}(y_1, \dots, y_m) = p(\{w \mid (Y_1(w), \dots, Y_m(w)) \in C\}).$$

En particular,

$$\sum_{(y_1, \dots, y_m) \in \zeta} p_{Y_1, \dots, Y_m}(y_1, \dots, y_m) = 1$$

Si  $\underline{\Xi} := (\Xi_1, \dots, \Xi_n)$  es una conjunto de  $n$  variables aleatorias, donde para cada  $j = 1, \dots, n$ ,  $\Xi_j$  es una variable aleatoria continua, entonces  $\underline{\Xi}$  es un vector aleatorio si para cada conjunto  $A \subset \mathbb{R}^n$  tal que  $1_{A \cap [N, N]^n}$  es integral de Riemann para cada  $N \in \mathbb{Z}_+$ .

$$\{w \mid (\Xi_1(w), \dots, \Xi_n(w)) \in A\} \in \mathcal{F}$$

hay una función integral de Riemann  $\pi_{\Xi_1, \dots, \Xi_n} : \mathbb{R}^n \rightarrow \mathbb{R}_+$  donde  $\mathbb{R}_+$  denota los números reales positivos tal que para cada conjunto  $A \subset \mathbb{R}^n$  tal que  $1_{A \cap [N, N]^n}$  es integral de Riemann para cada  $N \in \mathbb{Z}_+$ .

$$p(\{w \mid (\Xi_1(w), \dots, \Xi_n(w)) \in A\}) = \int_A \pi_{\Xi_1, \dots, \Xi_n}(x_1, \dots, x_n) d_{x_1}, \dots, d_{x_n}.$$

En particular

$$\int_{\mathbb{R}^n} \pi_{\Xi_1, \dots, \Xi_n}(x_1, \dots, x_n) d_{x_1}, \dots, d_{x_n} = 1$$

### 2.1.4 Distribución marginal.

Sea  $X = (X_1, \dots, X_n)$  un vector aleatorio discreto, con función de probabilidad conjunta  $p_{X_1}, \dots, p_{X_n}$  la distribución de probabilidad para  $(X_{j_1}, \dots, X_{j_m})$  donde  $m \leq n$  y  $1 \leq j_1 < \dots < j_m < n$  se conoce como la distribución marginal y la función de probabilidad marginal es definida como:

$$p_{X_{j_1}, \dots, X_{j_m}}(x_1, \dots, x_m) = \sum_{(y_1, \dots, y_n | y_{j_1}, \dots, y_{j_m}) = (x_1, \dots, x_m)} p_{X_1, \dots, X_n}(y_1, \dots, y_n)$$

En particular, para dos variables discretas  $X$  e  $Y$  tomando valores en los espacios  $\zeta_X$  y  $\zeta_Y$  respectivamente, con la función de probabilidad conjunta  $p_{X,Y}$  la función de probabilidad marginal de la variable aleatoria  $X$  es definida como:

$$p_X(x) = \sum_{y \in \zeta_Y} p_{X,Y}(x, y)$$

Y la función de probabilidad marginal de la variable aleatoria  $Y$  se define como:

$$p_Y(y) = \sum_{x \in \zeta_X} p_{X,Y}(x, y)$$

Si  $\Xi = (\Xi_1, \dots, \Xi_n)$  es un vector aleatorio continuo, con función de densidad conjunta  $\pi_{\Xi_1, \dots, \Xi_n}$  entonces la función de densidad marginal para  $\Xi_{j_1}, \dots, \Xi_{j_m}$ , donde  $\{j_1, \dots, j_m\} \subset \{1, \dots, n\}$  es definida como:

$$\pi_{\Xi_{j_1}, \dots, \Xi_{j_m}}(x_1, \dots, x_m) = \int_{\mathbb{R}^{n-m}} \pi_{\Xi_1, \dots, \Xi_n}(y_1, \dots, y_n | (y_{j_1}, \dots, y_{j_m}) = (x_1, \dots, x_m)) \prod_{k \notin \{j_1, \dots, j_m\}} dy_k.$$

### 2.1.5 Independencia.

Dos variables aleatorias discretas  $X$  y  $Y$  son independientes si y sólo si  $p_{X,Y} = p_X p_Y$

Un conjunto de  $d$  variables aleatorias  $\{X_1, \dots, X_d\}$  se dice que son conjuntamente independientes si para cualquier vector aleatorio  $(X_{i_1}, \dots, X_{i_m})$  donde  $i_j \neq i_k$  para todo  $j \neq k$

$$p_{X_{i_1}, \dots, X_{i_m}} = \prod_{j=1}^m p_{X_{i_j}}.$$

### 2.1.6 Independencia condicional.

La Independencia condicional es la notación clave probabilística en redes bayesianas. Lo siguiente es un resumen de algunas propiedades básicas de la independencia condicional.

Sea  $(\underline{X}, \underline{Y}, \underline{Z})$  tres vectores aleatorios discretos, con función de probabilidad conjunta

$p_{\underline{X}, \underline{Y}, \underline{Z}}$ . Sea  $\mathcal{X}_X, \mathcal{X}_Y, \mathcal{X}_Z$  denotan los espacios de estado para  $\underline{X}, \underline{Y}$  y  $\underline{Z}$  respectivamente.

Los vectores  $\underline{X}$  e  $\underline{Y}$  se dice que son condicionalmente independientes dado  $\underline{Z}$  si para todo  $(\underline{x}, \underline{y}, \underline{z}) \in \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$ .

$$p_{\underline{X}, \underline{Y}, \underline{Z}}(\underline{x}, \underline{y}, \underline{z}) = p_{\underline{X}|\underline{Z}}(\underline{x} | \underline{z}) p_{\underline{Y}|\underline{Z}}(\underline{y} | \underline{z}) p_{\underline{Z}}(\underline{z}).$$

Esto se indicara por la notación

$$\underline{X} \perp \underline{Y} | \underline{Z}$$

La notación  $\underline{X} \perp \underline{Y}$  denota que  $\underline{X}$  y  $\underline{Y}$  son independientes es decir

$$p_{\underline{X} \perp \underline{Y}}(\underline{x}, \underline{y}) = p_{\underline{X}}(\underline{x}) p_{\underline{Y}}(\underline{y}) \text{ para todo } (\underline{x}, \underline{y}) \in \mathcal{X}_X \times \mathcal{X}_Y.$$

Lo siguiente es todo el equivalente a  $\underline{X} \perp \underline{Y} | \underline{Z}$  :

1. Para todo  $(\underline{x}, \underline{y}, \underline{z}) \in \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$  tal que  $p_{Y|Z}(\underline{y} | \underline{z}) > 0$  y  $p_Z(\underline{z}) > 0$ ,  

$$p_{X|Y,Z}(\underline{x} | \underline{y}, \underline{z}) = p_{X|Z}(\underline{x}, \underline{z}).$$
2. Entonces existe una función  $a: \mathcal{X}_X \times \mathcal{X}_Z \rightarrow [0,1]$  tal que para todo  
 $(\underline{x}, \underline{y}, \underline{z}) \in \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$  satisface  $p_{Y|Z}(\underline{y} | \underline{z}) > 0$  y  $p_Z(\underline{z}) > 0$ ,  

$$p_{X|Y,Z}(\underline{x} | \underline{y}, \underline{z}) = a(\underline{x}, \underline{z}).$$
3. Existen funciones  $a: \mathcal{X}_X \times \mathcal{X}_Z \rightarrow R$  y  $b: \mathcal{X}_Y \times \mathcal{X}_Z \rightarrow R$  tal que para todo  
 $(\underline{x}, \underline{y}, \underline{z}) \in \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$  satisface  $p_Z(\underline{z}) > 0$ ,  $p_{X|Y,Z}(\underline{x} | \underline{y}, \underline{z}) = a(\underline{x}, \underline{z})b(\underline{y}, \underline{z})$
4. Para todo  $(\underline{x}, \underline{y}, \underline{z}) \in \mathcal{X}_X \times \mathcal{X}_Y \times \mathcal{X}_Z$  tal que  $p_Z(\underline{z}) > 0$ .  

$$p_{X,Y,Z}(\underline{x}, \underline{y}, \underline{z}) = \frac{p_{X,Z}(\underline{x}, \underline{z})p_{Y,Z}(\underline{y}, \underline{z})}{p_Z(\underline{z})}.$$
5. Existen funciones  $a: \mathcal{X}_X \times \mathcal{X}_Z \rightarrow R$  y  $b: \mathcal{X}_Y \times \mathcal{X}_Z \rightarrow R$  tal que  

$$p_{X,Y,Z}(\underline{x}, \underline{y}, \underline{z}) = a(\underline{x}, \underline{z})b(\underline{y}, \underline{z}).$$

## 2.2 Actualización de la probabilidad de Bayes.

La probabilidad a priori es la probabilidad  $P$  sobre  $F$ , la probabilidad a priori es la probabilidad clásica. Considere una distribución a priori denotada por  $p(\cdot)$  y se tiene la información de que  $B \in F$  y se quiere actualizar las probabilidades sobre esta información. Esta actualización consiste en introducir una nueva distribución de probabilidad  $p^*$  en  $F$ . Supongamos que  $p(B) > 0$  entonces se sabe que  $B$  es seguro, la actualización requiere  $p^*(B) = 1$ .

Por lo tanto  $p^*(B^c) = 0$ , donde  $B^c = \Omega \setminus B$  es el complemento de  $B$  en  $\Omega$ . La probabilidad de actualización está construida de modo que la proporción de probabilidades para cualquier  $B_1 \subset B$  y  $B_2 \subset B$  no cambie es decir  $B_i \subset B$  donde  $i = 1, 2$ .

$$\frac{p^*(B_1)}{p^*(B_2)} = \frac{p(B_1)}{p(B_2)}$$

Para un arbitrario  $A \in F$  da lo siguiente:

$$p^*(A) = p^*(A \cap B) + p^*(A \cap B^c).$$

Dado  $p^*(B^c) = 0$  sigue que  $p^*(A \cap B^c) \leq p(B^c) = 0$  en otras palabras de manera arbitraria  $A \in F$ , dado  $p^*(B) = 1$

$$p^*(A) = \frac{p^*(A \cap B) + p^*(A \cap B^c)}{p^*(B)} = \frac{p^*(A \cap B)}{p^*(B)} = \frac{p(A \cap B)}{p(B)}.$$

La transformación  $p \rightarrow p^*$  se conoce como la actualización de probabilidad de Bayes.

Cuando la evidencia es obtenida es preciso que un evento  $B \in F$ , la regla de Bayes es aplicada para actualizar la distribución de probabilidad. La notación  $p(A|B)$  denota a su vez  $p^*(A)$  se tiene:

$$p(A|B) \stackrel{def}{=} p^*(A) = \frac{p(A \cap B)}{p(B)}. \quad (2.1)$$

Esto se conoce como la probabilidad condicional de  $A$  dado  $B$ .

### 2.3 Factor de Bayes.

Sea  $P$  y  $Q$  dos distribuciones sobre un algebra  $A$  para cualquiera de dos eventos  $A, B \in A$  el factor de Bayes definido como:

$$F_{p,q}(A, B) = \frac{q(A)/q(B)}{p(A)/p(B)}. \quad (2.2)$$

$P$  es una probabilidad antes de la actualización y  $Q$  es la probabilidad actualizada. El factor de Bayes indica sí o no la nueva información a incrementado la probabilidad de un evento  $A$  con respecto a  $B$ .

## 2.4 Regla de Bayes aplicado a las variables aleatorias.

Sean  $X$  e  $Y$  dos variables aleatorias discretas sea  $p_X, p_Y, p_{X|Y}, p_{Y|X}$  denotan las función de probabilidad de  $X, Y, X|Y$  e  $Y|X$ , respectivamente. De ello se deduce directamente de la Regla de Bayes que para todo  $x, y$ .

$$p_{X|Y}(x|y) = \frac{p_{Y,X}(y|x)p_X(x)}{p_Y(y)}. \quad (2.3)$$

Si  $X$  e  $Y$  son variables aleatorias continuas con función de densidad  $\pi_X$  y  $\pi_Y$  entonces la función de probabilidad condicional de  $X$  dado  $Y = y$  es:

$$\pi_{X|Y}(x|y) = \frac{\pi_{Y|X}(y|x)\pi_X(x)}{\pi_Y(y)}.$$

## 2.5 Regla de Bayes.

La regla de Bayes se obtiene teniendo en cuenta un conjunto exhaustivo y finito de hipótesis mutuamente excluyentes  $\{H_i\}_{i=1}^m$ . La ley de la probabilidad total dado para cualquier  $E$ .

$$p(E) = \sum_{i=1}^m p(H_i)p(E|H_i) \quad (2.4)$$

$$p(H_i|E) = \frac{p(E|H_i)p(H_i)}{\sum_{i=1}^m p(H_i)p(E|H_i)} \quad (2.5)$$

Aquí  $p(H_i)$  es la probabilidad inicial o a priori de una hipótesis, antes de cualquier evidencia  $E$  sea obtenido, entonces  $p(H_i|E)$  indica la forma  $p(H_i)$  se actualiza a una nueva probabilidad, llamada probabilidad a posteriori sobre la base de esta nueva evidencia.

## 2.6 Regla de Jeffrey.

Supongamos que nueva evidencia implica formar un conjunto exhaustivo de  $r$  de hipótesis mutuamente exclusiva  $(G_i)_{i=1}^r$ , las evidencias tienen probabilidades  $p^*(G_i)$ . La regla de Jeffrey puede aplicarse a la situación en la que puede suponer que para todos los eventos  $A \in F$  las probabilidades  $p(A|G_i)$  permanecen sin cambios. Solo la evolución de hipótesis mutuamente exclusiva  $(G_i)_{i=1}^r$ .

### Definición 2.6.1 (Actualización de Jeffrey)

La regla de Jeffrey para el cálculo de la actualización de la probabilidad para cualquier evento  $A$ , viene dada por:

$$p^*(A) = \sum_{i=1}^r p^*(G_i) p(A|G_i). \quad (2.6)$$

La regla de Jeffrey proporciona una función de probabilidad consistente, tal que  $p^*(A|G_i) = p(A|G_i)$  para toda  $i$ .

## 2.7 Grafo.

Como ya se menciona es necesario desarrollar algunas definiciones básicas sobre grafos. Donde cada nodo representa una de las variables  $\{\alpha_1, \dots, \alpha_d\}$ , y  $E$  el conjunto de aristas, cada arista representa relaciones de dependencia directas e indirectas entre las variables. Un grafo acíclico dirigido es una estructura grafica básica para, posteriormente, poder especificar mediante un grafo un problema con incertidumbre en el que se relacionan un conjunto de variables.

### Arista.

Una arista es lo que se denomina un arco dirigido, es decir muestra la relación de una variable con otra y la podemos representar de esta manera  $\alpha_i \rightarrow \alpha_j$  y se denomina arista dirigida pero una arista no dirigida se la puede representar de la siguiente manera  $\alpha_i - \alpha_j$ .

### Nodo.

Son variables aleatorias  $V = \{\alpha_1, \dots, \alpha_d\}$  donde cada variable  $\alpha_i$  representa un nodo.

**Nodo padre.**

$\alpha$  es un nodo padre de  $\beta$  sí y sólo si existe una arista de  $\alpha \rightarrow \beta$  entonces se dice también que  $\beta$  es el **nodo hijo** de  $\alpha$ .

**Definición 2.7.1 (Grafo, Grafo simple).**

Un grafo  $G = (V, E)$  consiste en un conjunto finito de nodos  $V$  y un conjunto de aristas  $E$ , donde se encuentra cada arista en  $V \times V$ . El conjunto de aristas por lo tanto consta de pares de nodos. Sea  $V = \{\alpha_1, \dots, \alpha_d\}$ . Un grafo se dice que es simple si  $E$  no contiene aristas de la forma  $(\alpha_j, \alpha_j)$  (que es un bucle desde el nodo a sí mismo) y cualquiera de las aristas  $(\alpha_j, \alpha_k) \in E$  aparece exactamente una vez. Es decir, no están permitidas múltiples aristas. Para cualquiera de los dos nodos distintos  $\alpha$  y  $\beta \in V$ , el par ordenado  $(\alpha, \beta) \in E$  si y sólo si existe una arista dirigida de  $\alpha$  a  $\beta$ . Una arista no dirigida se denominará  $\langle \alpha, \beta \rangle$ .

Para un grafo simple que puede contener tanto las aristas dirigidos y no dirigidos, el conjunto de aristas  $E$  se puede descomponer como  $E = D \cup U$ , donde  $D \cap U = \varnothing$ , el conjunto vacío. Los conjuntos  $U$  y  $D$  son definidas por:

$$\langle \alpha, \beta \rangle \in U \Leftrightarrow (\alpha, \beta) \in E \text{ y } (\beta, \alpha) \in E$$

$$(\alpha, \beta) \in D \Leftrightarrow (\alpha, \beta) \in E \text{ y } (\beta, \alpha) \notin E$$

**Definición 2.7.2 (Padre, Hijo, Vecino Dirigido y no Dirigido, Familia).**

Considere un grafo  $G = (V, E)$ , donde  $V = (\alpha_1, \dots, \alpha_d)$  y sea  $E = D \cup U$ , donde  $D$  es el conjunto de aristas dirigidas y  $U$  el conjunto de aristas no dirigidas.

$$\langle \alpha, \beta \rangle \in U \Leftrightarrow (\alpha, \beta) \in E \text{ y } (\beta, \alpha) \in E$$

$$(\alpha, \beta) \in D \Leftrightarrow (\alpha, \beta) \in E \text{ y } (\beta, \alpha) \notin E$$

Sea  $(\alpha_j, \alpha_k) \in G$ , si  $(\alpha_j, \alpha_k) \in D$ , entonces  $\alpha_k$  se conoce como un hijo de  $\alpha_j$  y  $\alpha_j$  como padre de  $\alpha_k$ .

Para cualquier nodo  $\alpha \subseteq V$ , el conjunto de padres se define como:

$$\Pi(\alpha) = \{\beta \in V \mid (\beta, \alpha) \in D\}. \quad (2.7)$$

el conjunto de hijos se define como:

$$Ch(\alpha) = \{\beta \in V \mid (\alpha, \beta) \in D\}. \quad (2.8)$$

Para cualquier subconjunto  $A \subseteq V$ , el conjunto de los padres de  $A$  se define como:

$$\Pi(A) = \bigcup_{\alpha \in A} \{\beta \in V \setminus A \mid (\beta, \alpha) \in D\}. \quad (2.9)$$

El conjunto de vecinos dirigidos de un nodo  $\alpha$  se define como:

$$N_{(d)}(\alpha) = \Pi(\alpha) \cup Ch(\alpha).$$

y el conjunto de vecinos no dirigidos de  $\alpha$  se define como:

$$N_{(u)}(\alpha) = \{\beta \in V \mid \langle \alpha, \beta \rangle \in U\}. \quad (2.10)$$

Para cualquier subconjunto  $A \subseteq V$ , el conjunto de los vecinos no dirigidos de  $A$  se define como:

$$N_{(u)}(A) = \bigcup_{\alpha \in A} \{\beta \in V \setminus A \mid \langle \alpha, \beta \rangle \in U\}. \quad (2.11)$$

Para un nodo  $\alpha$ , el conjunto de vecinos  $N(\alpha)$  se define como:

$$N(\alpha) = N_{(u)}(\alpha) \cup N_{(d)}(\alpha).$$

La familia de un nodo  $\beta$  es el conjunto que contiene el nodo  $\beta$ , junto con sus padres y vecinos no dirigidos. Se denota:

$$F(\beta) = \{\beta\} \cup \Pi(\beta) \cup N_{(u)}(\beta) = \{\text{familia\_de\_}\beta\}.$$

Cuando  $G$  es no dirigido, esto se reduce a  $F(\beta) = \{\beta\} \cup N(\beta)$ .

**Definición 2.7.3 (Grafo Dirigido y no Dirigido).**

Si todas las aristas de un grafo son no dirigidas, entonces el grafo  $G$  se dice que es no dirigido. Si todas las aristas son dirigidas, entonces se dice que el grafo es dirigido.

**Definición 2.7.4 (Senda).**

Sea  $G = (V, E)$  un grafo, donde  $E = D \cup U, D \cap U = \emptyset$ . Una senda  $\tau$  entre dos nodos  $\alpha \in V$  y  $\beta \in V$  es un conjunto de nodos  $\tau = (\tau_1, \dots, \tau_m)$ , donde  $\tau_i \in V$  para cada  $i = 1, \dots, m$ ,  $\tau_1 = \alpha$  y  $\tau_m = \beta$  y tal que para cada  $i = 1, \dots, m-1, \tau_i \sim \tau_{i+1}$ . Es decir, para cada  $i = 1, \dots, m-1$ , ya sea  $(\tau_i, \tau_{i+1}) \in D$  o  $(\tau_{i+1}, \tau_i) \in D$  o  $\langle \tau_i, \tau_{i+1} \rangle \in U$ .

**Definición 2.7.5 (Sub-grafo, Sub-grafo inducido).**

Sea  $A \subseteq V$  y  $E_A \subseteq E \cap A \times A$ . Entonces  $F = (A, E_A)$ . Es un sub-grafo de  $G$ . Si  $A \subset V$  y  $E_A \subseteq E \cap A \times A$  entonces  $G_A = (A, E_A)$  es el sub-grafo inducido por  $A$ .

**Definición 2.7.6 (Grafo conectado, Componente conectado).**

Un grafo se dice que está conectado si entre dos nodos cualesquiera  $\alpha_j \in V$  y  $\alpha_k \in V$  hay una senda. Un componente conectado de un grafo  $G = (V, E)$  es un sub-grafo inducido  $G_A$  tal que  $G_A$  está conectado y de tal manera que si  $A \neq V$ , entonces para cualquier par de nodos  $(\alpha, \beta) \in V \times V$  tal que  $\alpha \in A$  y  $\beta \in V \setminus A$ , no hay senda entre  $\alpha$  y  $\beta$ .

**Definición 2.7.7 (Camino, Camino dirigido).**

Sea  $G = (V, E)$  denota un grafo simple, donde  $E = D \cup U$ . Es decir,  $D \cap U = \emptyset$ ,  $D$  denota aristas dirigidas y  $U$  denota las aristas no dirigidas. Un camino de longitud  $m$  de un nodo  $\alpha$  a un nodo  $\beta$  es una secuencia de distintos nodos  $(\tau_0, \dots, \tau_m)$  de tal manera que  $\tau_0 = \alpha$  y  $\tau_m = \beta$  tal que  $(\tau_{i-1}, \tau_i) \in E$  para cada  $i = 1, \dots, m$ . Es decir, para cada  $i = 1, \dots, m$ , o bien  $(\tau_{i-1}, \tau_i) \in D$  o  $\langle \tau_{i-1}, \tau_i \rangle \in U$ .

