

# Uso de factores para inferencia exacta en redes bayesianas de más de diez nodos

Verónica Esther Arriola Ríos

Facultad de Ciencias

UNAM

v.arriola@ciencias.unam.mx

## 1. Introducción

El uso de inferencia en redes bayesianas en sistemas inteligentes que trabajan bajo incertidumbre es ya una práctica común. Los sistemas comerciales que hacen uso de ellos requieren trabajar con redes de cientos de nodos. Para resolver problemas sobre ellas, en la práctica, no basta con aplicar manualmente el teorema de Bayes en su formulación para dos variables. En este escenario el uso de la estructura matemática *factores* resulta una herramienta indispensable, que permite sistematizar los cálculos numéricos involucrados, facilitando la implementación de los algoritmos en programas para inferencia. En este artículo revisaremos la definición y uso de esta estructura, además del análisis de dependencias e independencias probabilísticas encontradas en redes bayesianas, que permitirán reducir el número de operaciones requeridas.

La inferencia en redes de Bayes ha sido aplicada a un gran número de campos. Por ejemplo, en computación cognitiva, el proyecto BAM-BI modeló los procesos cognitivos bajo incertidumbre con inferencia probabilística iniciando por la escala bioquímica, y luego con cascadas de señales celulares que podrían realizar los cálculos probabilísticos necesarios [4]. En genómica, [2] estudiaron los efectos de las variaciones genómicas entre genes humanos en la regulación de la expresión de los genes para diferentes tejidos utilizando las técnicas presentadas en [5, 9] para analizar la expresión de los genes en múltiples tejidos con un modelo bayesiano jerárquico. Así mismo, los modelos bayesianos son aplicados frecuentemente al tratamiento de señales obtenidas de sensores como en [3] o incluso para programar robots que aprenden de su medio ambiente [8].

El método para realizar inferencia exacta, que expondré a continuación, surgió del curso en línea *Probabilistic Graphical Models* con Daphne Koller de la Universidad de Stanford [6], quien a su vez tiene un libro publicado en el tema [7]. [1] es también una referencia relevante.

## 2. El método

El método a exponer a continuación permite:

Resolver y programar algoritmos de inferencia exacta en redes de Bayes con  $n$  nodos, donde  $n \gg 10$ , realizando optimizaciones a las fórmulas manualmente.

Existen también algoritmos para realizar las optimizaciones de manera automática, pero eso queda fuera del alcance de este artículo.

### 2.1 Prerrequisitos

Esta exposición presupone familiaridad con los conceptos y habilidades siguientes:

**Sumas  $\sum$ :** Para trabajar ágilmente con estos temas es muy importante trabajar con fluidez con sumas utilizando el símbolo  $\sum$ , multiplicar y factorizar tomando en cuenta estos términos para trabajar con las operaciones de marginalización y producto de probabilidades, como se verá más adelante. Para quienes han tenido poca experiencia con esta notación, es común caer en errores al simplificar divisiones.

**Antecedentes de probabilidad y estadística:** Asumimos familiaridad con los problemas de estimación de probabilidad mediante muestreo simple, como los de bolsas de dulces utilizados para ilustrar la definición de probabilidad y primeros métodos de cálculo aproximado por muestreo; así como con los conceptos de evento, espacio de eventos, muestras, universo, tamaño de un evento, ilustración con diagramas de Venn, variables aleatorias discretas, distribución de probabilidad, los axiomas de la probabilidad, probabilidad condicional, independencia, independencia condicional y teorema de Bayes.

**Notación:** Es importante también hacer notar la relación entre algunas formas de notación utilizadas a partir de la lógica computacional, que son originalmente inconsistentes con las definiciones y notación utilizada en las definiciones formales de probabilidad, pero que se hayan frecuentemente en textos de inteligencia artificial como [10].

Para formalizar su compatibilidad, establecemos que al escribir  $P(\neg lluvia)$  nos referimos a la probabilidad de que la proposición  $\neg lluvia$  sea verdadera o, que la proposición  $lluvia$  sea falsa; también, que  $Lluvia$  es una variable aleatoria. Dado que las variables aleatorias tienen como rango a los reales, formalmente definimos un mapeo del espacio de los símbolos que tienen significado a los reales, para que nuestro uso de la probabilidad esté bien justificado. Escribiremos los nombres de las variables aleatorias comenzando con mayúscula y sus valores posibles con minúsculas. En este ejemplo en que la variable se corresponde con una proposición, se trata de una variable binaria y podemos establecer que si llueve  $Lluvia = 1$  y si no llueve  $Lluvia = 0$ , entonces la probabilidad anterior se escribiría formalmente como  $P(Lluvia = 0)$ ; pero las otras notaciones son utilizadas por conveniencia. Se pueden tener variables aleatorias de más valores, por ejemplo, digamos que la variable aleatoria  $Clima$  puede tomar los valores  $\{soleado, lluvioso, nublado\}$ , en este caso estos valores podrían ser mapeados a los reales:  $\{soleado \rightarrow 0, lluvioso \rightarrow 1, nublado \rightarrow 2\}$ ; para decir que la variable toma uno de estos valores podemos abreviar  $P(Clima = soleado)$ ; si es evidente por el contexto que trabajamos con la variable  $Clima$  podemos abreviar  $P(soleado)$  o  $P(lluvioso)$ , etc.

## 2.2 Factores

**Definición 2.1** (Factor). Sea  $\mathbf{D} = \{X_1, \dots, X_k\}$  un conjunto de variables aleatorias, donde cada  $X_i$  toma valores de un conjunto  $\mathbf{D}_{X_i}$ ; y el conjunto de todas las combinaciones posibles de asignaciones a las variables en  $\mathbf{D}$  es  $Val(\mathbf{D})$ . Definimos al factor  $\phi$  como una función de  $Val(\mathbf{D})$  a  $\mathbb{R}$ . El conjunto de variables  $\mathbf{D}$  es denominado el *alcance* del factor y se denota  $Alcance[\phi]$ . [7]

Se utilizarán factores para representar las distribuciones de probabilidad, distribuciones de probabilidad conjuntas y condicionales, en lugar de las tablas de dos dimensiones que utilizan otros libros como [10]. El objetivo de introducir a los factores será el automatizar las operaciones de marginalización, normalización y reducción de probabilidades, de modo que será posible programarlas en una clase de cualquier lenguaje orientado a objetos, en lugar de desarrollarlas manualmente. En este sentido, en ocasiones es útil relacionar este concepto con el de tablas de verdad que se utilizan en lógica, pues también se asocia un valor a cada posible combinación de asignaciones a un conjunto de variables; en lo que respecta a la estructura, se puede comparar a los factores tanto con las tablas de verdad como con las matrices.

La figura 1 muestra un primer ejemplo de factor y su uso para representar una distribución de probabilidad discreta. A la izquierda se muestra un factor que corresponde a la distribución de probabilidad condicional  $P(Lluvia|Estación)$ . Obsérvese cómo, para cada posible asignación al par  $(Estación, Lluvia)$ , el factor asigna la probabilidad correspondiente a ese evento. Tratándose de una distribución de probabilidad condicional, aquellos renglones para los cuales las variables evidencia tienen el mismo valor suman uno.

Nótese que la definición formal de factor no establece algún orden particular para las variables en su alcance o para los renglones que lo constituyen, por lo que, desde el punto de vista teórico es posible realizar permutaciones. Sin embargo, al momento de programarlos la implementación será más eficiente si se establece alguna convención para el orden. Esto permite no almacenar en memoria los valores correspondientes a las variables en cada renglón, pues estos quedarán implícitos por la convención; solo será necesario almacenar las variables en el alcance y los valores asociados. Digamos, por ejemplo, que se fija el orden de las variables  $X_i$  en el alcance al orden en el que fueron especificadas, asociando un índice a esta posición; que los valores posibles de cada variable  $D_{x_i}$  también se listan en un orden fijo con su índice correspondiente y que las posibles combinaciones de todas ellas  $Val(\mathbf{D})$  llevarán el orden creciente de un reloj digital, se asocia otro índice a cada una de estas combinaciones y se indican los valores  $\phi(x_1, \dots, x_k)$  que les corresponden asignándoles el índice correspondiente. En ese caso es posible utilizar un *polinomio de direccionamiento* para obtener en tiempo  $\mathcal{O}(k)$ <sup>1</sup> la posición  $P(x_1, \dots, x_k)$  en la que se encuentra el valor asociado. El polinomio tiene la forma siguiente:

$$P(x_1, \dots, x_k) = \sum_{i=0}^{k-1} \left( \prod_{j=i+1}^{k-1} |D_{X_j}| \right) pos(x_i) \\ = ((pos(x_0)|D_{X_1}| + pos(x_1)|D_{X_2}| + \dots)|D_{X_k}| + pos(x_k)),$$

donde  $x_1, \dots, x_k$  son los valores que toman las variables y  $pos(x_i)$  es el índice asignado a ese valor dentro de la lista de valores posibles. En este punto hacemos explícito que  $|Val((D))| = \prod_{i=0}^k |D_{X_i}|$ , lo cual indica el número total de renglones que tiene el factor. Considerando que  $|Val((D))| \gg k$  el tiempo que toma determinar así la posición de cada asignación es aceptable.