El camino es un camino dirigido si  $(\tau_{i-1}, \tau_i) \in D$  para cada  $i = 1, \dots, m$ . Es decir, no hay ninguna arista no dirigida a lo largo del camino dirigido.

**Definición 2.7.8 (Descendiente y Antepasado).**

Sea  $G=(V, E)$  un grafo. Un nodo  $\alpha$  es un descendiente de un nodo  $\beta$  si y sólo si existe un camino dirigido desde  $\alpha$  a  $\beta$ . Un nodo  $\gamma$  es un antepasado de un nodo  $\alpha$  si y sólo si es un camino dirigido de  $\gamma$  a  $\alpha$ . El conjunto de los descendientes de  $D(\alpha)$  de un nodo  $\alpha$  se define como:

$$D(\alpha) = \left\{ \beta \in V \mid \exists \tau = (\tau_0, \dots, \tau_k) : \tau_0 = \alpha, \tau_k = \beta, (\tau_j, \tau_{j+1}) \in D, j = 0, 1, \dots, k \right\} \quad (2.12)$$

El conjunto de antepasados  $A(\alpha)$  de un nodo  $\alpha$  se define como:

$$A(\alpha) = \left\{ \beta \in V \mid \exists \tau = (\tau_0, \dots, \tau_k) : \tau_0 = \beta, \tau_k = \alpha, (\tau_j, \tau_{j+1}) \in D, j = 0, 1, \dots, k \right\} \quad (2.13)$$

En ambos casos, los caminos son dirigidos; contiene solo aristas dirigidas.

**Definición 2.7.9 (Ciclo).**

Sea  $G=(V, E)$  un grafo. Un  $m$ -ciclo en  $G$  es una secuencia de distintos nodos.

$$\tau_0, \dots, \tau_{m-1}$$

De tal manera que  $\tau_0, \dots, \tau_{m-1}, \tau_0$  es un camino.

**Definición 2.7.10 (Grafo acíclico dirigido (DAG)).**

Un grafo  $G=(V, E)$  se dice que es un grafo acíclico dirigido si cada arista es dirigida (es decir,  $G$  es un grafo simple de tal manera que para cada par

$$(\alpha, \beta) \in V \times V, (\alpha, \beta) \in E, (\beta, \alpha) \notin E$$

y para cualquier nodo  $\alpha \in V$  no existe ningún conjunto de nodos distintos  $\tau_1, \dots, \tau_m$  tal que  $\alpha \neq \tau_i$  para todo  $i = 1, \dots, m$  y  $(\alpha, \tau_1, \dots, \tau_m, \alpha)$  forma un camino dirigido. Es decir, no hay  $m$ -ciclos en  $G$  para cualquier  $m \geq 1$ .

**Definición 2.7.11 (Árbol).**

Un árbol es un grafo  $G = (V, E)$  que está conectado y tal que para cualquier nodo  $\alpha \in V$ , no hay senda entre  $\alpha$  y  $\alpha$  y para cualquiera de los dos nodos  $\alpha$  y  $\beta$  en  $V$  con  $\alpha \neq \beta$ , hay una única senda. Una hoja de un árbol es un nodo que está conectado a exactamente otro nodo.

**Definición 2.7.12 (Bosque).**

Un bosque es un grafo en el que todos sus componentes están conectados, estos son árboles.

**2.8 Grafos aciclicos dirigidos y distribucion de probabilidad.**

Ahora consideramos un vector aleatorio  $\underline{X} = (X_1, \dots, X_d)$ . Es decir  $V = (X_1, \dots, X_d)$  denota el conjunto de nodos del grafo, que es un conjunto de variables aleatorias.

**2.8.1 Factorización de una Distribución de Probabilidad.**

Para cualquier conjunto de variables aleatorias  $X_1, \dots, X_d$ , la función de probabilidad  $p_{X_1, \dots, X_d}$  se puede escribir como.

$$P_{X_1, \dots, X_d} = P_{X_1} P_{X_2|X_1} P_{X_3|X_1, X_2} \cdots P_{X_d|X_1, \dots, X_{d-1}}$$

La forma de escribir la factorización de una distribución de probabilidad es la siguiente:

$$P_{X_1, \dots, X_d} = P_{X_{\sigma(1)}} P_{X_{\sigma(2)}|X_{\sigma(1)}} P_{X_{\sigma(3)}|X_{\sigma(1)}, X_{\sigma(2)}} \cdots P_{X_{\sigma(d)}|X_{\sigma(1)}, \dots, X_{\sigma(d-1)}}.$$

Un grafo acíclico dirigido puede ser utilizado para indicar que ciertas variables son condicionalmente independientes de otras variables, lo que indica cómo una factorización puede ser simplificada.

**Definición 2.8.1.1 (Factorización a lo largo de un Grafo Acíclico Dirigido).**

Una función de probabilidad  $p_{X_1, \dots, X_d}$  de variables  $X_1, \dots, X_d$ . Se dice que a lo largo de la factorización de un grafo acíclico dirigido  $G$  debemos seguir lo siguiente: hay un orden de las variables  $X_{\sigma(1)}, \dots, X_{\sigma(d)}$  de tal manera que,

- $\Pi(X_{\sigma(1)}) = \Pi_{\sigma(1)} = \varphi$  es decir  $X_{\sigma(1)}$  no tienes padres.
- Para cada  $j$ ,  $\Pi(X_{\sigma(j)}) = \Pi_{\sigma(j)} \subset \{X_{\sigma(1)}, \dots, X_{\sigma(j-1)}\}$
- $P_{X_{\sigma(j)} | X_{\sigma(1)}, \dots, X_{\sigma(j-1)}} = P_{X_{\sigma(j)} | \Pi_{\sigma(j)}}$

Para cada orden de las variables, hay un grafo acíclico dirigido que indica cómo factorizar la distribución de probabilidad y las variables que pueden ser excluidos.

**Instancia.** Cuando el estado de la variable es conocido, se dice que la variable esta instanciada. En un grafo acíclico dirigido, hay tres formas básicas en las que es posible que dos variables pueden conectarse a través de una tercera variable y todo el grafo es construido a partir de estas conexiones.

**2.8.2 Tipos de conexiones.**

**Conexión cadena.** Considere una situación con tres variables aleatorias  $(X_1, X_2, X_3)$ , donde  $X_1$  influye en  $X_2$ , que a su vez influye en  $X_3$ , pero no hay ninguna influencia directa de  $X_1$  a  $X_3$ . Si el estado de  $X_2$  es desconocido, entonces la información sobre  $X_1$  influirá en la distribución de probabilidad de  $X_2$ , que a su vez influye en la distribución de probabilidad de  $X_3$ . Del mismo modo, la información sobre  $X_3$  influye en la distribución de probabilidad de  $X_1$  a través de  $X_2$ .

Si el estado de  $X_2$  es conocido, el canal está bloqueado y  $X_1$  y  $X_3$  se independizan dado  $X_2$ . El grafo aciclico dirigido indica que la distribución de probabilidad de  $(X_1, X_2, X_3)$ , se puede factorizar como:

$$P_{X_1, X_2, X_3} = P_{X_1} P_{X_2|X_1} P_{X_3|X_2}$$

Si  $X_2 = x_2$  es conocida, entonces

$$P_{X_1, X_2, X_3}(\cdot, x_2, \cdot) = (P_{X_1}(\cdot) P_{X_2|X_1}(x_2|\cdot)) (P_{X_3|X_2}(\cdot|x_2))$$

Las variables  $X_1$  y  $X_3$  son condicionalmente independientes, dado  $X_2$ , a partir del grafo aciclico dirigido las variables  $X_1$  y  $X_3$  se dice que son d-separados dado  $X_2$ .

**Conexión Tenedor.** La influencia puede pasar entre todos los hijos de  $X_1$  a menos que el estado de  $X_1$  sea conocida. Si  $X_1$  es conocida, entonces las variables  $X_2$  y  $X_3$  se dice que son d-separados dado  $X_1$ . La evidencia puede ser transmitida a través de un nodo tenedor a menos que esta sea instanciada.

El grafo aciclico dirigido para la conexión tenedor indica que la distribución de probabilidad se puede factorizar como:

$$P_{X_1, X_2, X_3} = P_{X_1} P_{X_2|X_3} P_{X_3|X_1}$$

Esto implica que si el estado de  $X_1$  es conocido, entonces

$$P_{X_2, X_3, X_1}(\cdot, \cdot | x_1) = P_{X_2|X_1}(\cdot | x_1) P_{X_3|X_1}(\cdot | x_1).$$

De manera que  $X_2$  y  $X_3$  son condicionalmente independientes.

**Conexión Colisionador.** Si no se sabe nada acerca de  $X_1$  salvo que puede inferirse a partir del conocimiento acerca de sus padres, entonces los padres son independientes, la información sobre uno de ellos no afectará a la evaluación de los valores de probabilidad de los otros. La Información acerca de una posible causa de un evento no da información acerca de otras posibles causas, si no hay información de que si o no el caso realmente sucedió. Pero si hay alguna información relativa acerca del evento, entonces la información sobre una posible causa puede dar información sobre las otras causas. La información sólo se puede transmitir a través de una conexión colisionador si la información ha sido recibida ya sea por la variable en la conexión o de uno de sus descendientes. La factorización de la distribución  $P_{X_1, X_2, X_3}$  correspondiente al grafo aciclico dirigido para la conexión colisionador es:

$$P_{X_1, X_2, X_3} = P_{X_2} P_{X_3} P_{X_1|X_2, X_3}$$

Es evidente que  $X_2$  y  $X_3$  no son condicionalmente independientes dada  $X_1$ , pero si  $X_1$  es desconocida, entonces,  $X_2$  y  $X_3$  son independientes; para cualquier  $(x_2^i, x_3^j) \in \mathcal{X}_2 \times \mathcal{X}_3$ , donde  $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$  son los espacios del estado para las variables aleatorias  $X_1, X_2$  y  $X_3$  respectivamente.

$$\begin{aligned} P_{X_2, X_3}(x_2^{(i)}, x_3^{(j)}) &= \sum_{y \in \mathcal{X}_1} P_{X_2}(x_2^{(i)}) P_{X_3}(x_3^{(j)}) P_{X_1|X_2, X_3}(y | x_2^{(i)}, x_3^{(j)}) \\ &= P_{X_2}(x_2^{(i)}) P_{X_3}(x_3^{(j)}) \sum_{y \in \mathcal{X}_1} P_{X_1|X_2, X_3}(y | x_2^{(i)}, x_3^{(j)}) \\ &= P_{X_2}(x_2^{(i)}) P_{X_3}(x_3^{(j)}) \end{aligned}$$

**Definición 2.7.13(Senda S-Activo).**

Sea  $G=(V, E)$  un grafo acíclico dirigido. Sea  $S \subset V$  y sea  $X, Y \in V \setminus S$ . Una senda  $\tau$  entre las dos variables  $X$  e  $Y$  se dice que es S-activo si:

1. Cada nodo colisionador de  $\tau$  está en  $S$ , o tiene un descendiente en  $S$ .
2. Cada nodo está fuera de  $S$ .

**Definición 2.7.14(Senda bloqueada).**

Una senda entre  $X$  e  $Y$  que no es S-activa se dice que es bloqueado por  $S$ .

**Definición 2.7.15(D-separación).**

Sea  $G=(V, E)$  un grafo acíclico dirigido, donde  $V = \{X_1, \dots, X_d\}$  es un conjunto de variables aleatorias. Sea  $S \subset V$  tal que todas las variables en  $S$  son instancias y todas las variables en  $V \setminus S$  no son instanciadas. Dos distintas variables  $X_i, X_j$  no están en  $S$  son d-separados por  $S$ , si todos las sendas entre  $X_i$  y  $X_j$  son bloqueados por  $S$ .

Sea  $C$  y  $D$  denota dos conjuntos de variables. Si todas las sendas de cualquier variable en  $C$  a cualquier variable en  $D$  es bloqueado por  $S$ , entonces los conjuntos  $C$  y  $D$  se dice que son d-separados por  $S$ . Es decir

$$C \perp D \parallel_G S. \quad (2.14)$$

**Definición 2.7.16 (D-Conectado).**

Si dos variables  $X$  e  $Y$  no son d-separados, se dice que son d-conectados.

**2.9 Potencial.**

Sea  $V = \{X_1, \dots, X_d\}$  denota un conjunto de variables aleatorias, donde la variable  $X_j$  tiene espacio de estados  $\mathcal{X}_j = (x_j^{(1)}, \dots, x_j^{(k_j)})$  para  $j = 1, \dots, d$ . Sea  $\mathcal{X} = \times_{j=1}^d \mathcal{X}_j$  denota el espacio de estados para  $\underline{X}$ . Sea  $\tilde{V} = \{1, \dots, d\}$  denota el conjunto indexado para las variables. Para  $D \subset \tilde{V}$ , donde  $D = \{j_1, \dots, j_m\}$  sea  $\mathcal{X}_D = \times_{j \in D} \mathcal{X}_j$  y sea  $\underline{X}_D = \{X_{j_1}, \dots, X_{j_m}\}$ . Sea  $\underline{x} \in \mathcal{X}$  denota un elemento genérico de  $\mathcal{X}$  y sea  $\underline{x}_D = (x_{j_1}, \dots, x_{j_m}) \in \mathcal{X}_D$ , cuando  $\underline{x} = (x_1, \dots, x_d) \in \mathcal{X}$ .

Por otra parte, para  $W \subset V$ , sea  $\tilde{W}$  denota el conjunto indexado para  $W$ . La notación  $\mathcal{X}_W$  se utiliza para denotar  $\mathcal{X}_{\tilde{W}}$ ,  $\underline{X}_W$  para denotar  $\underline{X}_{\tilde{W}}$  y  $x_W$  para denotar  $x_{\tilde{W}}$ .

**Definición 2.7.17 (Potencial).**

Un potencial  $\phi$  sobre un dominio  $\mathcal{X}_D$  es definido como una función no negativa  $\phi: \mathcal{X}_D \rightarrow R_+$ . El espacio  $\mathcal{X}_D$  es conocido como el dominio del potencial. Si el dominio es un vector aleatorio  $\underline{X}_D$ , entonces el vector aleatorio  $\underline{X}_D$  también puede conocerse como el dominio del potencial.

Un potencial sobre un dominio  $\mathcal{X}_D$  tiene  $\prod_{j \in D} k_j$  entradas. Para  $W \subset V$ , el dominio de un potencial  $\mathcal{X}_W$  también puede ser denotado por el conjunto de variables aleatorias  $W$ .

**Definición 2.7.18 (Suma de Potenciales).**

Dos potenciales  $\phi$  y  $\phi'$  definidas sobre el mismo dominio  $\mathcal{X}_D$ . Su suma se define, como la suma de coordenadas-Wise<sup>2</sup>, para cada  $\underline{x}_D \in \mathcal{X}_D$ .

$$(\phi + \phi')(\underline{x}_D) = \phi(\underline{x}_D) + \phi'(\underline{x}_D)$$

**Definición 2.7.19 (Multiplicación de potenciales).**

Dos potenciales  $\phi$  y  $\phi'$  se puede multiplicar juntos para obtener el potencial  $\phi \cdot \phi'$  si ambos son definidas sobre el mismo dominio.

**Definición 2.7.20 (División de Potenciales).**

Dos potenciales  $\phi$  y  $\phi'$  se puede dividir si se definen en el mismo dominio. La división de un potencial  $\phi$  por  $\phi'$  para dar el potencial de  $\phi / \phi'$  se define por la división de coordenadas-Wise<sup>2</sup> donde la definición es  $b = 0 \rightarrow \frac{a}{b} = 0$ .

**2.10 Red Bayesiana.** Se define de la siguiente manera:

**Definición 2.7.21 (Red Bayesiana).**

Una red bayesiana es un par  $(G, p)$ , donde  $G = (V, E)$  es un grafo acíclico dirigido con el conjunto de nodos  $V = \{1, \dots, d\}$  para algún  $d \in \mathbb{N}$ ,  $E$  es el conjunto de aristas, y  $p$  distribución de probabilidad o una familia de distribuciones de probabilidad, indexados por un conjunto de parámetros  $\Theta$ , sobre  $d$  variables aleatorias discretas,  $\{X_1, \dots, X_d\}$ .

El par  $(G, p)$  satisface los siguientes criterios:

- Para cada  $\theta \in \Theta$ ,  $p(\cdot | \theta)$  es una función de probabilidad con el mismo espacio de estado  $\mathcal{X}$ , donde  $\mathcal{X}$  tiene un número finito de elementos. Es decir, para cada  $\theta \in \Theta$ ,  $p(\cdot | \theta): \mathcal{X} \rightarrow [0, 1]$  y  $\sum_{\underline{x} \in \mathcal{X}} p(\underline{x} | \theta) = 1$ .

---

<sup>2</sup> <http://www.enclikipediaofmath.org/index.php?title=coordinate-wise>

- Para cada nodo  $X_v \in V$  con variables no padres, hay asignado un potencial denotado por  $p_{X_v}$ , dado la distribución de probabilidad de la variable aleatoria  $X_v$ . Para cada variable  $X_v \in V$  con un conjunto de nodos padres no vacío

$$\Pi_v = \left( X_{b_1^{(v)}}, \dots, X_{b_m^{(v)}} \right)$$

hay un potencial asignado.

$$p_{X_v|\Pi_v}$$

- Contiene la función de probabilidad condicional de  $X_v$  dado las variables  $\{X_{b_1^{(v)}}, \dots, X_{b_m^{(v)}}\}$ . Si  $X_v$  no tiene padres, el conjunto  $\Pi_v = \emptyset$  el conjunto vacío, de modo que  $p_{X_v} = p_{X_v|\Pi_v}$ . La función de probabilidad conjunta de  $P$  puede ser factorizado usando las potenciales  $p_{X_v|\Pi_v}$  definidas así:

$$p_{X_1, \dots, X_d} = \prod_{v=1}^d p_{X_v|\Pi_v}.$$

- La factorización es mínima en el sentido de que para un ordenamiento de las variables tal que  $\Pi_j \subseteq \{X_1, \dots, X_{j-1}\}$ ,  $\Pi_j$  es el conjunto más pequeño de variables tal que  $X_j \perp \Pi_j^c \mid \Pi_j$ . Es decir:

$$\Pi_j = \cap \left\{ A \subseteq \{X_1, \dots, X_{j-1}\} \text{ tal que } X_j \perp A^c \mid A \right\}.$$

## Capítulo 3

### *Inferencia de una Red Bayesiana Discreta*

Cuando se realiza la propagación de probabilidades en una Red Bayesiana se le llama inferencia en Redes Bayesianas aquí se plantea la obtención de las probabilidades a posteriori es así que en el presente capítulo veremos la construcción de la red árbol que nos permitirá actualizar la probabilidad de manera eficiente a la luz de la nueva información, existen varias maneras de actualizar la probabilidad, pero nuestra atención se enfocara a la construcción de la redes bayesianas como un árbol. Mencionaremos algunos conceptos de grafos que son necesarios para la construcción de la red árbol.

Para la actualización de probabilidades en necesario conocer el tipo de evidencia que será introducida, por tal motivo, existen diferentes evidencias en una red bayesiana.

#### **3.1 Evidencia.**

Sea  $X = (X_1, \dots, X_d)$  denota un vector aleatorio, con espacio de estados finitos

$\mathcal{X} = \times_{i=1}^d \mathcal{X}_i$  donde  $\mathcal{X}_j = (x_j^{(1)}, \dots, x_j^{(k_j)})$  es el espacio de estados de la variable aleatoria  $X_j$ .

Los tipos de evidencias existentes en la red bayesiana son las siguientes:

Un hallazgo difícil, es una instancia,  $\{X_i = x_i^l\}$  para un valor particular  $i \in \{1, \dots, d\}$  y un valor particular de  $l \in \{1, \dots, k_i\}$ . Esto especifica que la variable  $X_i$  se encuentra en estado  $x_i^{(l)}$ . Se expresa como un

$k_1 \times \dots \times k_d$  potencial  $e$  donde:

$$e(x_1^{(p_1)}, \dots, x_d^{(p_d)}) = \begin{cases} 1 \rightarrow p_i = l \\ 0 \rightarrow p_i \neq l \end{cases}$$

Es decir, las entradas correspondientes a las configuraciones que contienen las instanciaciones son 1 y las entradas correspondientes a todas las otras configuraciones son 0.

**Evidencia difícil.** Es un conjunto de hallazgos difíciles. Esto se da por un conjunto de potenciales  $e = (e_1, \dots, e_m)$  donde  $e_m$  es un hallazgo difícil en una de las variables.

Un hallazgo suave, en una variable  $X_j$  especifica la distribución de probabilidad de la variable  $X_j$ . Es decir, el potencial  $p_{X_j|\Pi_j}$  se sustituye por un potencial  $p_{X_j}^*$  con dominio  $\mathcal{X}_j$ .

**Evidencia suave.** Es un conjunto de hallazgos suaves.

Un hallazgo virtual, en la variable  $X_j$  es un conjunto de los cocientes de probabilidad  $\{L(x_j^{(m)}), m = \{1, \dots, m\}\}$  tal que la actualización de probabilidad condicional potencial para  $X_j | \Pi_j =: \pi_j^{(n)}$  esto para  $m = 1, \dots, k_j$

$$p_{X_j|\Pi_j}^*(x_j^{(m)} | \pi_j^{(n)}) = \frac{1}{\sum_{q=1}^{k_j} p_{X_j|\Pi_j}(x_j^{(q)} | \pi_j^{(n)}) L(x_j^{(q)})} p_{X_j|\Pi_j}(x_j^{(m)} | \pi_j^{(n)}) L(x_j^{(m)}) \quad (3.1)$$

**Evidencia Virtual.** Es un conjunto de hallazgos virtuales.

Evidencias suaves y evidencias virtuales son los dos tipos de evidencia diferentes. Cuando la evidencia suave es recibida en una variable, el enlace entre la variable y sus padres son limitados, si la evidencia suave es recibido en una variable  $X_j$  entonces el potencial  $p_{X_j|\Pi_j}$  es reemplazado por un nuevo potencial  $p_{X_j}^*$ . Cuando la evidencia virtual es recibida, los enlaces son preservados. La evidencia virtual puede ser modelada por adicionar un nuevo nodo a la red.

### 3.2. Evidencia Difícil.

Considere una situación donde el dominio de variables son  $\{X, Y\}$ ,  $\mathcal{X}_X = \{x_1, x_2, x_3\}$  y  $\mathcal{X}_Y = \{y_1, y_2, y_3\}$ .

#### Introduciendo evidencias difíciles.

Sea  $p_{\underline{X}}$  denota una función de probabilidad conjunta sobre un vector aleatorio  $\underline{X}$  y sea  $e$  un hallazgo que corresponde a un elemento de evidencia. La evidencia es introducida en  $p_{\underline{X}}$  para multiplicar los potenciales. Sea  $p_{\underline{X};e}$  denota el resultado:

$$p_{\underline{x};e} = p_{\underline{x}} \cdot e \quad (3.2)$$

Supongamos que varios hallazgos difíciles son recibidos, expresados en potenciales  $(e_1, \dots, e_k)$ . A continuación, la evidencia difícil  $e = (e_1, \dots, e_k)$  es introducida por:

$$p_{\underline{x};e} = p_{\underline{x}} \cdot \prod_{j=1}^k e_j,$$

Donde la multiplicación es en el sentido de potenciales, ha sido extendido para el dominio  $\chi$ .

El potencial  $\prod_{j=1}^k e_j$  (donde los dominios ha sido extendido a  $\chi$ )

El potencial  $p_{\underline{x};e}$  no es una función de probabilidad, en el sentido de que las entradas no suman 1. Para calcular la probabilidad potencial condicional, condicionado en la evidencia recibida, es necesario calcular la probabilidad de la evidencia. Esta dado por:

$$p(e) = \sum_{\underline{x} \in \chi} p_{\underline{x};e}(\underline{x})$$

y la probabilidad potencial condicional está dada por:

$$p_{\underline{x}|e} = \frac{p_{\underline{x};e}}{p(e)}. \quad (3.3)$$

Donde la división se toma en el sentido de los potenciales.

### 3.3 Evidencias suaves y evidencias virtuales.

Un hallazgo virtual en una variable  $X_j$  afecta la probabilidad potencial  $p_{X_j|\Pi_j}$  sin afectar la probabilidad potencial condicional posterior. El hallazgo virtual puede ser considerado como una variable adicional V, esto es adicionado en la red, solo con una flecha V, para la variable afectada por la evidencia suave.

### 3.4 Construcción de una Unión de Árboles.

#### 3.4.1 Grafo Completo, Subconjunto Completo.

Un grafo  $G$  es completo si todos los pares de nodos son unidos por una arista no dirigida. Es decir, para cada  $(\alpha, \beta) \in V \times V$  con  $\alpha \neq \beta$ ,  $(\alpha, \beta) \in E$  y  $(\beta, \alpha) \in E$ . En otras palabras,  $\langle \alpha, \beta \rangle \in U$  donde  $U$  denota el conjunto de aristas no dirigidas.

Un subconjunto de los nodos se llama completo si induce a un sub-grafo completo.

### 3.4.2 Clique.

Un clique es un sub-grafo completo que es máxima con respecto a  $\subseteq$ . En otras palabras, un clique no es un sub-grafo de cualquier otro grafo completo.

### 3.4.3 Nodo Simplicial.

Para un grafo no dirigido, la familia de un nodo  $\beta$  es  $F(\beta) = \{\beta\} \cup N(\beta)$ , donde  $N(\beta)$  denota el conjunto de vecinos de  $\beta$ . Un nodo  $\beta$  en un grafo no dirigido es llamado Simplicial si su familia  $F(\beta)$  es un clique.

**3.4.4 Conexión y Componentes Fuertes.** Sea  $G = (V, E)$  un grafo simple, donde  $E = U \cup D$ . Sea  $\alpha \rightarrow \beta$  denota que hay un camino de  $\alpha$  a  $\beta$ . Si hay ambos  $\alpha \rightarrow \beta$  y  $\beta \rightarrow \alpha$  entonces  $\alpha$  y  $\beta$  se dice que están conectados. Es decir:

$$\alpha \leftrightarrow \beta$$

Esto es claramente una relación de equivalencia. La clase de equivalencia de  $\alpha$  es denotada por  $[\alpha]$ . En otras palabras,  $\beta \in [\alpha]$  si y solo si  $\beta \leftrightarrow \alpha$ . Estas clases de equivalencia se llaman componentes fuertes de  $G$ .

Tenga en cuenta que un grafo es conectado si entre dos nodos existe una senda, pero cualquiera de los dos nodos  $\alpha$  y  $\beta$  solo son conectados si hay un camino de  $\alpha$  a  $\beta$  y un camino entre  $\beta$  y  $\alpha$ .

**3.4.5 Separador.** Un subconjunto  $S \subseteq V$  es llamado separador  $(\alpha, \beta)$  si todo camino de  $\alpha$  a  $\beta$  intersecta  $S$ . Sea  $A \subseteq V$  y  $B \subseteq V$ , tal que  $A, B$  y  $S$  son disjuntos. El subconjunto  $S$  se dice separador  $A$  de  $B$  si es un separador  $(\alpha, \beta)$  para todo  $\alpha \in A$  y  $\beta \in B$ .

**3.4.6 Separador mínimo.** Sea  $A \subseteq V$ ,  $B \subseteq V$  y  $S \subseteq V$  tres subconjuntos disjuntos de  $V$ . Sea  $S$  separador  $A$  de  $B$ . El separador  $S$  se dice que es un separador mínimo de  $A$  y  $B$ , si ningún subconjunto propio de  $S$  es en sí un separador  $(\alpha, \beta)$  para cualquier  $(\alpha, \beta) \in A \times B$ .

### 3.4.7 Árbol con raíz.

Un árbol con raíz  $T$  es un grafo árbol con un nodo  $\rho$  llamado raíz. Una hoja de un árbol es un nodo que es unido a un máximo de otro nodo.

### 3.4.8 Diámetro.

El diámetro de un árbol es la longitud de la senda más larga entre dos nodos hoja.

### 3.4.9 Grafo Moral.

Sea  $G$  un DAG. Entonces  $G$  se dice que es moralizada si todas las aristas no dirigidas entre todos los pares de padres de cada nodo que no ha sido unido, son adicionadas y entonces todas las aristas son sin dirección.

### 3.4.10 Acorde.

Sea  $G = (V, E)$  un grafo no dirigido. Sea  $\sigma$  un ciclo  $n$  en  $G$ . Un acorde de este ciclo es un par  $(\alpha_i, \alpha_j)$  de nodos no consecutivos en  $\sigma$  talque  $\alpha_i \sim \alpha_j$  en  $G$ .

### 3.4.11 Triangulado.

Un grafo no dirigido  $G = (V, E)$  es triangulado si cada uno de sus ciclos de longitud  $\geq 4$  posee un acorde.

**Lema 3.1.**

Si  $G = (V, E)$  es triangulado, entonces el grafo inducido por  $G_A$  también es triangulado.

**Demostración del lema 3.1.**

Considerar cualquier ciclo de longitud  $\geq 4$  en el grafo restringido. Todas las aristas conectadas permanecen en estos nodos. Si el ciclo posee un acorde en el grafo original, el acorde permanece en el grafo restringido.

**3.4.12 Descomposición.**

Una tripleta  $(A, B, S)$  de subconjuntos disjuntos del conjunto de nodos  $V$  de un grafo no dirigido se dice formar una descomposición de  $G$  o descompone  $G$  si,

- $V = A \cup B \cup S$  y
- $S$  separa  $A$  de  $B$ ,
- $S$  es un subconjunto completo de  $V$ .

$A, B$  o  $S$  puede ser el conjunto vacío. Si  $A$  y  $B$  no son vacíos, entonces la descomposición es propia.

**3.4.13 Grafo descompuesto.**

Un grafo no dirigido  $G$  es descompuesto si:

1. Es completa
2. Posee una descomposición adecuada  $(A \cup B \cup S)$  de forma que ambos sub grafos  $G_{A \cup S}$  y  $G_{B \cup S}$  son descompuestas.

**3.4.14 Grafo descompuesto y Triangulación del Grafo.**

Los grafos descompuestos proporcionan las bases para uno de los principales métodos para la actualización de una distribución de probabilidad descrita en términos de una red bayesiana. El grafo aciclico dirigido es moralizado y triangulado. El grafo triangular se descompone y es organizado para formar una unión de arboles.

Sea  $G=(V, E)$  un grafo no dirigido que cumple las siguientes condiciones:

1.  $G$  es descompuesta.
2.  $G$  es triangulado.
3. Por cada par de nodos  $(\alpha, \beta) \in V \times V$ , el separador mínimo es completo.

**3.4.15. Eliminación de nodos de secuencia perfecta.** Sea  $V = \{\alpha_1, \dots, \alpha_d\}$ , denota el conjunto de nodos de un grafo  $G$ . Una eliminación de nodos de secuencia perfecta de un grafo  $G$  es un orden del conjunto de nodos  $\{\alpha_1, \dots, \alpha_d\}$  tal que para cada  $j$  en  $1 \leq j \leq d-1$ ,  $\alpha_j$  es un nodo simplicial del sub-grafo de  $G$  inducida por  $(\alpha_j, \alpha_{j+1}, \dots, \alpha_d)$ .

**Lema 3.2.** Cada grafo triangulado  $G$  tiene un nodo simplicial. Por otra parte, si  $G$  no es completo, entonces tiene dos nodos simplicial no adyacentes.

**Demostración del lema 3.2.** El lema es trivial si  $G$  está completo, o bien  $G$  tiene dos o tres nodos. Supongamos que  $G$  no es completa. Supongamos que el resultado es cierto para todos los grafos con menos nodos que  $G$ . Consideremos dos nodos no adyacentes  $\alpha$  y  $\beta$ . Sea  $S$  denota el separador mínimo de  $\alpha$  y  $\beta$ . Entonces  $G_A$  denota el componente conectado más grande de  $G_{V \setminus S}$  tal que  $\alpha \in A$  y sea  $B = V \setminus (A \cup S)$ , de modo que  $\beta \in B$ .

Por inducción,  $G_{A \cup S}$  es completo, si no tiene dos nodos simpliciales no adyacentes. Entonces  $G_S$  es completo, se deduce que al menos uno de los dos nodos simpliciales está en  $A$ . Un nodo es por lo tanto también simplicial de  $G$ , porque ninguno de sus vecinos está en  $B$ . Si  $G_{A \cup S}$  es completo, entonces cualquier nodo de  $A$  es un nodo simplicial de  $G$ .

En todos los casos, hay un nodo simplicial de  $G$  en  $A$ . Del mismo modo, hay un nodo simplicial en  $B$ . Estos dos nodos son entonces los nodos simpliciales no adyacentes de  $G$ .

**Teorema 3.1.** Un grafo  $G$  es triangulado si y sólo si tiene una eliminación de nodos de secuencia perfecta.

### Demostración del Teorema 3.1

Supongamos que  $G$  es triangulado. Suponer también que cada grafo triangulado con menos nodos que  $G$  tiene una secuencia de eliminación perfecta. Por el lema 3.2,  $G$  tiene un nodo simplicial  $\alpha$ . Removiendo  $\alpha$  retorna un grafo triangulado. (Considere cualquier ciclo de longitud  $\geq 4$  con un acorde. Si el ciclo se mantiene después de que el nodo es eliminado, entonces, el acorde no se elimina). Procediendo por inducción, se deduce que  $G$  tiene una secuencia de eliminación perfecta.

Por el contrario, supongamos que  $G$  tiene una secuencia perfecta, por ejemplo  $\{\alpha_1, \dots, \alpha_d\}$ . Considere cualquier ciclo de longitud  $\geq 4$ . Sea  $j$  el primer índice tal que  $\alpha_j$  está en el ciclo. Sea  $V(C)$  denotar el conjunto de nodos del ciclo y sea  $V_j = \{\alpha_1, \dots, \alpha_d\}$ . Entonces  $V(C) \subseteq V_j$ . Por tanto  $\alpha_j$  es simplicial en  $G_{V_{j+1}}$ , los vecinos de  $\alpha_j$  en el ciclo son adyacentes, por lo tanto, el ciclo tiene un acorde. Entonces  $G$  es triangulado.

**3.4.16 Eliminación de un nodo.** Sea  $G=(V, E)$  un grafo no dirigido. Un nodo  $\alpha$  se elimina de un grafo no dirigido  $G$  de la siguiente manera:

1. Para todos los pares de vecinos  $(\beta, \gamma)$  de  $\alpha$  adiciona un enlace si  $G$  ya no contiene ni una. El enlace adicionado son llamados relleno.
2. Retire  $\alpha$ .

El resultado del grafo se denota por  $G^{-\alpha}$ .

**3.4.17 Eliminación de Secuencia.** Una eliminación de secuencia de  $G$  es una ordenación lineal de sus nodos.

Sea  $\sigma$  una eliminación de secuencia y sea  $\Lambda$  denota el relleno producido por la eliminación de un nodo de  $G$  en el orden de  $\sigma$ . Denotemos por  $G^\sigma$  el grafo  $G$  extendido por  $\Lambda$ .

**3.4.18 Eliminación de dominio.** Considere una eliminación de secuencia  $\sigma$ , es decir un orden lineal de los nodos, de modo que para cualquier nodo  $\alpha$ ,  $\sigma(\alpha)$  denota el número asignado a  $\alpha$ . Un nodo  $\beta$  se dice que es una eliminación de orden superior a  $\alpha$  si  $\sigma(\beta) > \sigma(\alpha)$ . La eliminación del dominio de un nodo  $\alpha$  es el conjunto de vecinos de  $\alpha$  para la eliminación de orden superior.

En  $G^\sigma$ , cualquier nodo  $\alpha$  junto con sus vecinos de eliminación de orden superior forma un subconjunto completo. Los vecinos de  $\alpha$  para la eliminación de orden superior es denotada por  $N_{\sigma(\alpha)}$ . Los conjuntos  $N_{\sigma(\alpha)}$  son la eliminación de dominios correspondientes a la eliminación de secuencia  $\sigma$ .

Un algoritmo eficiente claramente trata de minimizar el número de rellenos. Si es posible, uno debe encontrar una eliminación de secuencia que no introduce rellenos.

**Proposición 3.1.** Todos los cliques en un  $G^\sigma$  son un  $N_{\sigma(\alpha)}$  para algún  $\alpha \in V$ .

**Demostración.** Sea  $C$  un clique en  $G^\sigma$  y sea  $\alpha$  una variable en  $C$ , el orden de eliminación más bajo, entonces  $C = N_{\sigma(\alpha)}$ .

**3.4.19 Unión de árboles.** El propósito de esta sección es definir la unión de árboles y mostrar la forma de construir. La misma ofrece una herramienta clave para la actualización de una red bayesiana.

Sea  $\zeta$  una colección de subconjuntos de un conjunto finito  $V$  y  $T$  sea un árbol con  $\zeta$  como su conjunto de nodos. Entonces  $T$  se dice que es una unión de árboles si cualquier intersección  $C_1 \cap C_2$  de un par  $C_1, C_2$  de los conjuntos en  $\zeta$  está contenido en cada nodo

del único camino en  $T$  entre  $C_1$  y  $C_2$ . Sea  $G$  un grafo no dirigido y  $\zeta$  la familia de sus cliques. Si  $T$  es una unión de árboles con  $\zeta$  como su conjunto de nodos, entonces  $T$  es conocida como unión de árboles para el grafo  $G$ .

**Teorema 3.2.** Existe una unión de árboles  $T$  de cliques para el grafo  $G$  si y sólo si  $G$  es descompuesto.

**Demostración del Teorema 3.2** La prueba es por construcción, una secuencia es establecida de la siguiente manera. En primer lugar, un nodo simplicial  $\alpha$  es elegido;  $F_\alpha$  es un clique. El algoritmo continúa eligiendo los nodos de  $F_\alpha$  que sólo tienen los vecinos en  $F_\alpha$ . Sea  $i$  el número de nodos en  $F_\alpha$  que sólo tienen los vecinos en  $F_\alpha$ . El conjunto de nodos  $F_\alpha$  es designada por  $V_i$  y el conjunto de los nodos en  $F_\alpha$  que no tiene vecinos en  $F_\alpha$  designada por  $S_i$ . Este conjunto es un separador.

Ahora remueva los nodos en  $F_\alpha$  que no tienen vecinos fuera  $F_\alpha$  y el nuevo grafo  $G'$ . Elegir un nuevo nodo  $\alpha$  en el grafo  $G'$  de tal manera que  $F_\alpha$  es un clique. Repita el proceso, con el índice  $j$ , donde  $j$  es el índice anterior, más el número de nodos en la actual  $F_\alpha$  que sólo tienen los vecinos en  $F_\alpha$ .

Cuando las partes se han establecido, cada separador  $S_i$  es entonces conectado a un clique  $V_j$  con  $j > i$  y tal que  $S_i \subset V_j$ . Esto es posible siempre, porque  $S_i$  es un conjunto completo y en la secuencia de eliminación descrita arriba, el primer punto de  $S_i$  es eliminado cuando se trata de un clique de índice mayor que  $i$ .

Es necesario probar que a estructura es un árbol y que esto tiene la propiedad de una unión de árboles

En primer lugar, cada clique tiene a lo sumo uno de los padres, por lo que no hay múltiples caminos. La estructura es por lo tanto un árbol.

Para probar la condición de unión de árboles, considere dos cliques,  $V_i$  y  $V_j$  con  $i > j$ , y sea  $\alpha$  un miembro de ambos. Hay un único camino entre  $V_i$  y  $V_j$ .

Porque  $\alpha$  no es eliminado cuando se trata de  $V_i$ , este es un miembro de  $S_i$ . Es también un miembro de los padres de  $V_i$ , se dice  $V_k$  es miembro de los padres de  $V_k$  y por inducción esta es también un miembro de  $V_j$  y, por supuesto, todos los separadores en medio.

Una Red Bayesiana presenta una factorización de distribución de probabilidad de acuerdo al grafo aciclico dirigido, como ya anteriormente desarrollamos las definiciones para obtener una unión de árboles ahora desarrollaremos el método para actualizar las distribuciones de probabilidad.

Para la actualización de distribuciones de probabilidad de una unión de árboles debe considerar la evidencia difícil mencionada anteriormente y la tarea es calcular la distribución condicional dado la evidencia difícil la probabilidad dado la evidencia se define de esta manera

$$p_{\underline{x}|e} = \frac{p_{\underline{x},e}}{p(e)}$$

Donde

$$p(e) = \sum_{\underline{x} \in \mathcal{X}} p_{\underline{x},e}(\underline{x})$$

y

$$p_{\underline{x},e} = p_{\underline{x}} \prod_{j=1}^m e_j$$

El objetivo es seguir el siguiente procedimiento

1. Moralizar el grafo aciclico dirigido de la red bayesiana
2. Triangular el grafo moralizado
3. Sean los cliques del grafo triangulado sean los nodos del árbol, que es una unión de árboles.

Para obtener un resultado eficiente el paso más importante de la construcción es la triangulación del grafo moralizado.

### 3.5 Potencial y Propiedad distributiva.

**Notación.** En este caso repetiremos varias notaciones. Sea  $\tilde{V} = \{1, \dots, d\}$  denota la indexación de conjuntos para los  $d$  nodos para el grafo  $G = (V, E)$ . Para cada nodo  $j \in \tilde{V}$ , es asociado un estado de espacios finitos  $\mathcal{X}_j = \left( x_j^{(1)}, \dots, x_j^{(k_j)} \right)$ , el conjunto de los estados posibles de la variable aleatoria  $X_j$  el espacio de estados de  $\underline{X} = (X_1, \dots, X_d)$  es denotado por  $\mathcal{X} = \times_{j=1}^d \mathcal{X}_j$ . Sea  $D \subset \tilde{V}$ , denota un subconjunto de los nodos. La notación

$$\underline{x}_D = \times_{v \in D} x_v$$

Es usado para denotar una configuración (o un conjunto de salidas) en los nodos en  $D$ . El espacio de estados de las variables en  $D$  es denotado por:

$$\mathcal{X}_D = \times_{v \in D} \mathcal{X}_v.$$

Supóngase  $D \subseteq W \subseteq \tilde{V}$  y  $\underline{x}_W \in \mathcal{X}_W$ . Es decir  $\underline{x}_W = \times_{v \in W} x_v$ . Entonces ordenando las variables de  $W$  tal que  $\mathcal{X}_W = \mathcal{X}_D \times \mathcal{X}_{W \setminus D}$ , la proyección de  $\underline{x}_W$  sobre  $D$  es definido como la variable  $\underline{x}_D$  que satisface,

$$\underline{x}_W = (\underline{x}_D, \underline{x}_{W \setminus D})$$

**3.6 Potencial y carga.** Sea  $\Phi = \{\phi_1, \dots, \phi_m\}$  una conjunto de funciones de valores reales no negativos en  $\mathcal{X}$ . Las funciones  $\phi_j \in \Phi$  son llamados potenciales. El conjunto de potenciales  $\Phi$  es conocido como carga. Para cada  $j = 1, \dots, m$ ,  $\mathcal{X}_{D_j}$  denotara el estado de espacio para el potencial  $\phi_j$  mientras el conjunto  $D_j \subset \tilde{V}$  denota el conjunto de índices para las variables de argumento de  $\phi_j$ . El conjunto  $\mathcal{X}_{D_j}$  es llamado el dominio de variables de  $\phi_j$ .

La función de probabilidad conjunta  $p_{X_1, \dots, X_d}$  es en sí mismo un potencial con dominio  $\chi$ . Si la función de probabilidad conjunta puede ser factorizada de acuerdo a un grafo aciclico dirigido  $G = (V, E)$  la descomposición de escribe como:

$$p_{X_1, \dots, X_d} = \prod_{j=1}^d p_{X_j | \Pi_j}$$

Entonces para cada  $j = 1, \dots, d$ ,  $\phi_j$  definido por  $\phi_j(x_j, \underline{x}_{\tilde{\Pi}_j}) = p_{X_j | \Pi_j}(x_j | \underline{x}_{\tilde{\Pi}_j})$  es un potencial con dominio  $\chi_{D_j} = \chi_j \times \chi_{\Pi_j}$  y  $D_j = \{j\} \cup \tilde{\Pi}_j$ .