**Definición 2.2** (Reducción). De acuerdo con [7]:

Sea  $\phi(\mathbf{Y})$  un factor y  $\mathbf{U} = \mathbf{u}$  una asignación para las variables  $\mathbf{U} \subseteq \mathbf{Y}$ . Definimos la *reducción* del factor  $\phi$  al contexto  $\mathbf{U} =$

<sup>1</sup>Siendo  $k$  el número de variables en el alcance.

**Variables aleatorias:**

$$D_{Estaci3n} = \{Primavera, Verano, Oto\~{n}o, Invierno\}$$

$$D_{Lluvia} = \{0, 1\}$$

|                 |               | <b>Factor</b>        |  |                |
|-----------------|---------------|----------------------|--|----------------|
| <i>Estaci3n</i> | <i>Lluvia</i> | $P(Lluvia Estaci3n)$ |  |                |
| Primavera       | 0             | 0.75                 |  | } $\Sigma = 1$ |
| Primavera       | 1             | 0.25                 |  |                |
| Verano          | 0             | 0.30                 |  | } $\Sigma = 1$ |
| Verano          | 1             | 0.70                 |  |                |
| Oto\~{n}o       | 0             | 0.70                 |  | } $\Sigma = 1$ |
| Oto\~{n}o       | 1             | 0.30                 |  |                |
| Invierno        | 0             | 0.8                  |  | } $\Sigma = 1$ |
| Invierno        | 1             | 0.2                  |  |                |

**Tabla bidimensional**

|                 |               | $P(Lluvia Estaci3n)$ |              |  |
|-----------------|---------------|----------------------|--------------|--|
| <i>Estaci3n</i> | <i>Lluvia</i> |                      |              |  |
|                 | 0             | 1                    |              |  |
| Primavera       | 0.75          | 0.25                 | $\Sigma = 1$ |  |
| Verano          | 0.30          | 0.70                 | $\Sigma = 1$ |  |
| Oto\~{n}o       | 0.70          | 0.30                 | $\Sigma = 1$ |  |
| Invierno        | 0.8           | 0.2                  | $\Sigma = 1$ |  |

**Figura 1.** Uso de un factor para representar la distribuci3n de probabilidad condicional para las variables *Estaci3n* y *Lluvia* (izquierda), en contraste con una tabla bidimensional (derecha). Se muestra al factor con las variables de su alcance en las columnas de la izquierda, las combinaciones de su valores posibles  $D_{Estaci3n} \times D_{Lluvia}$  se enumeran expl3citamente debajo; los valores asociados a cada una de ellas se listan en la columna de la derecha. Aunque en un inicio la tabla bidimensional es m3s f3cil de leer, el factor es inmediatamente generalizable a  $n$  dimensiones, mientras que la tabla se volver3 cada vez m3s confusa.

$\mathbf{u}$ , denotado  $\phi[\mathbf{U} = \mathbf{u}]$  (y abreviado  $\phi[\mathbf{u}]$ ), como el factor sobre el alcance  $\mathbf{Y}' = \mathbf{Y} - \mathbf{U}$ , tal que

$$\phi[\mathbf{u}](\mathbf{y}') = \phi(\mathbf{y}', \mathbf{u}).$$

Para  $\mathbf{U} \not\subset \mathbf{Y}$ , definimos que  $\phi[\mathbf{u}]$  sea  $\phi[\mathbf{U}' = \mathbf{u}']$ , donde  $\mathbf{U}' = \mathbf{U} \cap \mathbf{Y}$  y  $\mathbf{u}' = \mathbf{u}\langle \mathbf{U}' \rangle$ , donde  $\mathbf{u}\langle \mathbf{U}' \rangle$  denota a las asignaciones en  $\mathbf{u}$  a las variables en  $\mathbf{U}'$ .

En la figura 2 se utiliza la operaci3n de reducci3n para seleccionar solamente aquellos renglones que indican la probabilidad de lluvia dado cuando es primavera. La complejidad de esta operaci3n es proporcional al n3mero de renglones que permanecen en el factor.

| <i>Estación</i>  | <i>Lluvia</i>  | $P(\textit{Lluvia}, \textit{Estación})$ |
|------------------|--|---|
| <i>Primavera</i> | 0  | 0.1875                                  |
| <i>Primavera</i> | 1  | 0.0625                                  |
| Verano           | 0  | 0.075                                   |
| Verano           | 1  | 0.175                                   |
| Otoño            | 0  | 0.175                                   |
| Otoño            | 1  | 0.075                                   |
| Invierno         | 0  | 0.2                                     |
| Invierno         | 1  | 0.05                                    |
| $\sum = 1$       |  |   |
| $\Rightarrow$    |  |   |
| <i>Lluvia</i>    | $P(\textit{Lluvia}, \textit{Estación} = \textit{primavera})$ |   |
| 0                | 0.1875   |   |
| 1                | 0.0625   |   |
| $\sum = 0.25$    |  |   |

**Figura 2.** Ejemplo de reducción  $\textit{Estación} = \textit{Primavera}$ . El factor resultante solo conserva aquellos renglones donde la variable reducida tiene el valor indicado, la variable reducida desaparece del alcance. Esta operación requeriría además una normalización de los valores en el factor resultante para ser equivalente a la operación de reducción en probabilidad.

**Definición 2.3** (Normalización). Dado un factor  $\phi$  con  $n = |\textit{Val}(D)|$  renglones sea

$$\alpha = \sum_{i=1}^n \phi_i$$

con  $\phi_i$  el valor asociado al renglón  $i$ ,  $\alpha$  es la suma de los valores de todos los renglones. Entonces:

$$\textit{normalización}(\phi) = \frac{\phi}{\alpha}$$

donde el valor en cada renglón de  $\phi$  ha sido dividido entre  $\alpha$ .

Obsérvese en el ejemplo de la figura 3, que el alcance de ambos factores (antes y después de la normalización) solo es la variable  $\textit{Lluvia}$ , por lo que solo tienen dos renglones. Sin embargo, cada factor corresponde a distribuciones de probabilidad diferentes. El primero contiene dos renglones de la distribución de probabilidad conjunta  $P(\textit{Lluvia}, \textit{Estación})$ ; mientras que el otro, tras la normalización, contienen el de la condicional  $P(\textit{Lluvia} | \textit{Estación} = \textit{primavera})$ .

**Definición 2.4** (Marginalización). Según [7]:

Sea  $\mathbf{X}$  un conjunto de variables y  $Y \notin \mathbf{X}$  una variable. Sea  $\phi(\mathbf{X}, Y)$  un factor. Definimos la *marginalización* de  $Y$  del

|               |                                   |               |
|---------------|-----------------------------------|---------------|
| <i>Lluvia</i> | $P(Lluvia, Estación = primavera)$ |               |
| 0             | 0.1875                            | $\Rightarrow$ |
| 1             | 0.0625                            |               |
|               | $\sum = 0.25$                     |               |
| <i>Lluvia</i> | $P(Lluvia Estación = primavera)$  |               |
| 0             | 0.75                              |               |
| 1             | 0.25                              |               |
|               | $\sum = 1$                        |               |

**Figura 3.** Para normalizar los valores en un factor se suman los valores en todos los renglones y se divide cada valor entre esa suma. Esta operación se define para facilitar el trabajo con distribuciones de probabilidad.

|                 |               |                       |               |
|-----------------|---------------|-----------------------|---------------|
| <i>Estación</i> | <i>Lluvia</i> | $P(Lluvia, Estación)$ |               |
| Primavera       | 0             | 0.1875                | $\Rightarrow$ |
| Primavera       | 1             | 0.0625                |               |
| Verano          | 0             | 0.075                 |               |
| Verano          | 1             | 0.175                 |               |
| Otoño           | 0             | 0.175                 |               |
| Otoño           | 1             | 0.075                 |               |
| Invierno        | 0             | 0.2                   |               |
| Invierno        | 1             | 0.05                  |               |
|                 |               | $\sum = 1$            |               |

|               |             |
|---------------|-------------|
| <i>Lluvia</i> | $P(Lluvia)$ |
| 0             | 0.63750     |
| 1             | 0.36250     |
|               | $\sum = 1$  |

**Figura 4.** Ejemplo donde la variable *Estación* es marginalizada, que también se corresponde con la operación de marginalización en probabilidad. Cada renglón en el factor resultado es la suma sobre todos los valores posibles de *Estación*: *Primavera, Verano, Otoño e Invierno*, manteniendo a la variable *Lluvia* con valor constante.

factor  $\phi$ , denotada  $\sum_Y \phi$  como el factor  $\psi$  sobre  $\mathbf{X}$  tal que:

$$\psi(\mathbf{X}) = \sum_Y \phi(\mathbf{X}, Y).$$

Esto es, se suman aquellas entradas de la tabla donde los valores de  $\mathbf{X}$  coinciden, pero  $Y$  cambia. Como hay que recorrer todos los renglones solicitándolos en este orden, se requieren  $\mathcal{O}(k|Val(\mathbf{D})|)$  operaciones.

La figura 4 muestra el uso de la operación de marginalización para calcular la probabilidad  $P(Lluvia)$  a partir de la distribución de probabilidad conjunta  $P(Lluvia, Estación)$ .

**Definición 2.5** (Multiplicación). Según [7]:

Sean  $\mathbf{X}$ ,  $\mathbf{Y}$  y  $\mathbf{Z}$  tres conjuntos disjuntos de variables y  $\phi_1(\mathbf{X}, \mathbf{Y})$  y  $\phi_2(\mathbf{Y}, \mathbf{Z})$  dos factores. Definimos el *producto de factores*  $\phi_1 \times \phi_2$  como el factor  $\psi : Val(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \mapsto \mathbb{R}$  como sigue:

|     | <i>Estación</i> | <i>Lluvia</i> | $P(Lluvia Estación)$  |        |
|-----|-----------------|---------------|-----------------------|--------|
| (1) | Primavera       | 0             | 0.75                  |        |
| (2) | Primavera       | 1             | 0.25                  |        |
| (3) | Verano          | 0             | 0.30                  |        |
| (4) | Verano          | 1             | 0.70                  | ×      |
| (5) | Otoño           | 0             | 0.70                  |        |
| (6) | Otoño           | 1             | 0.30                  |        |
| (7) | Invierno        | 0             | 0.80                  |        |
| (8) | Invierno        | 1             | 0.20                  |        |
|     | $\sum = 1$      |               |                       |        |
|     | <i>Estación</i> | $P(Estación)$ |                       |        |
| (a) | Primavera       | 0.25          |                       |        |
| (b) | Verano          | 0.25          |                       | =      |
| (c) | Otoño           | 0.25          |                       |        |
| (d) | Invierno        | 0.25          |                       |        |
|     | $\sum = 1$      |               |                       |        |
|     | <i>Estación</i> | <i>Lluvia</i> | $P(Lluvia, Estación)$ |        |
|     | Primavera       | 0             | (1) × (a) =           | 0.1875 |
|     | Primavera       | 1             | (2) × (a) =           | 0.0625 |
|     | Verano          | 0             | (3) × (b) =           | 0.075  |
|     | Verano          | 1             | (4) × (b) =           | 0.175  |
|     | Otoño           | 0             | (5) × (c) =           | 0.175  |
|     | Otoño           | 1             | (6) × (c) =           | 0.075  |
|     | Invierno        | 0             | (7) × (d) =           | 0.2    |
|     | Invierno        | 1             | (8) × (d) =           | 0.05   |
|     | $\sum = 1$      |               |                       |        |

**Figura 5.** Este ejemplo muestra el cálculo de  $P(Lluvia, Estación)$  despejando de la definición de probabilidad condicional, de modo que:  $P(Lluvia, Estación) = P(Lluvia|Estación)P(Estación)$ . La multiplicación de  $P(Lluvia|Estación)$  por  $P(Estación)$  se realiza utilizando factores. Obsérvese el uso de las variables comunes a ambos factores y sus valores para determinar qué renglón se multiplica con cuál.

$$\psi(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \phi_1(\mathbf{X}, \mathbf{Y}) \cdot \phi_2(\mathbf{Y}, \mathbf{Z})$$

Es decir, para multiplicar dos factores  $\phi_1$  y  $\phi_2$  se multiplican aquellos renglones donde los valores de las variables en común  $\mathbf{Y}$  coinciden. Si la solución se construye recorriendo los renglones de  $\psi$ , solicitando los renglones correspondientes en  $\phi_1$  y  $\phi_2$  para calcularlos, la complejidad de la multiplicación es  $\mathcal{O}(k_\psi k_{\phi_1} k_{\phi_2} |Val(\mathbf{D}_\psi)|) = \mathcal{O}(|Val(\mathbf{D}_\psi)|^2)$ . La figura 5 ilustra el funcionamiento de esta operación.



### 2.3 Distribuciones de probabilidad

De entre las distribuciones de probabilidad, se enfatiza la importancia de la *tabla de distribución de probabilidad conjunta completa*, donde aparecen todas las *variables del sistema* y a partir de la cual se puede realizar cualquier consulta ( $P(X, Y, Z)$ ,  $P(X|Y, Z)$ ,  $P(Y)$ , etc). Destaca también el problema de la complejidad en tiempo y espacio si se quiere generar esta tabla para sistemas con muchas variables, pues el número de renglones que tiene este factor es el producto del número de valores posibles para cada variable en el sistema. Aún si el sistema utiliza solo variables binarias, el crecimiento de la tabla es exponencial como función del número de variables, pues el número de renglones es  $n = 2^d$  donde  $d$  es el número de variables.

Como se vio en los ejemplos del apartado anterior, las operaciones de marginalización, reducción, normalización y multiplicación en probabilidad se pueden implementar directamente con factores. Los factores pueden representar distribuciones de probabilidad con cualquier número de variables. Sin embargo, es importante recalcar que se puede operar del mismo modo con factores aunque no correspondan a distribuciones de probabilidad. En particular aparece la relación:

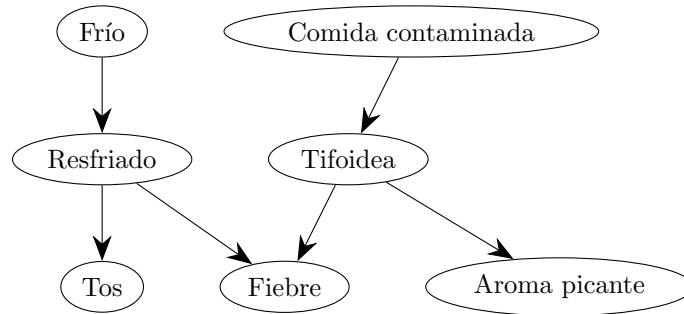
Reducción en probabilidad = Reducción en factores + Normalización

Utilizando esta fórmula es posible simplificar algunas operaciones al realizar inferencia, al posponer la normalización de factores reducidos. En medio de este proceso los factores podrían no estar representando una distribución de probabilidad, sino solo términos intermedios en el cálculo.