**3.7 Contracción.** Una contracción de unacarga, o conjunto de potenciales es una operación de multiplicación y división de potenciales,

**3.8 Marginalizacion.** Sea  $V = \{X_1, \dots, X_d\}$  denota el conjunto de  $d$  variables aleatorias, el conjunto indexado  $\tilde{V} = \{1, \dots, d\}$ . Sea  $U \subseteq W \subseteq V$  y sea  $\phi$  un potencial definido sobre  $\chi_W$ . la siguiente expresion  $\sum_{\chi_{W \setminus U}} \phi$  denota el margen o suma de marginales de  $\phi$  sobre  $\chi_U$  y es definida por  $\underline{x}_U \in \chi_U$

$$\left( \sum_{W \setminus U} \phi \right) (\underline{x}_U) = \left( \sum_{\underline{z} \in \chi_{W \setminus U}} \phi \right) (\underline{z}, \underline{x}_U)$$

La siguiente notación tambien es usada:

$$\phi^{\downarrow U} = \left( \sum_{W \setminus U} \phi \right)$$

**3.8.1 Marginalización y Propiedad distributiva.** La propiedad distributiva es usada cuando se marginaliza un producto de potenciales. Esto puede ser escrito como sigue: sea  $\phi_1$  sea un potencial con dominio  $\chi_{D_1}$  y sea  $\phi_2$  un potencial con dominio  $\chi_{D_2}$ ,

supóngase que  $A \subset D_1 \cup D_2$  y el producto  $\phi_1 \phi_2$  es marginalizado sobre  $\chi_A$ . Si  $A \cap D_i = \varnothing$ , entonces,

$$\sum_{\chi_A} \phi_1 \phi_2 = \phi_1 \sum_{\chi_A} \phi_2.$$

Supóngase que  $\phi_1$  tiene dominio  $\chi_{D_1 \cup D_3}$  y  $\phi_2$  tiene dominio  $\chi_{D_2 \cup D_3 \cup D_4}$ , donde  $D_1, D_2, D_3$  y  $D_4$  son disjuntos. La propiedad distributiva puede ser escrita como,

$$\sum_{\underline{x}_2 \in \chi_{D_2}} \phi_1(\underline{x}_1, \underline{x}_3) \phi_2(\underline{x}_2, \underline{x}_3, \underline{x}_4) = \phi_1(\underline{x}_1, \underline{x}_3) \sum_{\underline{x}_2 \in \chi_{D_2}} \phi_2(\underline{x}_2, \underline{x}_3, \underline{x}_4).$$

### 3.9 Eliminación y Dominio del Grafo.

Sea  $U$  denota un subconjunto de  $V$  y el conjunto  $W = V \setminus U$ , Sea  $\tilde{U}$ ,  $\tilde{W}$  y  $\tilde{V}$  denota la indexación de conjuntos  $U$ ,  $W$  y  $V$  considerar el cálculo de:

$$\Phi^{\downarrow U}(\underline{x}) = \left( \sum_{\underline{x}_{V \setminus U} \in \chi_{V \setminus U}} \Phi(\underline{x}_{V \setminus U}, \underline{x}_U) \right)$$

Por contracción

$$\Phi(\underline{x}) = \prod_{j=1}^m \phi_j(\underline{x}_{D_j})$$

Recordemos que la operación  $\Phi^{\downarrow U}(\underline{x})$  significa marginalizando  $\Phi$  sobre todas las variables que no están en el conjunto  $U$ . La variable  $x_v$  con índice  $v \in \tilde{W} = \tilde{V} \setminus \tilde{U}$  es eliminado de  $\sum_{\underline{x}_{V \setminus U} \in \chi_{V \setminus U}} \Phi(\underline{x}_{V \setminus U}, \underline{x}_U)$  por el siguiente procedimiento:

1. Sea  $\Phi_v$  o  $\Phi_{x_v}$ , denota la contracción de los potenciales in  $\Phi$  que tiene  $X_v$  en su dominio ; eso es :

$$\Phi_v = \prod_{j \in D_j} \phi_j.$$

2. Sea  $\phi^{(v)}$  o  $\phi^{(X_v)}$  denota la función  $\sum_{x_v \in \mathcal{X}_v} \Phi_v$ .

3. Encontrar un nuevo conjunto de potenciales  $\Phi^{-v}$  o  $\Phi^{-X_v}$

$$\Phi^{-v} = (\Phi \cup \{\phi^{(v)}\}) \setminus \Phi_v.$$

**Proposición 3.3.** Sea  $\Phi$  una contracción sobre un dominio  $\mathcal{X}_W$  y sea  $U \subset W$ ,

$$\Phi^{\downarrow U}(\underline{x}_U) = \left( \sum_{\underline{x}_{W \setminus U} \in \mathcal{X}_{W \setminus U}} \Phi(\underline{x}_{W \setminus U}, \underline{x}_U) \right)$$

Puede ser calculado a través de eliminación sucesiva de las variables.

$$X_v \in W \setminus U.$$

**Demostración de la Proposición 3.3.** Por la propiedad conmutativa de la multiplicación y adición junto con la propiedad distributiva, la eliminación de  $X_v$  dada

$$\Phi^{\downarrow U}(\underline{x}_U) = \sum_{\underline{x}_{W \setminus \{X_v\}} \in \mathcal{X}_{W \setminus \{X_v\}}} \Phi^{-X_v}(\underline{x}_U, \underline{x}_{W \setminus \{X_v\}}).$$

Por el argumento dado arriba,  $\Phi^{-X_v}$  es la contracción de un potencial sobre  $\mathcal{X}_{W \setminus \{X_v\}}$  y esto es claro, por inducción, que la marginal puede ser calculado a través de la eliminación sucesiva.

**3.10 Dominio del Grafo.** El dominio del grafo para el conjunto de potenciales en  $\Phi$  es un grafo no dirigido con las variables como nodos y el enlace entre cualquier par de variables que son miembros del mismo dominio.

**3.11 Eliminando un Nodo.** Sea  $G = (V, E)$  un grafo no dirigido, donde  $V = \{X_1, \dots, X_d\}$ .

Cuando un nodo  $X_v$  es eliminado del grafo  $G$  el resultado del grafo es denotado por  $G^{-X_v}$ .

Si  $G$  es el dominio del grafo para el conjunto de potenciales  $\Phi$  entonces es claro que el grafo  $G^{-X_v}$  es el dominio del grafo para el conjunto de potenciales  $\Phi^{-X_v}$ .

**3.12 Unión de árboles.** Sea  $\Phi = \{\phi_1, \dots, \phi_m\}$  un conjunto de potenciales en  $\mathcal{X}$  y sea  $G$  un grafo triangulado, una unión de árboles para  $\Phi$  es una unión de árboles para  $G$  tal que:

- Cada  $\phi_j$  está asociado con un clique  $C_j = \{X_{j1}, \dots, X_{jk}\}$  tal que  $\mathcal{X}_{C_j}$  el dominio de  $\phi_j$  es el espacio de estados de las variables  $\{X_{j1}, \dots, X_{jk}\}$ .
- Cada arista es asignada por un separador consistente de los nodos quedando después de la eliminación en un clique por una eliminación de secuencia.

**Resumen:** Construir un algoritmo de inferencia.

- Tome una red bayesiana y encontrar su dominio de grafo  $G$ .
- Triangular el dominio del grafo :  $G \rightarrow G'$ .
- Encontrar una eliminación de secuencia en  $G'$ .
- La eliminación de secuencia determina la eliminación de los dominios.
- La eliminación de dominios son cliques.
- Organizar los cliques en una unión de árbol.

Queda por describir un esquema de propagación para la marginalización para calcular:

$$P_{\underline{X}|e_U}(\underline{x})^{\downarrow A} = \left( \sum_{\underline{X}_{\tilde{V}|A} \in \mathcal{X}_{\tilde{V}|A}} P_{\underline{X}|e_U}(\underline{x}_A, \underline{x}_{\tilde{V}|A}) \right)$$

La tarea se hace menos complicada usando Hugin Lite, Hugin<sup>3</sup> propaga las probabilidades usando la formula Aalborg,

---

<sup>3</sup> <http://www.hugin.com/>

$$p_{\underline{x}}(x_1, \dots, x_d) = \frac{\prod_{C \in \zeta} \phi_C(\underline{x}_C)}{\prod_{S \in \delta} \phi_S(\underline{x}_S)}$$

Donde  $\zeta$  son cliques del grafo moral triangulado y  $\delta$  separadores de una unión de árboles.

**3.13 Factorización a lo largo de un grafo no dirigido.** Sea  $G = (V, E)$  un grafo no dirigido donde  $V = \{X_1, \dots, X_d\}$  es un conjunto de variables discretas.

Una probabilidad conjunta  $p_{\underline{x}}$  sobre un vector aleatorio  $\underline{X} = \{X_1, \dots, X_d\}$  es decir se factoriza de acuerdo a  $G$  si existe potenciales o factores  $\phi_A$  definidos sobre  $\times_{v \in \tilde{A}} \mathcal{X}_v$  donde  $A$  es un conjunto completo de nodos en  $G$  y  $\tilde{A}$  es un conjunto índice para  $A$ , tal que:

$$p_{\underline{x}}(\underline{x}) = \prod_A \phi_A(\underline{x}_A)$$

Donde la notación es clara, el producto es sobre todos los potenciales.

La distribución de probabilidad conjunta factorizado a lo largo de un grafo son conocidos como distribución de probabilidad de Markov, si sostiene la factorización entonces claramente para cualquier  $W \subset V$  tenemos

$$p_{\underline{x}_W}(\underline{x}_W) = \sum_{\underline{x}_{V \setminus W} \in \mathcal{X}_{V \setminus W}} \prod_A \phi_A(\underline{x}_W, \underline{x}_{V \setminus W})$$

Donde, para cada  $A$ , el dominio de  $\phi_A$  primero ha sido extendido a  $\mathcal{X}$ .

**Proposición 3.4.** Sea  $G$  un grafo no dirigido descompuesto y sea  $(A, B, S)$  descompone  $G$ , entonces siguiendo las dos observaciones son equivalentes:

1.  $p$  factoriza a lo largo de  $G$  y
2. Ambos  $p_{A \cup S}$  y  $p_{B \cup S}$  factoriza a lo largo de  $G_{A \cup S}$  y  $G_{B \cup S}$  respectivamente y

$$p(\underline{x}) = \frac{p_{A \cup S}(\underline{x}_{A \cup S}) p_{B \cup S}(\underline{x}_{B \cup S})}{p_S(\underline{x}_S)}$$

**Demostración de la proposición 3.4.** Dado que el grafo es descompuesto, sus cliques pueden ser organizados como una unión de árbol, por lo tanto, sin pérdida de generalidad, la factorización puede ser tomada de la forma,

$$p(\underline{x}) = \prod_{K \in \zeta} \phi_K(\underline{x}_K).$$

Donde el producto está sobre los cliques de  $G$  por tanto  $(A, B, C)$  descompone  $G$  cualquier clique es tomada como un subconjunto de  $A \cup B$  o como un subconjunto de  $B \cup S$ . Por lo tanto  $S$  es un subconjunto estricto de cualquier clique de  $A \cup S$  contiene  $S$  y  $S$  es un subconjunto estricto de cualquier clique de  $B \cup S$  que contiene  $S$ , entonces  $K$  denota un clique, donde:

$$p(\underline{x}) = \prod_{K \subseteq A \cup S} \phi_K(\underline{x}_K) \prod_{K \subseteq B \cup S} \phi_K(\underline{x}_K)$$

Puesto que  $S$  es en sí mismo completo, este es un subconjunto de cualquier clique que contiene  $S$ , tal que no se observa el clique en la descomposición en ambos  $A \cup S$  y  $B \cup S$ .

$$h(\underline{x}_{A \cup S}) = \prod_{K \subseteq A \cup S} \phi_K(\underline{x}_K)$$

y

$$k(\underline{x}_{B \cup S}) = \prod_{K \subseteq B \cup S} \phi_K(\underline{x}_K).$$

Entonces

$$p(\underline{x}) = h(\underline{x}_{A \cup S}) k(\underline{x}_{B \cup S})$$

Y la distribución marginal es dada por

$$p_{A \cup S}(\underline{x}_{A \cup S}) = \sum_{\chi_B} h(\underline{x}_{A \cup S}) k(\underline{x}_{B \cup S}).$$

La propiedad distributiva ahora da

$$p_{A \cup S}(\underline{x}_{A \cup S}) = h(\underline{x}_{A \cup S}) \sum_{\chi_B} k(\underline{x}_{B \cup S}) = h(\underline{x}_{A \cup S}) k(\underline{x}_S)$$

Donde

$$k(\underline{x}_S) := \sum_{\chi_B} k(\underline{x}_{B \cup S})$$

Similarmente

$$p_{B \cup S}(\underline{x}_{B \cup S}) = k(\underline{x}_{B \cup S}) h(\underline{x}_S)$$

Donde

$$h(\underline{x}_S) := \sum_{\chi_A} h(\underline{x}_{A \cup S}).$$

Esto sigue que

$$p(\underline{x}) = h(\underline{x}_{A \cup S}) k(\underline{x}_{B \cup S}) = \frac{p(\underline{x}_{A \cup S}) p(\underline{x}_{B \cup S})}{k(\underline{x}_S) h(\underline{x}_S)}$$

Es más

$$p_S(\underline{x}_S) = \sum_{\chi_{A \cup S}} h(\underline{x}_{A \cup S}) k(\underline{x}_{B \cup S}) = \sum_{\chi_A} h(\underline{x}_{A \cup S}) \sum_{\chi_B} k(\underline{x}_{B \cup S}) = h(\underline{x}_S) k(\underline{x}_S)$$

Es por tanto

$$p(\underline{x}) = \frac{p_{A \cup S}(\underline{x}_{A \cup S}) p_{B \cup S}(\underline{x}_{B \cup S})}{p_S(\underline{x}_S)}.$$

En el transcurso de la prueba, también se ha demostrado que  $p_{A \cup S}$  y  $p_{B \cup S}$  son factorizables a lo largo del grafo correspondiente.

**3.14 Propiedad de Markov.** Por una aplicación recursiva de la preposición, junto con

$$p(\underline{x}) = \prod_{C \subseteq A \cup S} \phi_C(\underline{x}_C) \prod_{C \subseteq B \cup S} \phi_C(\underline{x}_C)$$

$$p(\underline{x}) = \frac{\prod_{C \in \zeta} p_C(\underline{x}_C)}{\prod_{S \in \delta} p_S(\underline{x}_S)}$$

$\zeta$  denota el conjunto de cliques y  $\delta$  denota el conjunto de separadores.

**3.15. Factorizar a lo largo de una Unión de Árboles.**

Sea  $p$  una distribución de probabilidad factorizada a lo largo de un grafo acíclico dirigido  $G = (V, E)$ . La factorización está dada por:

$$P_{\underline{x}}(\underline{x}) = \prod_{v=1}^d p_{X_v, \Pi_v}(x_v | \underline{x}_{\tilde{\Pi}_v})$$

Donde  $\tilde{\Pi}_j$  denota el conjunto indexado para el conjunto de padres  $\Pi_j$ . Está claro que esto puede ser expresado como una factorización de acuerdo con el grafo moralizado  $G^{mor}$  Es no dirigido:

$$p_{\underline{x}}(\underline{x}) = \prod_{v=1}^d \phi_{A_v}(\underline{x}_{A_v})$$

Donde  $A = \{X_v\} \cup \Pi_v$  y  $\phi_{A_v}(\underline{x}_{A_v}) = p_{X_v, \Pi_v}(x_v | \underline{x}_{\tilde{\Pi}_v})$ .

Por lo tanto, una distribución de probabilidad factorizada a lo largo de un DAG es también factorizada a lo largo del grafo moral  $G^{mor}$ . El problema es que puede que no sea posible la representación de los conjuntos  $(A_v)_{v=1}^d$  en un árbol. Para permitir esto,

$G^{mor}$  es triangulada para dar  $(G^{mor})^t$ . Hay que recordar que  $(G^{mor})^t$  es descompuesta y sus cliques pueden ser organizados en una unión de árboles  $T$ . Entonces:

$$p_{\underline{x}}(\underline{x}) = \prod_{C \in \zeta} \phi_C(\underline{x}_C)$$

Donde  $\phi_C(\underline{x}_C)$  es el producto de todas las  $p(x_v | \underline{x}_{\bar{v}})$  cuyos argumentos pertenecen a  $C$ . Para cada moralización, siempre hay un clique. Tenga en cuenta que esta factorización no es necesariamente única. Resulta que:

$$p_{\underline{x}}(\underline{x}) = \frac{\prod_{C \in \zeta} p_C(\underline{x}_C)}{\prod_{S \in \delta} p_S(\underline{x}_S)} \quad (3.4)$$

Donde  $\zeta$  denota el conjunto de cliques y  $\delta$  denota el conjunto de los separadores de  $(G^{mor})^t$ .

**3.16 Factorización a lo largo de una unión de árboles y carga marginal.** Sea  $p_{\underline{x}}$  es una distribución de probabilidad sobre un vector aleatorio  $\underline{X} = (X_1, \dots, X_d)$ . Supóngase que las variables puede ser organizada como una unión de árboles, con cliques  $C$  y separadores  $\delta$  donde  $p_C$  y  $p_S$  denota la función de probabilidad marginal sobre una variable de cliques  $C \in \zeta$  y variables de separadores  $S \in \delta$  respectivamente. La siguiente ecuación es conocida como la factorización a lo largo de una unión de árboles,

$$p_{\underline{x}}(\underline{x}) = \frac{\prod_{C \in \zeta} p_C(\underline{x}_C)}{\prod_{S \in \delta} p_S(\underline{x}_S)}$$

Y la carga

$$\Phi = \{p_s : S \in \delta, p_c : C \in \zeta\}$$

es conocida como la carga marginal.

### 3.17 Introduciendo evidencia.

$$p_{\underline{x}}(\underline{x}) = \frac{\prod_{C \in \zeta} p_C(\underline{x}_C)}{\prod_{S \in \delta} p_S(\underline{x}_S)}$$

Esta ecuación expresa la distribución a priori en términos de potenciales sobre los cliques y separadores de  $(G^{mor})^t$  o la unión de árboles. Supóngase que la nueva evidencia difícil  $e_U$  es obtenida en las variables  $U$  es decir:

$$U \subseteq V, (\underline{x}_U = \underline{y}_U)$$

y la probabilidad sobre las variables  $V \setminus U$ :

$$p_{\underline{x}; e_U}(\underline{x}) = \begin{cases} \frac{\prod_{C \in \zeta} p_C(\underline{x}_C)}{\prod_{S \in \delta} p_S(\underline{x}_S)} & \underline{x}_U = \underline{y}_U \\ 0 & \underline{x}_U \neq \underline{y}_U \end{cases}$$

Ahora el conjunto,  $\phi_C(\underline{x}_C) = p_C(\underline{x}_{C \setminus U}, \underline{y}_{C \cap U})$  y  $\phi_S(\underline{x}_S) = p_S(\underline{x}_{S \setminus U}, \underline{y}_{S \cap U})$

Entonces

$$p_{\underline{x}_{V \setminus U}, \underline{x}_U}(\underline{x}_{V \setminus U} | \underline{y}_U) = \frac{\prod_{C \in \zeta} \phi_C(\underline{x}_C)}{\prod_{S \in \delta} \phi_S(\underline{x}_S)} \quad (3.5)$$

La distribución posterior

$$p_{\underline{x}_{V \setminus U} | \underline{x}_U}(\cdot | \underline{y}_U) = \frac{p_{\underline{x}_{V \setminus U}, \underline{x}_U}(\cdot, \underline{y}_U)}{p_{\underline{x}_U}(\underline{y}_U)}$$

Entonces es calculado por hallar una representación de la función  $p_{\underline{x}_{V \setminus U}, \underline{x}_U}(\cdot, \underline{y}_U)$  sobre el dominio  $\mathcal{X}_{V \setminus U}$  usando el algoritmo definido abajo. El algoritmo para cualquier función  $f: \mathcal{X} \rightarrow \mathbb{R}_+$  (no necesariamente una función de probabilidad) es expresado como,

$$f(\underline{x}) = \frac{\prod_{C \in \zeta} \phi_C(\underline{x}_C)}{\prod_{S \in \delta} \phi_S(\underline{x}_S)} \quad (3.6)$$

Para un conjunto de potenciales  $\Phi = \{\phi_C, C \in \zeta, \phi_S, S \in \delta\}$  donde  $\zeta$  y  $\delta$  son los cliques y separadores de una unión de árboles, el algoritmo actualizado  $\Phi$  para un conjunto de potenciales

$$\Phi^* = \{\phi_C^*, C \in \zeta, \phi_S^*, S \in \delta\}$$

$$\phi_C^*(\underline{x}_C) = \sum_{\underline{z} \in \mathcal{X}_{V \setminus C}} f(\underline{z}, \underline{x}_C)$$

$$\phi_S^*(\underline{x}_S) = \sum_{\underline{z} \in \mathcal{X}_{V \setminus S}} f(\underline{z}, \underline{x}_S)$$

Para cada  $C \in \zeta$  y cada  $S \in \delta$  entonces la probabilidad de la evidencia es:

$$p_{\underline{x}_U}(\underline{y}_U) = \sum_{\underline{z} \in \mathcal{X}_{V \setminus (U \cap C)}} \phi_C^*(\underline{z}, \underline{y}_{U \cap C}) = \sum_{\underline{z} \in \mathcal{X}_{S \setminus (U \cap S)}} \phi_S^*(\underline{z}, \underline{y}_{U \cap S})$$

Para todo  $S \in \delta$  y  $C \in \zeta$ . La distribución de probabilidad condicional sobre las variables  $V \setminus U$  puede ser calculada por marginalizar el clique o separador con el dominio más pequeño, dando una representación de la distribución condicional en términos de distribución marginal sobre los cliques y separadores.

### 3.18 Flujo de mensajes.

Considerar una función no negativa con dominio  $X \times Y \times Z$ ,  $F : X \times Y \times Z \rightarrow R_+$ , puede ser escrito como:

$$F(x, y, z) = \frac{f(x, z)g(y, z)}{h(z)}. \quad (3.7)$$

Para potenciales  $f : X \times Z \rightarrow R_+$ ,  $g : Y \times Z \rightarrow R_+$  y  $h : Z \rightarrow R_+$

El siguiente algoritmo devuelve una representación de  $F(x, y, z) = \frac{F_1(x, z)F_2(y, z)}{F_3(z)}$ ,

donde

$$F_1(x, z) = \sum_{y \in Y} F(x, y, z), \quad F_2(y, z) = \sum_{x \in X} F(x, y, z), \quad F_3(z) = \sum_{(x, y) \in X \times Y} F(x, y, z).$$

$$\text{Primeramente } F_1(x, z) = \sum_{y \in Y} F(x, y, z) = \sum_y \frac{f(x, z)g(y, z)}{h(z)} = \frac{f(x, z)}{h(z)} \sum_{y \in Y} g(y, z).$$

Define la función auxiliar  $h^*(z) = \sum_y g(y, z)$  y la actualización

$$f^*(x, z) = f(x, z) \frac{h^*(z)}{h(z)} \text{ entonces claramente } f^*(x, z) = f(x, z) \frac{h^*(z)}{h(z)} = F_1(x, z)$$

El cálculo de la función marginal  $F_1(x, z)$  por medio de la función auxiliar  $h^*(z)$  describe como transmite un flujo de mensaje local de  $ZY$  a  $XZ$  a través de un separador  $Z$ . El factor

$$\frac{h^*(z)}{h(z)}$$

Es la probabilidad actualizada:

$$F(x, y, z) = \frac{f(x, z)g(y, z)}{h(z)} = \frac{f(x, z)g(y, z)h^*(z)}{h^*(z)h(z)} = F_1(x, z) \frac{1}{h^*(z)} g(y, z).$$

La transmisión del flujo ha dado lugar a una nueva representación de  $F(x, y, z)$ , similar a la original, pero donde uno de los factores es una función marginal. Del mismo modo, un mensaje puede ser transmitida en otra dirección, es decir de  $XZ$  a  $ZY$  Usando el mismo procedimiento, establecido como:

$$\tilde{h}(z) = \sum_{x \in X} F_1(x, z) = \sum_{(x, y) \in X \times Y} F(x, y, z) = F_3(z).$$

Sea  $\tilde{g}(y, z) = g(y, z) \frac{\tilde{h}(z)}{h^*(z)}$ .

De esto se deduce que  $\tilde{g}(y, z) = F_2(y, z)$

Esto es así porque  $F(x, y, z) = F_1(x, z) \frac{1}{h^*(z)} \tilde{g}(y, z) = F_1(x, z) \frac{1}{F_3(z)} \tilde{g}(y, z)$

y por lo tanto, ya que  $F_3(z) = \sum_{x \in X} F_1(x, z)$  se deduce que

$$F_2(y, z) = \sum_{x \in X} F(x, y, z) = \tilde{g}(y, z) \sum_{x \in X} F_1(x, z) \frac{1}{F_3(z)} = \tilde{g}(y, z)$$

La transmisión de mensajes en ambas direcciones resulta en una nueva representación general de la función  $F(x, y, z)$ ;

$$\begin{aligned} F(x, y, z) &= f^*(x, z) \frac{1}{h^*(z)} g(y, z) = f^*(x, z) \frac{1}{h^*(z)} \frac{h^*(z)}{\tilde{h}(z)} \tilde{g}(y, z) \\ &= f^*(x, z) \frac{1}{\tilde{h}(z)} \tilde{g}(y, z) \\ &= F_1(x, z) \frac{1}{F_3(z)} F_2(y, z) \end{aligned}$$

La representación original haciendo uso de los potenciales se ha transformado en una nueva representación donde todos los potenciales son funciones marginales.

**3.19 Cálculo local de la Unión de Árboles.** Consideramos una unión de árboles  $T$  con nodos  $\zeta$  y  $\delta$  separadores, y sea  $\Phi$  una carga:

$$\Phi = \{\phi_C : C \in \zeta, \phi_S : S \in \delta\}, \quad (3.8)$$

Es un conjunto de potenciales tal que  $\phi_C = \chi_C \rightarrow R_+$  y  $\phi_S = \chi_S \rightarrow R_+$  para cada  $C \in \zeta$  y cada  $S \in \delta$ .

**3.20 Contracción de unacarga sobre una Unión de Árboles.** La contracción de unacarga sobre la unión de árboles se define como:

$$f(\underline{x}) = \frac{\prod_{C \in \zeta} \phi_C(\underline{x}_C)}{\prod_{S \in \delta} \phi_S(\underline{x}_S)} \quad (3.9)$$

**3.21 Transmisión del mensaje local.** Sea  $C_1$  y  $C_2$  dos nodos vecinos adyacentes en  $T$  y separados por  $S^0$ :

$$\phi_{S^0}^*(\underline{x}_{S^0}) = \sum_{\underline{z} \in \chi_{C_1|S^0}} \phi_{C_1}(\underline{z}, \underline{x}_{S^0}) \quad (3.10)$$

Y el conjunto

$$\lambda_{S^0} = \frac{\phi_{S^0}^*(\underline{x}_{S^0})}{\phi_{S^0}(\underline{x}_{S^0})} \quad (3.11)$$

Note que directamente de la definición de la división de potenciales  $\lambda_{S^0} = 0$  para  $\phi_{S^0} = 0$  entonces la actualización de probabilidad es definida como la cantidad  $\lambda_{S^0}$ . La transmisión de mensaje esta definido como la operación de actualización  $\phi_{S^0}$  a  $\phi_{S^0}^*$  y  $\phi_{C_2}$  a  $\phi_{C_2}^* = \lambda_{S^0} \phi_{C_2}$ .

**Lema 3.3.** Sea  $f = \chi \rightarrow R_+$  la contracción de una carga  $\Phi = \{\phi_S, S \in \delta, \phi_C, C \in \zeta\}$  en una unión de árboles donde  $\zeta$  es el conjunto de cliques y  $\delta$  el conjunto de separadores. Un flujo no cambia la contracción de la carga.

**Demostración del lema 3.3.** La contracción inicial es dada por

$$f(\underline{x}) = \frac{\prod_{C \in \zeta} \phi_C(\underline{x}_C)}{\prod_{S \in \delta} \phi_S(\underline{x}_S)}. \quad (3.12)$$

Por la definición de división de potenciales,  $f(\underline{x}) = 0$  para todo  $\underline{x}$  tal que  $\phi_S(\underline{x}_S) = 0$  para algún  $S \in \delta$  esta es parte de la definición en la hipótesis que  $f$  tiene una representación de la forma dada en la ecuación (3.12). La carga, después el flujo de  $C_1$  a  $C_2$  será denotado por

$$\Phi^* = \{\phi_C^* : C \in \zeta, \phi_S^* : S \in \delta\}$$

Y la contracción

$$f^*(\underline{x}) := \frac{\prod_{C \in \zeta} \phi_C^*(\underline{x}_C)}{\prod_{S \in \delta} \phi_S^*(\underline{x}_S)}. \quad (3.13)$$

Note que,

$$f^*(\underline{x}) := \frac{\phi_{C_2}^*(\underline{x}_{C_2}) \prod_{C \in \zeta, C \neq C_2} \phi_C(\underline{x}_C)}{\phi_{S^0}^*(\underline{x}_{S^0}) \prod_{S \in \delta, S \neq S^0} \phi_S(\underline{x}_S)}. \quad (3.14)$$

Hay tres casos para considerar

- Para  $\underline{x}$  tal que  $\phi_{S^0}(\underline{x}_{S^0}) > 0$  y  $\phi_{S^0}^*(\underline{x}_{S^0}) > 0$

$$\frac{\phi_{C_2}^*}{\phi_{S^0}^*} = \frac{\phi_{C_2} \lambda_{S^0}}{\phi_{S^0}^*} = \frac{\phi_{C_2} \left( \frac{\phi_{S^0}^*}{\phi_{S^0}} \right)}{\phi_{S^0}^*} = \frac{\phi_{C_2}}{\phi_{S^0}}$$

- Para el segundo caso y el tercer caso, Para  $\underline{x}$  tal que  $\phi_{S^0}(\underline{x}_{S^0})=0, f(\underline{x})=0$  de la definición. Por lo tanto de la ecuación (3.11),  $\lambda_{S^0}=0$  sigue de la división entre cero en el sentido de la división de los potenciales y por lo tanto, la definición de  $\phi_{C_2}^*$  sigue que  $\phi_{C_2}^*=0$ , esto sigue de la ecuación (3.13) donde  $f^*(\underline{x})=0$  tal que  $0=f^*(\underline{x})=f(\underline{x})$ .
- Para  $\underline{x}$  tal que  $\phi_{S^0}(\underline{x}_{S^0})>0$  pero  $\phi_{S^0}^*(\underline{x}_{S^0})=0$  esto sigue directamente de la ecuación (3.14) (Usando la definición de división entre 0 en el sentido de potenciales) entonces  $f^*(\underline{x})=0$  esto queda para mostrar que  $f(\underline{x})=0$  de la definición:

$$0 = \phi_{S^0}^*(\underline{x}_{S^0}) = \sum_{\underline{z} \in \mathcal{X}_{C_1|S^0}} \phi_{C_1}(\underline{z}, \underline{x}_{S^0})$$

- Por tanto  $\phi_{C_1}(\underline{x}_{C_1}) \geq 0$  para todo  $\underline{x}_{C_1} \in \mathcal{X}_{C_1}$  sigue que  $\phi_C(\underline{z}, \underline{x}_{S^0})=0$  para todo  $\underline{z} \in \mathcal{X}_{C_1|S^0}$  desde:

$$f(\underline{x}) = \phi_{C_1}(\underline{x}_{C_1}) \frac{\phi_{C_2}(\underline{x}_{C_2}) \prod_{C \in \zeta, C \neq C_1, C_2} \phi_C(\underline{x}_C)}{\phi_{S^0}(\underline{x}_{S^0}) \prod_{S \in \delta, S \neq S^0} \phi_S(\underline{x}_S)}.$$

Esto sigue directamente del hecho de que los dominios de los cliques  $C_1$  y  $C_2$  y otros separadores  $S^0$  no incluye  $\mathcal{X}_{S^0}$ , y

$$\frac{\phi_{C_2}(\underline{x}_{C_2})}{\phi_{S^0}(\underline{x}_{S^0})} < +\infty$$

tal que  $f(\underline{x})=0$  por tanto  $f(\underline{x})=f^*(\underline{x})$

En todos los casos, esto indica que un flujo no cambia la contracción de unacarga.

### 3.22 Programa.

El objetivo es construir una serie de transmisiones entre los cliques de un unión de árboles, para actualizar un conjunto de potenciales, cuya contracción es una distribución de probabilidad posterior sobre los cliques y separadores.

### 3.23 Sub-árbol y Vecino clique.

Un sub-árbol  $T'$  de una unión de árboles  $T$  es un conjunto relacionado de nodos de  $T$  junto con las aristas en  $T$  entre ellos.

Un clique  $C$  de una unión de árboles  $T$  es un vecino de un sub-árbol  $T'$  si el correspondiente nodo de  $T$  no es un nodo de  $T'$  pero está relacionado a  $T'$  por una arista de  $T$ .

### 3.24 Propagación floja.

Es el método para actualizar una distribución que es factorizada sobre una unión de árboles conocido como propagación floja.

**Proposición 3.6.** Para cualquier árbol  $T$ . Existe un Programa activo completamente.

**Demostración de la Proposición 3.6.** Si sólo hay un clique, la proposición es clara, son transmisiones no necesarias, asume que hay más que un clique. Sea  $C_0$  denota una hoja en  $T$ . Sea  $T_0$  un subárbol de  $T$  obtenido por remover  $C_0$  y la correspondiente arista  $S_0$ . Asume que la proposición es verdad para  $T_0$  adicionando la arista.

$$C_0 \rightarrow S_0 \rightarrow T_0$$

Al principio del programa

$$C_0 \leftarrow S_0 \leftarrow T_0$$

para el final del programa proporciona un programa completamente activo para  $T$ .

### 3.25 Base de un sub-árbol y restricción de unacarga.

Sea  $T'$  un sub-árbol de  $T$ , con nodos  $\zeta' \subseteq \zeta$  y aristas  $\delta' \subseteq \delta$ . La base de  $T'$  es definido como el conjunto de variables.

$$U' = \cup_{C \in \zeta'} C$$

Sea

$$\Phi = \{\phi_C : C \in \zeta, \phi_S : S \in \delta\}$$

Unacarga  $T$  y Su restricción para  $T'$  es definido como

$$\Phi_{T'} = \{\phi_C : C \in \zeta', \phi_S : S \in \delta'\}$$

La contracción de  $\Phi_{T'}$ , es definido como:

$$\frac{\prod_{C \in \zeta'} \phi_C(\underline{x}_C)}{\prod_{S \in \delta'} \phi_S(\underline{x}_S)}$$

**Proposición 3.7.** Sea  $\Phi^0 = \{\phi_C^0 : C \in \zeta, \phi_S^0 : S \in \delta\}$  denota unacarga inicial para una función  $f$  que tiene factorización,

$$f(\underline{x}) = \frac{\prod_{C \in \zeta} \phi_C^0(\underline{x}_C)}{\prod_{S \in \delta} \phi_S^0(\underline{x}_S)}$$

Donde  $\zeta$  y  $\delta$  son los conjuntos de cliques y separadores para la unión de árboles  $T$ . Supóngase que  $\Phi^0$  es modificada por una secuencia de flujos en función de algún programa, la contracción de la carga para  $T'$  es el margen de la contracción  $f$  de la carga para  $T$  en  $U'$ .

**Corolario 3.1.** Sea  $\{\phi_C : C \in \zeta, \phi_S : S \in \delta\}$  denota el potencial actual sobre cliques y separadores. Para cualquier conjunto  $A \subseteq V$  sea  $f_A = \sum_{x_{V \setminus A}} f$  el marginal sobre  $A$  siempre que un clique  $C$  está activo, su potencial es  $\phi_C = f_C = \sum_{x_{V \setminus C}} f$ .

**Demostración del corolario 3.1.** Un clique simple es un subárbol, el resultado se lo muestra en la proposición 3.7.

**Corolario 3.2.** El flujo activo se ha transmitido en ambas direcciones cruzando una arista en  $T'$ , el potencial para el separador asociado es  $\phi_S = f_S = \sum_{x_{V \setminus S}} f$ .

**Demostración del corolario 3.2.** El potencial  $\phi_S$  para el separador asociado es por definición de la actualización

$$\phi_S = \sum_{x_{C \setminus S}} \phi_C$$

Tal que

$$\sum_{\mathcal{X}_{C|S}} \phi_C = \sum_{\mathcal{X}_{C|S}} f_C = f_S,$$

Porque  $\phi_C$  es  $f_C$  por el corolario anterior.

### **El resultado principal.**

Después de transmitir un programa completamente activo el resultado de la carga es la carga consistente de la marginal sobre los cliques y separadores y su contracción representa  $f$  en otras palabras, la siguiente formula, conocida como la fórmula de Aalborg

$$f(\underline{x}) = \frac{\prod_{C \in \mathcal{C}} f_C(\underline{x}_C)}{\prod_{S \in \mathcal{S}} f_S(\underline{x}_S)}.$$

**Demostración de la Proposición 3.7** Esto sigue de los últimos dos corolarios y el lema

3.3, que indica que la contracción no se altera por los flujos.

### **3.26 Unión de árboles delgado.**

A veces los cliques en una unión de árbol puede contener muchas más variables, que pueden ser problemáticos si se está utilizando métodos numéricos para calcular distribuciones marginales. Este problema es abordado por R.G. Cowell, A. P. Dawid, S.L.Lauritzen y D.J. Spiegelhalter en [67], que discuten los métodos de romper la distribución de probabilidad de los cliques en pequeñas distribuciones marginales.

### **3.27 Usando una Unión de Árboles con Evidencia Virtual y Evidencia Suave.**

Los métodos descritos hasta ahora pueden extenderse al problema de la actualización de la evidencia virtual y la evidencia suave.

El trato con las evidencias virtuales es simple, por cada hallazgo virtual se suma un nodo virtual, que será instanciado de acuerdo al hallazgo virtual. Simplemente se agrega el nodo virtual para encontrar el clique que contiene la variable para la cual es un hallazgo virtual.

Incorporando la evidencia suave no se puede llevar a cabo de una manera simple, porque cuando hay un hallazgo suave en una variable, el DAG se ve alterado por la eliminación de las aristas dirigidas de los nodos de los padres a la variable. La entrada es una red bayesiana con un conjunto de hallazgos suaves y hallazgos difíciles. El método devuelve una distribución de probabilidad conjunta con dos propiedades:

1. Los hallazgos son la distribución marginal para la distribución actualizada.
2. La distribución actualizada es el más cercano a la distribución marginal (donde se usa la divergencia de Kullback-Leibler) que satisface esta restricción.

## *Capítulo 4*

### *Árbol de Chow Liu*

#### **4.1 Introducción**

En los anteriores capítulos ya se explicó la construcción de un árbol ahora el objetivo del mismo es encontrar la mejor estructura es decir aquella que nos ayude a tomar decisiones. El árbol de Chow Liu proporciona la red con el peso máximo de entre todas las redes.

#### **4.2 Asignación de una Distribución de Probabilidad a la Estructura de Grafo.**

Para obtener una red bayesiana con un grafo acíclico dirigido  $G=(V,E)$ , el conjunto de aristas  $E$  se refiere a menudo a la estructura de la red. Sea  $\tilde{\mathcal{E}}$  denota el conjunto de todos los posibles conjuntos de aristas que dan un grafo acíclico dirigido con conjunto de nodos  $V$ . Se supone que  $E$  es desconocido y tiene que deducirse de los datos, en los casos individuales que se supone que son observaciones sobre vectores aleatorios que son independientes, condicionado a la estructura de grafo  $E$  y los parámetros  $\underline{\theta}$ . La distribución a priori sobre los vectores de parámetro de  $\underline{\theta}_{j,l}$  se toman de la familia  $Dir(\alpha_{jl}, \dots, \alpha_{jk,l})$  para todos los nodos y las configuraciones de los padres  $(j,l)$ . Como

es habitual  $X = \begin{pmatrix} \underline{X}_{(1)} \\ \vdots \\ \underline{X}_{(n)} \end{pmatrix}$  denota la matriz donde cada fila es una copia independiente de

$X$  y  $E$  denota la estructura del grafo aciclico dirigido. Este tiene una distribución a priori  $p_{\mathcal{E}}$ , que es la función de probabilidad para una variable aleatoria  $\mathcal{E}$  que toma valores en  $\tilde{\mathcal{E}}$ , el conjunto de posibles estructuras de grafo.

**4.3 Distribución apriori de la estructura del grafo** Hay varias maneras posibles de construir una distribución a priori  $p_{\mathcal{E}}$ . Si se conoce a priori que la estructura del grafo se encuentra dentro de un subconjunto  $A \subseteq \tilde{\mathcal{E}}$ , entonces una elección obvia es la apriori uniforme sobre  $A$ :

$$p_\varepsilon(E) = \begin{cases} \frac{1}{|A|} & \text{si } E \in A \\ 0 & \text{en otro caso} \end{cases}$$

Donde  $|A|$  es el número de elementos de un subconjunto  $A \subseteq \tilde{\varepsilon}$ .

Otra opción sencilla es asignar a cada par de nodos  $X_i$  y  $X_j$  una distribución de probabilidad con tres valores de manera que:

$$p(\{(X_i, X_j) \in E\}) + p(\{(X_j, X_i) \in E\}) + p(\{(X_i, X_j) \notin E\}, \{(X_j, X_i) \notin E\}) = 1$$

Una vez más, la distribución uniforme en las tres posibilidades puede ser elegida. Entonces, para una estructura dada  $E$ , la probabilidad a priori se obtiene multiplicando las probabilidades apropiadas sobre todas las aristas en  $E$  y la normalización dada  $p_\varepsilon(E)$ .

**4.4 Maximización de la estructura del grafo.** La maximización de la estructura del grafo, teniendo en cuenta datos  $x$ , está dada por

$$p_{X|\varepsilon}(x|E) = \int p_{X|\underline{\theta}, \varepsilon}(x|\underline{\theta}, E) \pi_{\underline{\theta}|\varepsilon}(\underline{\theta}|E) d\underline{\theta},$$

Donde  $\pi_{\underline{\theta}|\varepsilon}(\cdot|E)$  denota la distribución a priori sobre los parámetros  $\underline{\theta}$ , condicionada a la estructura del grafo, y el uso de la estructura del producto a priori de la distribución de Dirichlet,

$$p_{X|\varepsilon}(x|E) = \int \prod_{k=1}^n p_{X|\underline{\theta}, \varepsilon}(x_{(k)}|\underline{\theta}, E) \prod_{j=1}^d \prod_{l=1}^{q_j} \phi(\underline{\theta}_{j,l}, \underline{\alpha}_{j,l}) d\underline{\theta}_{j,l},$$

Donde  $(\underline{\theta}_{j,l}, \underline{\alpha}_{j,l})$  es una forma compacta de la distribución de densidad de Dirichlet  $Dir(\alpha_{j1l}, \dots, \alpha_{jk,l})$ .

Debido a  $p_{X|\Theta,\varepsilon}(x|\underline{\theta}, E)$  tiene una forma de producto conveniente, calcular la integral de Dirichlet es sencilla y está dado por:

$$p_{X|\varepsilon}(x|E) = \prod_{j=1}^d \prod_{l=1}^{q_j} \frac{\Gamma\left(\sum_{i=1}^{k_j} \alpha_{jil}\right)}{\Gamma\left(n(\pi_j^l) + \sum_{i=1}^{k_j} \alpha_{jil}\right)} \prod_{i=1}^{k_j} \frac{\Gamma\left(n(x_j^i | \pi_j^l) + \alpha_{jil}\right)}{\Gamma(\alpha_{jil})} \quad (4.1)$$

Esta es la maximización de Cooper-Herskovitz de la estructura de grafo.

La regla de selección bayesiano para un grafo  $G = (V, E)$  utiliza el grafo que maximiza la probabilidad posterior

$$p_{\varepsilon|X}(E|x) = \frac{p_{X|\varepsilon}(x|E) p_{\varepsilon}(E)}{p_X(x)} \quad (4.2)$$

Donde  $p_{\varepsilon}$  es la probabilidad apriori sobre el espacio de los conjuntos de aristas. Las probabilidades apriori para dos conjuntos de aristas diferentes  $E_1$  y  $E_2$  se define como:

$$\frac{p_{\varepsilon}(E_1)}{p_{\varepsilon}(E_2)}$$

*Posterior odds ratio* es define como:

$$\frac{p_{\varepsilon|X}(E_1|x)}{p_{\varepsilon|X}(E_2|x)}$$

Pero esta ecuación puede ser expresada como:

$$\text{probabilidad aposterior} = \text{la verosimil} \times \text{probabilidad apriori}.$$

A primera vista, el procedimiento de aprendizaje puede parecer bastante sencillo, sólo hay un número finito de diferentes posibles grafos aciclicos dirigidos  $G = (V, E)$  con  $d$  nodos y por lo tanto, sólo un número finito de valores de  $p_{\varepsilon|X}(E|x)$  será calculada. Del conjunto de aristas  $E$  se elegirá el valor máximo. Sea  $N(d)$  el número de posibles

grafos acíclicos dirigidos en los  $d$  nodos. Entonces enumeramos todos los grafos posibles  $(E_r)_{r=1}^{N(d)}$  dado:

$$p_X(x) = \sum_{r=1}^{N(d)} p_{X|\varepsilon}(x|E_r) p_\varepsilon(E_r).$$

R.W. Robinson dio la siguiente función recursiva para calcular el número  $N(d)$  de grafos acíclicos dirigidos con  $d$  nodos:

$$N(d) = \sum_{i=1}^d (-1)^{i+1} \binom{d}{i} 2^{i(d-1)} N(d-1). \quad (4.3)$$

**4.5 Predicción del problema.** Dado una matriz de datos  $n \times d$  entonces  $x$  representa  $n$  instancias independientes de un vector aleatorio  $\underline{X} = (X_1, \dots, X_d)$ , el problema de predicción consiste en calcular la distribución  $p_{\underline{X}_{(n+1)}|X}(x|X)$ , la distribución de probabilidad condicional de la siguiente observación. En primer lugar, la estimación de máxima verosimilitud  $E(x)$  es definida como el valor de  $E$  que maximiza

$$\hat{p}_{X|\varepsilon}(x|E) = \prod_{j=1}^d \prod_{l=1}^{q_j} \prod_{i=1}^{k_j} \hat{\theta}_{jil}^{n_k(x_j^{(i)}|\pi_j^l)}$$

Dónde:

$$\hat{\theta}_{jil} = \frac{n(x_j^{(i)}, \pi_j^l)}{n(\pi_j^l)}$$

$n(x_j^{(i)}, \pi_j^l)$  Denota el número de veces que la configuración  $x_j^{(i)}, \pi_j^l$  aparece en el dato y

$n(\pi_j^l)$  denota el número de veces que la configuración del padre  $\pi_j^l$  aparece.