La regla de la cadena:

$$P(X_1, \dots, X_n) = P(X_1|X_2, \dots, X_n)P(X_2|X_3, \dots, X_n) \dots P(X_n) \quad (1)$$

será el punto de partida para el trabajo que vendrá a continuación con Bayes. Esta regla viabiliza el modelado de fenómenos mediante variables aleatorias y distribuciones de probabilidad, pues frecuentemente es más natural obtener estimaciones experimentales para las distribuciones de probabilidad condicional que para las conjuntas. En los ejemplos que se mostraron se puede ver que es más fácil modelar y obtener estadísticas para aproximar la probabilidad de lluvia dependiendo de la estación, que haber obtenido la distribución conjunta desde un inicio. Sin embargo, habiendo construido la conjunta es posible, mediante las operaciones ya definidas, recuperar cualquier otra conjunta o condicional; pues si se requiere la probabilidad conjunta de un subconjunto de variables basta con marginalizar de la conjunta completa aquellas



**Figura 6.** Construcción de una red de Bayes siguiendo relaciones de causalidad. El *Frío* puede provocar un *Resfriado*; *Comida contaminada*, *Tifoidea*. Síntomas típicos de un *Resfriado* son *Tos* y *Fiebre*; síntomas típicos de la *Tifoidea* son *Fiebre* y un *Aroma picante* en el enfermo. En un universo restringido a estas dos enfermedades, un paciente con *Fiebre* podría tener cualquiera de las dos, con distintas probabilidades.

variables que no son de interés:

$$P(X_1, \dots, X_n) = \sum_{Z_1, \dots, Z_m} P(X_1, \dots, X_n, Z_1, \dots, Z_m)$$

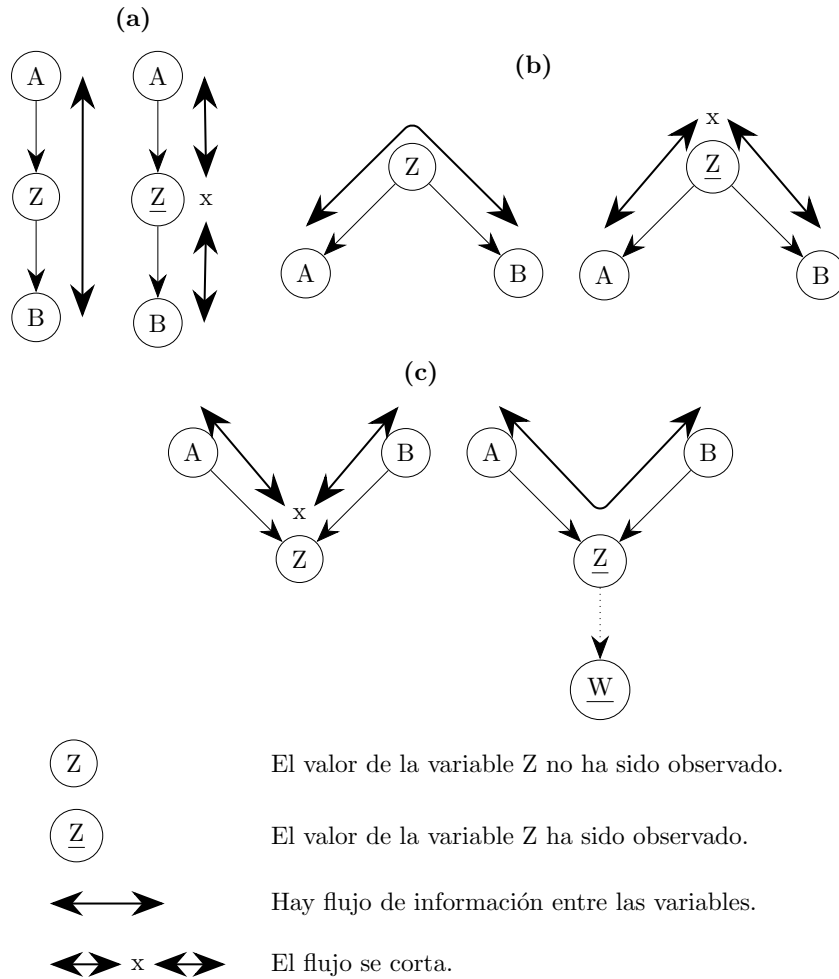
y para obtener una condicional se puede utilizar la definición:

$$P(X_1, \dots, X_n | Z_1, \dots, Z_m) = \frac{P(X_1, \dots, X_n, Z_1, \dots, Z_m)}{P(Z_1, \dots, Z_m)} \quad (2)$$

## 2.4 Gráficas bayesianas e independencia condicional

La técnica para modelar el comportamiento de sistemas utilizando redes de Bayes comienza por diseñar un grafo dirigido acíclico con las variables del sistema, escribiendo a los nodos que representan a variables *causa* en la parte de arriba y a sus *consecuencias* hacia abajo, de modo que *la gráfica refleje las relaciones de dependencia e independencia entre las variables de un dominio dado* [figura 6].

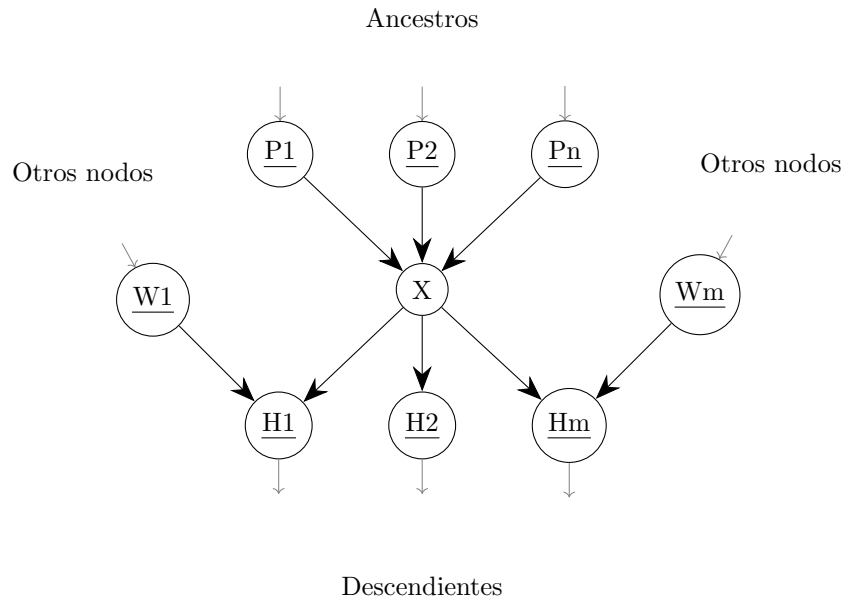
Sobre esta gráfica se estudian los conceptos de camino entre dos nodos (como si la gráfica no estuviera dirigida) y ruta activa e inactiva (tomando en cuenta la dirección de las aristas y el flujo de información probabilística entre las variables), dependiendo de las variables que han sido o no observadas. Se dice que una variable ha sido observada si estamos estudiando una probabilidad condicional. Las variables con respecto a las cuales se condiciona la consulta se dice que son variables *evidencia* y han sido *observadas*. Por ejemplo en  $P(A, B | Z, W)$  decimos que  $Z$  y  $W$  fueron observadas. Cuando no existe una ruta activa entre dos variables  $A$  y  $B$  afirmamos que son probabilísticamente independientes ( $P(A, B) = P(A)P(B)$ ); cuando la ausencia de



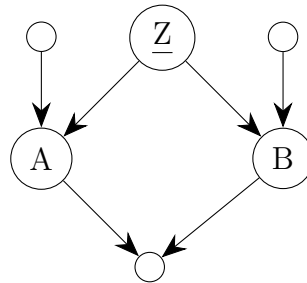
**Figura 7.** Reglas que indican cuándo fluye información probabilística entre dos nodos  $A$  y  $B$  separados por un tercero  $Z$ , es decir cuándo esos nodos son probabilísticamente dependientes. (a) La información fluye entre ascendientes y descendientes siempre que no haya sido observado algún nodo en medio del camino. (b) La información también fluye entre nodos hermanos, mientras el padre común no haya sido observado. (c) En el último caso, donde dos nodos son padres de un mismo nodo, el flujo se activa cuando el hijo común o algún descendiente suyo es observado; es costumbre referirse a la figura que forman los nodos  $A - \underline{Z} - B$  como estructura-v.

esta ruta se debe a que un conjunto de variables  $\mathbf{Z}$  fueron observadas, decimos que  $A$  y  $B$  son condicionalmente independientes dadas  $\mathbf{Z}$  ( $P(A, B|\mathbf{Z}) = P(A|\mathbf{Z})P(B|\mathbf{Z})$ ) [figura 7].