#### 4.6 NP-Hard.

Para el caso general se han desarrollado otros algoritmos. El problema de la propagación en redes bayesianas es NP Hard lo que significa que no es posible obtener un algoritmo de complejidad polinomial para el problema de la propagación en redes bayesianas con una topología general. Para intentar paliar esta complejidad se han desarrollado algoritmos de propagación aproximada, que también son NP-hard en la precisión de la estimación.

#### 4.7 Árbol de Chow-Liu

Encontrar la estructura óptima de entre todas las estructuras posibles es un problema NP-hard. Además, no sólo tiene uno para estimar la distribución de probabilidad de un número finito de muestras, también se tiene que almacenar el resultado de probabilidad potenciales en una cantidad limitada de memoria de la máquina. En general, no es posible considerar todas las posibles dependencias. Es por esta razón que el árbol de Chow Liu nos ayuda a reducir el grafo y encontrar la mejor solución.

**4.8 Divergencia de Kullback-Leibler.** La divergencia de Kullback-Leibler entre dos distribuciones de probabilidad discreta  $f$  y  $g$  con el mismo espacio de estado  $\mathcal{X}$  se define como

$$D(f|g) = \sum_{x \in \mathcal{X}} f(\{x\}) \log \frac{f(\{x\})}{g(\{x\})}$$

Si  $g(\{x\}) = 0$  para  $f(\{x\}) \neq 0$  entonces  $f(\{x\}) \cdot \log 0 = +\infty$ . Si  $D(f|g) = +\infty$ . Entonces hay menos una salida  $x$  tal que  $f$  y  $g$  pueden ser distinguidas sin error.

**Lema 4.1** Para cualquier par de distribuciones de probabilidad discreta  $f$  y  $g$ , se tiene que:

$$D(f|g) \geq 0$$

**Demostración Lema 4.1** La prueba utiliza la desigualdad de Jensen, es decir, que para cualquier función convexa  $\phi$ ,  $E[\phi(X)] \geq \phi(E[X])$ . Tenga en cuenta que  $f(\{x\}) \geq 0$  para todo  $x \in \mathcal{X}$  y que  $\sum_{x \in \mathcal{X}} f(\{x\}) = \sum_{x \in \mathcal{X}} g(\{x\}) = 1$ . El uso de este, junto con el hecho de que  $-\log$  es convexa, los rendimientos.

$$D(f|g) = -\sum_{x \in \mathcal{X}} f(\{x\}) \log \left( \frac{g(\{x\})}{f(\{x\})} \right) \geq -\log \left( \sum_{x \in \mathcal{X}} f(\{x\}) \frac{g(\{x\})}{f(\{x\})} \right) = -\log 1 = 0$$

El árbol de Chow-Liu asume que cada variable tiene a lo sumo uno de los padres. Supongamos que hay  $d$  variables y las posibles dependencias entre las variables. La aproximación de Chow-Liu encuentra  $d - 1$  dependencias de primer orden entre las variables. Es decir, la distribución de  $p$  sobre el vector aleatorio  $\underline{X} = (X_1, \dots, X_d)$  se considera que es de la forma,

$$p_{X_1, \dots, X_d} = \prod_{j=1}^d p_{X_{m_j} | X_{m(i)}}$$

Donde  $(m_1, \dots, m_d)$  es una permutación de  $(1, \dots, d)$  y  $0 \leq i(j) \leq j$ . Cada variable puede estar condicionada a más de una variable. Cuando estimamos el árbol, la divergencia Kullback-Leibler se utiliza para determinar comodas distribuciones sonel uno al otro. Para calcular el mejor ajustado del árbol Chow-Liu, los potenciales  $\hat{p}_{X_i, X_j}$ ,  $\hat{p}_{X_i}$  y  $\hat{p}_{X_j}$  deben ser calculados, donde  $\hat{p}_{X_i, X_j}$ ,  $\hat{p}_{X_i}$  y  $\hat{p}_{X_j}$  denotan las estimaciones empíricas de los potenciales,  $p_{X_i, X_j}$ ,  $p_{X_i}$  y  $p_{X_j}$  de datos. La definición de información mutua se basa en la divergencia Kullback-Leibler.

#### 4.9 Información mutua.

La información mutua entre las variables  $X$  y  $Y$  se define como:

$$I(X, Y) = \sum_{x, y} p_{X, Y}(x, y) \log \frac{p_{X, Y}(x, y)}{p_X(x) p_Y(y)}$$

Esto es, por supuesto, la divergencia de Kullback-Leibler entre  $p_{X,Y}$  y  $p_X, p_Y$  :

$$I(X, Y) = D_{KL}(p_{X,Y} | p_X p_Y).$$

Sea  $\hat{I}$  denota la estimación de la información mutua de los datos, es decir,

$$\hat{I}(X, Y) = \sum_{x,y} \hat{p}_{X,Y}(x, y) \log \frac{\hat{p}_{X,Y}(x, y)}{\hat{p}_X(x) \hat{p}_Y(y)}. \quad (4.4)$$

La idea es encontrar el árbol dependiente de peso máximo, es decir, un árbol  $\sigma$  tal que para cualquier otro árbol  $\sigma'$ .

$$\sum_{j=1}^d \hat{I}(X_j, X_{\sigma(j)}) \geq \sum_{j=1}^d \hat{I}(X_j, X_{\sigma'(j)})$$

Este utiliza el algoritmo de Kruskal, que se describe a continuación. Aquí  $(j, \sigma(j))_{j=1}^d$  denota el conjunto de aristas de peso máximo, del árbol  $(j, \sigma'(j))_{j=1}^d$  denota el conjunto de aristas de cualquier otro árbol admisible.

El procedimiento de optimización Algoritmo de Kruskal:

1. Las  $d$  variables de rendimiento  $d(d-1)/2$  aristas. Las aristas son indexadas en orden decreciente, de acuerdo con sus pesos  $b_1, b_2, b_3, \dots, b_{d(d-1)/2}$ .
2. Los bordes  $b_1$  y  $b_2$  son seleccionados. Luego, la arista  $b_3$  es añadido, si no forma un ciclo.
3. Esto se repite, a través de  $b_4, \dots, b_{d(d-1)/2}$ , en ese orden, adicionando aristas si no forman un ciclo y deshacerse de ellos si forman un ciclo.

Este procedimiento devuelve un árbol único, si los pesos son diferentes. Si dos pesos son iguales, se puede imponer un orden arbitrario. Desde el  $d(d-1)/2$  aristas, exactamente  $d-1$  es elegido.

**Lema 4.2** Algoritmo de Kruskal devuelve el árbol con el peso máximo.

**Demostración Lema 4.2** El resultado puede ser probado por inducción. Evidentemente, es cierto para dos nodos. Supongamos que es cierto para los nodos  $d$ , y considerar un conjunto de  $d+1$  nodos, etiquetados  $(X_1, X_2, \dots, X_{d+1})$  donde se ordenan de manera que para cada  $j=1, \dots, d+1$ , el árbol máximo de  $(X_1, X_2, \dots, X_j)$  da el árbol máxima de cualquier selección de  $j$  nodos de todo el conjunto de  $d+1$  nodos.

Sea  $b_{(i,j)}$  denota el peso de la arista  $(i, j)$  para  $1 \leq i < j \leq d+1$ . Las aristas serán consideradas no dirigidas. Sea  $T_j^{d+1}$  denota el árbol máximo obtenido por seleccionar  $j$  nodos de los  $d+1$  y considerar  $T_{d+1}^{(d+1)}$ .

Sea  $Z$  denota el nodo hoja en  $T_{d+1}^{(d+1)}$  tal que entre todos los nodos hoja en  $T_{d+1}^{(d+1)}$ . La arista  $(Z, Y)$  en  $T_{d+1}^{(d+1)}$  tiene el menor peso. Removiendo el nodo  $Z$  dado los arboles máximos en  $d$  nodos del conjunto de  $d+1$  nodos. Esto se ve de la siguiente manera. Claramente, no hay árbol con mayor peso que se pueden formar con estos  $d$  nodos, de lo contrario el árbol en  $d$  nodos con mayor peso con la adición de la hoja  $(Z, Y)$  sería un árbol en  $d+1$  nodos con un peso mayor que  $T_{d+1}^{d+1}$ . De ello resulta que  $Z = X_{d+1}$  y por lo tanto  $X_{d+1}$  es un nodo hoja de  $T_{d+1}^{(d+1)}$ .

Por hipótesis inductiva,  $T_d^{(d+1)}$  puede ser obtenido por el algoritmo de Kruskal a los pesos

$(b_{(i,j)})_{1 \leq i < j \leq d}$ . Ahora considere una aplicación del algoritmo de Kruskal para la pesos

$(b_{(i,j)})_{1 \leq i < j \leq d+1}$  y tenga en cuenta que para cualquier  $(i, j)$  con  $i < j$  tal que  $(X_i, X_j)$  la

arista no dirigido forma parte del árbol  $T_d^{(d+1)}$ ,  $b_{(i,d+1)} < b(i, j)$  y  $b_{(j,d+1)} < b(i, j)$ . Por lo

tanto, si las aristas  $(b_{(i,j)})_{1 \leq i < j \leq d+1}$  se enumeran de acuerdo a su peso y la prueba del

algoritmo de Kruskal fue aplicado, entonces todos las aristas utilizados en  $T_d^{(d+1)}$

aparecerá más arriba en la lista de cualquier arista  $\left(b_{(k,d+1)}\right)_{k=1}^d$  y por lo tanto todas las aristas  $T_d^{(d+1)}$  serán incluidos por el algoritmo antes de las aristas  $\left(b_{(k,d+1)}\right)_{k=1}^d$  son consideradas. De ello resulta que  $T_{d+1}^{(d+1)}$  es el grafo obtenido mediante la aplicación de algoritmo de Kruskal para los nodos  $(X_1, \dots, X_{d+1})$ .

**Aproximación predictiva del árbol de Chow Liu.** La aproximación predictiva considera la distribución a priori y un conjunto de parámetros, y lo utiliza para la construcción de la distribución posterior, condicionado en las observaciones.

Esto se interpreta como la función de verosimilitud de la estructura del árbol y se utiliza como parte de una predicción técnica para el árbol óptimo Chow-Liu.

$$p_{\underline{x}}(\cdot|T) = p_{X_1} \prod_{j=2}^d p_{X_j|X_{m(j)}}$$

Sea

$$\theta_j = p_{X_j|X_{m(j)}}(1|1) \quad j = 2, \dots, d$$

$$\phi_j = p_{X_j|X_{m(j)}}(1|0) \quad j = 2, \dots, d$$

y

$$\theta_1 = p_{X_1}(1)$$

Consideremos ahora  $t$  completa, instancias independientes  $\underline{x}_1, \dots, \underline{x}_t$  de las variables en el árbol de red, donde para cada  $j = 1, \dots, t$  el vector fila  $\underline{x}_j = (x_{j1}, \dots, x_{jd})$  denota instanciación  $j$ . sea la matriz:

$$x = \begin{pmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_t \end{pmatrix}$$

Denota el conjunto completo de instancias independientes y sea  $X$  la matriz aleatoria, donde cada fila  $\underline{X}_j$ ,  $j=1, \dots, t$  es una copia independiente de  $\underline{X}=(X_1, \dots, X_d)$ .

Entonces

$$\begin{aligned} p_X(x|T) &= \prod_{l=1}^t p_{\underline{X}}(x_l | T) \\ &= \theta_1^{n_1} (1-\theta_1)^{t-n_1} \prod_{j=2}^d \theta_j^{n_j(1,1)} (1-\theta_j)^{n_j(0,1)} \phi_j^{n_j(1,0)} (1-\phi_j)^{n_j(0,0)} \end{aligned}$$

Para  $j \geq 2$ ,

$$\begin{aligned} n_j(1,1) &= \sum_{l=1}^t x_{lj} x_{l,m(j)}, & n_j(1,0) &= \sum_{l=1}^t x_{lj} (1-x_{l,m(j)}) \\ n_j(0,1) &= \sum_{l=1}^t (1-x_{lj}) x_{l,m(j)}, & n_j(0,0) &= \sum_{l=1}^t (1-x_{lj}) (1-x_{l,m(j)}) \end{aligned}$$

y

$$N_j(1) = \sum_{l=1}^t x_{lj}, \quad N_j(0) = t - N_j(1) \quad j=1, \dots, d$$

Sea

$$\begin{aligned} n &= \left\{ \left( N_j(0), N_j(1) \right)_{j=1}^d, n_j(1,1)_{j=2}^d, n_j(1,0)_{j=2}^d, n_j(0,1)_{j=2}^d, n_j(0,0)_{j=2}^d \right\} \\ N_{m(j)}^d(0) &= n_j(0,0) + n_j(1,0) \quad \text{y} \quad N_{m(j)}^d(1) = n_j(1,1) + n_j(0,1) \end{aligned}$$

La interpretación de estas cantidades está claro, por ejemplo,  $n_j(1,1)$  cuenta el número de filas de  $x$  en el que la configuración de la familia  $(x_{kj}, x_{k,m(j)}) = (1,1)$  aparece. Considerado como una función de  $(\underline{\theta}, \underline{\phi})$  dónde  $\underline{\theta} = (\theta_1, \dots, \theta_d)$  y  $\underline{\phi} = (\phi_2, \dots, \phi_d)$ , esta es la función verosímil:

$$L(\underline{\theta}, \underline{\phi}; x) = \prod_{l=1}^t p_{\underline{X}}(x_l | T)$$

$$= \theta_1^{N_1} (1-\theta_1)^{t-N_1} \prod_{j=2}^d \theta_j^{n_j(1,1)} (1-\theta_j)^{n_j(0,1)} \phi_i^{n_i(1,0)} (1-\phi_i)^{n_i(0,0)} \quad (4.5)$$

Cuando los parámetros  $(\underline{\theta}, \underline{\phi})$  son incluidos en la notación, la distribución de probabilidad puede ser escrito como:

$$p_{X|\theta, \phi}(x) \prod_{j=1}^d p_{X_j|X_{m(j)}, \theta_j, \phi_j}(x_j | x_{m(j)}).$$

Para calcular la distribución predictiva de un árbol fijo, una distribución a priori  $g(\underline{\theta}, \underline{\phi})$  se requiere sobre el espacio de parámetros. Luego, usando  $T$  para indicar la estructura del árbol, la distribución a priori predictiva para  $X$

$$p_X(x|T) = \int \prod_{l=1}^t p_{X_l|\theta, \phi}(x_l) g(\underline{\theta}, \underline{\phi}) d\underline{\theta} d\underline{\phi}$$

Donde  $\Theta \times \Phi$  denota el espacio de parámetros de  $(\underline{\theta}, \underline{\phi})$ . Es conveniente elegir

$$g(\underline{\theta}, \underline{\phi}) = \prod_{i=1}^d h(\theta_i) \prod_{i=2}^d k(\phi_i)$$

## *Capítulo 5*

### *Aplicación de Redes Bayesianas discretas*

#### **5.1 Introducción.**

La Red Bayesiana es un modelo grafo acíclico dirigido, para la construcción de la red se hará uso del paquete gRain (Graphical Independence Network) es una implementación en R. La construcción de redes en gRain es únicamente para variables discretas. Adicionalmente al programa R se considera el software Hugin Lite que al igual que el programa R tiene limitaciones en cuanto a la cantidad de variables.

Para el desarrollo del quinto capítulo, la base de datos llamada “signosvital” tiene 12 variables relacionado con la enfermedad de la diarrea con 462 registros de pacientes de los hospitales de la ciudad de La Paz. Los cuales mencionamos a continuación:

1. Sales de Rehidratación oral (SRO) (recibioSRO)
2. Consistencia (consistencia)
3. Vómitos (vomitos)
4. Convulsiones (convulsiones)
5. Plan de Hidratación (planhidratacion)
6. Estado de Hidratación (estadohidratacion)
7. Acidosis Metabólica (acidosismetabolica)
8. Fiebre (fiebre)
9. Desequilibrio Electrolítico (desequilibrioelectrolitico)
10. Hipokalemia (hipokalemia)
11. Hiponatremia (hiponatremia)
12. Método de medición. (metododemedicion)

Como mencionamos anteriormente las Redes Bayesianas es una técnica para analizar un número elevado de variables por tal motivo al inicio del trabajo se propuso analizar 30

variables con 462 registros, pero la limitación fue por no contar con una maquina con la capacidad de memoria suficiente, fue esta la razón por lo que se decidió reducir la cantidad de variables para el análisis y esto se hizo posible con la ayuda de un medico, ya que el mismo selecciono las variables de acuerdo a su importancia, variables para el análisis de la diarrea.

**Diarrea.** La diarrea<sup>4</sup> es una infección del intestino producida por virus o bacterias. Se define como la disminución de la consistencia de la materia fecal por mayor contenido de agua y con aumento del número de deposiciones diarias, generalmente mayor de tres veces en 24 horas. Sin embargo, la disminución de la consistencia es incluso más importante que la frecuencia, esta enfermedad es causada por muy diversos patógenos bacterianos y parásitos intestinales. Un individuo puede padecer uno a cinco veces anuales, dependiendo del área geográfica y la situación socioeconómica a la que pertenezca.

La diarrea puede acompañarse de náuseas y vómitos. Si la diarrea persiste en el tiempo o es muy copiosa puede conllevar a la deshidratación, por tal motivo se debe consumir sales de Rehidratación Oral (SRO) en el caso de la misma podría tener diferentes complicaciones debido a la falta de minerales y vitaminas que consumimos al ingerir o beber un alimento contaminado.

## **5.2 Descripción de las variables.**

Las variables seleccionadas son un tanto desconocidas en nuestro ámbito por tanto es necesario hacer las definiciones de cada una de ellas, como sigue a continuación:

- 1. Las sales de rehidratación oral (SRO).** Deben ser administradas solo en el tratamiento de la deshidratación, en la infancia, pacientes vulnerables y ancianos. Sin embargo aquellos adultos sanos que son capaces de ingerir líquidos comunes vía oral, no requieren de las mismas.
- 2. Vómitos.** Es la expulsión de contenido gástrico por la boca de forma espontánea o inducida.

---

<sup>4</sup> Diarrea y estreñimiento en books.google.com

3. **Convulsiones.** Son movimientos tónico tónico es decir movimientos involuntarios que no obedecen a nuestro cerebro.
4. **Consistencia.** En el caso de la diarrea la consistencia se refiere a las heces que no tienen forma y la persona acude con frecuencia a sus deposiciones.
5. **Estado de Hidratación.** Está relacionada con el tratamiento, para una deshidratación leve tenemos el plan A, deshidratación moderada el plan B y para una deshidratación grave tenemos plan C.
6. **Plan de Hidratación.** De acuerdo al tratamiento en Bolivia, se clasifica en tres planes:
  - Plan A: Para prevenir la deshidratación y la desnutrición. Se aplica en pacientes con diarrea aguda NO deshidratados.
  - Plan B: Para tratar una deshidratación moderada y de acuerdo a su peso se proporciona la dosis por vía oral, por cada kilo son 50 ml.
  - Plan C: Para tratar una deshidratación severa o grave en este caso ya no se proporciona sales de hidratación oral, por ser una deshidratación grave es tratada a través de la vena con suero.
7. **Fiebre.** La fiebre es el aumento de la temperatura corporal si la misma está por encima de un parámetro normal establecido, hay varias maneras de medir la fiebre.
  - Vía oral con un termómetro cuya ojiva es más larga, la temperatura normal es 37.2° C
  - Vía Axilar. Con un termómetro cuya ojiva es más larga, la temperatura normal es 37.2° C
  - Vía Rectal 38° C con un termómetro cuya ojiva es pequeña.
8. **Acidosis Metabólica.** La Acidosis Metabólica es un Desequilibrio de pH (Potencial de Hidrogeno), cada parte de nuestro cuerpo tiene un pH diferente El valor del pH se puede medir de forma precisa mediante un potenciómetro, también conocido como pH-metro y su rango está de 1 a 14 donde el parámetro normal es un pH 7.4. esto es medido en la sangre. Cuando el Ph es elevado

significa que el organismo ha acumulado demasiado ácido y no tiene suficiente bicarbonato para neutralizar con eficacia los efectos del ácido.

- 9. Desequilibrio Electrolítico.** Es ocasionado por la pérdida de sodio, potasio y cloro. Pero el componente más importante en este caso es el sodio y si esta aumenta por encima de su valor normal se llama hipernatremia y si está por debajo de su valor normal se llama hiponatremia. Y si el potasio esta por encima de su valor normal se llama se llama hiperkalemia y si está por debajo se llama hipokalemia. Básicamente el desequilibrio electrolítico observa el balance entre el agua y los electrolíticos y a su vez los electrolíticos son el sodio y el potasio y la pérdida o el desbalance de las mismas es ocasionada por la diarrea y los vómitos.
- 10. Hipokalemia.** La hipokalemia es la falta de potasio en el cuerpo, puede ser causada por varias anomalías. Entre las más frecuentes están: diarrea, quemaduras, disminución de la ingesta, enfermedades intestinales.
- 11. Hiponatremia.** Es la falta o abundancia de sodio en el cuerpo. La hiponatremia puede producir daño cerebral permanente, demencia y muerte, pese a su sencilla detección hay muchos casos con hiponatremia grave.
- 12. Método de medición.** Son los tipos de medición de la fiebre ya mencionadas anteriormente en la variable fiebre.

Una vez habiendo definido las variables, procedemos a obtener el modelo de la red apriori con paquete deal que es una implementación en R.

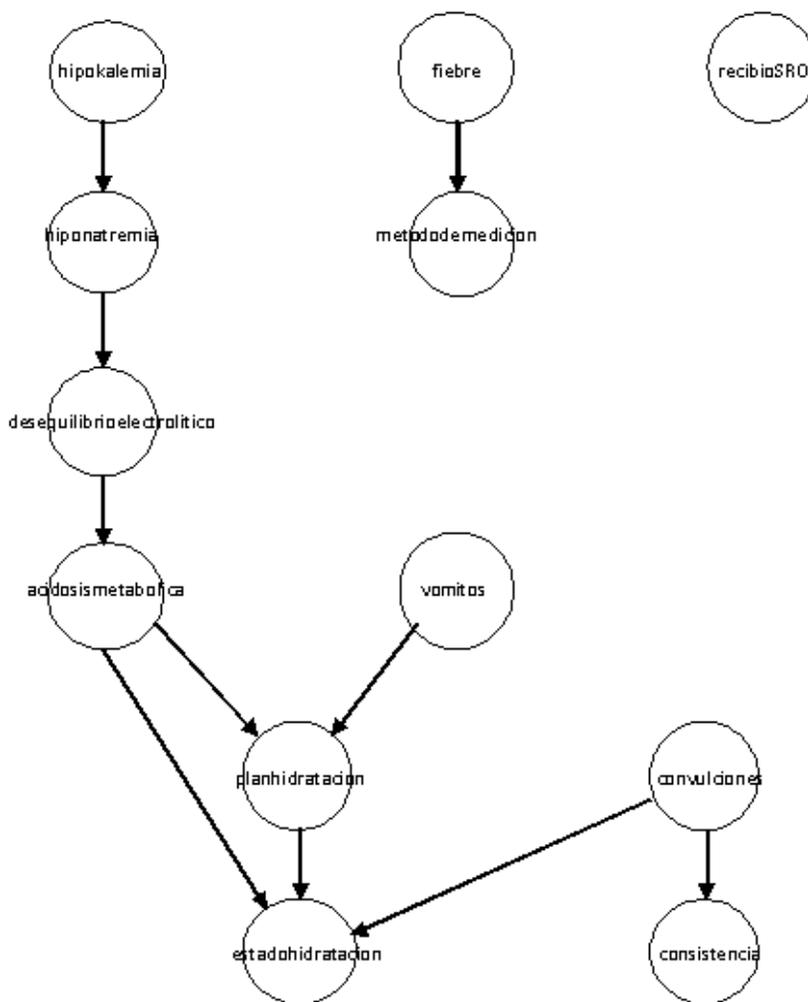
g<-

```
[~Hipokalemia,~Hiponatremia|Hipokalemia,~Desequilibrioelectrolitico|Hiponatremia,~Acidosismetabolica|Desequilibrioelectrolitico,~Planhidratacion|Acidosismetabolica:Vomitos,~Estadohidratacion|Planhidratacion:Convulsiones,~Consistencia|Convulsiones,~Fiebre,~Metododemedicion|Fiebre,~RecibioSRO].
```

La cantidad de redes apriori generadas con las 12 variables es de “5.21939651343829e+26 networks”. Como se mencionó anteriormente en el capítulo 4, esta cantidad crece de manera exponencial. Las probabilidades apriori de las redes

obtenidas ya sea en el programa R o Hugin Lite por defecto son probabilidades uniformes, cuando se realiza la propagación con el algoritmo Chow Liu se obtiene las probabilidades a posteriori y las probabilidades condicionales, de esta manera se observa las relaciones directas e indirectas que tienen las variables o nodos. En el capítulo 3 se describe el procedimiento para la construcción de una unión de arboles. A continuación se desarrolla este procedimiento:

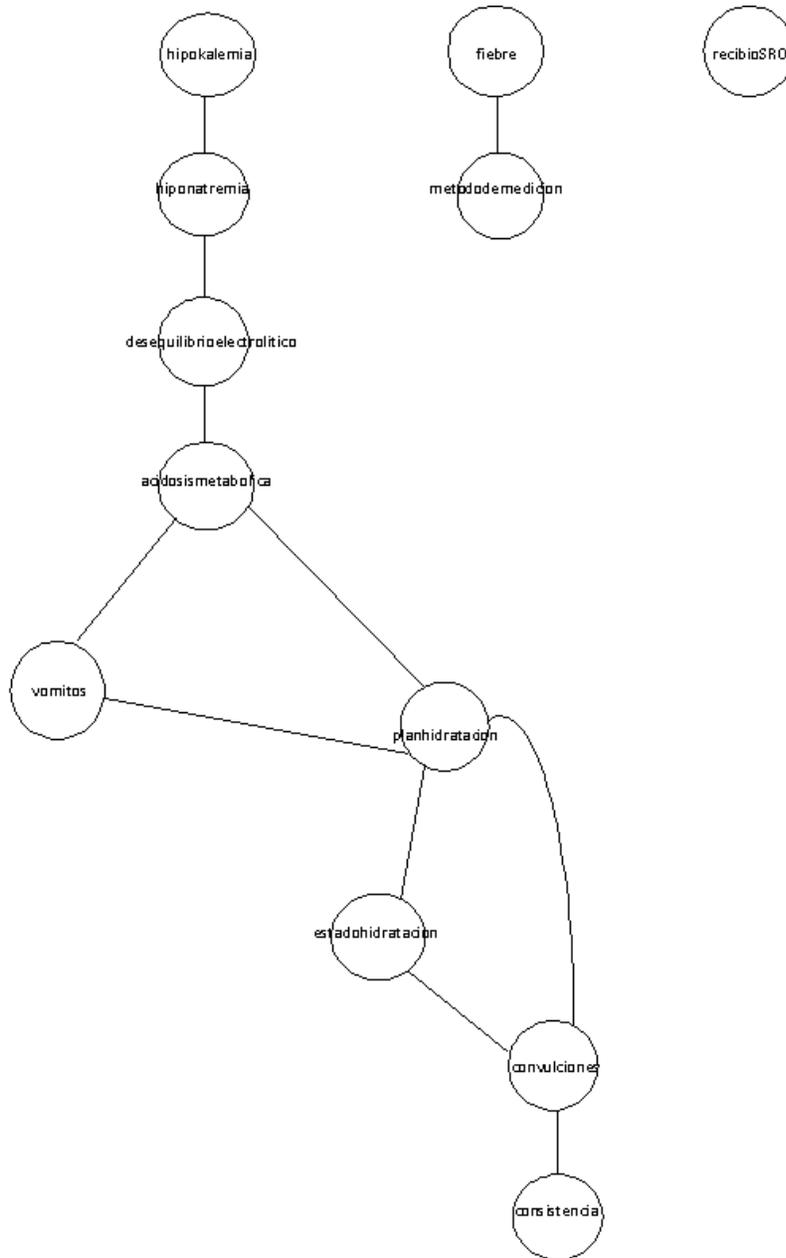
Para obtener el gráfico del modelo mencionado anteriormente 'g' se aplica el paquete gRain que es una implementación de R:



**Figura 5.1 Red apriori con el programa R**

Desarrollo del procedimiento para obtener una unión de arboles.

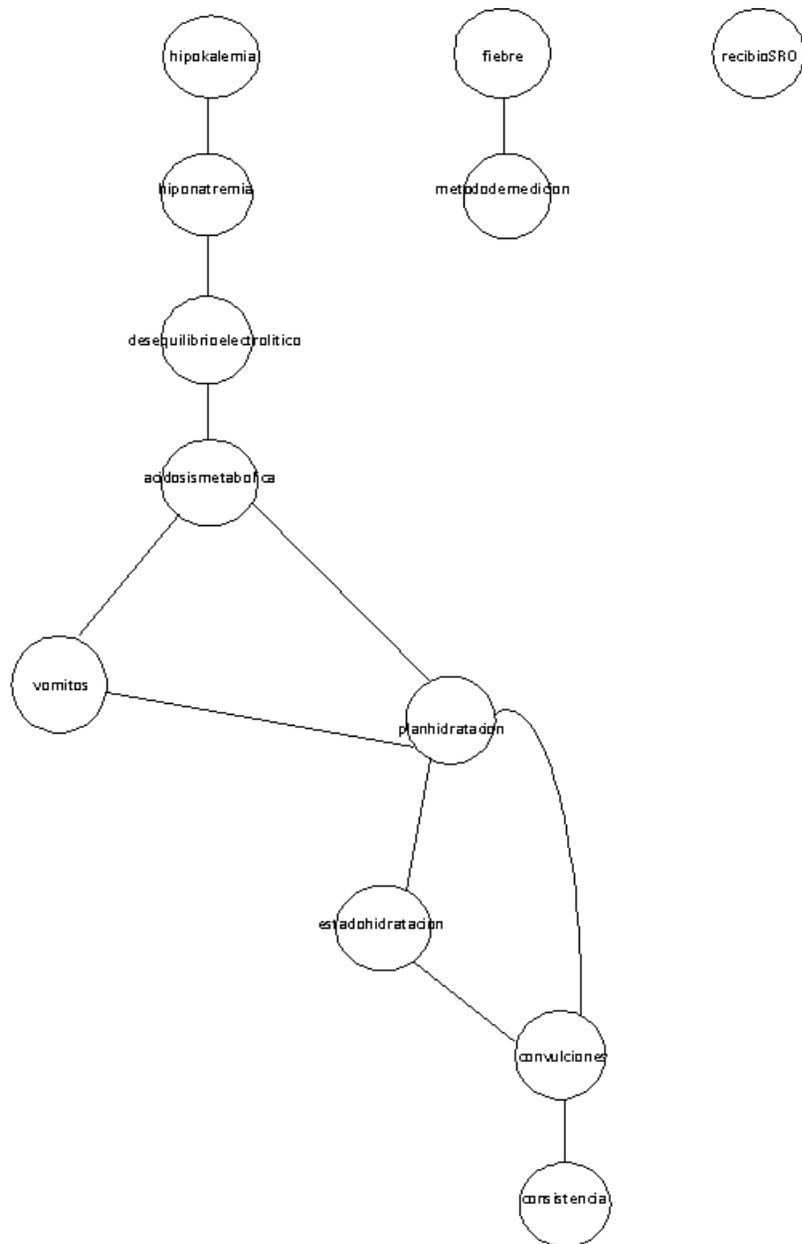
**Primero.-** Debemos moralizar el grafo aciclico dirigido de la red bayesiana.



**Figura 5.2 Grafo Moralizado con el programa R**

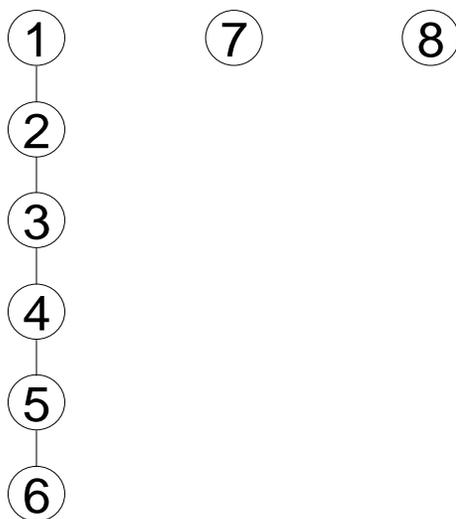
Es el grafo moralizado y cómo se observa tenemos un grafo acíclico no dirigido, Posteriormente se triangula el grafo de la siguiente manera:

**Segundo.-** Triangular el grafo moralizado obtenido



**Figura 5.3** Grafo triangulado con el programa R

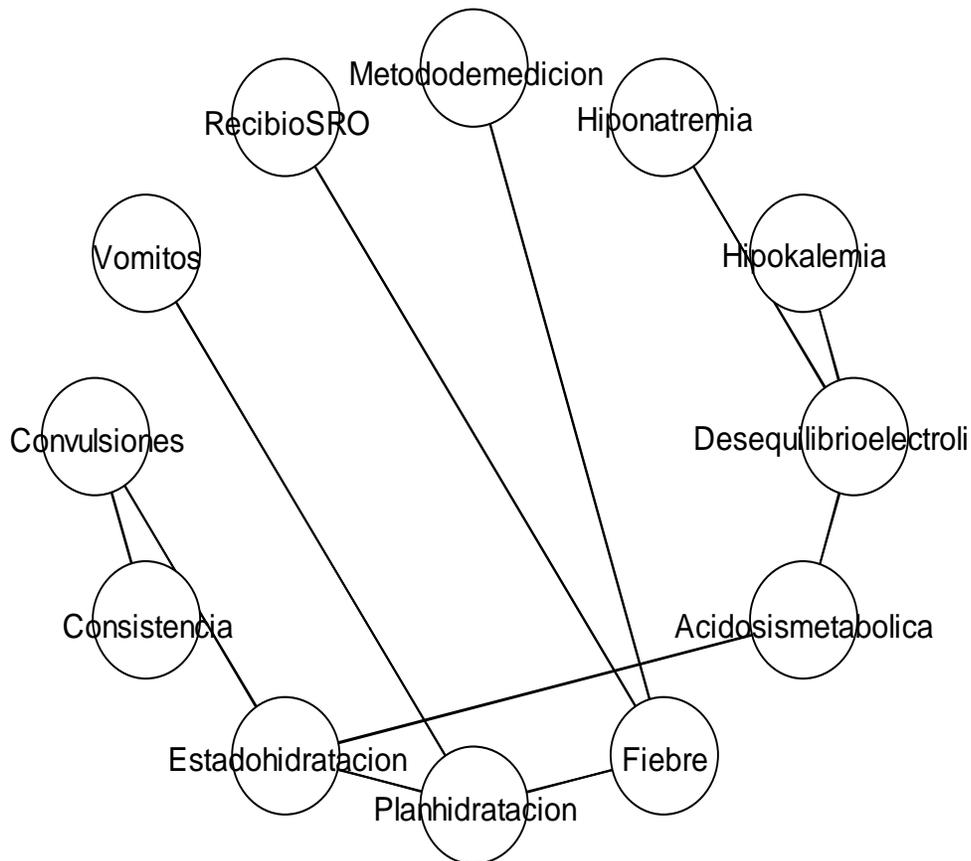
**Tercero.-** Habiendo obtenido el grafo moralizado y el grafo triangulado obtenemos los cliques del grafo triangulado, que es una unión de árboles.



**Figura 5.4 Cliques o grupo de nodos**

En la Figura 5.4, cada nodo representa un conjunto de nodos, este análisis se hizo con las probabilidades a priori, cada nodo representa la relación que tienen los diferentes nodos o variables, el nodo 1 conforma los nodos Hiponatremia e Hipokalemia, el nodo 2 conforma los nodos Desequilibrio electrolítico e Hiponatremia, el nodo 3 conforma los nodos Acidosis Metabólica y Desequilibrio Electrolítico, el nodo 4 conforma los nodos Convulsiones, Acidosis metabólica, Estado de Hidratación y Plan de Hidratación, el nodo 5 conforma los nodos Vómitos, Acidosis metabólica y Planhidratación, el nodo 6 conforma los nodos Consistencia y Convulsiones, el nodo 7 conforma los nodos Método de medición y Fiebre, en el nodo 8 conforma el nodo Sales de Rehidratación oral (SRO), se observa las posibles relaciones de manera a priori pero la mejor relación y estructura lo veremos con el algoritmo de Chow Liu, las Probabilidades Posteriores y

condicionales ser obtendrán con el programa Hugin Lite ya que esta fue la única manera de obtener las probabilidades requeridas.



**Figura 5.5** Árbol de Chow Liu con R

La figura 5.5 se obtuvo con el programa R es la misma relación que tenemos en la Figura 5.6, como mencionamos anteriormente hubo dificultades con el programa R para realizar diferentes salidas, por esta situación los resultados de probabilidades que tenemos en la Figura 5.6 fue a partir del programa Hugin Lite. De acuerdo a la definición del algoritmo de Chow Liu este algoritmo proporciona un grafo acíclico no

dirigido pero el software Hugin Lite proporciona por defecto un Grafo Aciclico Dirigido, esto simplemente para un mejor entendimiento de la relación que tienen las variables.

### 5.3 Resultados con el software Hugin Lite.

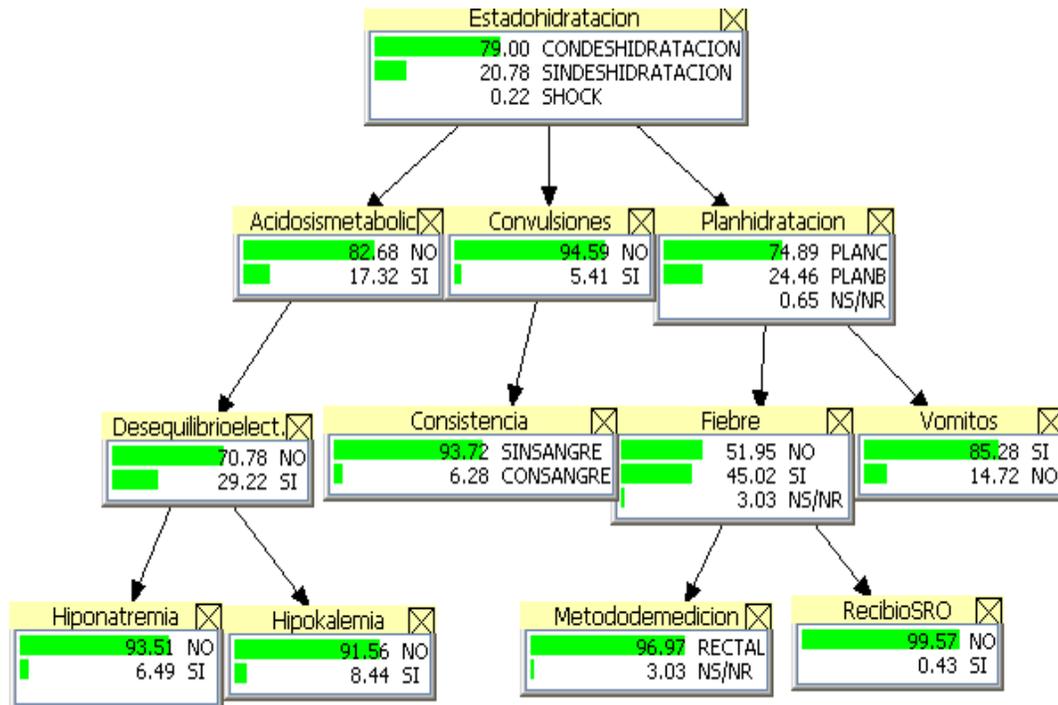


Figura 5.6 Árbol de Chow Liu con Hugin Lite

De acuerdo a los resultados del algoritmo del árbol de Chow Liu con Hugin Lite podemos observar que el 79% de la población en estudio, se encontraba con estado de deshidratación mientras que un 0.22% se encontraba en un estado de shock, la Acidosis Metabólica tiene una relación directa con el estado de Deshidratación, como se observa el 82.68% de los pacientes no tiene un pH menor o mayor al parámetro normal es por esta razón observamos un resultado negativo alto, en caso de tener un porcentaje alto en el estado de shock la Acidosis Metabólica daría positivo lo cual indicaría que el pH de los pacientes tendrían cierto Desequilibrio ya sea mayor o menor del parámetro normal, a su vez el Desequilibrio Electrolítico tiene una relación directa con la Acidosis

Metabólica, como el resultado de la Acidosis Metabólica dio negativo en un 82.68% el Desequilibrio Electrolítico debe ser también negativo como observamos en el resultado un 70.78% de los pacientes no han tenido un Desequilibrio Electrolítico lo que significa que no tuvieron una gran pérdida de Sodio, Potasio y Cloro pero esto a su vez indica que en pacientes no habría problemas de Hiponatremia, que es la falta o abundancia de Sodio en el cuerpo o Hipokalemia que es la falta de Potasio en el cuerpo, como podemos observar los resultados tenemos una respuesta negativa considerable en ambos casos con un 93.51% y 91.56% respectivamente.

El Estado de Hidratación tiene una relación directa con las convulsiones donde el 94.59% de los pacientes no han tenido convulsiones, cabe mencionar que las convulsiones se dan en cualquier caso, no importa el tipo de enfermedad que se esté tratando, pero en nuestro caso podemos observar que las Convulsiones tienen algún tipo de relación debido a que el 94.59% de los pacientes indican que no tuvieron convulsiones, durante su enfermedad lo cual indica que durante sus deposiciones, los pacientes no tuvieron pérdida de sangre y esto se puede observar en un 93.72% de los pacientes.

El Plan de Hidratación tiene una relación directa con el Estado de Hidratación, el 71.84% de los pacientes se encuentran con un tratamiento de un plan C, lo que significaría que están en una situación riesgosa debido a que este tratamiento ya no se realiza con las sales de hidratación oral sino que se lo hace con sondas en las venas y un 51.95% indica que ha tenido fiebre durante el tratamiento y la medición de la fiebre se hizo vía rectal lo que no es muy común en las personas mayores. Durante el tratamiento del Plan de Hidratación no se hizo a través del suero de rehidratación oral como ya observamos anteriormente la mayoría de los pacientes estaban en un estado crítico debido a que un 74.89% se encontraban con el tratamiento de un plan C es por esta razón que un 99.57% de los pacientes no siguieron el tratamiento con la sal de rehidratación oral pues era necesario seguir el tratamiento con suero, el vómito tiene una relación directa con el Plan de Hidratación ya que el 85.28% tuvo vómitos durante su enfermedad, por los resultados negativos de la Hipokalemia y la Hiponatremia podríamos afirmar que los vómitos no fueron con mucha frecuencia, pues cuando uno

expulsa el contenido gástrico por la boca expulsa muchos Electrolitos y por su puesto el Potasio y el Sodio que son factores muy importantes para nuestro salud.

#### 5.4 Relación de dependencia entre variables.

Las siguientes probabilidades condicionales fueron obtenidas con el programa Hugin Lite aquí se observa la relación que tiene los diferentes signos y síntomas vitales, durante la diarrea.

**Tabla 5.1.1 desequilibrio electrolíticosegún Hiponatremia**

Desequilibrio Electrolítico	Hiponatremia	
	no	si
No	1	0,777778
Si	0	0,222222
Experiencia	327	135

**Fuente: Elaboración propia**

De los pacientes que indican que no haber sufrido Desequilibrio Electrolítico también indican que no han tenido Hiponatremia, este resultado es de 327 pacientes de un total de 462 pacientes pero los 135 pacientes restantes dicen haber tenido Hiponatremia con un probabilidad del 0.778, lo cual no sería posible ya que el Desequilibrio Electrolítico es la causa directa de la Hiponatremia, es decir si uno no sufre de Desequilibrio Electrolítico tampoco debería sufrir Hiponatremia, por tanto debe hacerse un estudio minucioso de esta causa para saber la razón del porque está pasando esta anomalía o podría ser que los médicos dieron un incorrecto tratamiento.