A partir de aquí se determinará cuándo, de acuerdo con la gráfica, es posible afirmar que dos variables son independientes, condicionalmente independientes o independientes. Conceptos como el Manto de Márkov,



**Figura 8.** Formaci\u00f3n del manto de M\u00e1rkov. Al observar a las variables en los nodos padre  $P_i$  de un nodo  $X$  se corta el flujo hacia otros ancestros. Igualmente, al observar a los hijos  $H_i$  se corta el flujo hacia otros descendientes. Sin embargo, al haber observado a los hijos se activa el flujo hacia los padres  $W_i$  de los hijos, pues se han creado estructuras-v, por ello estos tambi\u00e9n forman parte del manto. Al observar a estos \u00faltimos se impide el flujo hacia cualquier otra parte de la gr\u00e1fica de Bayes.



**Figura 9.** Cuando no existen rutas activas entre dos variables de una red de Bayes se dice que est\u00e1n D-separadas. En este ejemplo  $A$  y  $B$  est\u00e1n D-separadas, dado que  $Z$  fue observada; por ello es posible escribir  $P(A, B|Z) = P(A|Z)P(B|Z)$ , siendo este un ejemplo de independencia condicional.

que es el conjunto m\u00ednimo de nodos alrededor de una variable  $X$  que se requiere observar para determinar la probabilidad de que  $X$  tome un valor concreto  $x_i$ , independientemente del resto de la gr\u00e1fica [figura 8]

y la D-separación, cuando dos variables en la gráfica son independientes entre sí, emergerán de este estudio [figura 9]. Los ejemplos que se presentan a continuación incluyen el uso de este cálculo de flujos.

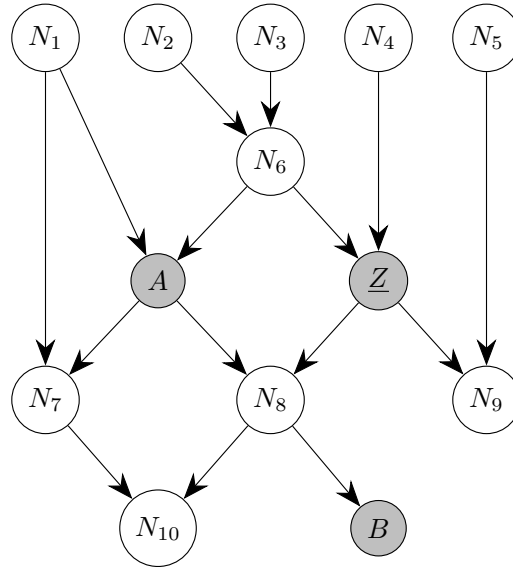
## 2.5 Inferencia exacta en redes de Bayes

Para apreciar las ventajas de la metodología que se ha venido construyendo y se desarrolla en esta sección es útil introducir un ejercicio con aproximadamente 12 nodos. Para ilustrar mejor las aplicaciones se podría plantear una red que represente alguna situación real, se podría elegir algún ejemplo de entre: problemas policíacos, fiestas con amigos, diagnóstico de enfermedades, etc. Sin embargo aquí solo se presentarán las reglas en su forma abstracta y se ilustrarán algunos conceptos utilizando los grafos de las figuras 6 y 10. El objetivo será explicar cómo se realizan consultas a bases de conocimiento con varias variables aleatorias, relacionadas mediante una gráfica de Bayes.

### 2.5.1. Consultas

Una consulta a una red de Bayes tendrá la forma del cálculo de una probabilidad. Para abreviar cuando se desea realizar consultas para todos los posibles valores de una variable se acostumbra utilizar el nombre de la variable con mayúscula, mientras que si solo se quiere un valor en particular este se especifica. Por ejemplo:

- ¿Cuál es la probabilidad de que un paciente tenga tifoidea dado que presenta fiebre y despidе un aroma picante?  $P(Tifoidea = 1 | Fiebre = 1, Aroma = 1)$
- ¿Cuál es la probabilidad de que un paciente tenga fiebre?  $P(Fiebre = 1)$
- ¿Cuál es la probabilidad de que un paciente no tenga fiebre?  $P(Fiebre = 0)$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga fiebre?  $P(Fiebre)$
- ¿Cuál es la probabilidad de que un paciente tenga tos y fiebre?  $P(Tos = 1, Fiebre = 1)$
- ¿Cuál es la probabilidad de que un paciente tenga tos junto con cada valor posible de fiebre?  $P(Tos = 1, Fiebre)$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga tos junto con cada valor posible de fiebre?  $P(Tos, Fiebre)$
- ¿Cuál es la probabilidad de que un paciente tenga/no tenga tos si se observa si tiene o no fiebre?  $P(Tos | Fiebre)$



Nodo sombreado, la variable  $X$  aparece en la consulta.



Variable subrayada, el valor de la variable  $X$  ha sido observado.

**Figura 10.** Ejemplo de una red bayesiana con más de 10 nodos. Para el ejemplo se quiere calcular una distribución de probabilidad que incluye a  $A, B$  y la variable evidencia  $Z$ .

### 2.5.2. Cálculo

El proceso para calcular la consulta consta de los pasos siguientes. Para ejemplificar cada paso se utilizará la red en figura 10. Se hace notar que el ejemplo de este documento es uno de los más complejos que se presentarían en un curso a nivel licenciatura.

1. Identificar a las variables que aparecen en la consulta ¿nos interesa solo un valor en particular? ¿o queremos la tabla para los eventos con todos sus valores posibles?

*Ej. Obtener  $P(A, B|Z = 2)$ , en este caso queremos las respuestas para todos los valores posibles de  $A$  y  $B$ , condicionados a un único valor de  $Z$ . En la figura 10 los nodos participantes han sido resaltados.*

2. Al estudiar este tema por primera vez, si la distribución a calcular es condicional, es conveniente comenzar por identificar a las probabilidades conjuntas requeridas según (2).

Se podrá trabajar con las distribuciones en el numerador y el denominador de forma independiente, por lo que en los pasos siguientes se asumirá que se quiere calcular una distribución de probabilidad conjunta, lo cual implica también que ya no hay nodos observados.

*En el ejemplo:*

$$P(A, B|Z = 2) = \frac{P(A, B, Z = 2)}{P(Z = 2)} \quad (3)$$

3. Determinar si hay rutas activas entre las variables participantes: los nodos que aparezcan a lo largo de estas rutas tendrán que ser incluidos en el cálculo. Si no hay rutas activas, es posible asumir independencia entre las variables desconexas y realizar factorizaciones de la forma:

$$P(X_1, \dots, X_n, Y_1, \dots, Y_m) = P(X_1, \dots, X_n)P(Y_1, \dots, Y_m)$$

donde las variables  $X_1, \dots, X_n$  se encuentran conectadas entre sí por rutas activas y las  $Y_1, \dots, Y_m$  pertenecen a otro grupo, desconectado del anterior.

*En el ejemplo las variables  $N_6$  y  $N_8$  deberán ser agregadas al cálculo [figura 11].*

4. Identificar a los ancestros de las variables participantes.

*En el ejemplo los ancestros de  $A, B$  y  $Z$  sobre rutas activas son agregados. [figura 12].*

5. Se deberá marginalizar a aquellas variables que no aparecían explícitamente en la pregunta.

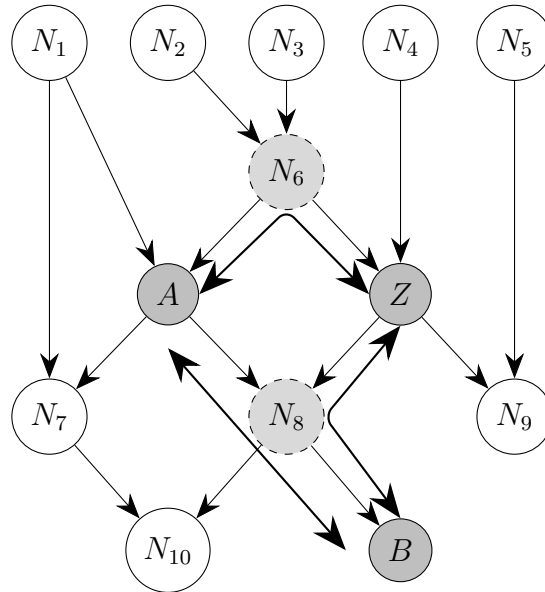
$$P(X_1, \dots, X_n, W_1, \dots, W_n) = \sum_{W_i} P(X_1, \dots, X_n, W_1, \dots, W_n)$$

donde las  $W_i$  representan a variables que no estaban en la pregunta original y se marginalizarán realizando una suma sobre todos los posibles valores de esas variables.

Utilizar la regla de la cadena para factorizar las distribuciones conjuntas a calcular. Se deberán incluir a todos los nodos identificados en los pasos anteriores. Al elegir en qué orden acomodar a los nodos en la regla de la cadena, tener cuidado de poner a los nodos en la parte de abajo de la gráfica condicionados a sus ancestros y no a los ancestros condicionados a sus descendientes.

$$P(X_1, \dots, X_n, W_1, \dots, W_n) = \sum_{W_i} P(X_1|X_2, \dots, W_n)P(X_2|X_3, \dots, W_n) \dots P(W_n)$$

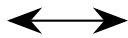
*En el ejemplo se marginaliza a los nodos  $N_i$  que fueron agregados:*



Nodo sombreado, la variable  $X$  aparece en la consulta.



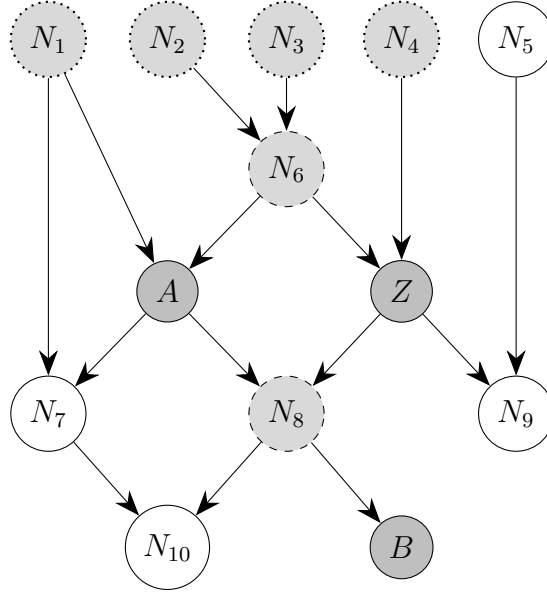
Nodo sombreado con borde segmentado, forma parte de una o más rutas activas.






Hay flujo de información entre las variables.

**Figura 11.** Hay una ruta activa entre  $A$  y  $B$ , pasando por  $N_8$  y otra entre  $Z$  y  $B$  pasando por el mismo nodo. También hay otra ruta entre  $A$  y  $Z$  pasando por  $N_6$ . Por lo anterior, todos los nodos sombreados deben ser incluidos en el cálculo.





-   $X$     Nodo sombreado, la variable  $X$  aparece en la consulta.
-   $X$     Nodo sombreado con borde segmentado, forma parte de una o más rutas activas.
-   $X$     Nodo sombreado con borde punteado, es ancestro sobre flujo activo.

**Figura 12.** Se añade a los ancestros de  $A, B, Z$  y otros nodos agregados en el paso anterior.

$$\begin{aligned}
 P(A, B|Z = 2) &= \frac{P(A, B, Z = 2)}{P(Z = 2)} \\
 P(A, B, Z = 2) &= \sum_{\substack{N_1, N_2, N_3, \\ N_4, N_6, N_8}} P(A, B, Z = 2, N_1, N_2, N_3, N_4, N_6, N_8) \quad (4) \\
 &= \sum_{\substack{N_1, N_2, N_3, \\ N_4, N_6, N_8}} P(B|A, Z = 2, N_1, N_2, N_3, N_4, N_6, N_8) \\
 &\quad P(N_8|A, Z = 2, N_1, N_2, N_3, N_4, N_6) \\
 &\quad P(A|Z = 2, N_1, N_2, N_3, N_4, N_6) \\
 &\quad P(Z = 2|N_1, N_2, N_3, N_4, N_6) \\
 &\quad P(N_6|N_1, N_2, N_3, N_4) \\
 &\quad P(N_1|N_2, N_3, N_4) \\
 &\quad P(N_2|N_3, N_4) \\
 &\quad P(N_3|N_4) \\
 &\quad P(N_4)
 \end{aligned}$$

6. Utilizar la red de Bayes para eliminar de las condiciones a los nodos con los cuales hay una relación de independencia condicional. Es decir, este es el paso donde se utiliza que basta con que los nodos participantes aparezcan condicionados a los valores de sus nodos padres, pues el flujo hacia sus demás ancestros ha sido cortado, llegando a lo que se conoce como la *regla de la cadena para redes de Bayes*:

$$P(X_1, \dots, X_n) = \prod_{X_i} P(X_i | \text{padres}(X_i))$$

*En el ejemplo, como cada variable ha quedado condicionada a sus ancestros podemos aplicar la fórmula de independencia condicional*

$$P(X, W_1, \dots, W_n | Z_1, \dots, Z_n) = \\ P(X | Z_1, \dots, Z_n) P(W_1, \dots, W_n | Z_1, \dots, Z_n)$$

*junto con una aplicación de la regla de la cadena (1)*

$$P(X, W_1, \dots, W_n | Z_1, \dots, Z_n) = \\ P(X | Z_1, \dots, Z_n, W_1, \dots, W_n) P(W_1, \dots, W_n | Z_1, \dots, Z_n)$$

despejando  $P(W_1, \dots, W_n | Z_1, \dots, Z_n)$  de ambas ecuaciones obtenemos:

$$P(X | Z_1, \dots, Z_n, W_1, \dots, W_n) = P(X | Z_1, \dots, Z_n)$$

donde los  $Z_i$  representan a los padres y los  $W_j$ , a cualquier otro ancestro. Con ello, el desarrollo anterior se simplifica a:

$$\begin{aligned}
 P(A, B|Z = 2) &= \frac{P(A, B, Z = 2)}{P(Z = 2)} \\
 P(A, B, Z = 2) &= \sum_{\substack{N_1, N_2, N_3, \\ N_4, N_6, N_8}} P(A, B, Z = 2, N_1, N_2, N_3, N_4, N_6, N_8) \\
 &= \sum_{\substack{N_1, N_2, N_3, \\ N_4, N_6, N_8}} P(B|N_8) \\
 &\quad P(N_8|A, Z = 2) \\
 &\quad P(A|N_1, N_6) \\
 &\quad P(Z = 2|N_4, N_6) \\
 &\quad P(N_6|N_2, N_3) \\
 &\quad P(N_1) \\
 &\quad P(N_2) \\
 &\quad P(N_3) \\
 &\quad P(N_4)
 \end{aligned} \tag{5}$$

Una vez comprendido este paso, es más rápido pasar directamente de (4) a (5).

7. Factorizar sumas y multiplicaciones tratando de reducir lo más posible el número de operaciones a realizar. Inicialmente basta con hacer este trabajo a mano, posteriormente se podrán introducir algoritmos de teoría de gráficas para realizar este paso de forma automática pero, como se mencionó, ese procedimiento está fuera del alcance de este artículo.

Este paso es crucial para que problemas que serían intratables de ser resueltos por fuerza bruta, puedan ser calculados en un tiempo razonable. El motivo es que cada vez que se realiza una marginalización el número de términos a considerar es menor. En la estructura factores se nota pues que el número de renglones en el factor resultante es el número de renglones del factor original entre el número de valores posibles para la variable marginalizada. El libro de [10] muestra cómo resolver ejemplos realizando las sustituciones numéricas sobre estas fórmulas, aquí veremos cómo sustituir por operaciones entre factores. Numéricamente se realizan exactamente las mismas operaciones, pero este método es más sistemático.