**Tabla 5.1.2 Consistencia según Convulsiones**

	Convulsiones	
Consistencia	no	si
Sinsangre	0,947368	0,76
Consangre	0,052632	0,24
Experiencia	437	25

**Fuente: Elaboración propia**

Como ya habíamos mencionado las convulsiones son independientes ya que esta ocurre en diferentes enfermedades, pero en nuestro caso las convulsiones tiene relación con la consistencia, 437 pacientes dicen haber tenido convulsiones durante sus deposiciones la respuesta fue dada con una probabilidad 0.947 lo cual es significativamente alta.

**Tabla 5.1.3 Acidosis Metabolica según Desequilibrio Electrolítico**

	Desequilibrioelectrolítico	
Acidosis metabólica	No	Si
No	0,793194	0,3
Si	0,206806	0,7
Experiencia	382	80

**Fuente: Elaboración propia**

Si un paciente no sufre Desequilibrio Electrolítico no debería sufrir Acidosis Metabólica como se observa en la tabla 5.1.3, los pacientes dicen no haber sufrido de Desequilibrio Electrolítico por tanto dicen también no haber sufrido de Acidosis Metabólica esto ocurre con una probabilidad de 0.7931, mientras que los pacientes que afirman haber tenido Desequilibrio Electrolítico como también afirman haber tenido Acidosis Metabólica con una probabilidad de 0.7 y de acuerdo al medico el resultado es correcto, como se vio anteriormente ambos nodos tienen una relación directa.

**Tabla 5.1.4 Plan de Hidratacion según Estado de Hidratación**

Plan de hidratación	Estado de hidratación		
	Con deshidratación	Sin deshidratación	Shock
Plan C	0,854795	0,34375	1
Plan B	0,139726	0,645833	0
Ns/Nr	0,005479	0,010417	0
Experiencia	365	96	1

**Fuente: Elaboración propia**

De los pacientes que dicen haber sufrido de deshidratación indican que estaban con un tratamiento de plan C esto ocurre con una probabilidad de 0.85, mientras que un paciente que se encontraba en un estado de shock indica que estaba con un tratamiento de plan C lo cual no es el tratamiento indicado cuando un paciente se encuentra en esa situación, pues ese tipo de tratamiento no hace ningún efecto y necesariamente requiere seguir un tratamiento con suero. De los pacientes que siguieron un tratamiento con el plan B y dicen no haber sufrido deshidratación están con una probabilidad de 0.64, esto podría indicar que estas personas estaban con una deshidratación leve es por esta razón que ellos indicaron que no tenían deshidratación. Mencionar una vez más, que el plan B es para pacientes que tienen deshidratación moderada por tanto no es grave.

**Tabla 5.1.5 Vomitos según Plan de Hidratacion**

Vomitos	Plan de hidratación		
	Plan C	Plan B	Ns/Nr
Si	0,884393	0,778761	0
No	0,115607	0,221239	1
Experiencia	346	113	3

**Fuente: Elaboración propia**

De los pacientes que recibieron un tratamiento de plan C dicen haber sufrido de vómitos con un probabilidad de 0.8843 esto ocurre en 346 pacientes, mientras que los que recibían un tratamiento de plan B tuvieron vómitos con una probabilidad 0.7787, este

número también es significativo, lo cual el medico debería hacer un estudio minucioso de lo que está pasando, esto podría indicar que ellos sufren de otra anomalía debido a que no es común que las pacientes con deshidratación leve sufran de vómitos.

**Tabla 5.1.6 Convulsiones según Estado de Hidratación**

Convulsiones	Estado de hidratación		
	Con deshidratación	Sin deshidratación	Shock
No	0,983562	0,802083	1
Si	0,016438	0,197917	0
Experiencia	365	96	1

**Fuente: Elaboración propia**

En el tabla 5.1.6 de un total de 365 pacientes con deshidratación respondieron no haber sufrido convulsiones con una probabilidad de 0.9853 y de los pacientes que no tenían deshidratación de igual manera respondieron no haber sufrido convulsiones con una probabilidad de 0.80 y de los pacientes con deshidratación y sin deshidratación que si sufrieron de convulsiones fue mínima debido a las probabilidades obtenidas 0.016 y 0.197 respectivamente. Incluso aquella persona que estaba en un estado de shock no tuvo convulsiones. Lo que afirma las conclusiones del médico, donde las convulsiones pasan sin importar que enfermedad se esté tratando, pero en el caso de la diarrea las convulsiones es un caso raro.

**Tabla 5.1.7 Método de medición según Fiebre**

Método de medición	Fiebre		
	No	Si	Ns/Nr
Rectal	1	1	0
Ns/Nr	0	0	1
Experiencia	240	208	14

**Fuente: Elaboración propia**

Uno de los mejores métodos para medir la fiebre es Vía Rectal, pero este método no es utilizado con frecuencia en los Hospitales Públicos o Centros de Salud, como podemos ver en tabla 5.1.7, todos los pacientes que tenían y no tenían fiebre dicen haberse medido la fiebre Vía Rectal, que es el método más efectivo en este caso.

**Tabla 5.1.8 Fiebre según Plan de Hidratación**

	Plan de hidratación		
Fiebre	Plan C	Plan B	Ns/Nr
No	0,508571	0,557522	0,333333
Si	0,465318	0,415929	0
Ns/Nr	0,026012	0,026549	0,666667
Experiencia	346	113	3

**Fuente: Elaboración propia**

En tabla 5.1.8 los pacientes que están con un plan de hidratación C de un total de 346 pacientes dicen haber tenido y no tenido fiebre con una probabilidad de 0.50 y 0.46 respectivamente, donde podemos concluir que no necesariamente debemos estar deshidratados para tener fiebre, como en el caso de plan B del total de 113 pacientes indican no haber tenido fiebre con una probabilidad de 0.55 y 0.41 respectivamente, casi podemos afirmar que la fiebre es independiente de la diarrea debido a las probabilidades obtenidas.

**Tabla 5.1.9 Hiponatremia según Desequilibrio Electrolítico**

	Desequilibrio electrolítico	
Hiponatremia	No	Si
No	1	0,711111
Si	0	0,288889
Experiencia	327	135

**Fuente: Elaboración propia**

En el tabla 5.1.9 los pacientes que decían no sufrir Desequilibrio electrolítico tampoco sufrían de hipokalemia con una probabilidad segura. Mientras que los pacientes que no padecían de desequilibrio dicen haber tenido hipokalemia lo cual no es posible, como ya mencionamos anteriormente la hipokalemia es una consecuencia del desequilibrio electrolítico que es el balance entre el agua y los electrolíticos y a su vez los electrolíticos son el sodio y el potasio y la pérdida o el desbalance de las mismas, por tanto es necesario hacer un análisis del porqué la probabilidad es 0.71, ver que es lo que está pasando o que tratamiento se está brindando a los pacientes.

**Tabla 5.1.10 Estado de Hidratación según Acidosis Metabólica**

Estado de hidratación	Acidosis metabólica		
	Con deshidratación	Sin deshidratación	shock
Si	0,783552	1	0
No	0,216438	0	1
Experiencia	365	96	1

**Fuente: Elaboración propia**

Los pacientes que estaban con deshidratación indican no haber sufrido acidosis metabólica, la respuesta fue con una probabilidad 0.78 y los que sufrían de acidosis metabólica la respuesta fue con una probabilidad de 0.216. el paciente que se encontraba en un estado de shock indica haber tenido acidosis metabólica; lo cual es posible, debido a la pérdida de bastantes defensas durante sus deposiciones.

**Tabla 5.1.11 Estado de Hidratación**

	Estado de hidratación
Con deshidratación	0,790043
Sin deshidratación	0,207792
Shock	0,002165
Experiencia	462

**Fuente: Elaboración propia**

En el tabla 5.1.10 de un total de 462 pacientes obtuvimos una probabilidad de 0.79, indicando que si estaban en un estado de deshidratación, mientras que la probabilidad de 0.20 representa a las personas que estaban sin deshidratación o deshidratación muy leve y por ultimo las personas que se encontraban en un estado de shock es un valor muy pequeño 0.0021. Lo que se concluye que los pacientes estaban con cierta gravedad pero no en una situación crítica.

En conclusión los pacientes que son objetos de estudio, no se encuentran en una situación complicada ya que la mayoría no tiene complicaciones como la perdida de electrolitos. Tampoco tiene problemas con los vómitos y lo más sorprendente es el método que prefirieron para saber si tenían o no tenían fiebre (Vía Rectal), este método como se mencionó anteriormente no se practica en la Hospitales Públicos y/o Centros de Salud. Algo que podemos resaltar también es el hecho de no aplicarse las sales de rehidratación oral como normalmente se hace o se conoce en el caso de la diarrea.

## *Capítulo 6*

### *Conclusiones Recomendaciones y*

#### **6.1 Conclusiones.**

Después de haber realizado el trabajo de Redes bayesianas discretas se llega a las siguientes conclusiones:

- La teoría de las redes bayesianas discretas permitio construir la estructura o topología de la s variables de estudio, como se planteo en los objetivos específicos.
- Esta técnica permite analizar la relación directa e indirecta que existe entre las variables.
- La propagación de probabilidades es importante por que permite actualizar los resultados si se conoce una nueva evidencia
- El algoritmo de Chow Liu permite estimar las probabilidades con una gran cantidad de parámetros.
- El tema está abierto a mucha más investigación ya que existen diversos tipos de redes probabilistas que por supuesto son temas de investigacion bastante importante en el campo de la estadística.

#### **6.2 Recomendaciones.**

Se recomienda los siguientes aspectos para futros trabajos:

- Muchas instituciones internacionales opinan que el análisis probabilístico es una de las mejores técnicas de análisis que permite estimar las probabilidades con gran cantidad de parámetros. Se recomienda continuar con el tema de redes bayesianas para enriquecer nuestros conocimientos con temas nuevos e importantes en el área científica.
- La interpretación de las redes bayesianas es necesario hacerlo con la ayuda de un experto en el área de investigación respectivo.

## *Apéndice A*

### *Redes bayesianas con R*

Uno de las herramientas más usadas para el análisis de redes bayesianas dentro del programa R es el paquete `deal`<sup>5</sup>. Nos permite realizar el análisis de variables discretas y continuas, el paquete incluye procedimientos para obtener todas las combinaciones de redes posibles y la red a priori con la mejor, se mencionó anteriormente que las 12 variables genera una cantidad aproximada de redes  $5.21939651343829e+26$ , este tipo de análisis se ve limitado en grandes cantidades de variables si el investigador no cuenta con un equipo de computación de última generación, lo que dificulta el análisis de investigaciones complejas. El paquete `deal` puede ser descargado de la página <http://cran.R-project.org/>, dentro del paquete se encuentra las diferentes funciones como ser `network`, `joint prior`, entre otros que son necesarios para el análisis de redes discretas, para obtener todas las posibles redes a priori a continuación definiremos lo siguiente :

**df** : Es una base de datos para obtener todas las posibles redes a priori (signos vitales)

**nw**: Es un objeto clase red

**Specifygraph, inspectprob, doprob**: objetos definidos dentro de R para clasificar las variables como continuos o discretos.

```
network
function (df, specifygraph = FALSE, inspectprob = FALSE, doprob = TRUE,
  yr = c(0, 350), xr = yr)
{
  if (length(dim(df)) < 1)
    stop("Can't handle networks with one node, sorry\n")
  nw <- list()
  nw$n <- ncol(df)
```

---

<sup>5</sup>Ihaka and Gentleman, 1996 "R"

```

nw$discrete <- c()
nw$continuous <- c()
nw$nodes <- list()
unit <- 2 * pi/nw$n
xc <- mean(xr)
yc <- mean(yr)
  for (i in 1:nw$n) {
    pos <- c(cos(unit * i + pi/4), sin(unit * i + pi/4)) *
      xc * 0.8 + c(xc, yc)
    if (is.factor(df[, i])) {
      nw$nodes[[i]] <- node(i, c(), "discrete", names(df)[i],
        length(levels(df[, i])), levels(df[, i]), position = pos)
      nw$discrete <- c(nw$discrete, i)
    }
    else {
      nw$nodes[[i]] <- node(i, c(), "continuous", names(df)[i],
        position = pos)
      nw$continuous <- c(nw$continuous, i)
    }
  }
nw$nd <- length(nw$discrete)
nw$nc <- length(nw$continuous)
stopifnot(nw$nd + nw$nc == nw$n)
names(nw$nodes) <- names(df)
class(nw) <- "network"
if (specifygraph) {
  nw <- drawnetwork(nw, nocalc = TRUE)$nw
}
if (doprob)
  nw <- prob(x = nw, df = df)
if (inspectprob)
  nw <- inspectprob(nw)
nw

```

A continuación la función `jointprior` nos permite conocer las probabilidades a priori de todas las combinaciones de redes obtenidas con la función `network`. Por tanto definiremos lo siguiente:

**nw**: Es un objeto clase red.

**N**: proporciona una base de datos imaginario.

**Phiprior** : es una cadena que especifica como el a priori phi es calculado

**Timetrace**: si es verdadero, muestra en la pantalla algunas de las redes generadas de manera a priori.

`Jointprior`

```

function (nw, N = NA, phiprior = "bottcher", timetrace = FALSE)
{
  if (timetrace) {
    t1 <- proc.time()
    cat("[Jointprior ")
  }
  if (nw$nd > 0) {
    jointprob <- jointdisc(nw, timetrace = timetrace)
    minN <- min(2/jointprob)
    if (is.na(N))
      N <- minN
    if (N < minN) {
      cat("Warning: Your choice of imaginary sample size is very
low\n")
      cat("We advise you to set the imaginary sample size to more
than",
          minN, "\n")
    }
    cat("Imaginary sample size:", N, "\n")
    jointalpha <- jointprob * N
    jointnu <- jointalpha
    jointrho <- jointalpha
  }
  else {
    jointnu <- N
    jointrho <- N
    jointalpha <- N
  }
  if (nw$nc > 0) {
    NN <- prod(dim(jointalpha))
    if (nw$nd > 0) {
      Dim <- dim(jointalpha)
      dparents <- nw$discrete
      lvek <- c()
      for (i in 1:NN) {
        cf <- findex(i, Dim, FALSE)
        label <- ""
        for (j in 1:ncol(cf)) {
          label <- paste(label,
nw$nodes[[dparents[j]]]$levelnames[cf[1,
j]], sep = ":")
        }
        lvek <- c(lvek, label)
      }
    }
    jointmu <- matrix(NA, NN, nw$nc)
    jointsigma <- list()
    jointphi <- list()
    jcont <- jointcont(nw, timetrace = timetrace)
    jointmu <- jcont$mu
    jointsigma <- jcont$sigma2
    dnames <- colnames(jointmu)
    for (i in 1:NN) {
      if (hiprior == "bottcher") {
        jointphi[[i]] <- jointsigma[[i]] * (jointnu[i] -
1)
      }
      else {

```

```

        if (phiprior == "heckerman") {
            jointphi[[i]] <- (jointrho[i] - 2)/(jointnu[i] +
                1) * jointnu[i] * jointsigma[[i]]
        }
        else stop("No such phiprior implemented")
    }
    colnames(jointmu) <- dnames
    colnames(jointsigma[[i]]) <- dnames
    rownames(jointsigma[[i]]) <- dnames
    colnames(jointphi[[i]]) <- dnames
    rownames(jointphi[[i]]) <- dnames
}
if (nw$nd > 0) {
    names(jointsigma) <- lvek
    names(jointphi) <- lvek
    rownames(jointmu) <- lvek
}
}
else {
    jointphi <- NA
    jointmu <- NA
    jointsigma <- NA
}
if (timetrace) {
    t2 <- proc.time()
    cat((t2 - t1)[1], "\n")
}
list(jointalpha = jointalpha, jointnu = jointnu, jointrho = jointrho,
    jointmu = jointmu, jointsigma = jointsigma, jointphi = jointphi)
}

```

**La función Heuristic permite obtener redes sin ciclos.**

```

function (initnw, data, prior = jointprior(network(data)), maxiter = 100,
    restart = 10, degree = size(initnw), trylist = vector("list",
    size(initnw)), trace = TRUE, timetrace = TRUE, removecycles =
FALSE)
{
    if (timetrace) {
        t1 <- proc.time()
        cat("[Heuristic ")
    }
    if (timetrace)
        s1 <- proc.time()[3]
    nwl <- autosearch(initnw, data, prior, maxiter, trylist,
        trace = trace, timetrace = TRUE, removecycles = removecycles)
    nw <- nwl$nw
    trylist <- nwl$trylist
    table <- nwl$table
    if (timetrace) {
        s2 <- proc.time()[3]
        sauto <- s2 - s1
        spert <- 0
        suniq <- 0
    }
    if (restart > 0) {
        for (i in 1:restart) {

```

```

    if (timetrace)
      s3 <- proc.time()[3]
    nw <- perturb(initnw, data, prior, degree, trylist = trylist,
      timetrace = TRUE)
    trylist <- nw$trylist
    nw <- nw$nw
    ms <- modelstring(nw)
    if (timetrace) {
      s4 <- proc.time()[3]
      spert <- spert + s4 - s3
    }
    if (!is.na(match(ms, table[, 1])))
      next
    table <- rbind(table, cbind(ms, nw$score))
    if (trace) {
      plot(nw)
      title("New network")
    }
    if (timetrace)
      s5 <- proc.time()[3]
    newnwl <- autosearch(nw, data, prior, maxiter, trylist =
trylist,
      trace = trace, timetrace = TRUE, removecycles =
removecycles)
    trylist <- newnwl$trylist
    table <- rbind(table, newnwl$table)
    if (timetrace) {
      s6 <- proc.time()[3]
      sauto <- sauto + s6 - s5
    }
    if (timetrace)
      s7 <- proc.time()[3]
    table <- table[!duplicated(table[, 1]), ]
    table <- table[sort.list(-as.numeric(table[, 2])),
]
    if (timetrace) {
      s8 <- proc.time()[3]
      suniq <- suniq + s8 - s7
    }
  }
}
if (initnw$n < 15)
  antal <- paste(numbermixed(initnw$nc, initnw$nd))
else antal <- "many"
cat("Tried", nrow(table), "out of approx.", antal, "networks\n")
if (timetrace) {
  t2 <- proc.time()
  cat((t2 - t1)[1], "]\n")
  cat("Perturb:", spert, ",Autosearch:", sauto, ",Unique:",
  suniq, "\n")
}
thebest <- as.network(table[1, ], initnw)
thebest <- learn(thebest, data, prior)$nw
list(nw = thebest, table = table, trylist = trylist)
}

autosearch(initnw,data,prior=jointprior(network(data)),maxiter=50,
  trylist= vector("list",size(initnw)),trace=TRUE,

```

```

        timetrace=TRUE, showban=FALSE, removecycles=FALSE)

heuristic(initnw, data, prior=jointprior(network(data)),
          maxiter=100, restart=10, degree=size(initnw),
          trylist= vector("list", size(initnw)), trace=TRUE,
          timetrace=TRUE, removecycles=FALSE)

g<-print(modelstring(getnetwork(nw))

```

El siguiente procedimiento es para hallar la unión de arboles.

El procedimiento es el siguiente: instalar los diferentes paquetes descargadas de la dirección <http://cran.R-project.org/>

```

utils:::menuInstallLocal()
> library(BiocGenerics)
> library(graph)
> library(MASS)
> library(RBGL)
> library(Rcpp)
> library(RcppEigen)
> library(igraph)
> library(gRbase)
> library(gRain)
Dag<-dagList(g)
Plot(dag)
moralize(dag)
triangulate(moralize(dag))
clique(triangulate(moralize(dag)))

```

El arbol de Chow Liu se obtiene con el commando `chow.liu(signosvital)`, es la base de datos de variables discretos y por ultimo `print(chow.liu(signosvital))`.

## *Apéndice B*

### *Software Hugin Lite*

El programa Hugin Lite ha existido desde 1989 y fue un software desarrollado para la inteligencia artificial y decisiones avanzadas en modelos estadísticos complejos como en el diagnóstico médico y análisis de riesgo. El software Hugin Lite emplea los algoritmos exactos más eficientes para actualizar las probabilidades.

Una red bayesiana consiste en un número de eventos, donde los eventos son variables aleatorias que es conectado con una arista para describir la relación entre las variables o la posible dependencia entre las mismas, por ejemplo una arista podría describir una relación de dependencia entre el síntoma causada por la enfermedad, así las aristas pueden ser usadas para representar la relación causa efecto.

Las redes bayesianas pueden y han sido usadas como componentes bajo incertidumbre en investigaciones grandes y complejos como es el área científica.

El software Hugin Lite puede ser instalado en Microsoft Windows, el programa puede ser descargado desde el internet ya que la misma es un software libre a su vez la misma puede interactuar con el programa R con algunos comandos y librerías descargas desde la página <http://cran.R-project.org/>.

Para esta investigación el programa Hugin fue necesario para obtener las probabilidades apriori y posteriori no fue posible hacerlo con el programa R debido a que no contábamos con el equipo de computación necesario para este tipo de trabajo.

El manejo de este programa es amigable ya que no hay dificultad en el aprendizaje del software y esta disponible en la página <http://www.hugin.com>.

## Bibliografía

- [1] Castillo E. Sistemas Expertos y Modelos de Redes Probabilísticas, Universidad de Cantabria y Universidad de Cornell, España.
  
- [2] Carmen L.R (2002) Explicación en Redes Bayesianas Causales, Aplicaciones Medicas, Departamebto de Inteligencia artificial, universidad Nacinal de Educacion a distancia, Madrid.
  
- [3] Daniel G.G. (2010) Desarrollo de un entorno de usuario para aplicación de redes bayesianas dinámicas a problemas de fusión de información, Madrid
  
- [4] García R. S. (2007) Análisis de Sensibilidad en Redes Bayesianas Gaussianas, Universidad Complutense De Madrid, Madrid.
  
- [5] Jensen F. and Nielsen T. (2007) Bayesian Network and Decision Graphs, Springer, Berlin, Heidelberg, New York, Hong Kong, London, Milan, Paris, Tokyo.
  
- [6] JoanM.X (2009) El Estreñimiento y la Diarrea, Amat S.L. Barcelona.
  
- [7] Koski T. and Noble J. (2009) Bayesian Network An Introduction, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19, United Kingdom.
  
- [8] K. R. Koch (2007) Introduction to Bayesian Statistics Second edition, Verlag Berlin Heidelberg.

[9] Matilde I.C. Apuntes Redes Bayesianas.

[10] Niels J. Hugin Expert White Paper.

[11] Sucar L. E. Redes Bayesianas, Puebla México.

[12] Susanne G. B. A Package for Learning Bayesian Networks, Aalborg.

[13] Steffen L. (2012) Graphical Models with R, Springer, New York.