*En el ejemplo se procede a factorizar tantos términos como sea viable de las sumas sobre valores posibles de las variables:*

Formalmente:

$$P(A, B, Z = 2) = \left( \sum_{N_1} P(N_1) \left( \sum_{N_2} P(N_2) \left( \sum_{N_3} P(N_3) \left( \sum_{N_4} P(N_4) \right. \right. \right. \right. \right. \right. \left. \left. \left. \left. \left( \sum_{N_6} P(N_6|N_2, N_3) P(A|N_1, N_6) P(Z = 2|N_4, N_6) \right. \right. \right. \right. \right. \left. \left. \left. \left. \left( \sum_{N_8} P(B|N_8) P(N_8|A, Z = 2) \right) \right) \right) \right) \right) \right) \right) \quad (6)$$

Dado que tantos paréntesis dificultan el trabajo, se suele simplificar la notación asumiendo que cada símbolo de suma afecta a todas las sumas a su derecha, de acuerdo a la convención que da prioridad a la multiplicación sobre la suma, a menos que se agreguen paréntesis para modificar esta precedencia. La misma ecuación entonces se escribe:

$$P(A, B, Z = 2) = \sum_{N_1} P(N_1) \sum_{N_2} P(N_2) \sum_{N_3} P(N_3) \sum_{N_4} P(N_4) \quad (7)$$

$$\sum_{N_6} P(N_6|N_2, N_3) P(A|N_1, N_6) P(Z = 2|N_4, N_6)$$

$$\sum_{N_8} P(B|N_8) P(N_8|A, Z = 2)$$

Repitiendo los mismos pasos también se podría calcular  $P(Z = 2)$ :

$$P(Z = 2) = \sum_{N_2, N_3, N_4, N_6} P(Z = 2, N_6, N_2, N_3, N_4)$$

$$= \sum_{N_2} P(N_2) \sum_{N_3} P(N_3) \sum_{N_4} P(N_4)$$

$$\sum_{N_6} P(Z = 2|N_6, N_4) P(N_6|N_2, N_4) \quad (8)$$

para luego dividir  $P(A, B, Z = 2)$  entre  $P(Z = 2)$ . Pero con el uso de factores existe otra alternativa, como se verá en el paso siguiente.

8. Utilizar a los factores para resolver las multiplicaciones, marginalizaciones y reducciones que indiquen las fórmulas obtenidas en el paso anterior de acuerdo a la regla siguiente: Cada vez que se elige un valor concreto de una variable, se aplica una reducción;

cada producto se convierte en una multiplicación de factores; cada suma, en una marginalización de la variable correspondiente.

*En el ejemplo, al haber elegido  $Z = 2$  indicamos una reducción. A cualquier factor que contenga la variable  $Z$  se le aplicará la reducción sobre esta variable para este valor. Por ejemplo, el factor que representa a  $P(Z = 2|N_6, N_4)$  quedará solo con los renglones donde  $Z = 2$ .*

*Tras haber realizado todas las operaciones indicadas por la ecuación (7), se obtendrá un factor con la distribución de probabilidad  $P(A, B, Z = 2)$ , es decir, tendrá tantos renglones como combinaciones de valores posibles haya para  $A, B$ , todas cuando  $Z = 2$ . Para convertir estos valores en una distribución de probabilidad condicionada a  $Z = 2$ , se necesitaría que la suma de estas probabilidades fuera uno. El número entre el cual es necesario dividir es  $P(Z = 2)$ , como especifica (3).*

*Curiosamente, podemos observar que sumar todos los renglones del factor  $P(A, B, Z = 2)$  es lo mismo que marginalizar a  $A$  y  $B$  de la distribución conjunta, esto es:  $\sum_{A,B} P(A, B, Z = 2) = P(Z = 2)$ . De este modo, calcular  $P(Z = 2)$  no requirió ya mucho esfuerzo adicional al trabajo ya realizado. Por ello es que, para obtener  $P(A, B|Z = 2)$  basta con normalizar  $P(A, B, Z = 2)$  y ya no es necesario ejecutar las operaciones indicadas en (8).*

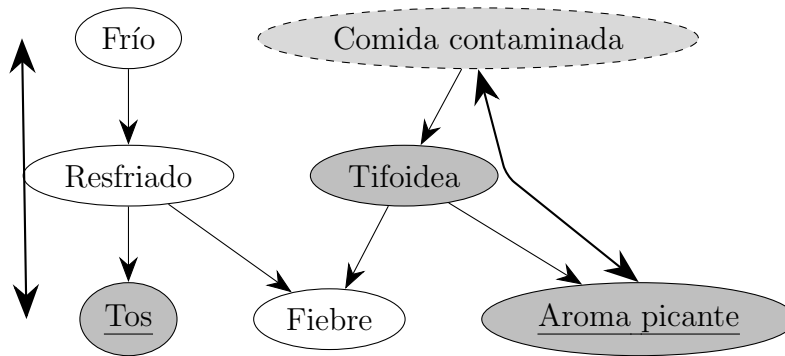
*Obsérvese que, si se hubiera pedido calcular  $P(A = 1, B = 0|Z = 2)$  y se hubieran hecho las reducciones correspondientes, no se contaría con el factor con todos los valores posibles de  $A$  y  $B$  y por ende sí sería necesario calcular  $P(Z = 2)$  aparte.*

Por ello hay dos estrategias para calcular:  $P(X_1 = x_1, \dots, X_n = x_n|Z_1 = z_1, \dots, Z_m = z_m)$ :

- (a) Aplicar las reducciones correspondientes solo a variables evidencia, obteniéndose el factor con la distribución de probabilidad conjunta  $P(X_1, \dots, X_n, Z_1 = z_1, \dots, Z_m = z_m)$ , normalizar y seleccionar al final el renglón de interés.
- (b) Aplicar todas las reducciones y calcular tanto el numerador como el denominador de la fórmula para la probabilidad condicional.

Cuál estrategia es más recomendable depende del número de variables en  $X$  y del número de valores en su dominio: si este número es alto, el número de renglones con combinaciones posibles crece y podría ser más recomendable solo realizar las operaciones con los renglones de interés. Por el contrario, si estos números son pequeños, es más sencillo obtener un solo factor y normalizarlo.

La estrategia utilizada en [10] corresponde a la primer opción: se calculan las probabilidades  $P(X_1, \dots, X_n)$  para luego normalizar



**Figura 13.**  $P(\text{tifoidea}|\text{tos}, \text{aroma picante})$ . Los nodos sombreados aparecen en la consulta, el de borde segmentado será introducido en el cálculo y luego, marginalizado, los nodos con fondo blanco no son requeridos. Las flechas con dos puntas indican los flujos activos.

dividiendo entre una constante  $\alpha$ , que es la suma de todas ellas; ahora se puede ver que esta corresponde en realidad a la probabilidad de la evidencia  $P(z_1, \dots, z_m)$  calculada indirectamente.

Una vez asimilado este mecanismo es posible esbozar una demostración de porqué es posible ignorar a las variables que no quedaron sobre los flujos activos, a partir de la distribución de probabilidad conjunta completa y la factorización de sus términos. Para ello basta analizar qué sucede con los términos que involucran a las hojas del grafo, que no estaban sobre flujos activos: factorizándolos correctamente se obtiene una suma sobre todos sus valores posibles para valores fijos de la evidencia, en caso de que la haya; luego entonces estos elementos se convierten en unos que multiplican al resto de la fórmula y pueden ser eliminados.

### 2.5.3. Ejemplo

Utilizando el diagrama de la figura 6 veamos cómo aplican los pasos anteriores para calcular la probabilidad de que alguien tenga tifoidea dado que tiene tos y aroma picante.

1. Para comenzar, esta consulta se expresa:

$$P(\text{tifoidea}|\text{tos}, \text{aroma picante})$$

Los nodos participantes se muestran en la figura 13.

2. Formalmente, la distribución de probabilidad condicional queda como:

$$P(\text{tifoidea}|\text{tos}, \text{aroma picante}) = \frac{P(\text{tifoidea}, \text{tos}, \text{aroma picante})}{P(\text{tos}, \text{aroma picante})}$$

sin embargo el paso siguiente tiene más información que aportar.

3. Al determinar los grupos de nodos con flujos activos es posible notar lo siguiente:
- Siendo *Aroma picante* descendiente directo de *Tifoidea*, siempre fluye información entre ellos. Siendo *Comida contaminada* ancestro de *Tifoidea*, también continúa el flujo en esa dirección. Intuitivamente esto coincide con el hecho de que, si la persona produce el aroma picante característico de la tifoidea, es probable que tenga esta enfermedad porque ingirió comida contaminada.
  - Entre *Tos* y las demás variables de interés no hay un flujo activo. De *Tos* hacia sus ancestros *Resfriado* y *Frío* sí lo hay, pero hacia *Fiebre* aparece una estructura-v, por lo que ya no hay paso. Lo que esto nos indica es que la tos en el paciente no está relacionada con que tenga *Tifoidea*, por lo que no nos brinda información al respecto. En todo caso, podría ser que el desafortunado paciente tenga un resfriado y tifoidea; calcular esa probabilidad podría ser otro ejercicio.
  - Gracias a las dependencias e independencias identificadas es posible simplificar la fórmula de paso anterior:

$$\begin{aligned} P(\textit{tifoidea}|\textit{tos}, \textit{aroma picante}) &= P(\textit{tifoidea}|\textit{aroma picante}) \\ &= \frac{P(\textit{aroma picante}, \textit{tifoidea})}{P(\textit{aroma picante})} \end{aligned}$$

4. El nodo *Comida contaminada* será incluido en el cálculo por ser padre de *Tifoidea*.
5. (Directo al siguiente paso)
6. La distribución de probabilidad conjunta se calcula entonces:

$$\begin{aligned} P(\textit{aroma picante}, \textit{tifoidea}) &= \sum_{\substack{\textit{Comida} \\ \textit{contaminada}}} (P(\textit{Comida contaminada}) \\ &\quad P(\textit{tifoidea}|\textit{Comida contaminada}) \\ &\quad P(\textit{aroma picante}|\textit{tifoidea})) \end{aligned}$$

7. Factorizando las sumas:

$$\begin{aligned} P(\textit{aroma picante}, \textit{tifoidea}) &= P(\textit{aroma picante}|\textit{tifoidea}) \\ &\quad \sum_{\substack{\textit{Comida} \\ \textit{contaminada}}} (P(\textit{tifoidea}|\textit{Comida contaminada}) \\ &\quad P(\textit{Comida contaminada})) \end{aligned}$$

8. Requerimos las definiciones de los factores con:
- $P(\textit{Aroma\_picante}|\textit{Tifoidea})$ ,

- $P(Tifoidea|Comida\ contaminada)$  y
- $P(Comida\ contaminada)$

Para obtener el resultado final tomaremos el atajo de calcular los resultados para todos los valores de *Tifoidea*, la alternativa es obtener ahora la fórmula para  $P(aroma\ picante)$  repitiendo los pasos anteriores. Suponiendo que contamos con un programa que implementa las operaciones de factores, el código para resolver este ejercicio podría verse como el algoritmo 1.

---

**Algoritmo 1**  $P(tifoidea|tos, aroma\ picante)$

---

▷ Consideramos todos los valores posibles para Tifoidea  
 $f \leftarrow \text{MULTIPLICA}(pTifoidea\_Comidacontaminada, pComidacontaminada)$   
 $f \leftarrow \text{MARGINALIZA}(f, Comidacontaminada)$   
 ▷  $f$  contiene ahora  $P(Tifoidea)$   
 ▷ La variable Aromapicante vale 1  
 $pAromapicante\_Tifoidea \leftarrow \text{REDUCE}(pAromapicante\_Tifoidea, Aromapicante, 1)$   
 $f \leftarrow \text{MULTIPLICA}(pAromapicante\_Tifoidea, f)$   
 $resultado \leftarrow \text{NORMALIZA}(resultado)$   
 ▷ Resultado contiene ahora  $P(Tifoidea|aroma\ picante)$

---

### 3. Conclusiones

El uso de factores permite realizar inferencia exacta en redes bayesianas con un alto número de nodos al permitir la sistematización de las operaciones de marginalización, reducción, normalización y producto de distribuciones de probabilidad. Existe un método claro para realizar inferencias exactas que se puede aplicar para calcular el valor de consultas a la red de Bayes, evitando cometer errores al determinar qué factores incluir en el cálculo. Igualmente la determinación de flujos activos sobre la gráfica evita considerar variables que no contribuyen al resultado final, reduciendo el número de operaciones requeridas para obtener el resultado numérico.

### Bibliografía

- [1] E. Alpaydin, *Introduction to Machine Learning*, 2.<sup>a</sup> ed., The MIT Press, 2010.
- [2] K. G. Ardlie, D. S. Deluca, A. V. Segrè, T. J. Sullivan, T. R. Young, E. T. Gelfand, C. A. Trowbridge, J. B. Maller, T. Tukiainen, M. Lek, L. D. Ward, P. Kheradpour, B. Iriarte, Y. Meng, C. D. Palmer, T. onu Esko, W. Winckler, J. N. Hirschhorn,



- M. Kellis, D. G. MacArthur, G. Getz, A. A. Shabalín, G. Li, Y.-H. Zhou, A. B. Nobel, I. Rusyn, F. A. Wright, T. Lappalainen, P. G. Ferreira, H. Ongen, M. A. Rivas, A. Battle, S. Mostafavi, J. Monlong, M. Sammeth, M. Mele, F. Reverter, J. M. Goldmann, D. Koller, R. Guigó, M. I. McCarthy, E. T. Dermitzakis, E. R. Gamazon, H. K. Im, A. Konkashbaev, D. L. Nicolae, N. J. Cox, T. Flutre, X. Wen, M. Stephens, J. K. Pritchard, Z. Tu, B. Zhang, T. Huang, Q. Long, L. Lin, J. Yang, J. Zhu, J. Liu, A. Brown, B. Mestichelli, D. Tidwell, E. Lo, M. Salvatore, S. Shad, J. A. Thomas, J. T. Lonsdale, M. T. Moser, B. M. Gillard, E. Karasik, K. Ramsey, C. Choi, B. A. Foster, J. Syron, J. Fleming, H. Magazine, R. Hasz, G. D. Walters, J. P. Bridge, M. Miklos, S. Sullivan, L. K. Barker, H. M. Traino, M. Mosavel, L. A. Siminoff, D. R. Valley, D. C. Rohrer, S. D. Jewell, P. A. Branton, L. H. Sobin, M. Barcus, L. Qi, J. McLean, P. Hariharan, K. S. Um, S. Wu, D. Tabor, C. Shive, A. M. Smith, S. A. Buia, A. H. Undale, K. L. Robinson, N. Roche, K. M. Valentino, A. Britton, R. Burges, D. Bradbury, K. W. Hambricht, J. Seleski, G. E. Korzeniewski, K. Erickson, Y. Marcus, J. Tejada, M. Taherian, C. Lu, M. Basile, D. C. Mash, S. Volpi, J. P. Struewing, G. F. Temple, J. Boyer, D. Colantuoni, R. Little, S. Koester, L. J. Carithers, H. M. Moore, P. Guan, C. Compton, S. J. Sawyer, J. P. Demchok, J. B. Vaught, C. A. Rabiner, N. C. Lockhart, K. G. Ardlie, G. Getz, F. A. Wright, M. Kellis, S. Volpi y E. T. Dermitzakis, «The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans», *Science*, vol. 348, núm. 6235, 2015, 648–660.
- [3] M. Faix, J. Lobo, R. Laurent, D. Vaufreydaz y E. Mazer, «Stochastic Bayesian Computation for Autonomous Robot Sensorimotor System», en *Workshop on Unconventional Computing for Bayesian Inference (UCBI) at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015)*, septiembre 2015.
- [4] M. Faix, E. Mazer, R. Laurent, M. Othman Abdallah, R. Le Hy y J. Lobo, «Cognitive computation: A Bayesian machine case study», en *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC)*, 2015, 67–75.
- [5] T. Flutre, X. Wen, J. Pritchard y M. Stephens, «A Statistical Framework for Joint eQTL Analysis in Multiple Tissues», *PLOS Genetics*, vol. 9, núm. 5, 2013, 1–13.
- [6] D. Koller, «Probabilistic graphical models», Coursera, sep 2012, El curso original en su versión libre ya no se encuentra disponible, en su lugar hay ahora una especialidad que presenta el mismo contenido pero en tres módulos.
- [7] D. Koller y N. Friedman, *Probabilistic graphical models, principles and techniques*, MIT Press Cambridge, 2009.
- [8] O. Lebeltel, P. Bessiere, J. Diard y E. Mazer, «Bayesian Robot Programming», *Autonomous Robots*, vol. 16, 2004, 49–79.
- [9] G. Li, A. Shabalín, I. Rusyn, F. Wright y A. Nobel, «An Empirical Bayes Approach for Multiple Tissue eQTL Analysis», *Biostatistics (Oxford, England)*, vol. 19, 2013, 391–406.
- [10] S. Russell y P. Norving, *Artificial intelligence, a modern approach*, 2.<sup>a</sup> ed., Pearson Prentice Hall, 2010.